Impact of ensemble size and horizontal resolution on the performance of the NCEP GEFS

Juhui (Jessie) Ma, Yuejian Zhu and Panxing Wang*

EMC/NCEP
*NUIST, China

Acknowledgements: D. Hou, M. Pena, R. Wobus and M. Wei

07/24/2012

Outline

- Background and motivation
- Methodology for evaluation
- Study of ensemble size
- Comparison of ensemble size .vs horizontal resolution
- Summary and reference
- NCEP current operational GEFS

1. Background and Motivation

- The limited computational resources constrain increases of model resolution and ensemble size.

- When designing an effective operational ensemble forecast system, two important questions we should consider:
  - What is a reasonable ensemble size to represent forecast uncertainties better with limited computational resources?
  - Is there any relationship between resolution and ensemble size?

- The study is extremely necessary for further developing the NCEP GEFS.

Previous study

- Buizza, R., and T. Palmer, 1998: Impact of ensemble size on ensemble prediction. *Mon. Wea. Rev.*, **126**, 2503-2518.

- Buizza, R., T. Petroliagis, T. N. Palmer, J. Barkmeijer, M. Hamrud, A. Hollingsworth, A. Simmons, and N. Wedi, 1998: Impact of model resolution and ensemble size on the performance of an ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **124**, 1935-1960.

- Buizza, R., 2010: Horizontal resolution impact on short and long range forecast error. *Quart. J. Roy. Meteor. Soc.*, **136**, 1020-1035.

- Reynolds, C. A., J. G. McLay, J. S. Goerss, E. A. Serra, D. Hodyss, and C. R. Sampson, 2011: Impact of Resolution and Design On the U.S. Navy Global Ensemble Performance in the Tropics. *Mon. Wea. Rev.*, **139**, 2145-2155.

Differences to previous study

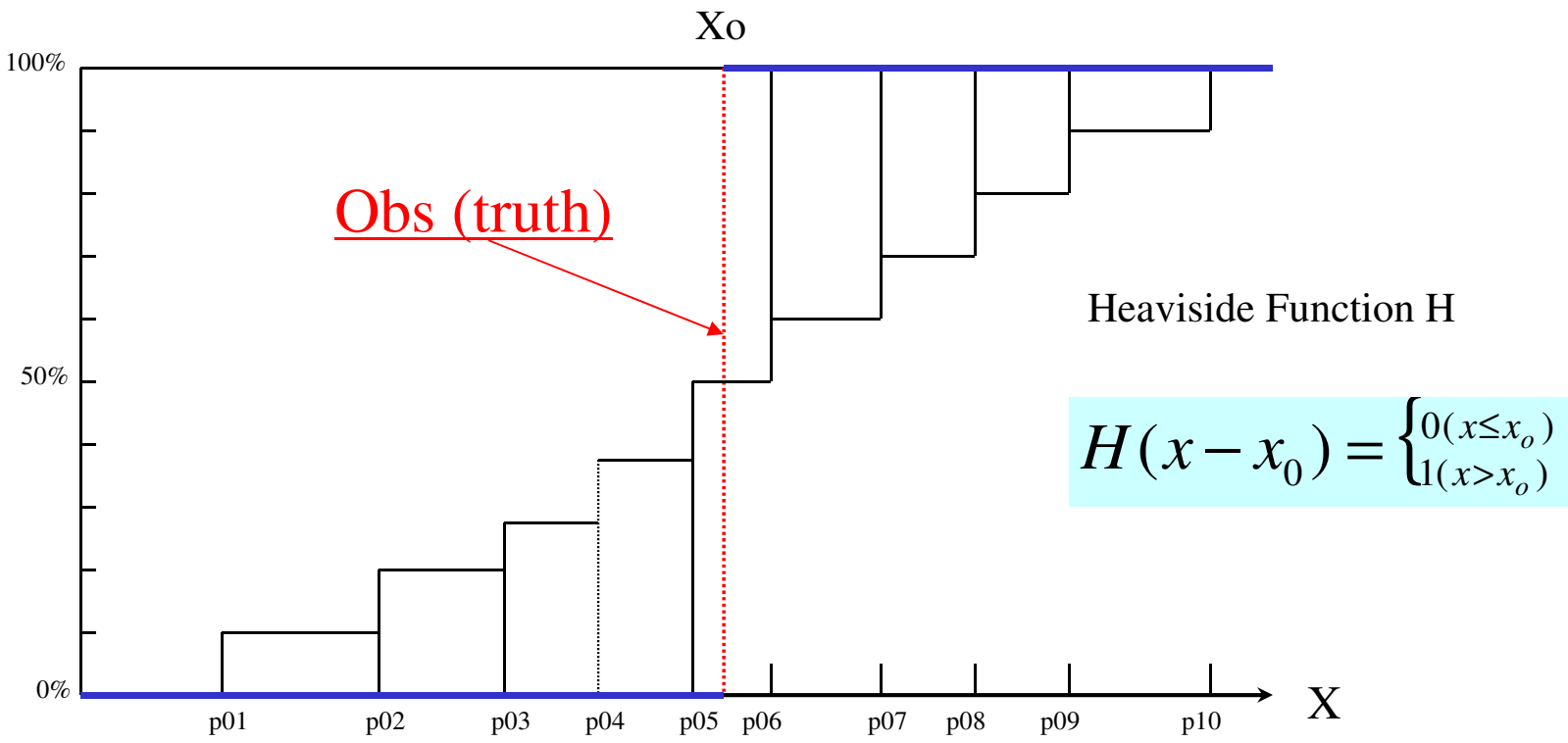
- Each model behaves differently. A similar study with recent versions of NCEP GEFS has not been carried out.
- This study employs with up to 80 ensemble members to explore the impact of ensemble size.
- This study compares two configurations based on equivalent computation resources to explore the relative impact of ensemble size and resolution.

