Inter-Comparison between ETR and EnKF using NCEP GFS Model

Mozheng Wei, Yuejian Zhu, Stephen Lord, William Lapenta
and Jeff Whitaker*
NOAA/NCEP/EMC, MD, USA
*****_NOAA Earth System Research Laboratory_, CO, USA

Acknowledgement: Richard Wobus, John Derber, Dave Parrish, Daryl Kleist and Jiayi Peng

Experimental Design

Experiments:

(a) id → x, NCO parallel, NCEP GSI analysis, ETR perturbations, scores are based on GSI analysis.

(b) id → r, EnKF ensemble mean as analysis, ETR perturbations, scores are based on EnKF analysis.

(c) id → k, EnKF ensemble mean as analysis, EnKF perturbations, scores are based on EnKF analysis.

if r is better (worse) than k, ETR perts are better (worse) than EnKF perts.
if r is better (worse) than x, EnKF analysis is better (worse) than GSI analysis.

Ensemble Methods: **(a) ETR (Ensemble Transform with Rescaling):** Operational at NOAA/NCEP since May 30, 2006 **(**Wei *et al.* 2005, 2008**).** **(b) EnKF (Ensemble Kalman Filter):** Running an ensemble Kalman filter with each member being updated by assimilating NCEP operational observations. Forward observational operator from GSI is used. (Whitaker and Hamill 2002, Whitaker *et al.* 2008).

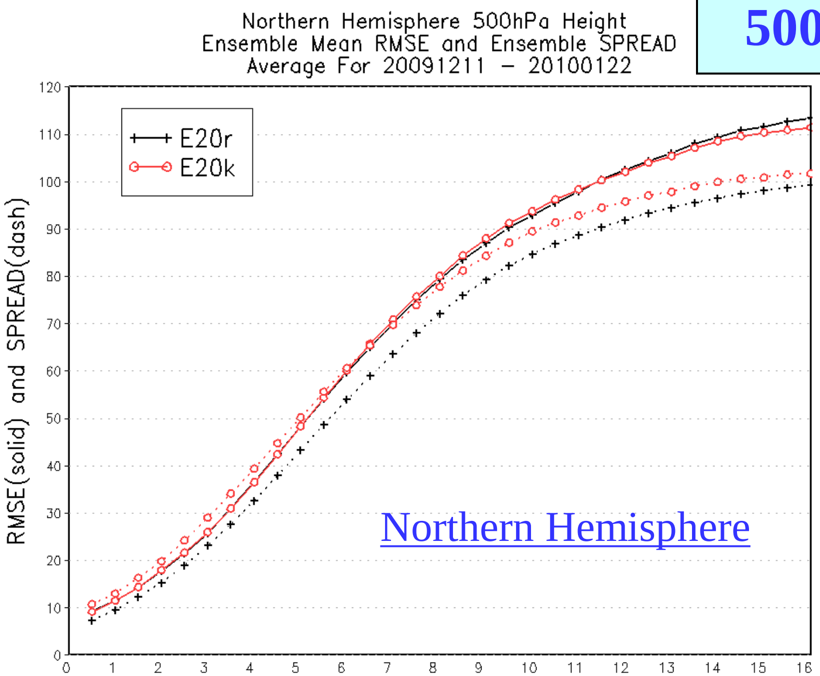
Test Period: 00Z Dec. 8, 2009 ---- 18Z Feb. 7, 2010.

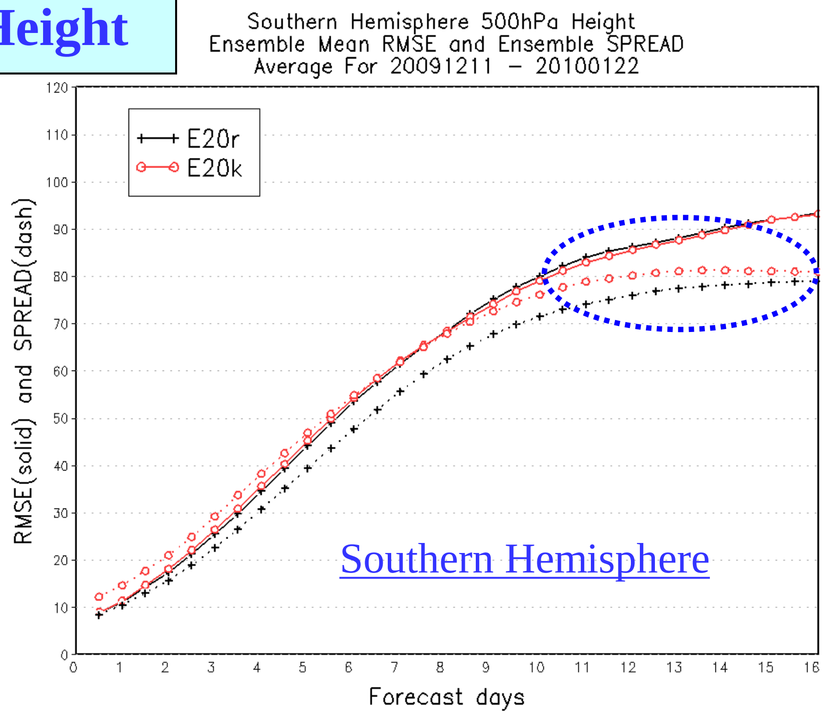
Verification Period: 00Z Dec. 11, 2009 ---- 00Z Jan. 22, 2010.

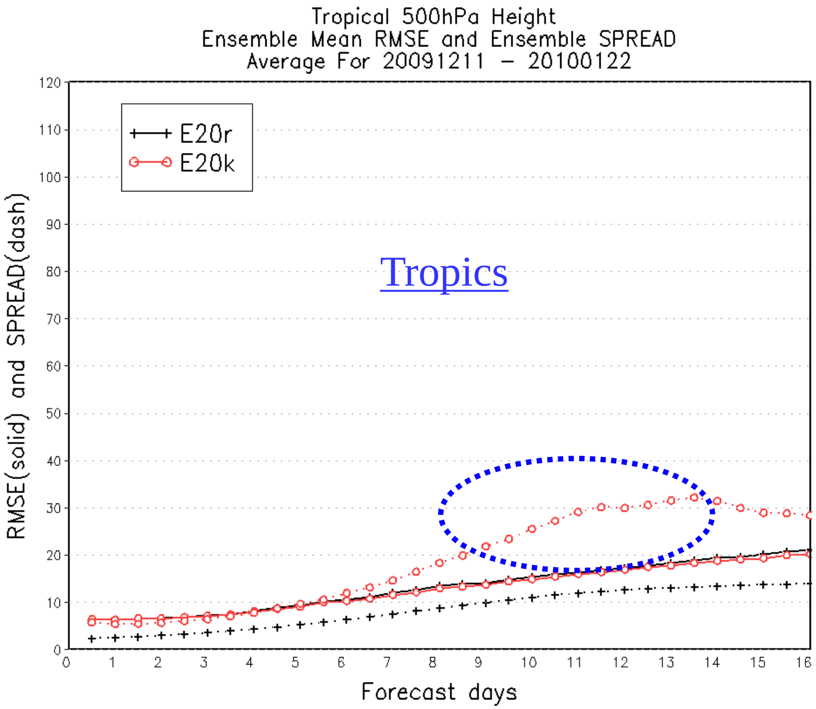
Model and Resolution: GFS, T190L28.

Ensemble Size: 20.

500hPa Height



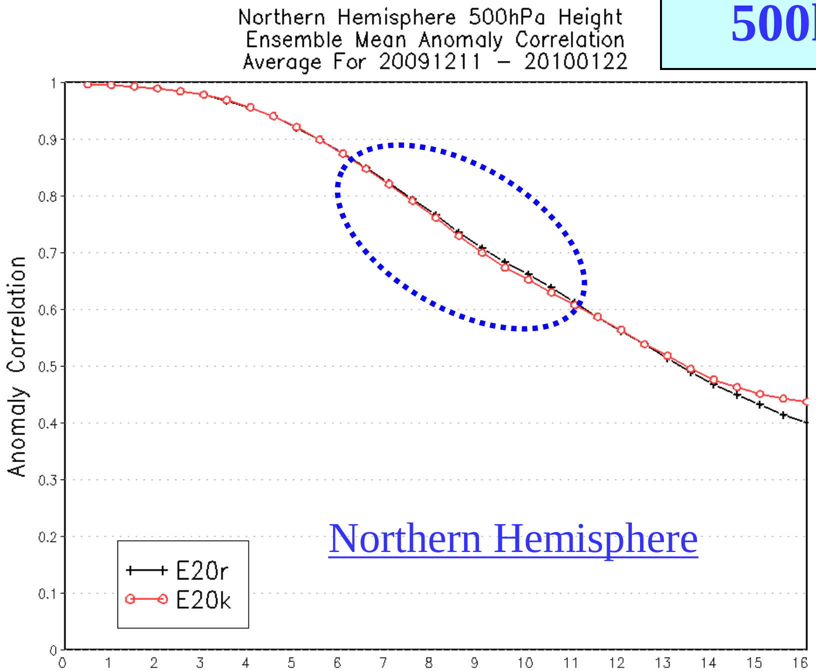


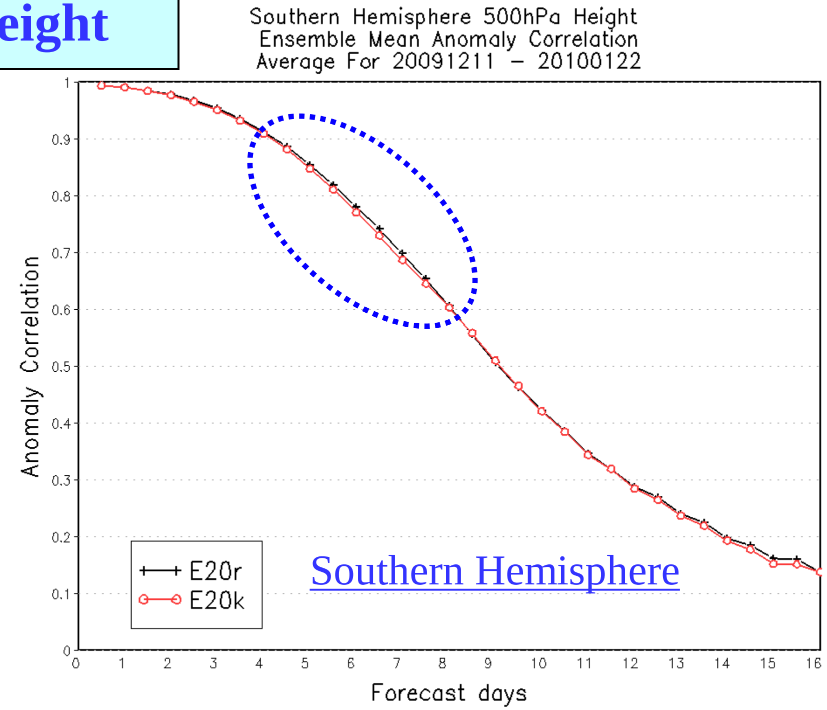


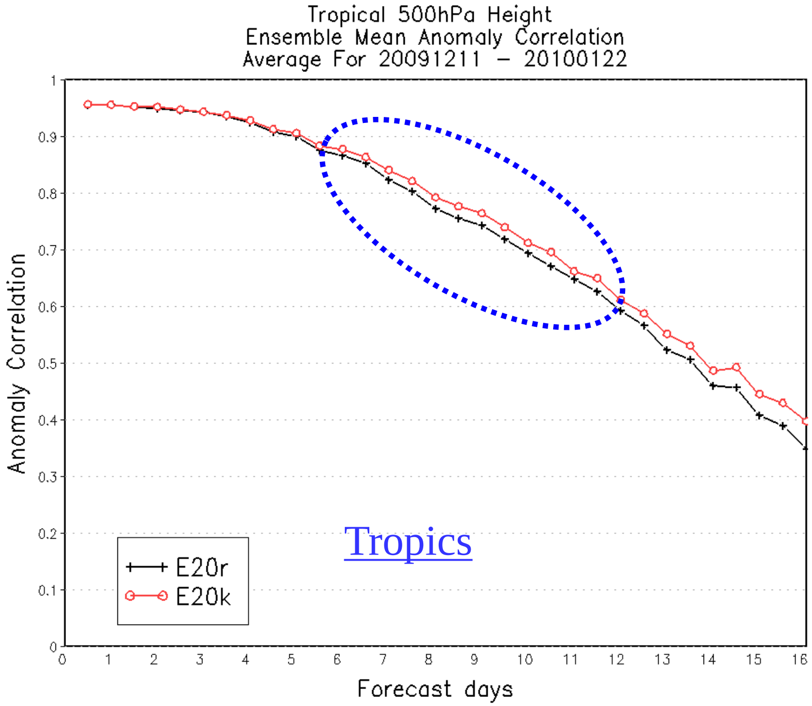
RMSE and Spread of Ensemble Mean

RMSE is similar, EnKF initial spread is notably larger than ETR. ETR spread grows faster than EnKF. Similar results are found for other variables.

500hPa height



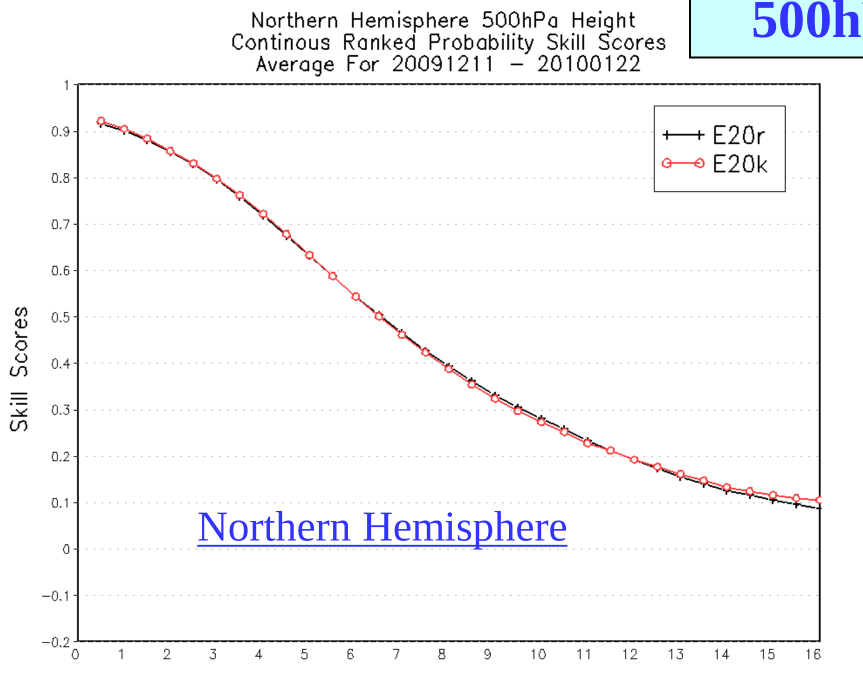


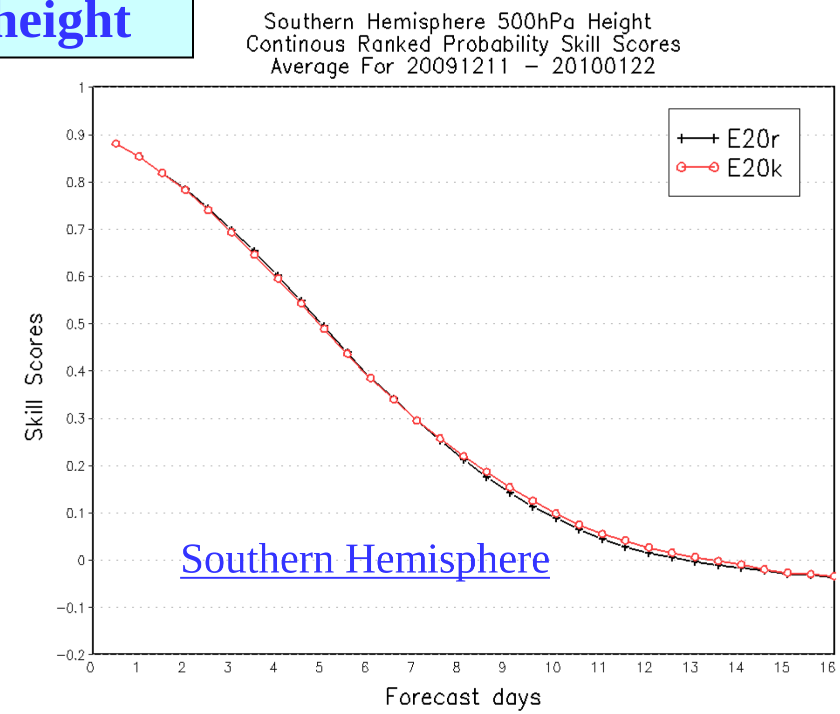


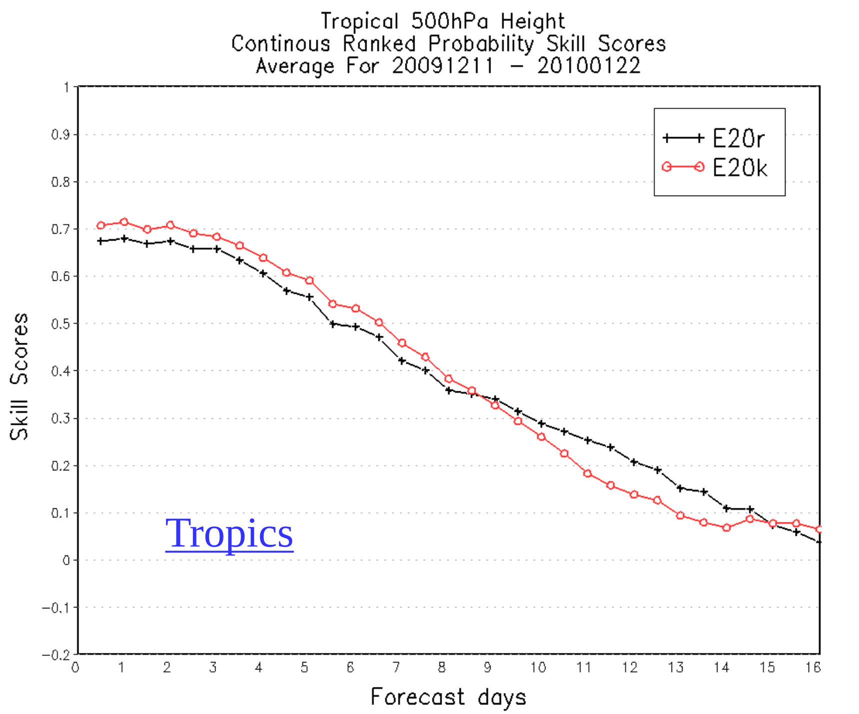
Ensemble Mean Anomaly Correlation

AC values are similar in NH and SH for shorter lead time, ETR is better for day 5-9, EnKF is better in the tropics. Results for Z1000 are very similar

500hPa height



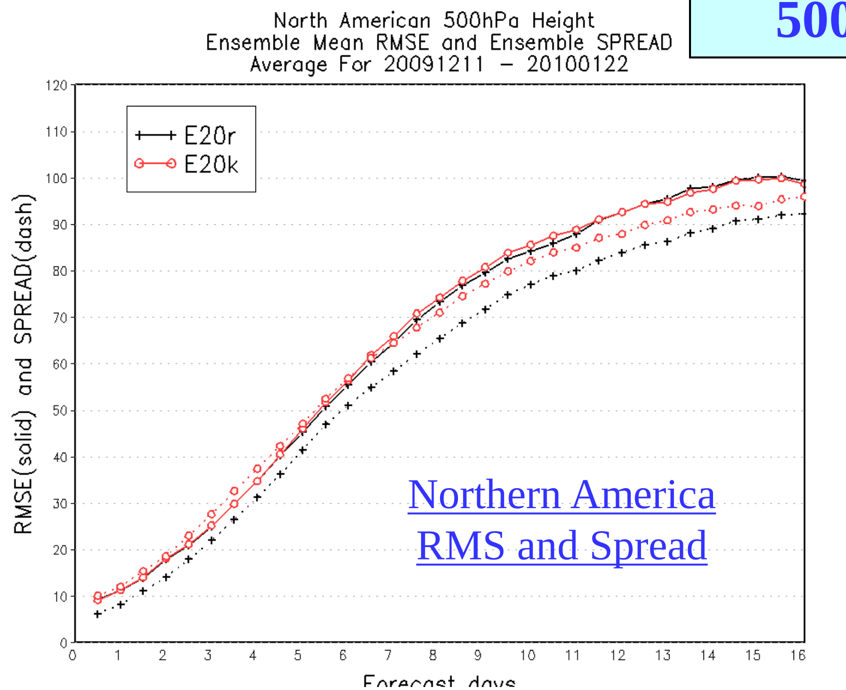


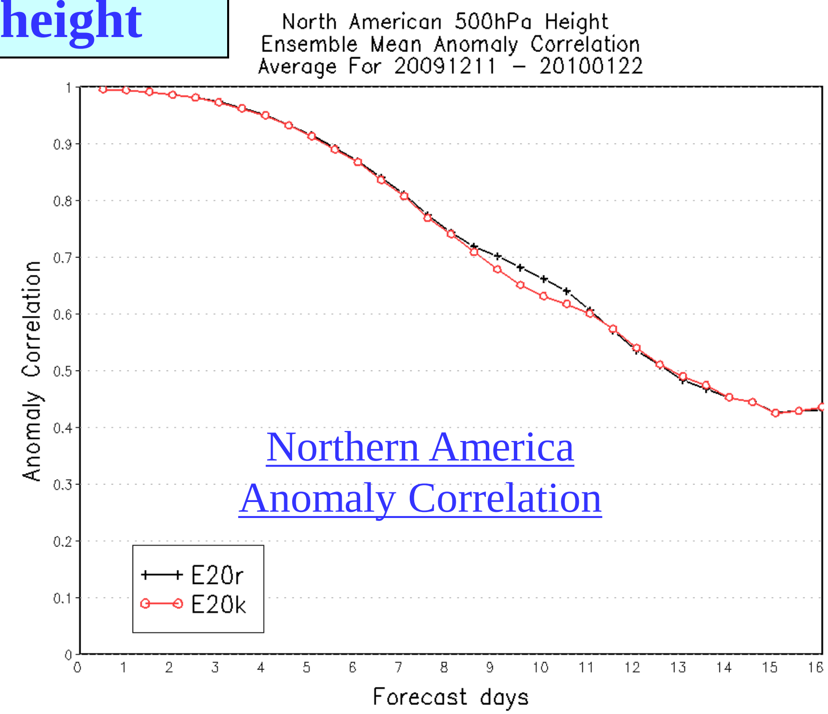


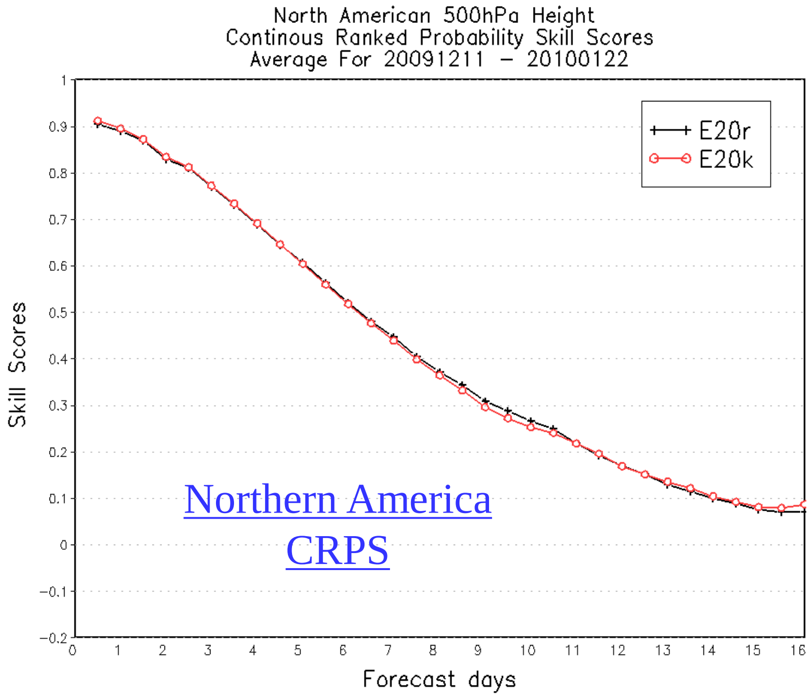
Probabilistic Scores: CRPSS

ETR and EnKF are similar over NH and SH
EnKF is better for short lead time and ETR is better for median lead time over the tropics.
This conclusion holds for Z1000.

500hPa height







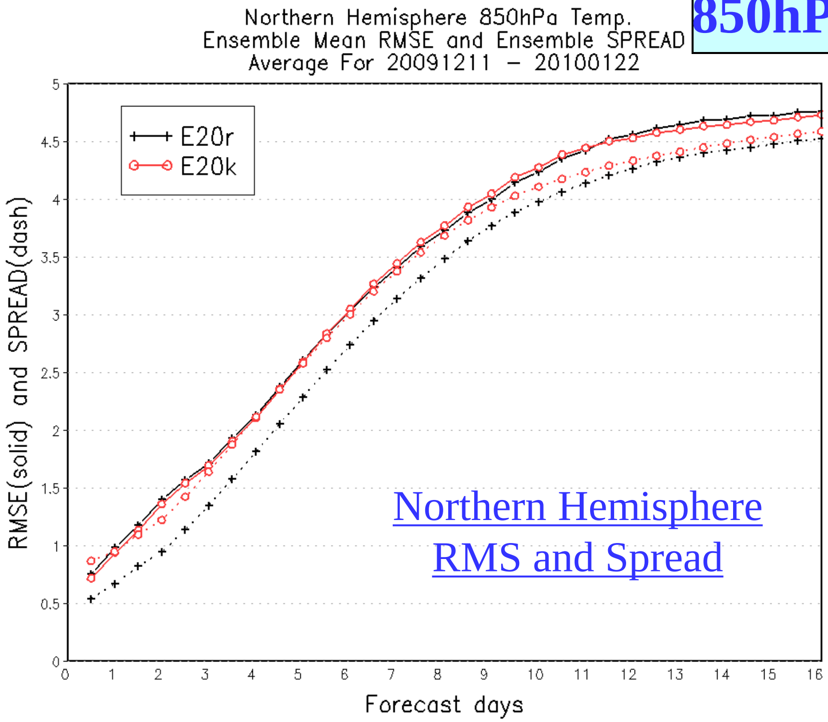
NA: RMS, Spread and CRPSS

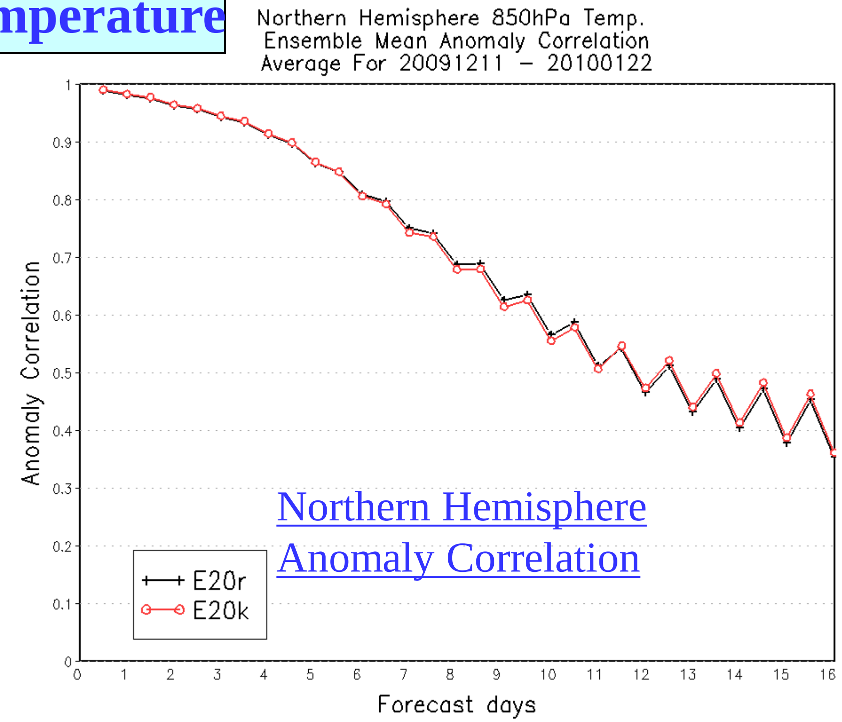
EnKF: RMSE is slightly larger, initial spread is notably larger than ETR, grows slower than ETR.

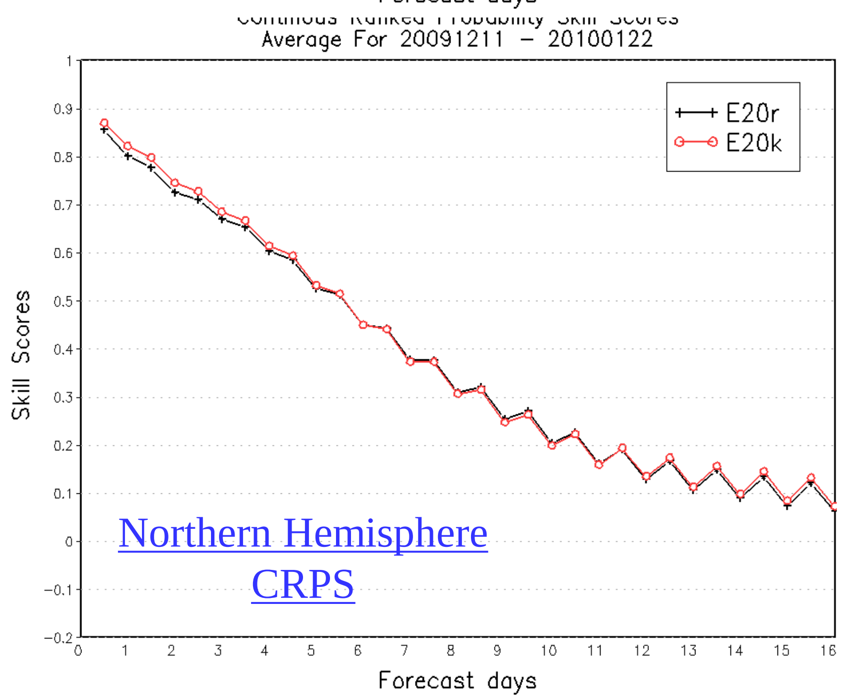
AC: similar, ETR better for lead time day 8-11.
CRPSS: similar, ETR better for lead time day 8-11

Europe: ETR is better (not shown)
Asia: EnKF is better (not shown)

850hPa temperature



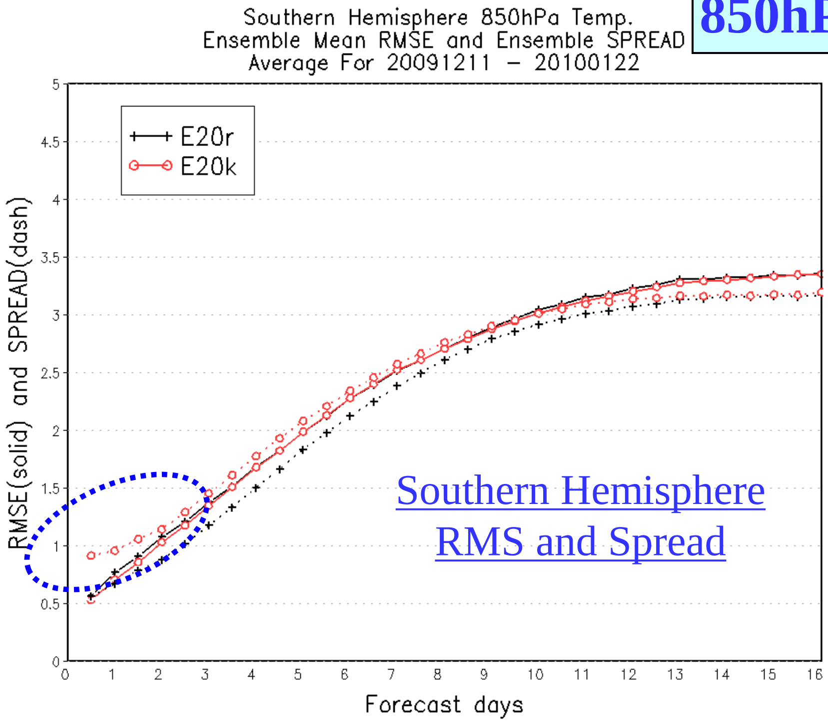


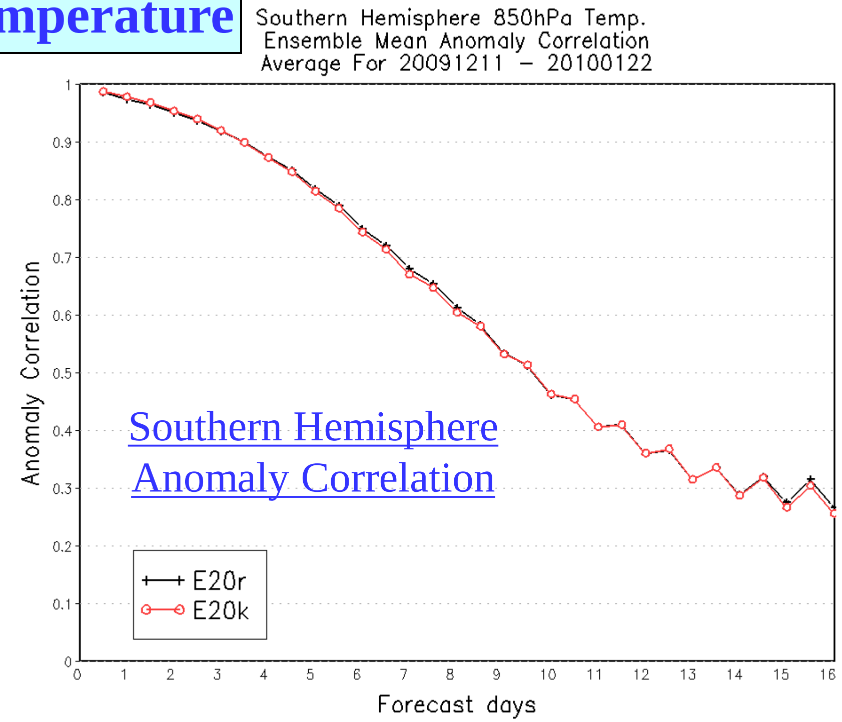


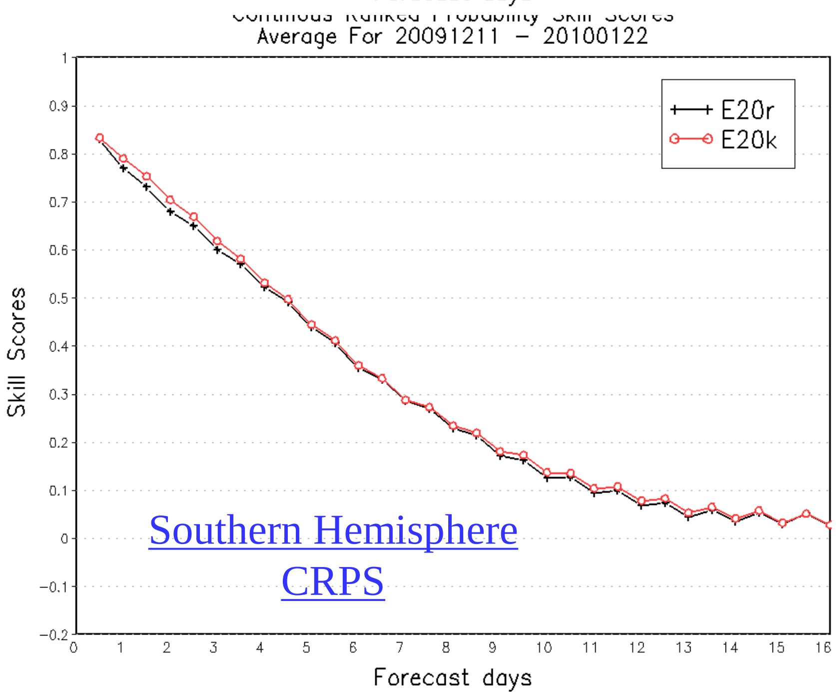
EnKF RMSE is slightly smaller for short lead time, slightly larger in median range,
EnKF initial spread is notably larger than ETR, grows slower than ETR.