2. GEFS Verification

- Reference (or truth): NCEP best analysis
- RMS error of ensemble mean (RMSE) (smaller, better)
  - It measures the distance between forecast and analyses.
- Pattern Anomaly Correlation score (PAC) (larger, better)
  - It calculates a correlation between the predicted anomaly and the observed anomaly with respect to their corresponding climatology.
  - It uses NCEP/NCAR 40-year reanalysis climatology.
- Continuous Rank Probability Score (CRPS) (smaller, better)
  - See next slide for detail calculation
- Brier Skill score (BSS) (larger, better) and its decompositions
  - Less than zero – no skill
  - Decomposition - reliability (REL) (smaller, better) and resolution (RES) (larger, better)
- Statistical significance testing— *t*-test (95% confidence interval)

Continuous Rank Probability Score

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



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

Brier Skill Score

See <<Statistical Methods in the Atmospheric Science>> by D. S. Wilks, Chapter 7: Forecast Verification

1. BS (Brier Score)

$$BS = \frac{1}{n}\sum_{k=1}^{n}(y_k - o_k)^2$$

*Where **y** is a forecast probability and **o** is an observation (probability), index **k** denotes a number of the **n** forecast event/pairs. **y** and **o** are limited from 0 to 1 in the probability sense. **BS**=0 is a perfect forecast, and **BS**=1 is missing everything*

2. BSS (Brier Skill Score)

$$BSS = \frac{BS_f - BS_{ref}}{BS_{perf} - BS_{ref}} = 1 - \frac{BS_f}{BS_{ref}}$$

***ref** is the reference which is mostly climatology, **BSperf**=0 for perfect forecast, **BSS** is ranged from 0-1.*

Brier Skill Score (and decomposition)