AC: similar, ETR better for lead time day 7-11.
CRPSS: similar, EnKF better for short lead time.

850hPa temperature



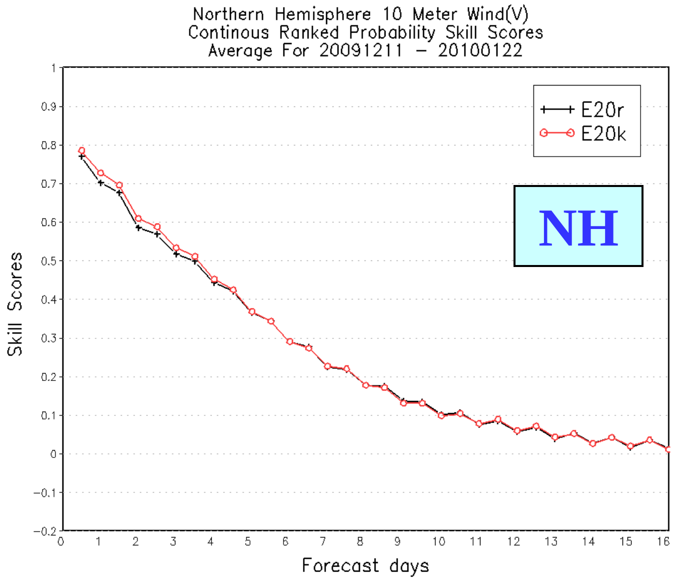


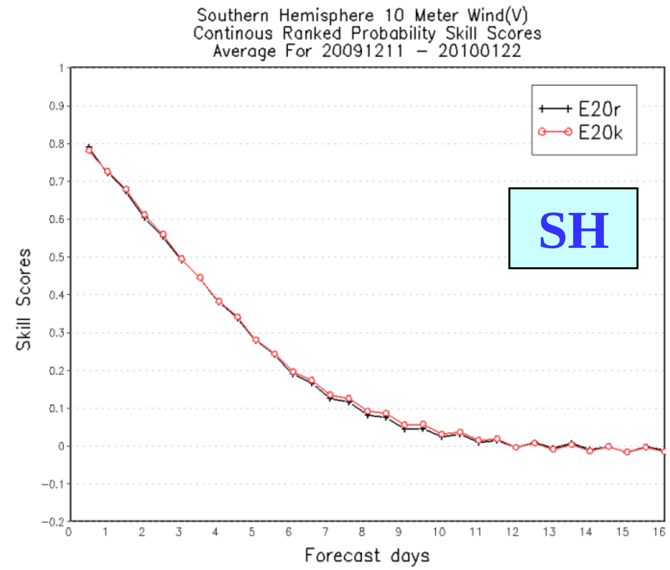


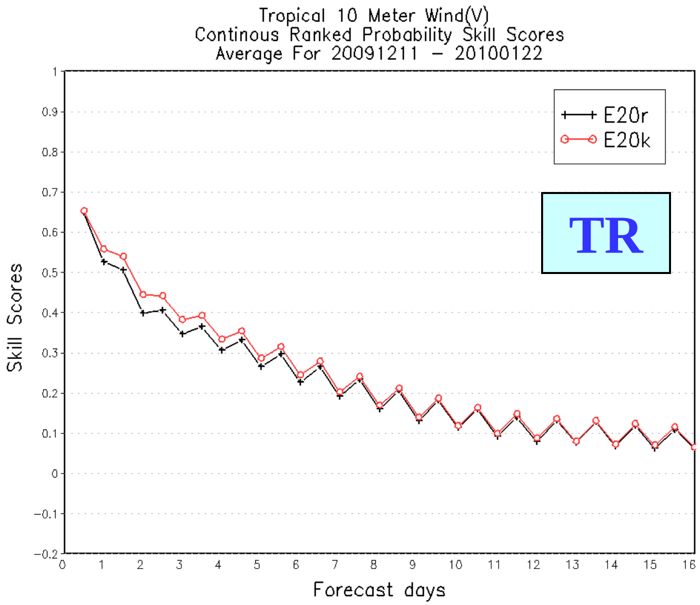
EnKF RMSE is slightly smaller for short lead time, slightly larger in median range,
EnKF initial spread is much larger than ETR, grows much slower than ETR.

AC: similar, ETR better for lead time days 5-8.
CRPSS: EnKF better for short lead time.

Probabilistic Score: CRPSS **10 meter wind (v)**

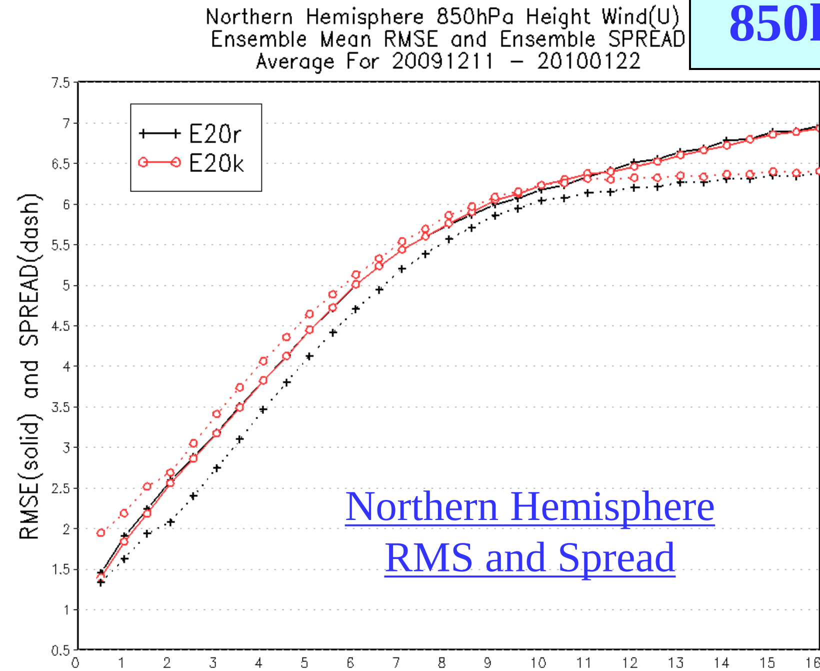


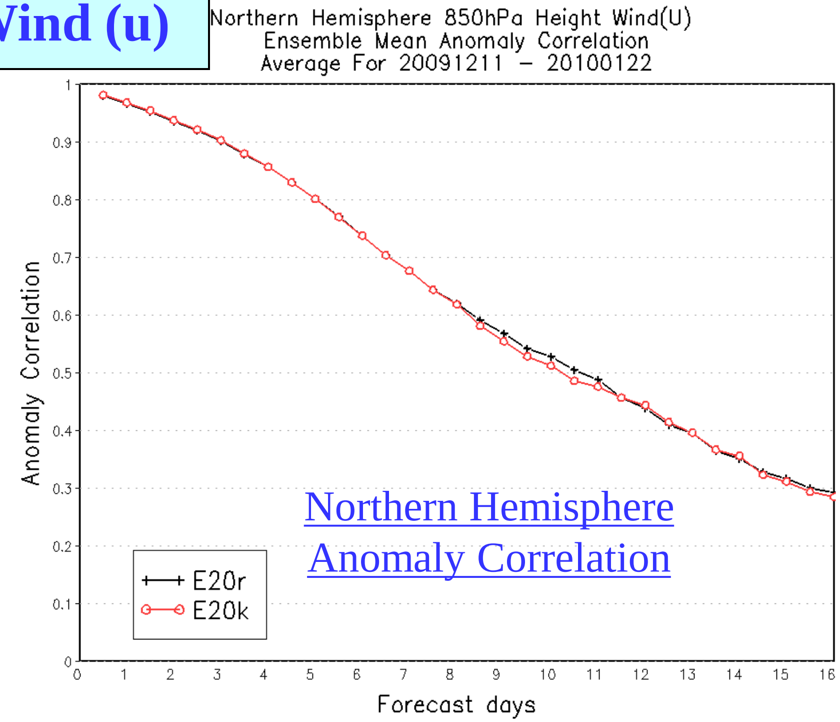


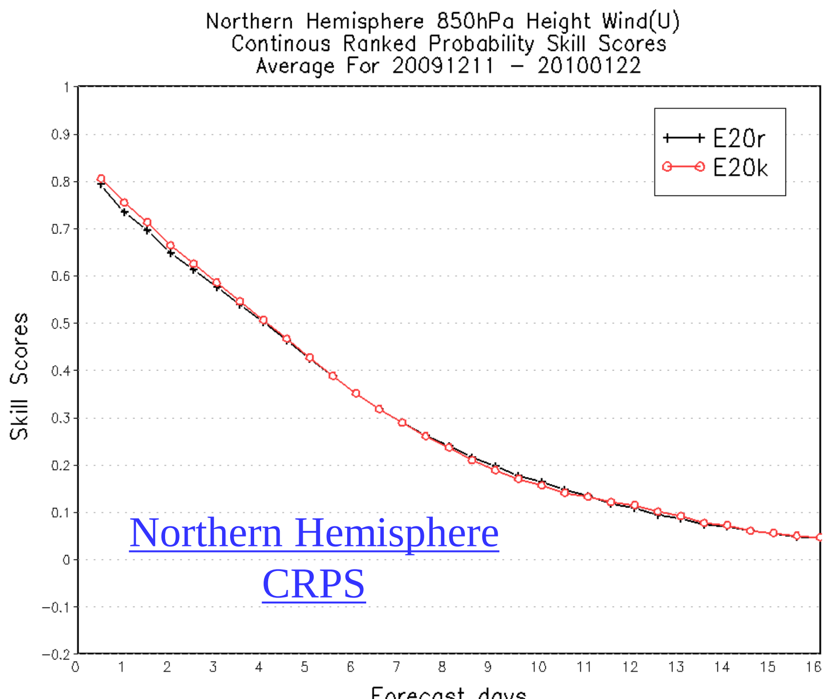


V10m : EnKF is slightly better for short lead time, particularly in the tropics, similar for median lead time.
This is also true for u10m, t2m, u850, v850, u200,v200

850hPa Wind (u)



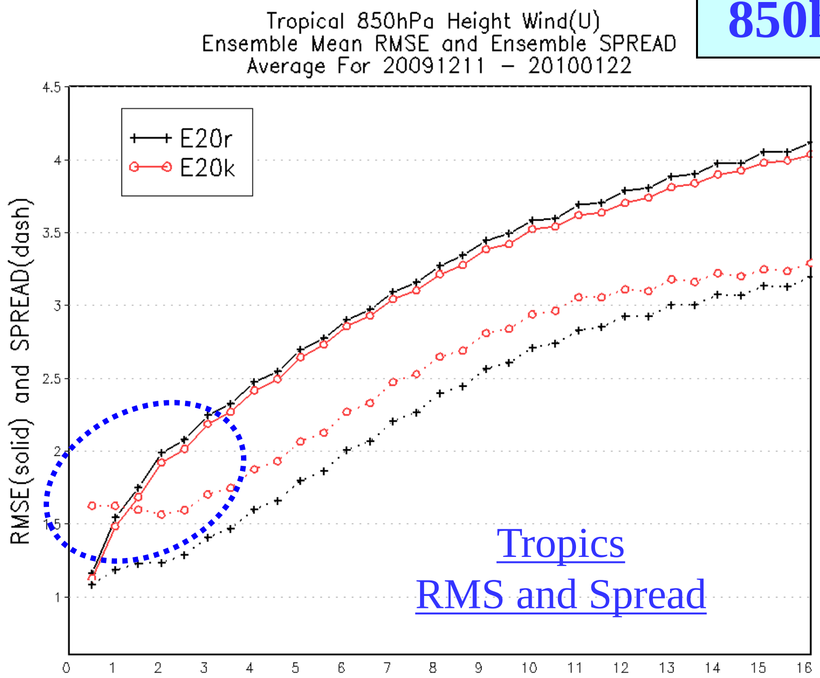


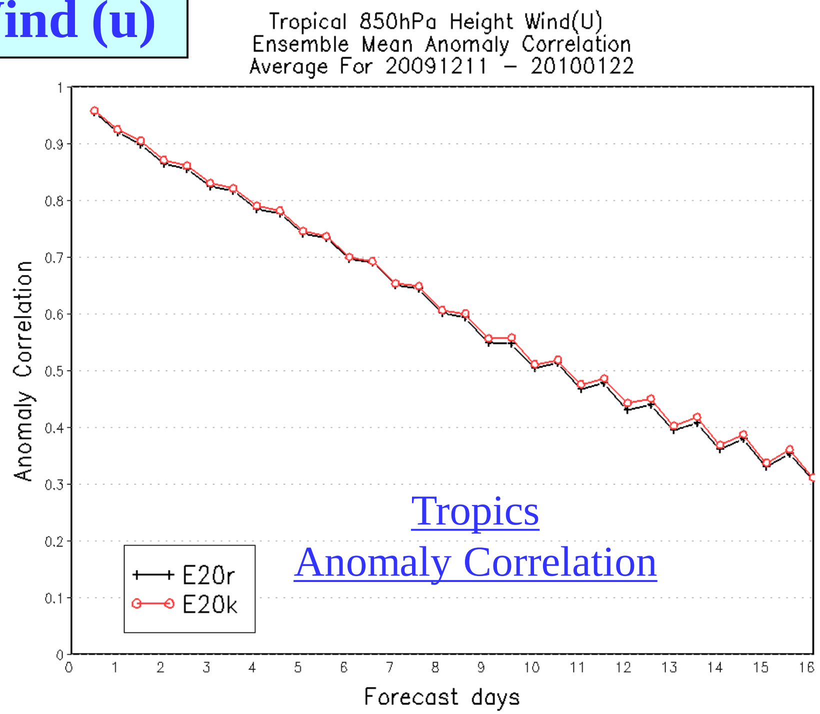


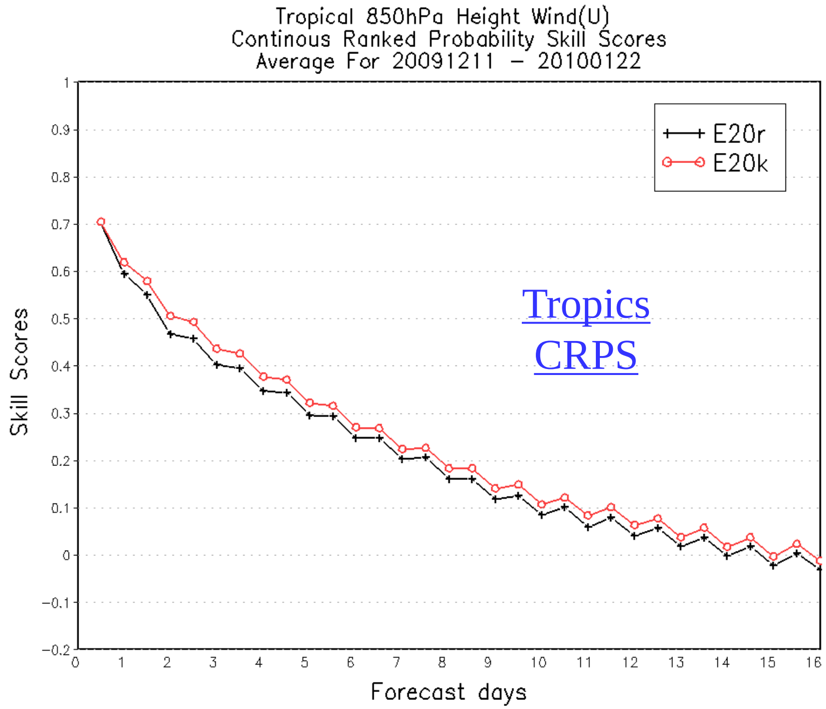
EnKF RMSE is slightly smaller for short lead time, slightly larger in median range,
EnKF initial spread is much larger than ETR, grows much slower than ETR.

AC: similar, ETR better for lead time days 8-11.
CRPSS: EnKF better for short lead time.

850hPa Wind (u)



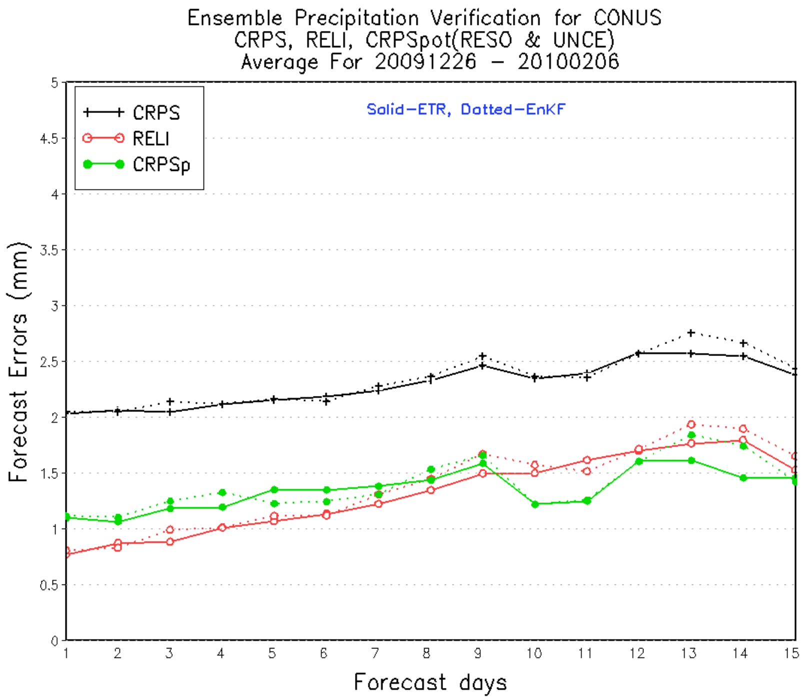


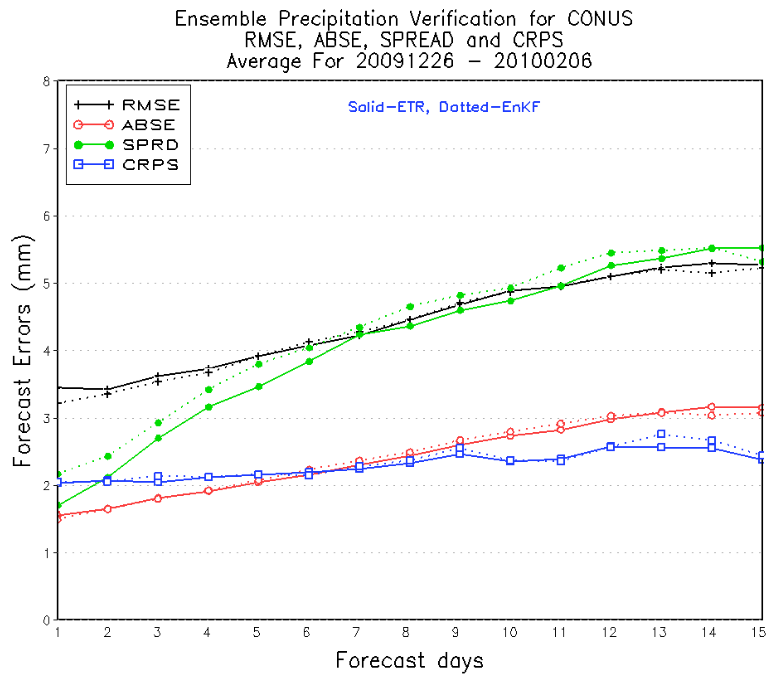


EnKF RMSE is smaller for all lead time,
EnKF initial spread is much larger than ETR,
grows much slower than ETR.

AC: similar, EnKF slightly better for larger lead time.
CRPSS: EnKF better.

Ensemble Precipitation for CONUS



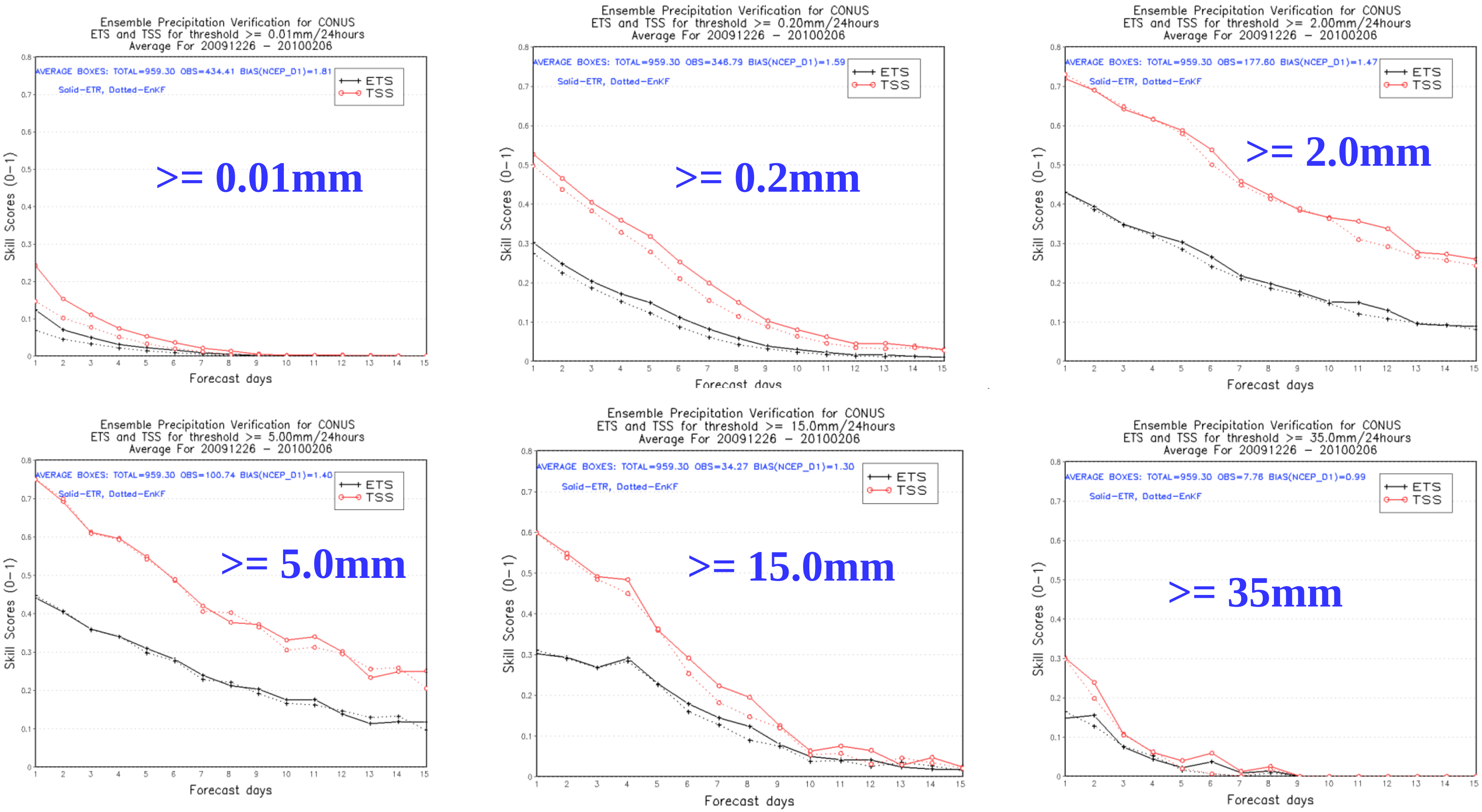


CRPS (Black): smaller-better
Reli (Red): smaller-better

ETR is slightly better

RMSE (black): EnKF is slightly smaller for short lead time,
Spread (green): EnKF initial spread is larger. ETR grows faster.
CRPS (blue): similar

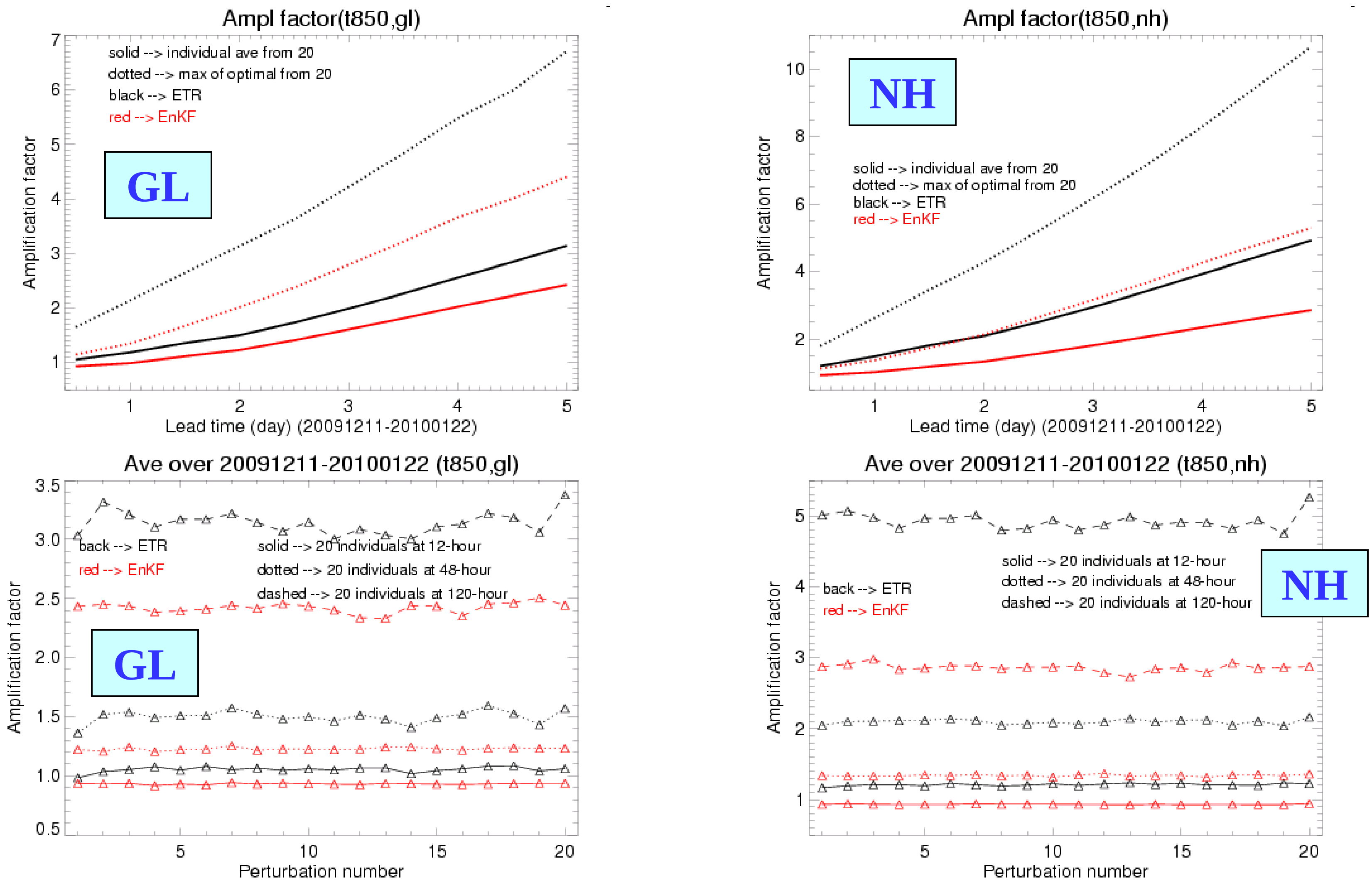
Ensemble Mean Precipitation for CONUS



ETS (Equitable Threat Score) and True Skill Score (TSS): ETR is clearly better for smaller threshold, as the threshold increases, the ETR advantage diminishes.

Amplification of Ensemble Perturbations

850hPa temperature

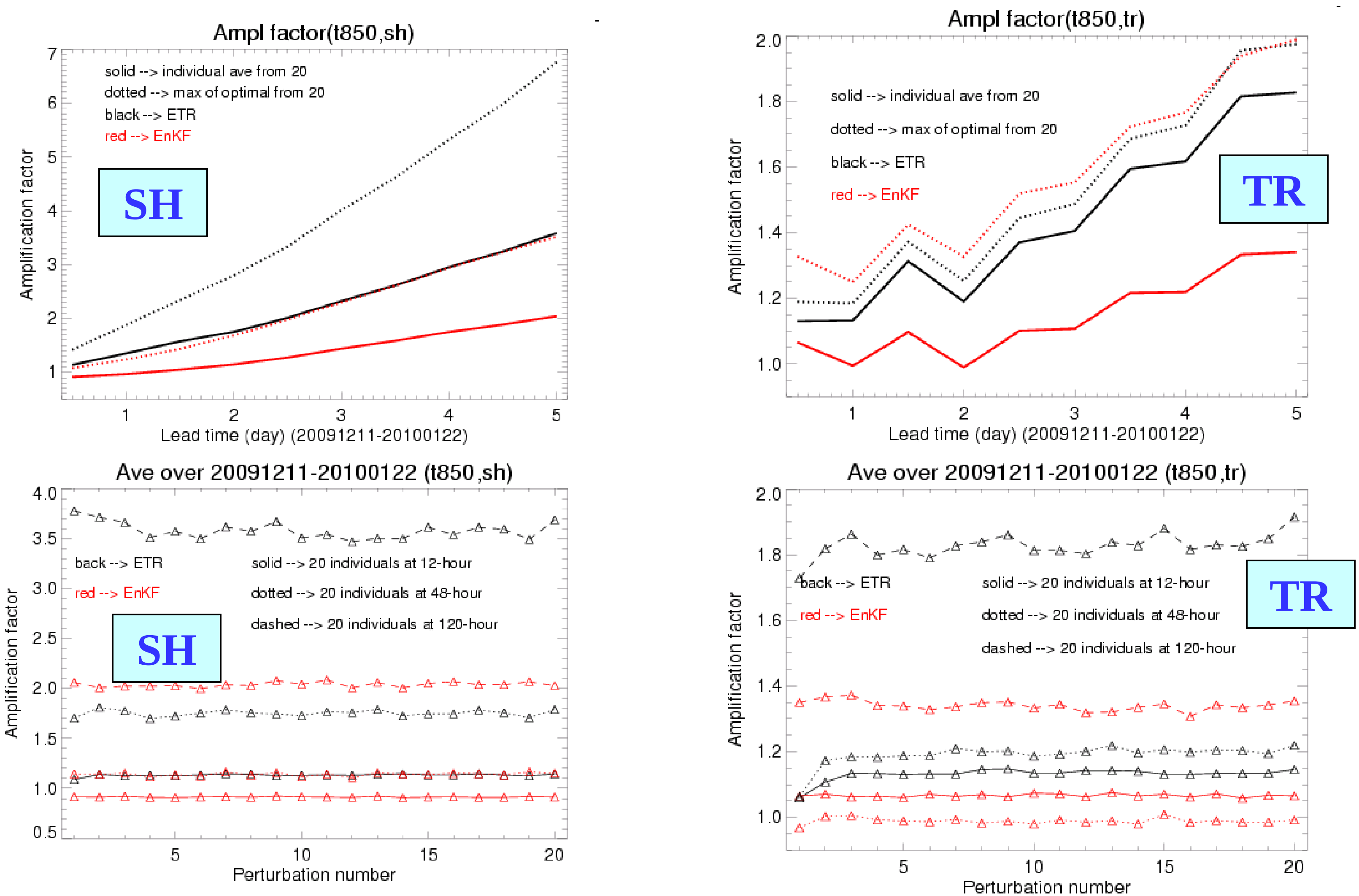


Top panel: solid line indicates the mean amplification factor of 20 perturbations from each ensemble. dotted line: all 20 ensemble perturbations are optimally combined such that the combined perturbation has the mathematically largest growth rate.

Amplification factor (AF) for t850: ETR perturbations much grow faster than EnKFs in GL and NH

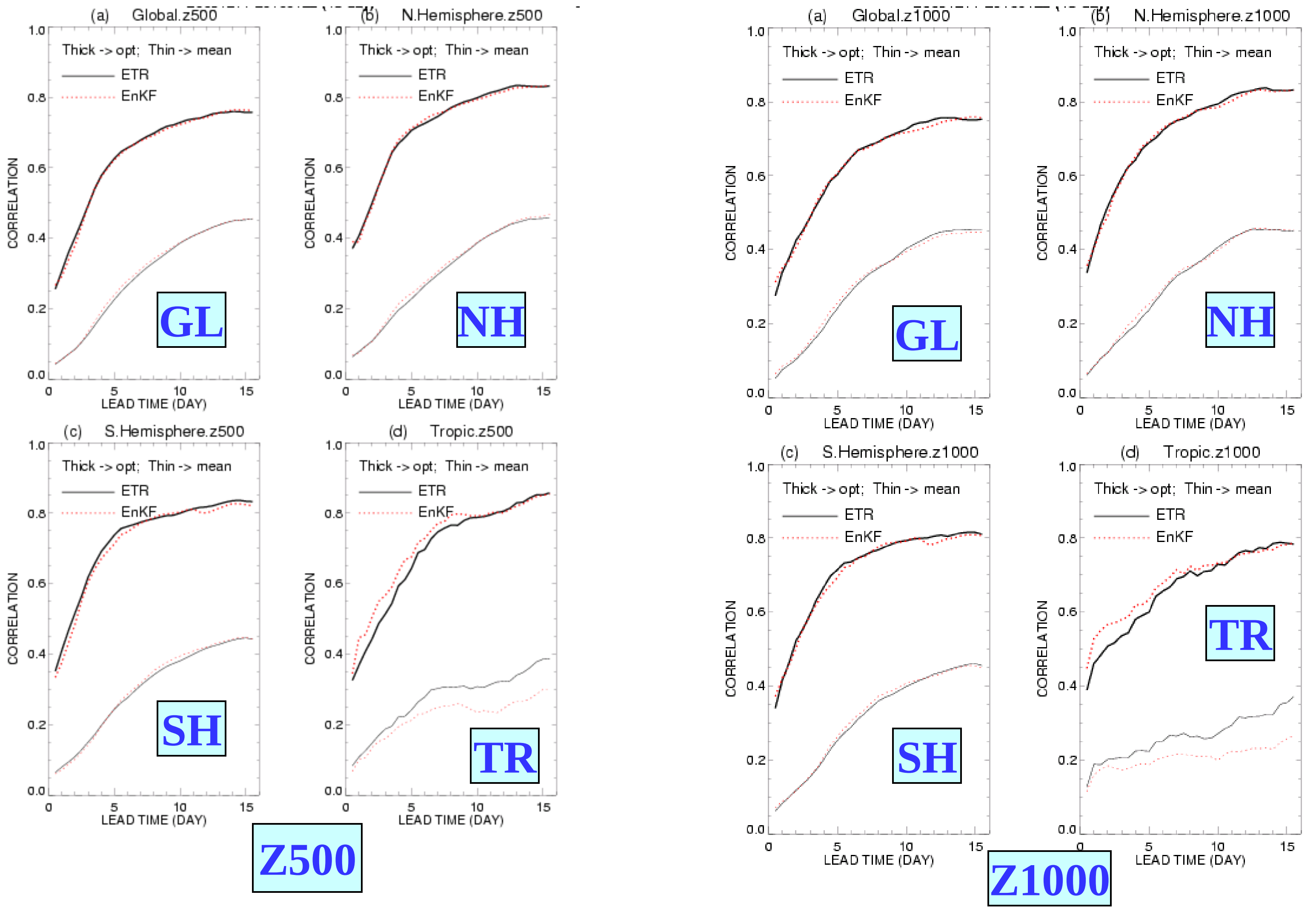
Amplification of Ensemble Perturbations

850hPa temperature



AF for t850: ETR perts grow faster than EnKF in SH. Over the tropics, individual perts of ETR grow faster, EnKF perts decay in the first 48 hours. But the optimally combined pert from EnKF has higher AF. AF for z500 (not shown): ETR perts grow faster than EnKF in all four domains.