3. Algebraic Decomposition of the Brier Score

After some algebra, the Brier Score can be expressed as three separated terms

Reliability *Resolution* *Uncertainty*

$$BS = \frac{1}{n}\sum_{i=1}^{I} N_i (y_k - \overline{o}_i)^2 - \frac{1}{n}\sum_{i=1}^{I} N_i (\overline{o}_i - \overline{o})^2 + \overline{o}(1-\overline{o})$$

where $n = \sum_{i=1}^{I} N_i$

Conditional probability of observed and sample climatology

$$\overline{o}_i = p(o_1 \mid y_i) = \frac{1}{N_i}\sum_{k \in N_i} o_k \quad \textit{and} \quad \overline{o} = \frac{1}{n}\sum_{k=1}^{n} o_k$$

4. Algebraic Decomposition of the Brier Skill Score

Resolution *Reliability*

$$BSS = 1 - \frac{BS_f}{BS_{ref}} = 1 - \frac{REL - RES + UNC}{UNC} = RESSS - RELSS$$

3. Ensemble size

- **Lorenz 96 model experiments**
  - -Design
  - -Results
- **NCEP GEFS model experiments**
  - -Design
  - -Results

Lorenz 96 model experiment

- Design

Model: Lorenz 96 model $\frac{dX_i}{dt} = (X_{i+1} - X_{i-2})X_{i-1} - X_i + F$

- i=1,2,…, N, N=1000 is the number of state vector variables.
- The magnitude of the forcing $F = 8$
- $\Delta t = 0.05$ (corresponds to approximately 6-h in the atmosphere)
- Fourth-order Runge-Kutta integration scheme

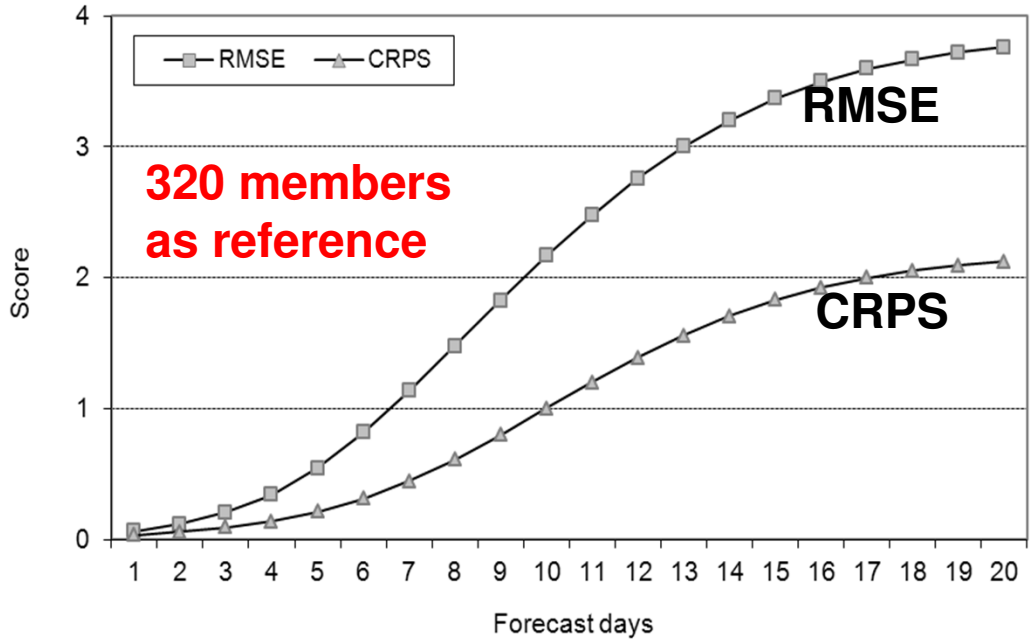
Analysis: 3DVAR method

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x}_b - \mathbf{x})^T \mathbf{B}^{-1}(\mathbf{x}_b - \mathbf{x}) + \frac{1}{2}(\mathbf{y} - H(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y} - H(\mathbf{x}))$$