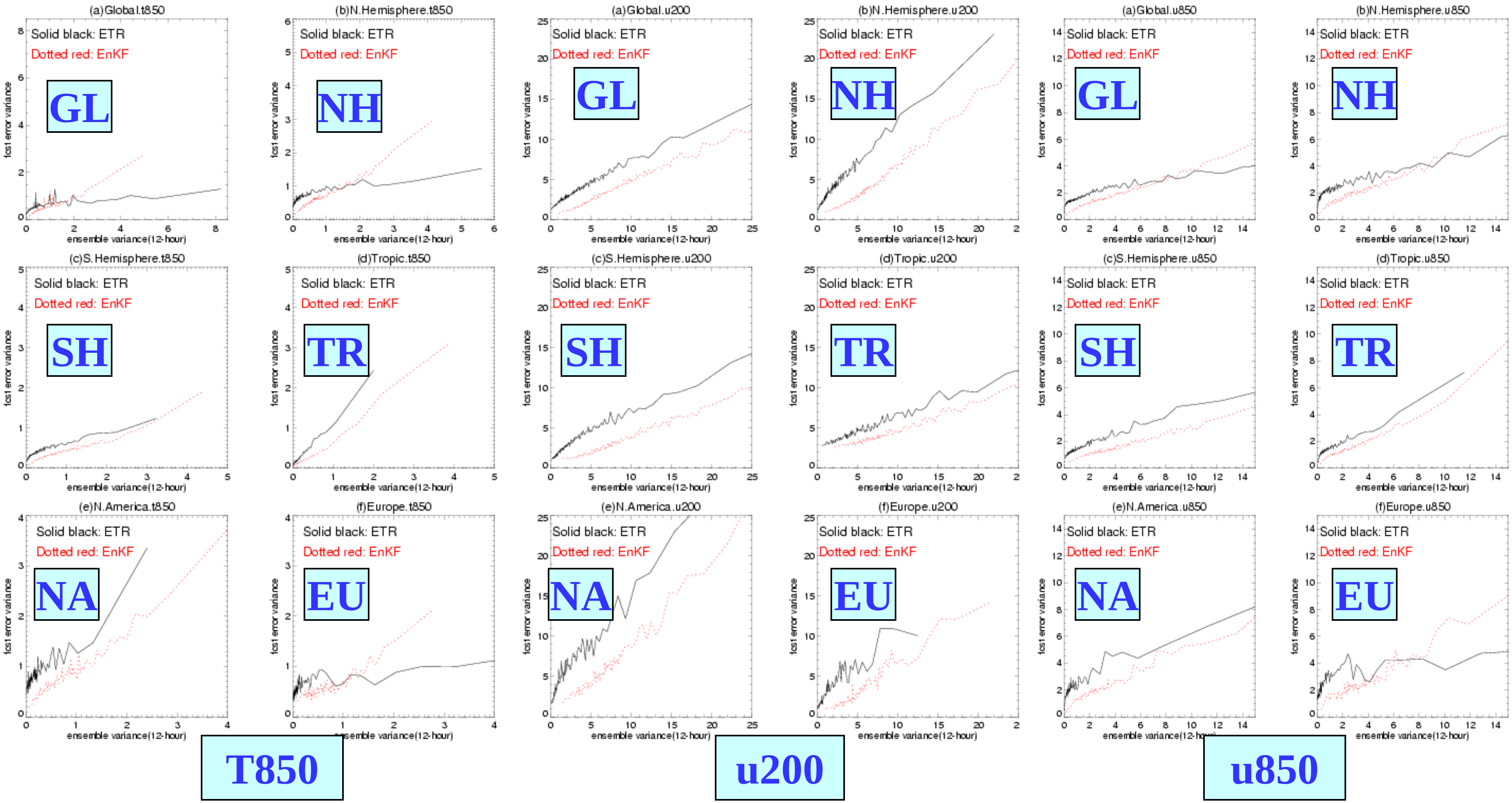
PECA (Pert versus error correlation analysis)



Thin line: mean correlation between each pert and the forecast error; thick line: all 20 ensemble perts are optimally combined in such a way so that the combined pert has the mathematically largest correlation with the forecast error.

ETR and EnKF have similar PECA values in GL,NH and NA, ETR has slightly higher value over SH, but lower in the tropics, slightly lower in India.

Ensemble variance versus forecast error variance



Ensemble variance versus forecast error variance: shows how well the forecast error variance can be explained by ensemble variance.

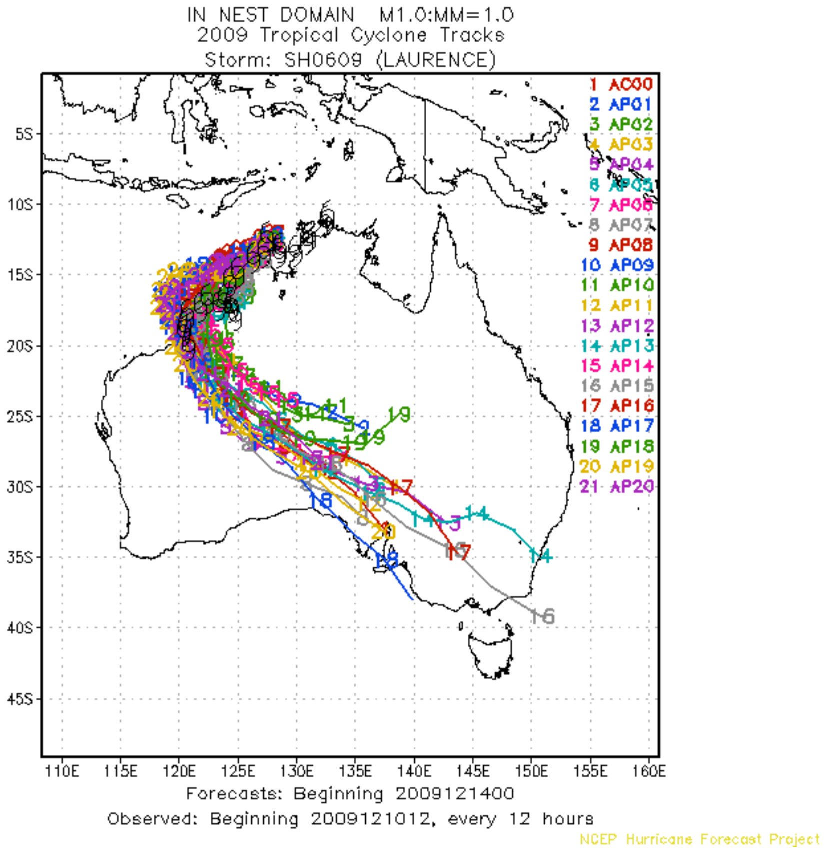
T850: ETR is slightly better for smaller variance, but worse for large variance over GL, NH and Europe. ETR is better over SH, TR and NA for all ranges.

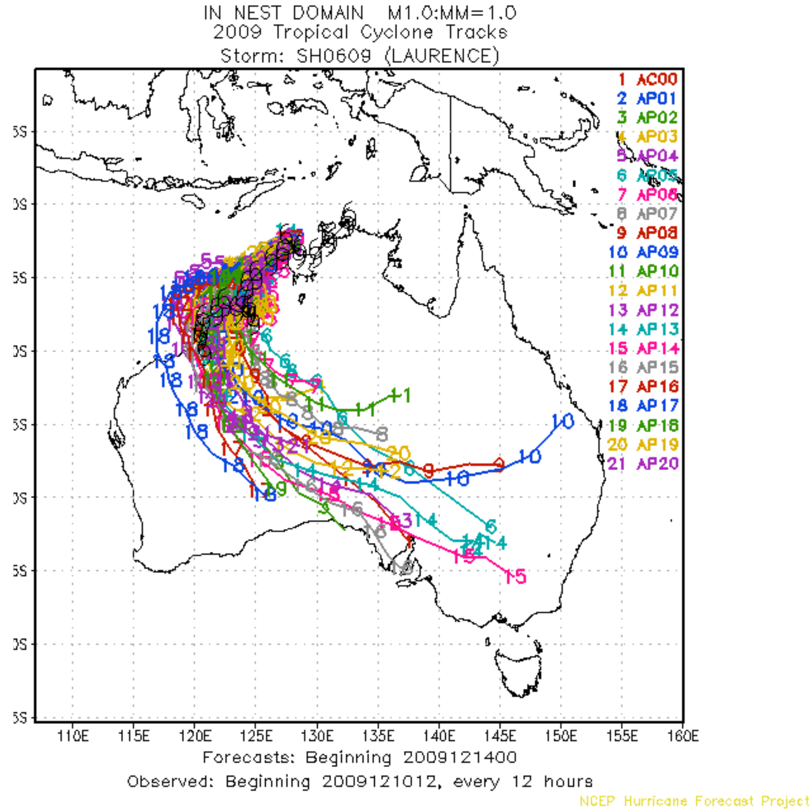
ETR is better over almost all regions for u200, v200, u850, v850, z500 and z1000 (not shown).

Number of bins=150 (GL), 130 (NH), 130(SH), 120(TR), 80(NA), 80 (EU).

Tropic Cyclone Tracks

Tropic Cyclone: Laurence





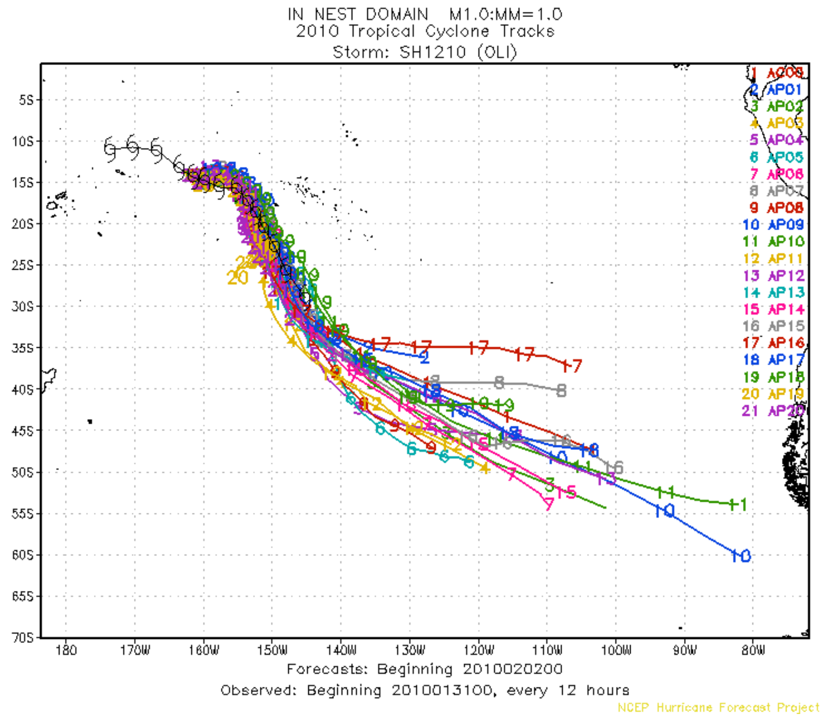
ETR

EnKF

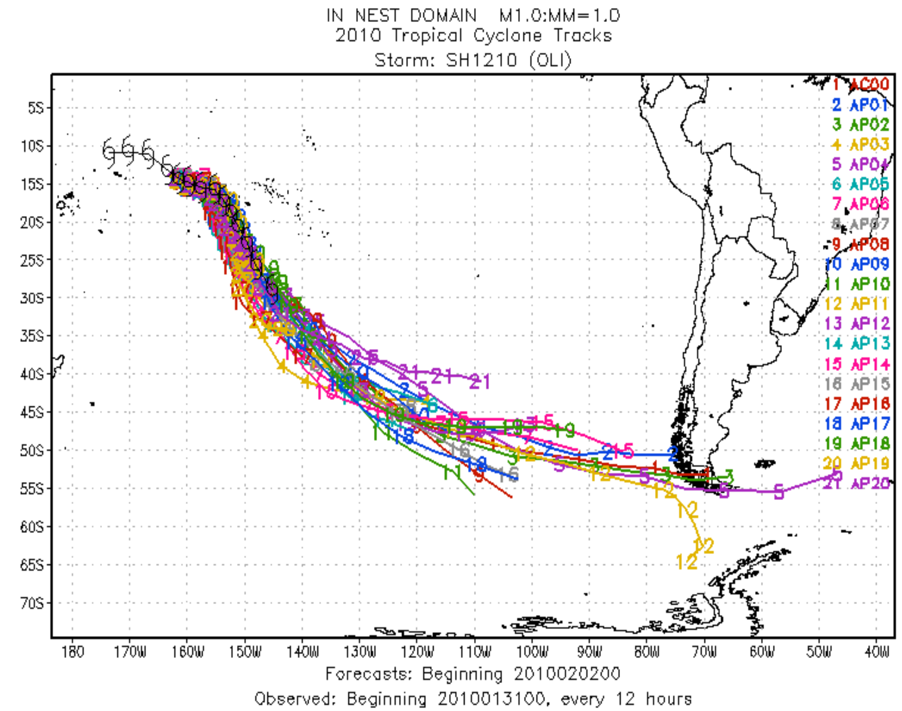
Tropic Cyclone Laurence started 2009121012, similar overall, ETR is slightly better.

Tropic Cyclone Tracks

Tropic Cyclone: Oli



ETR

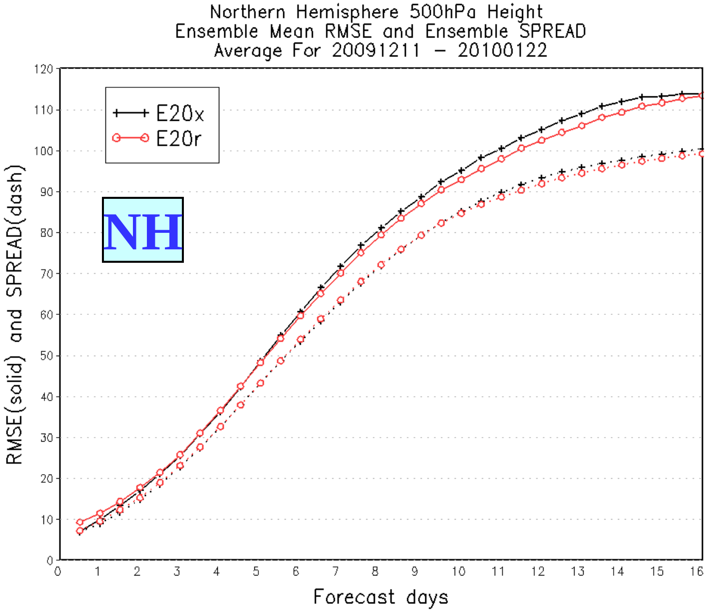


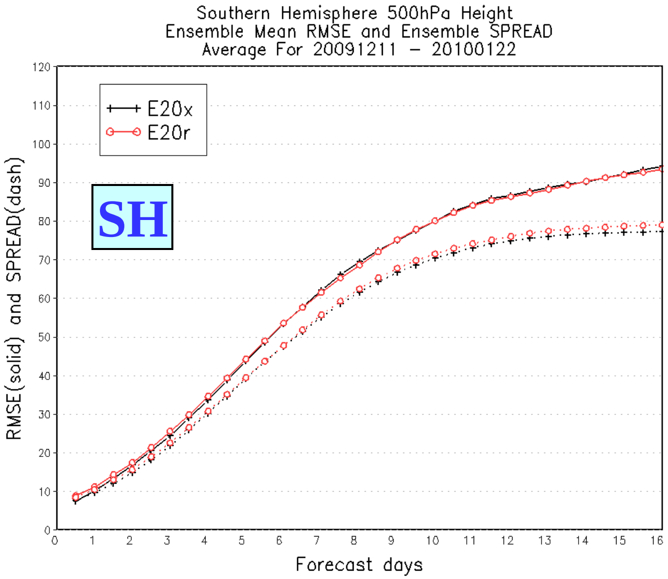
EnKF

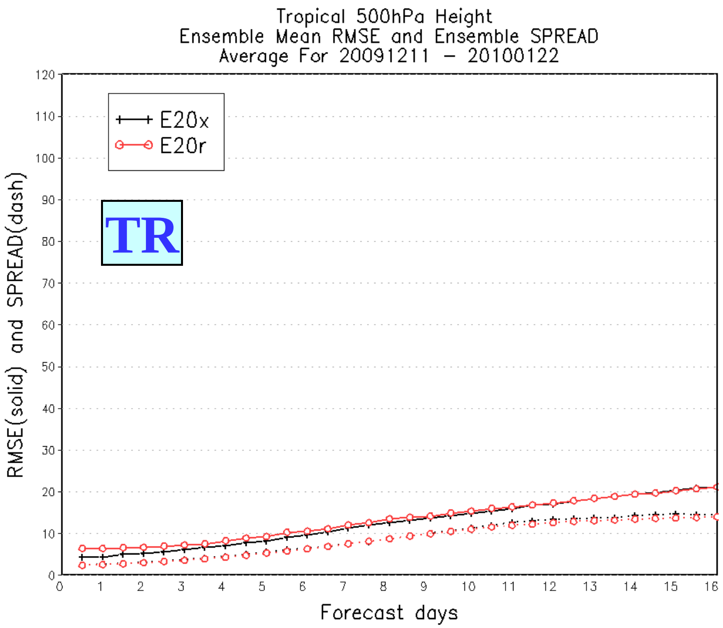
Tropic Cyclone Oli started 2010013100, similar overall, EnKF is slightly better.

GSI and EnKF, RMSE and Spread

500hPa height







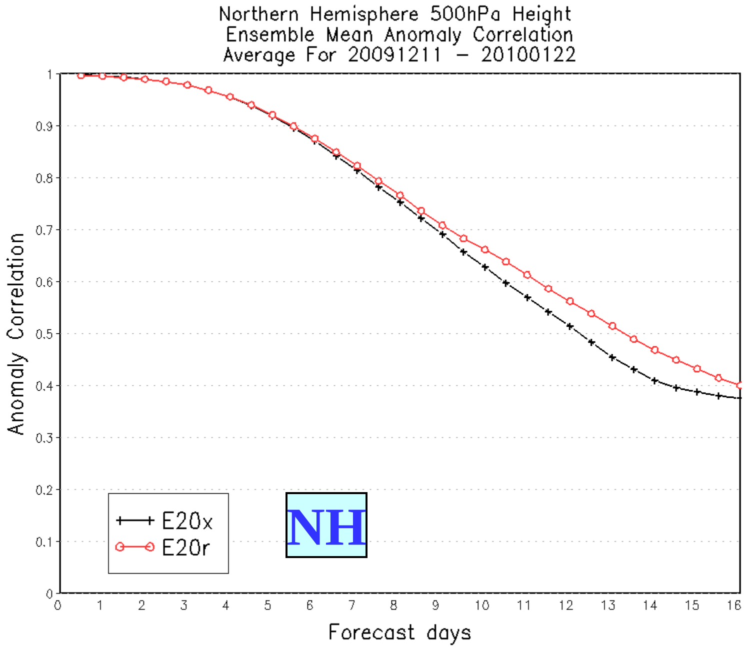
Z500: GSI and EnKF are similar overall, RMSE of GSI is slightly smaller for short lead time, slightly larger for longer lead time over NH and SH.
Over TR, GSI is smaller.

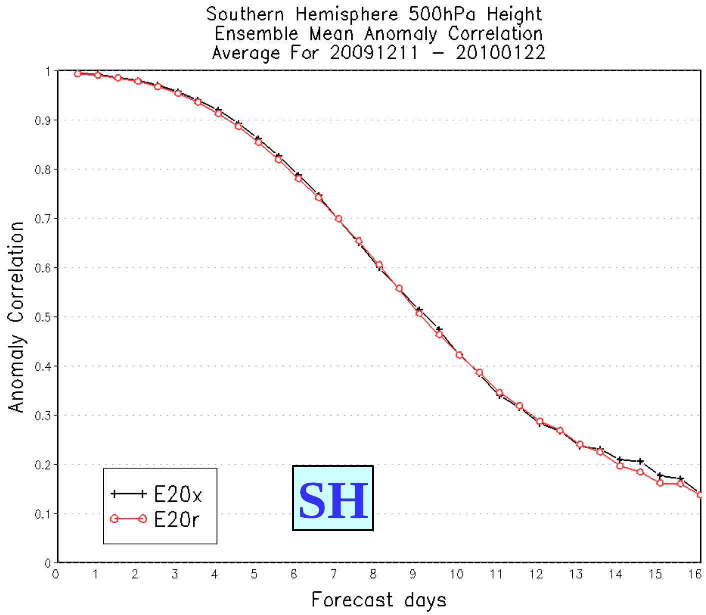
Spread is same, both use ETR perts.

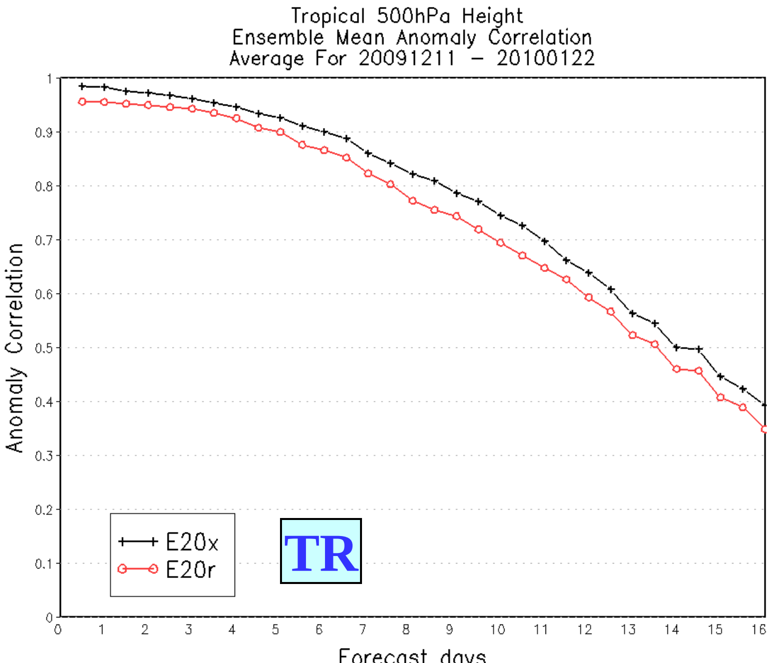
Z1000: similar results as Z500

GSI and EnKF, Ensemble Mean Anomaly Correlation

500hPa height







Z500: EnKF is better only for large lead time in NH. GSI is slightly better in SH and clearly better in TR.

A summary of comparison results

(1). RMSE: Z500, similar for ETR and EnKF, EnKF initial spread is notably larger than ETR. This results in better probabilistic scores of EnKF for short lead time. ETR spread grows faster than EnKF. Similar results are found for other variables.

(2). Anomaly Correlation: for Z500 AC values are similar in NH and SH, EnKF is better than in the tropics. Results for Z1000 are similar.
v10m: AC values for ETR are slightly higher in NH and SH, EnKF is slightly better than in the tropics. Results are similar for u10m, t850, u200, v200.

(3). CRPS: Z500, ETR and EnKF are similar over NH and SH, EnKF is better for short lead time and ETR is better for median lead time over the tropics. This conclusion holds for Z1000.
For V10m: EnKF is slightly better for short lead time, particularly in the tropics, similar for median lead time. This is also true for u10m, t2m, u850, v850, u200, v200

(4). ROC (not show): Z500, EnKF is better for short lead time. ETR is better for median range lead time.
It is true for Z1000.
For V10m: EnKF is better for short lead time, ETR is better for median range lead time
This is also true for u10m, u850, v850, u200 v200.
For t2m: EnKF is better for short lead time, similar for median range lead time.

A summary of comparison results

(5). PECA: ETR and EnKF have similar PECA values in GL and NH, NA, ETR has slightly higher value over SH, but lower in the tropics and slightly lower in India.

(6). Amplification factor: ETR perturbations grow much faster for almost all regions and all variables.

(7). Ensemble variance versus forecast error variance: ETR is better for most cases and most variables.

(8). Tropic Cyclone Tracks: either ETR or EnKF is better in a few cases.

(9). Precipitation for CONUS: ETR is slightly better in Reliability
RMSE: EnKF is slightly smaller for short lead time
Spread: EnKF initial spread is larger. ETR grows faster.
CRPS: similar
ETS (Equitable Threat Score) and True Skill Score (TSS): ETR is clearly better for smaller threshold, as the threshold increases, the ETR advantage is diminishes.

(10). Computing time: not tested, EnKF is expected to cost more.

A summary of comparison results

(11). GSI and EnKF: RMSE, Z500, GSI and EnKF are similar overall, GSI is slightly smaller for short lead time, slightly larger for longer lead time over NH and SH.
Over TR, GSI is smaller. Spread is same, both use ETR perts.
Similar results are found for Z1000.

(12). GSI and EnKF: Anomaly Correlation, Z500: EnKF is better than GSI only for large lead time in NH. GSI is slightly better in SH and clearly better in TR.

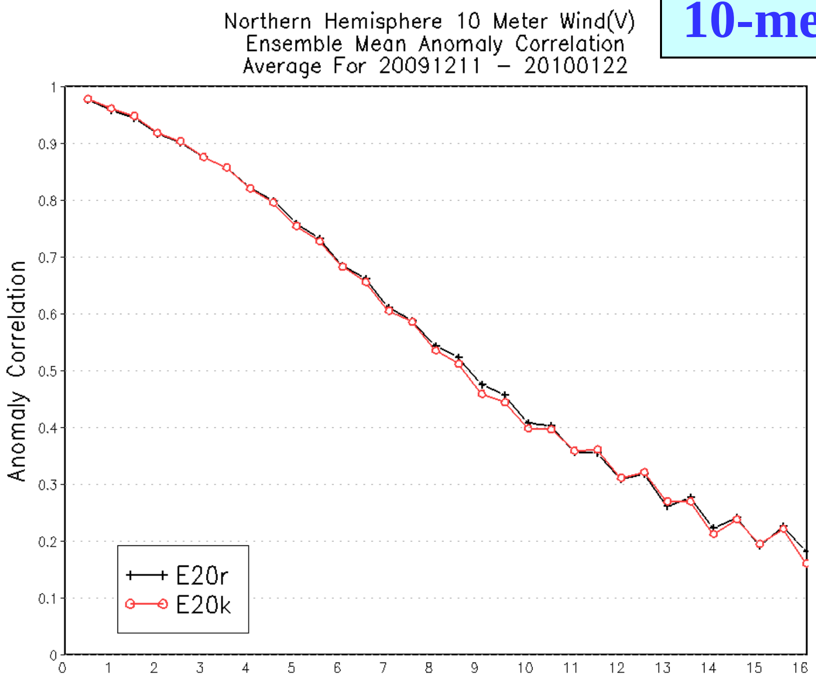
A summary of comparison results

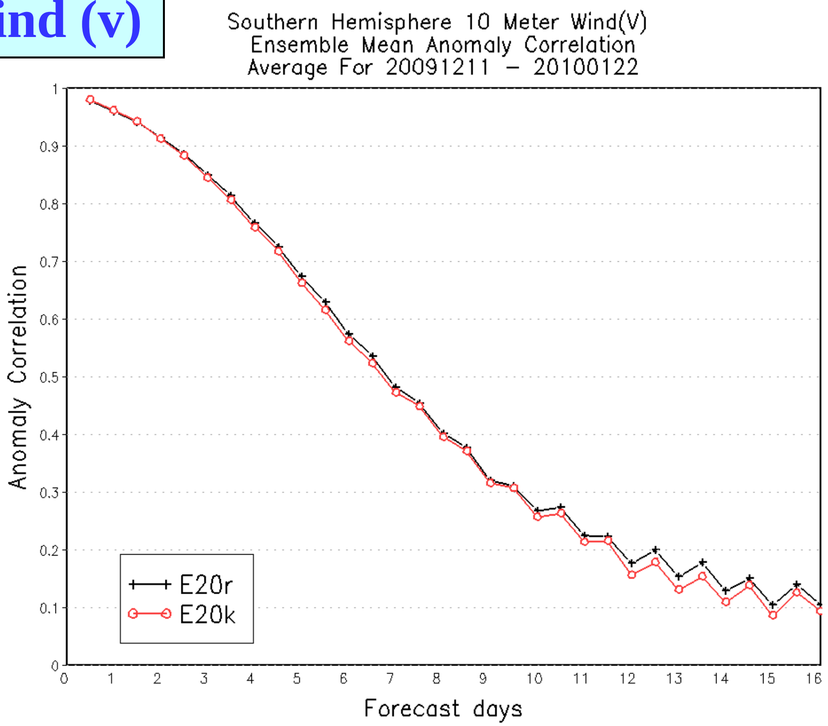
More results:

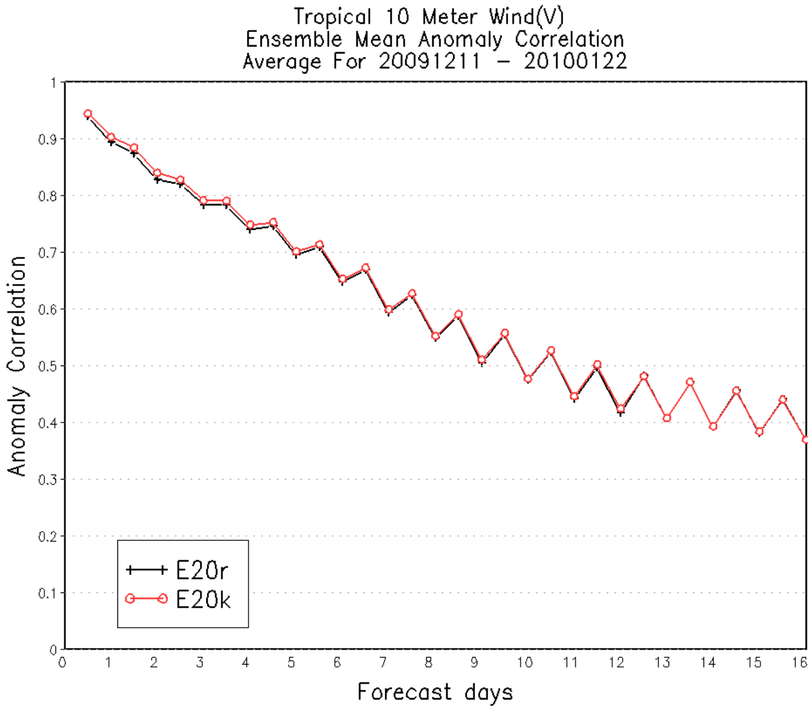
ETR perturbations (r) compared with EnKF perturbations (k)
http://www.emc.ncep.noaa.gov/gmb/wd20mw/html/ETR_EnKF_win0910.html
GSI analysis (x) compared with EnKF analysis (r):
http://www.emc.ncep.noaa.gov/gmb/wd20mw/html/GSI_EnKF_win0910.html
All three experiments on one figure (the same results from above two websites):
http://www.emc.ncep.noaa.gov/gmb/wd20mw/html/ETR_EnKF_GSI_win0910.html
ETR (r) compared with EnKF (k) for winds at 200hPa and 850hPa:
http://www.emc.ncep.noaa.gov/gmb/wd20mw/html/ETR_EnKF_200uv_win0910.html
Precipitation scores:
http://www.emc.ncep.noaa.gov/gmb/wd20mw/enkf/ETR_EnKF_prcp/
Amplification factors:
http://www.emc.ncep.noaa.gov/gmb/wd20mw/enkf/amplification/
PECA (Pert versus error correlation analysis) scores:
http://www.emc.ncep.noaa.gov/gmb/wd20mw/enkf/peca/
Forecast error variance explained by ensemble variance:
http://www.emc.ncep.noaa.gov/gmb/wd20mw/enkf/explained_variance/

Background Slides

10-meter wind (v)