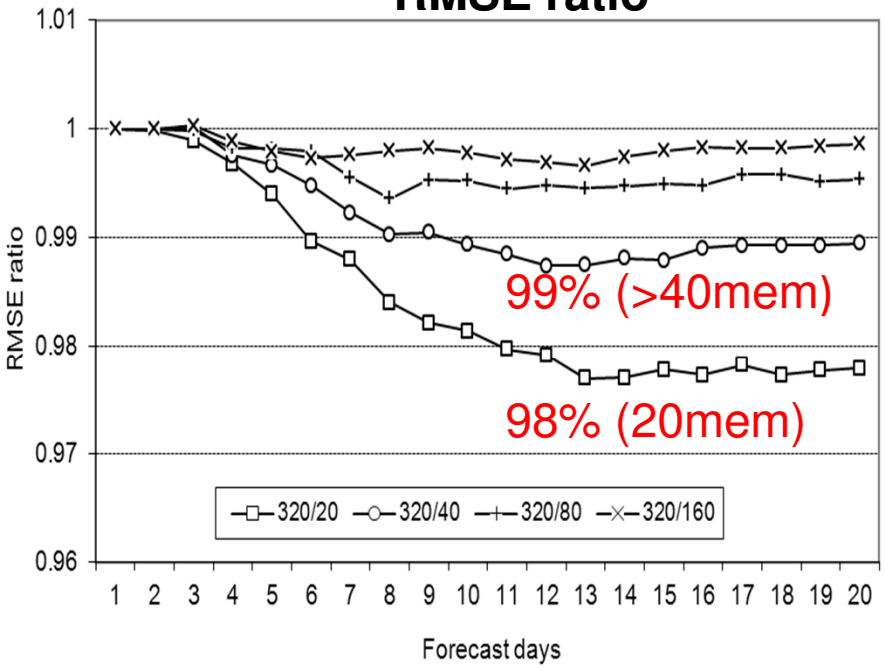
Initial perturbations : ETR based perturbation

Ensemble size: 20, 40, 80, 160, 320

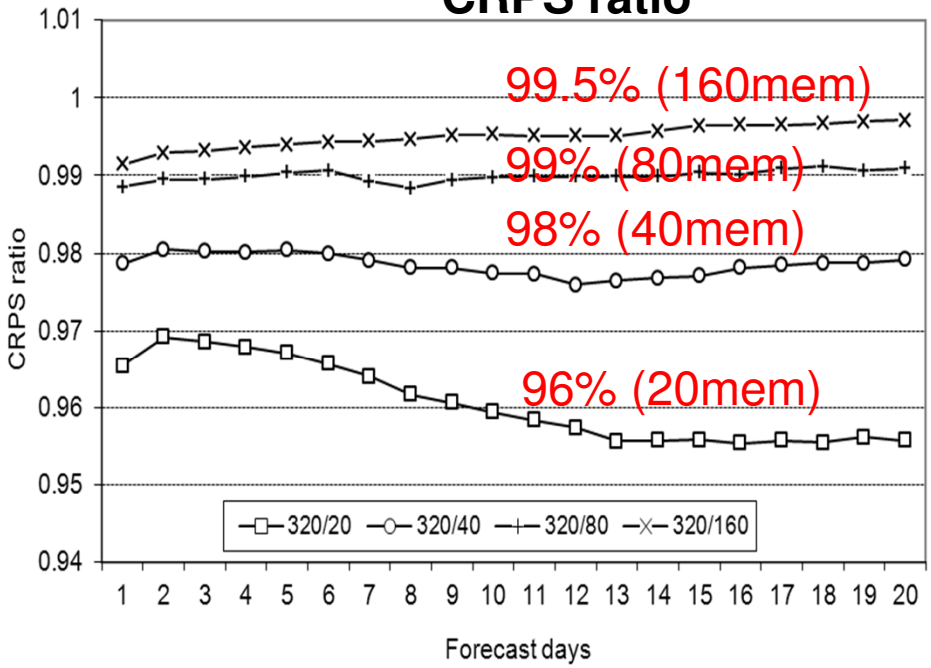
Verification: RMSE and CRPS



RMSE ratio



CRPS ratio



NCEP GEFS model experiments

- Design

Model: GFS v8.00 (prior to 2009)

Resolution: T126L28 (about 100km in horizontal)