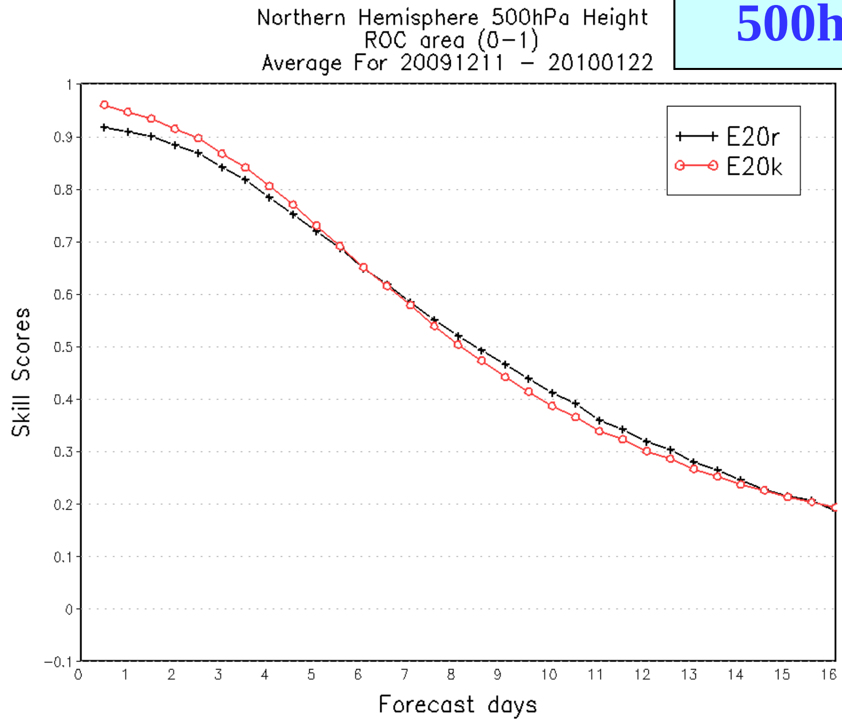


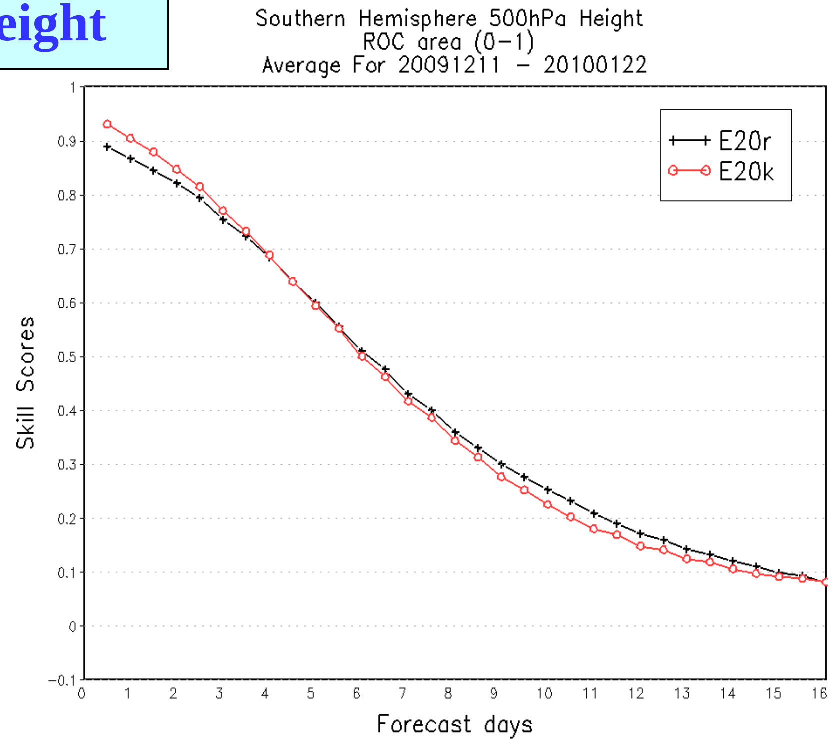


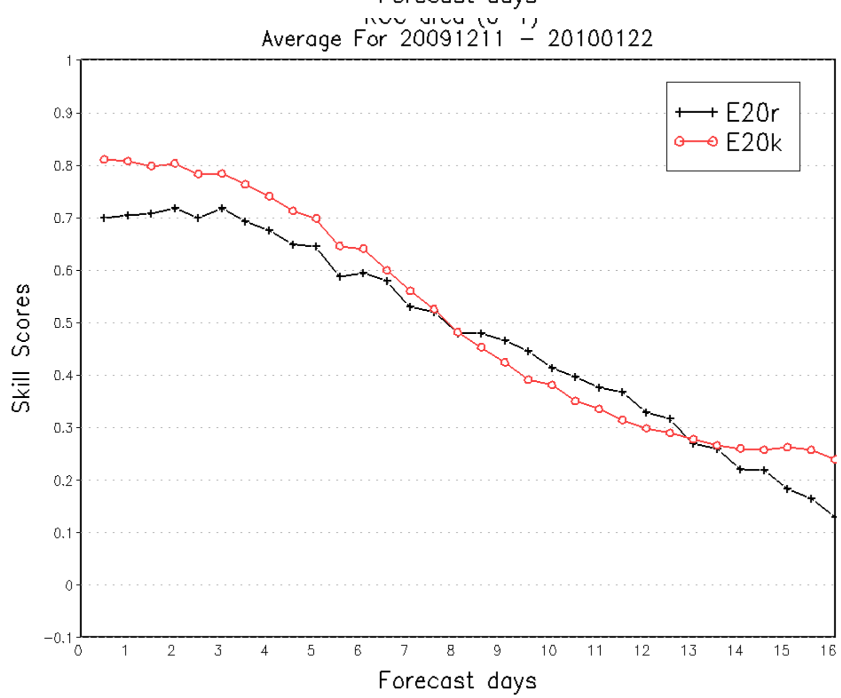
Ensemble Mean Anomaly Correlation

v10m: AC values for ETR are slightly higher in NH and SH,
EnKF is slightly better than in the tropics.
Results are similar for u10m, t850, u200, v200.

500hPa height



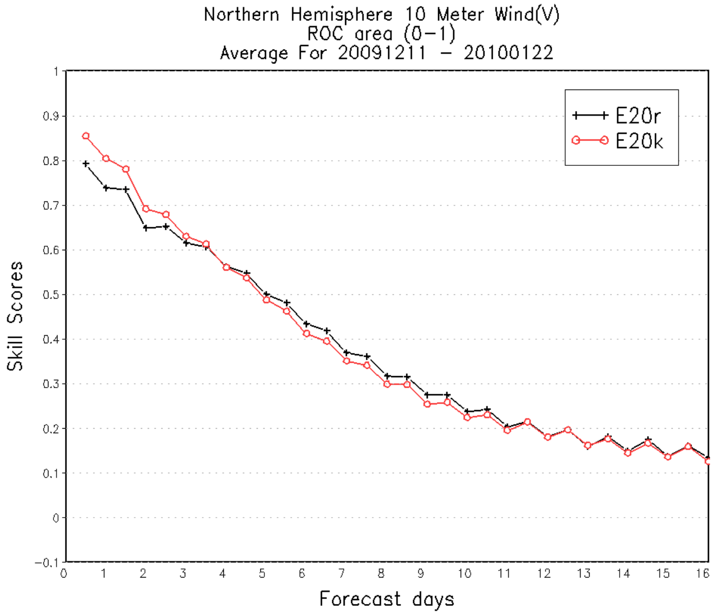


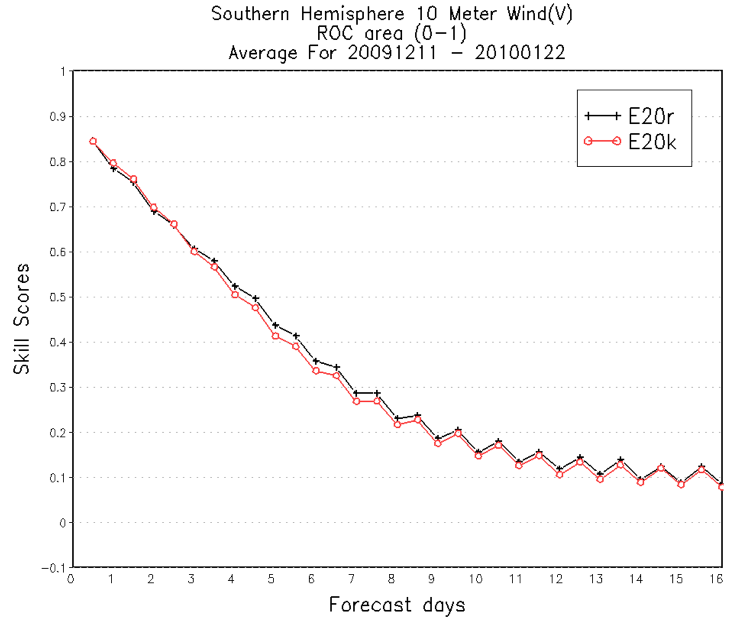


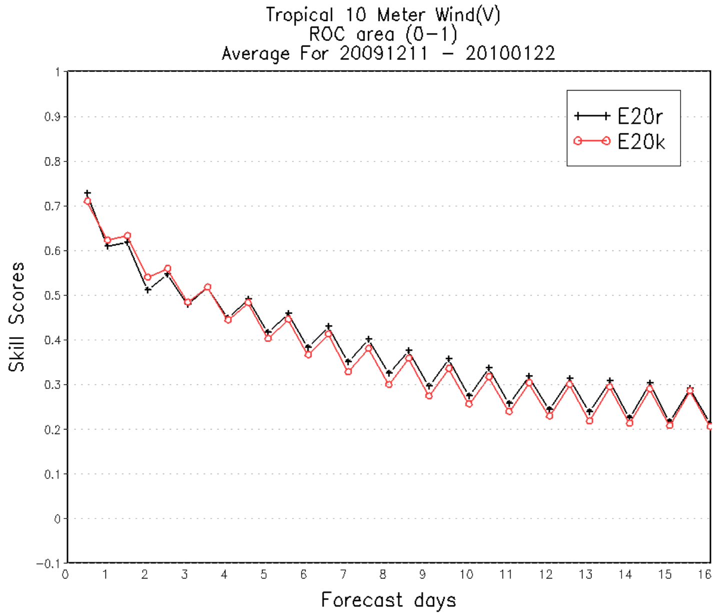
Probabilistic Score: ROC

Z500: EnKF is better for short lead time
ETR is better for median range lead time
It is true for Z1000

Probabilistic Score: ROC







V10m : EnKF is better for short lead time
ETR is better for median range lead time
This is also true for u10m, u850, v850, u200 v200

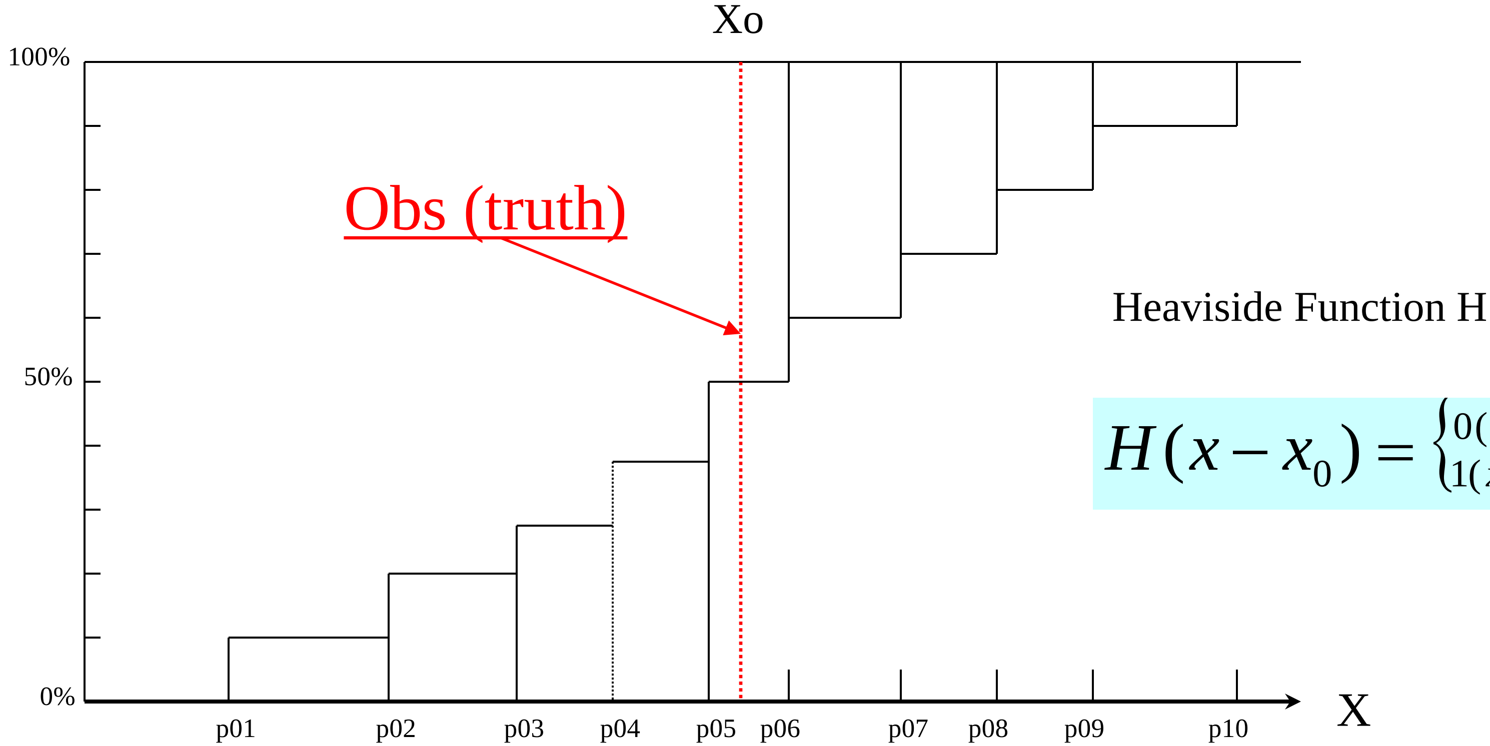
t2m: EnKF is better for short lead time, similar for median range lead time

Continuous Rank Probability Score

$$CRPS = \int_{-\infty}^{+\infty} [F(x) - H(x - x_0)]^2 dx$$

CRP Skill Score is

$$CRPSS = \frac{CRPS_c - CRPS_f}{CRPS_c}$$



$$H(x - x_0) = \begin{cases} 0 (x \le x_o) \\ 1 (x > x_o) \end{cases}$$

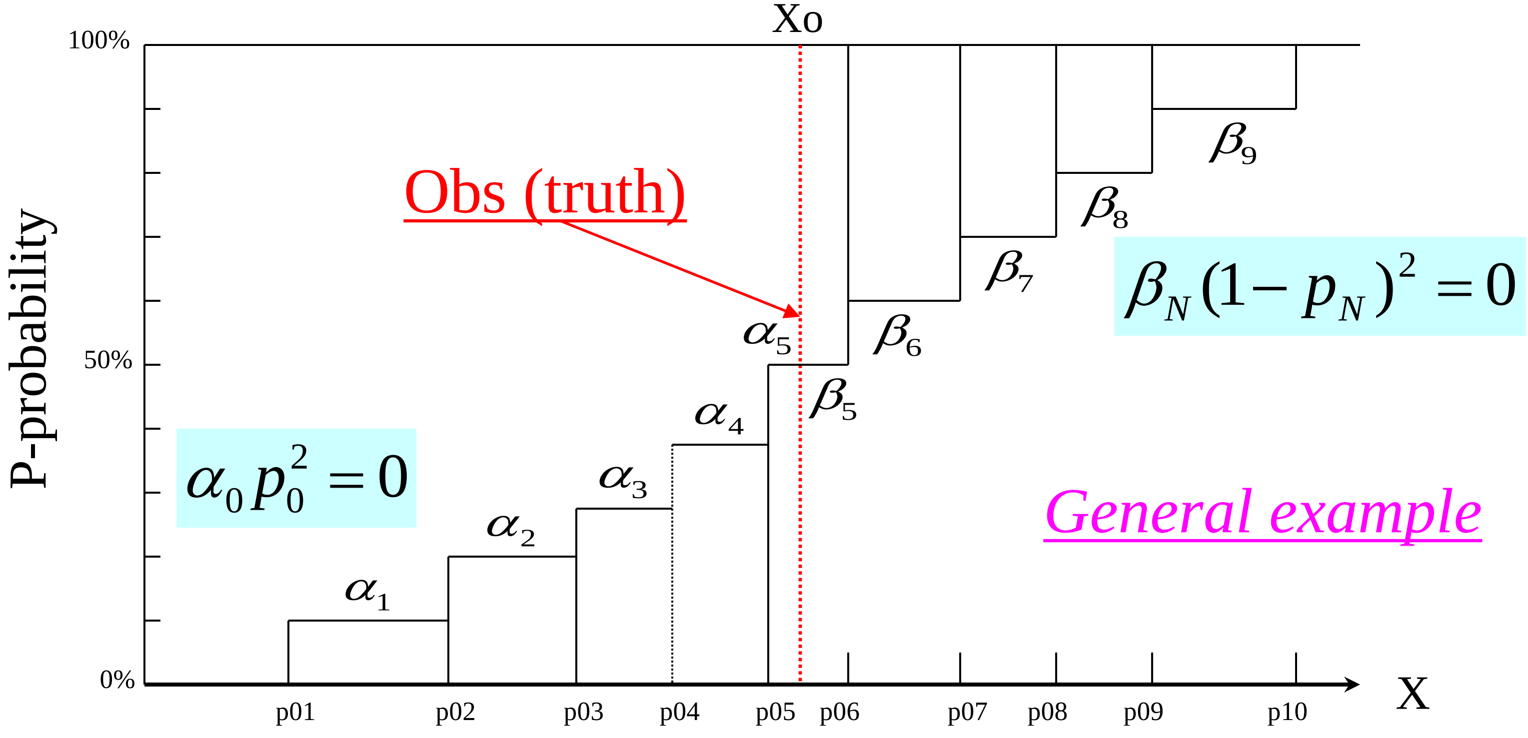
Order of 10 ensemble members (p01, p02,…,p10)

from Yuejian Zhu

CRPS Decomposition

$$CRPS = \int_{-\infty}^{+\infty} [F(x) - H(x - x_0)]^2 dx$$

$$\overline{CRPS} = \sum_{i=0}^{N} [\overline{\alpha}_i p_i^2 + \overline{\beta}_i (1 - p_i)^2]$$

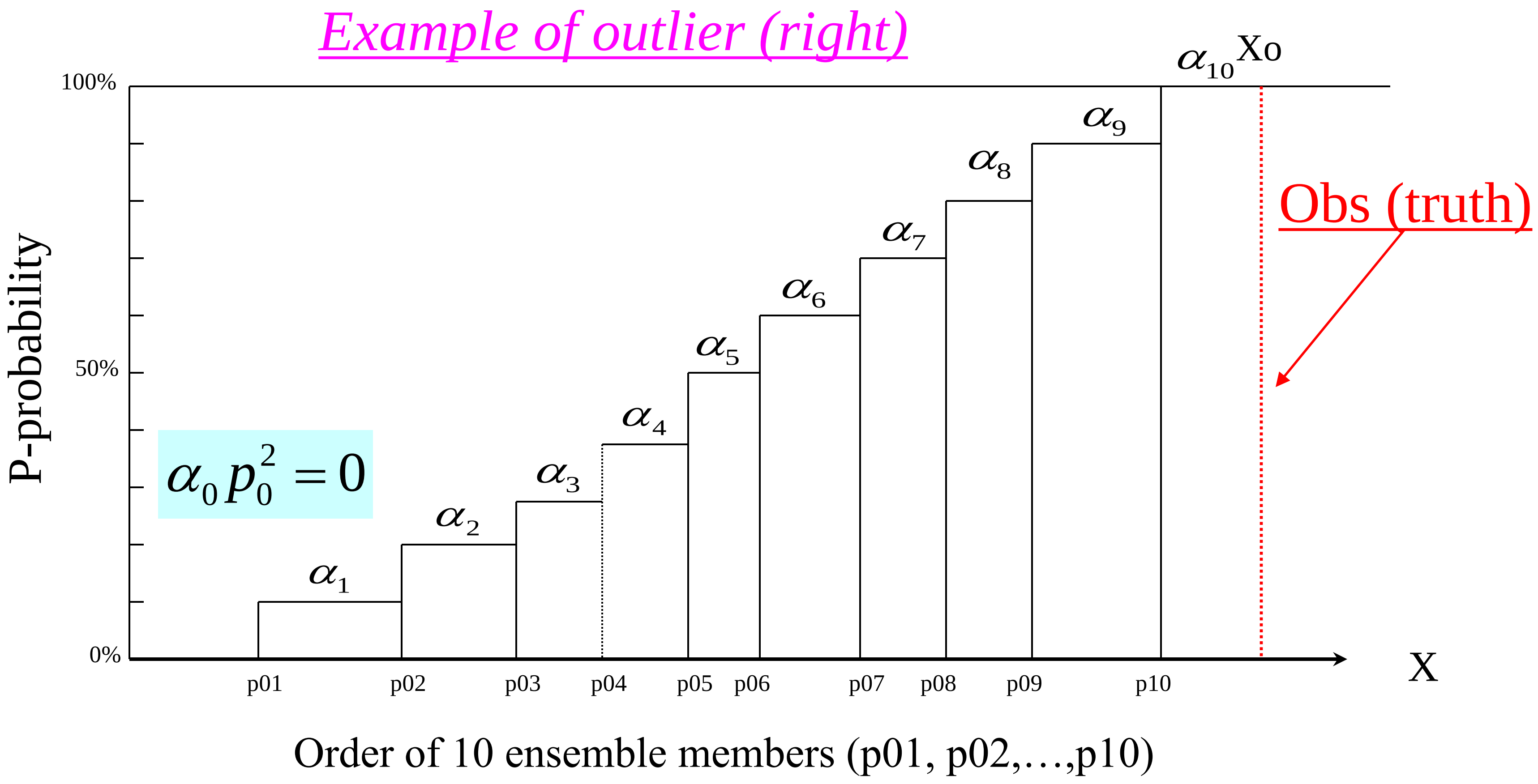


Order of 10 ensemble members (p01, p02,…,p10)

from Yuejian Zhu

CRPS Decomposition

$$\overline{CRPS} = \sum_{i=0}^{N} [\overline{\alpha}_i p_i^2 + \overline{\beta}_i (1 - p_i)^2]$$

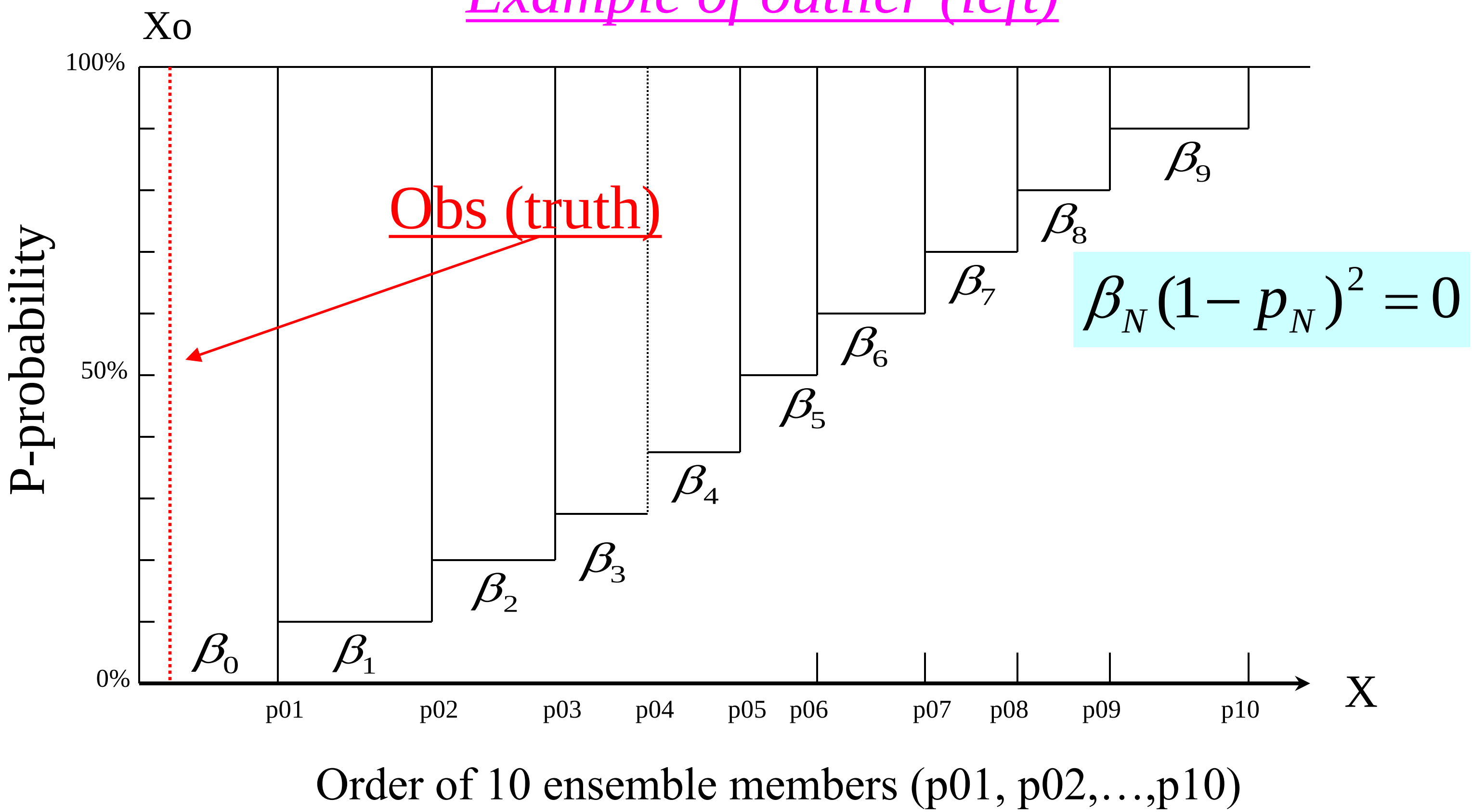


from Yuejian Zhu

CRPS Decomposition

$$\overline{CRPS} = \sum_{i=0}^{N} [\overline{\alpha}_i p_i^2 + \overline{\beta}_i (1 - p_i)^2]$$

Example of outlier (left)



from Yuejian Zhu

CRPS Decomposition

$$\overline{CRPS} = \sum_{i=0}^{N} [\overline{\alpha}_i p_i^2 + \overline{\beta}_i (1 - p_i)^2]$$

$$\overline{CRPS} = \sum_{i=0}^{N} \overline{g}_i [(1 - \overline{o}_i) p_i^2 + \overline{o}_i (1 - p_i)^2]$$

$$\overline{CRPS} = \overline{RELI} - \overline{RESO} + \overline{U}$$

$$\overline{RELI} = \sum_{i=0}^{N} \overline{g}_i (\overline{o}_i - p_i)^2$$

CRPSpot

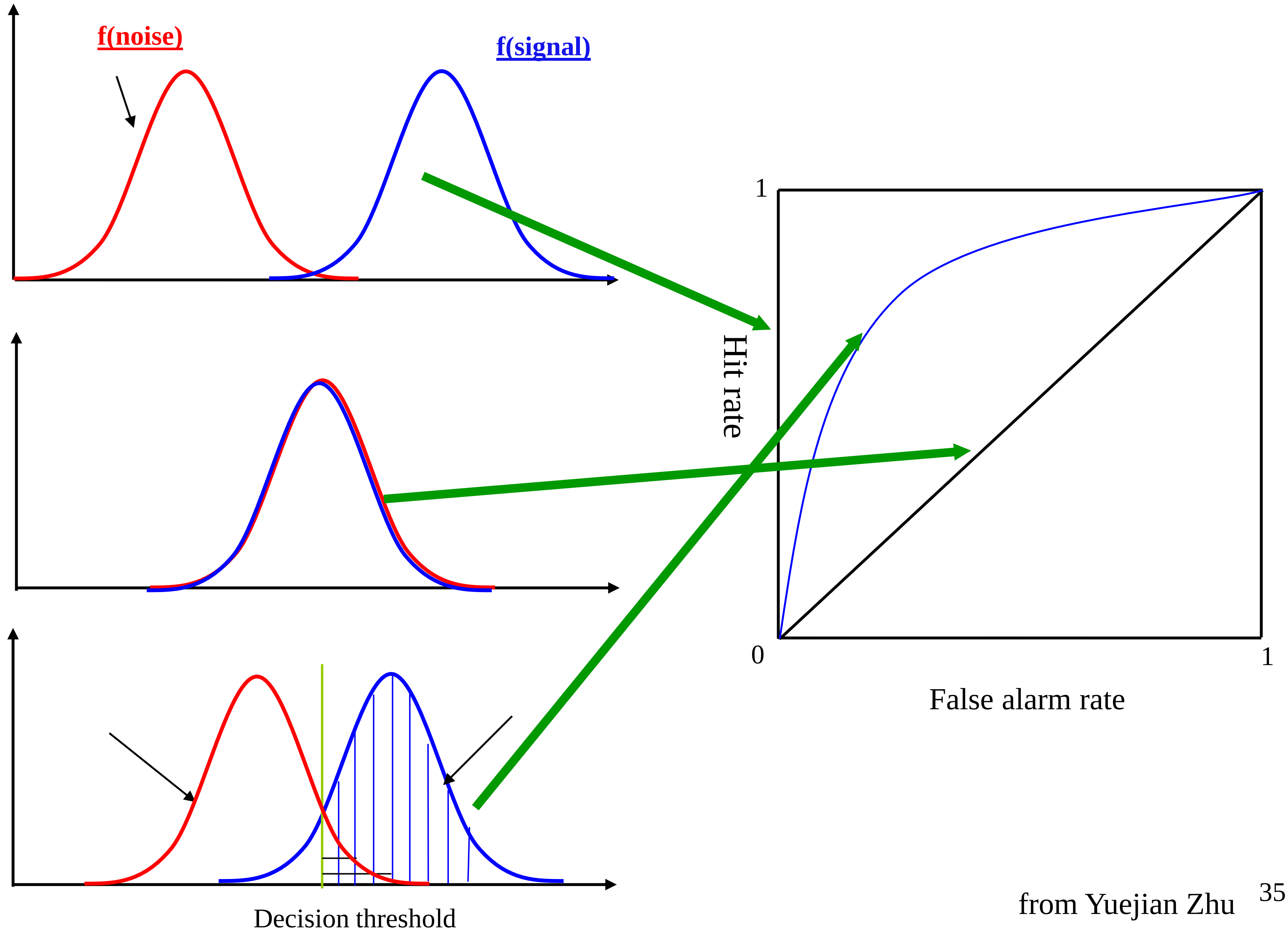
Where:

$$\overline{g}_i = \overline{\alpha}_i + \overline{\beta}_i$$

$$\overline{o}_i = \frac{\overline{\beta}_i}{\overline{\alpha}_i + \overline{\beta}_i}$$

from Yuejian Zhu

Relative Operating Characteristics area (ROC area)



from Yuejian Zhu

Prob. Evaluation **(cost-loss analysis)**

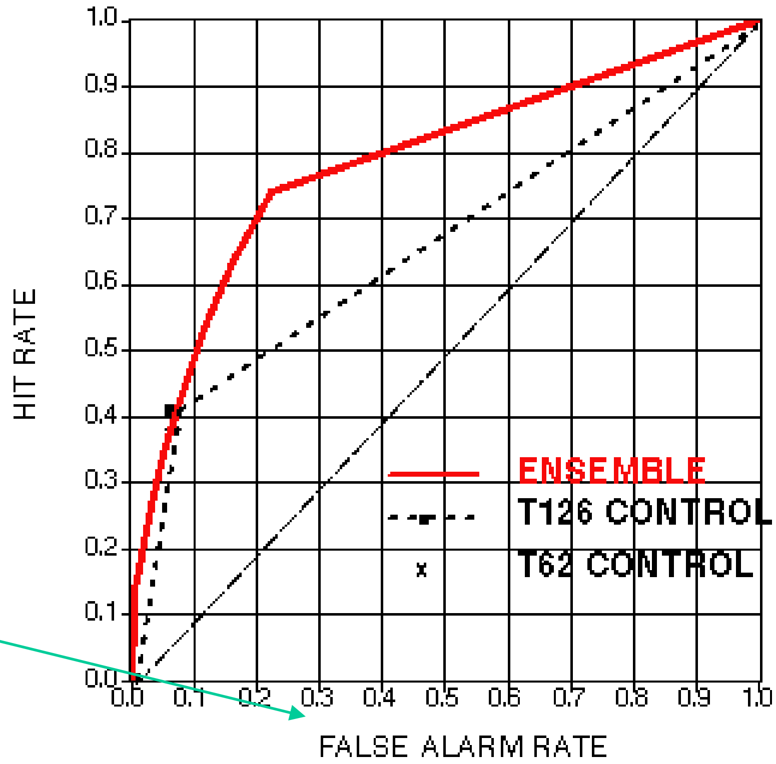
Based on hit rate (HR) and false alarm (FA) rate.

1. Relative Operating Characteristics (ROC) area - ***Appl. of signal detection theory for measuring discrimination between two alternative outcome.***

ROCarea = Intergrated area * 2 (0-1 normality)

h/(h+m)

f/(h+f)



Relative Operating Characteristics

o\f	y(f)	n(f)
y(o)	h	m
n(o)	f	c

ROC (Relative Operating Characteristics) curve for a 10-member T62 ensemble of forecasts and for T126 and T62 control forecasts for the 500 hPa height, ***NH extratropics****, March-May 1997. The closer a curve is to the upper left hand corner, the more ability the forecasting system has in delineating between cases when a certain event (in this case, the occurence of one of 10 climatologically equally likely bins) did or did not occur.*

from Yuejian Zhu

Category Forecast: Precipitation Evaluation

1. Frequency Bias (FBI)

$$FBI = \frac{(h+f)}{(h+m)}$$

2. ETS (equitable threat score)

$$ETS = \frac{h-R}{h+f+m-R}$$

where $R = \frac{(h+f)\cdot(f+m)}{(h+f+m+c)}$ *is randomly forecasting rate*

	OBS (Yes)	OBS (No)
FCST (Yes)	Hit (h)	False alarm (f)
FCST (No)	Miss (m)	Correct Reject (c)

3. TSS (true skill statistic or Hanssen-Kuipers discriminant)

$$TSS = \frac{h\cdot c - f\cdot m}{(h+m)\cdot(f+c)} = \frac{h}{h+m} - \frac{f}{f+c}$$

from Yuejian Zhu

PECA **(Perturbation vs. Error Correlation Analysis,** Wei and Toth 2003)

Goal: design an additional ensemble verification tool to measure performance of ensemble systems that

(a) are less sensitive to errors in analysis (and model).

(b) evaluates the degree of independency of ensemble members.

(c) measures how much forecast error can be explained by individual or optimally combined perturbations

(d) reflects more on the quality of ensemble method.

(e) higher PECA → more skillful ensemble.

Ensemble perturbations and forecast error:

$$\mathbf{p}_i(t) = \mathbf{f}_i(t) - \bar{\mathbf{f}}(t) \qquad \mathbf{e}(t) = \bar{\mathbf{f}}(t) - \mathbf{f}_{analysis}(t)$$

i=1, 2,....K, $\bar{\mathbf{f}}$ is ensemble mean

Obtain α_i by solving the least-square problem:

$$Min \| \mathbf{e} - \sum_{i=1}^{K} \alpha_i \mathbf{p}_i \|_{L2}$$

The optimally combined perturbation is defined as:

$$\mathbf{p}_{optimal} = \sum_{i=1}^{K} \alpha_i \mathbf{p}_i$$

PECA: correlation between forecast error and individual perturbations/optimally combined perturbation

Forecast error variance explained by ensemble variance

(Majumdar et al. 2001, 2002, Wei et al. 2006)

Goal: measures the range of forecast error variance explained by the ensemble variance for different variables and different domains.

Step 1: choose a variable for a particular domain.
Step 2: compute ensemble variance and squared forecast error at each grid point over this domain for a particular forecast lead time (6 hours or 12 hours).
step 3: draw a scatter plot using ensemble variance (abscissa) and squared forecast errors for all grid points.
step 4: divide the points into N equally populated bins in order of increasing ensemble variance.
step 5: ensemble variance and squared forecast errors are averaged within each bin.
step 6: draw a curve connecting the averaged value from each bin.

A better ensemble should explain larger range of forecast error variance. Thus the ensemble with steeper curve is considered better.

AF (Amplification factor, Wei et al. 2006)

Goal: **measures the growth rate of the individual perturbations and maximum growth rate of the optimally combined perturbation.**

AF of individual perturbations:

$$AF_i(t) = \frac{\| \mathbf{p}_i(t) \|_{L2}}{\| \mathbf{p}_i(0) \|_{L2}}$$

i=1, 2, ...K, t--> forecast lead time

Max AF of the optimally perturbations:

$$AF_{\max}(t) = \max\left[\frac{< \mathbf{Z}^f \mathbf{e}, \ \mathbf{Z}^f \mathbf{e} >}{< \mathbf{Z}^a \mathbf{e}, \ \mathbf{Z}^a \mathbf{e} >}\right]$$

$\mathbf{Z}^f = [\mathbf{p}_1^{\mathrm{f}}, \ \mathbf{p}_2^{\mathrm{f}}, \ \ldots\ldots, \mathbf{p}_{\mathrm{k}}^{\mathrm{f}}]$ ← Matrix formed by forecast perturbations

$\mathbf{Z}^a = [\mathbf{p}_1^{\mathrm{a}}, \ \mathbf{p}_2^{\mathrm{a}}, \ \ldots\ldots\ldots, \mathbf{p}_{\mathrm{k}}^{\mathrm{a}}]$ ← Matrix formed by initial perturbations

$\mathbf{e}^T = (e_1, \ \mathrm{e}_2, \ \ldots\ldots., \ \ \mathrm{e}_{\mathrm{K}})$ ← Coefficients assigned to different perturbations, obtained by solving above maximization problem.