Analysis: GDAS/GSI

Initial perturbations : ETR based perturbation (Wei and et al; 2006)

Cycling: 80 members every 6 hours

Ensemble sizes: 5, 10, 20, 40, 60 and 80

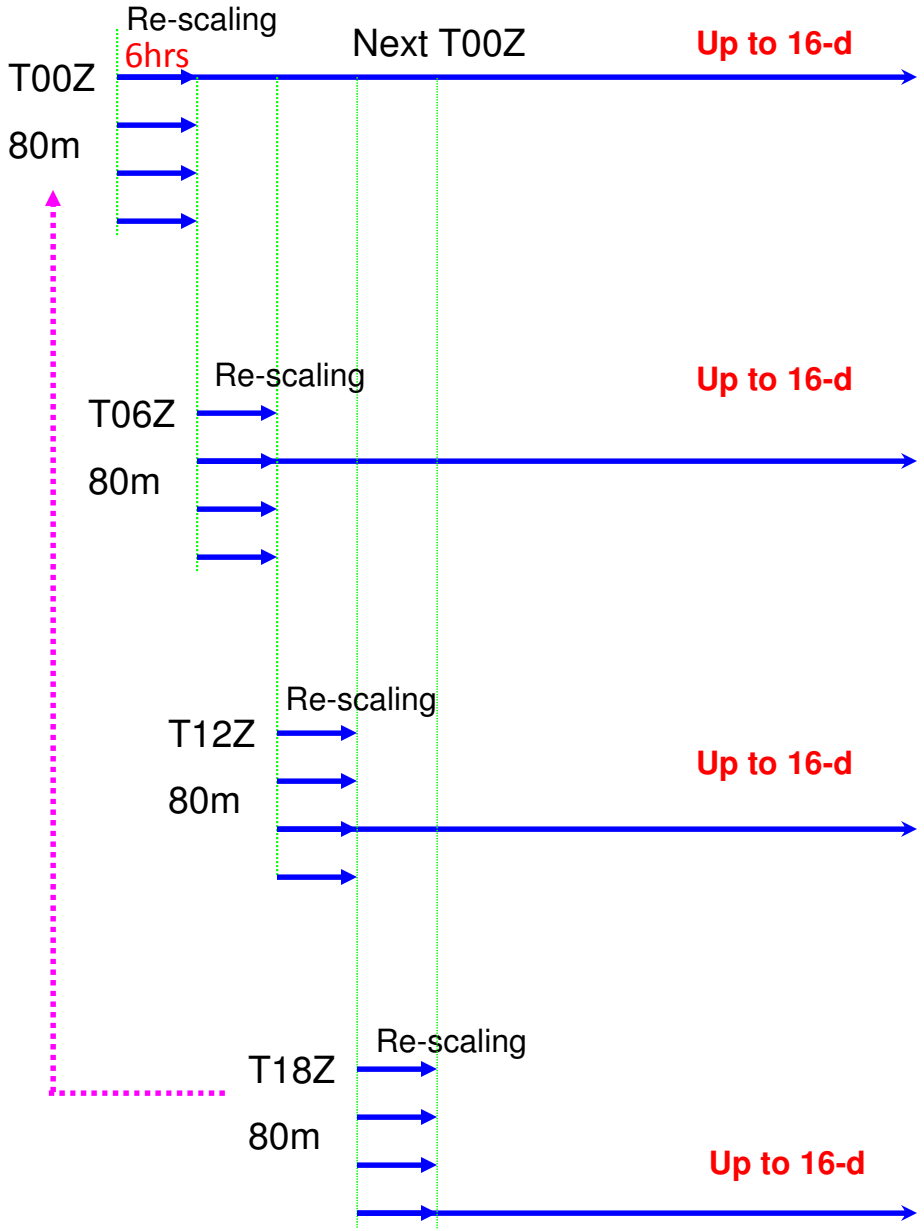
Forecast length: Out to 384 hours from each 00UTC

Period: December 1st, 2009 - January 31st, 2010

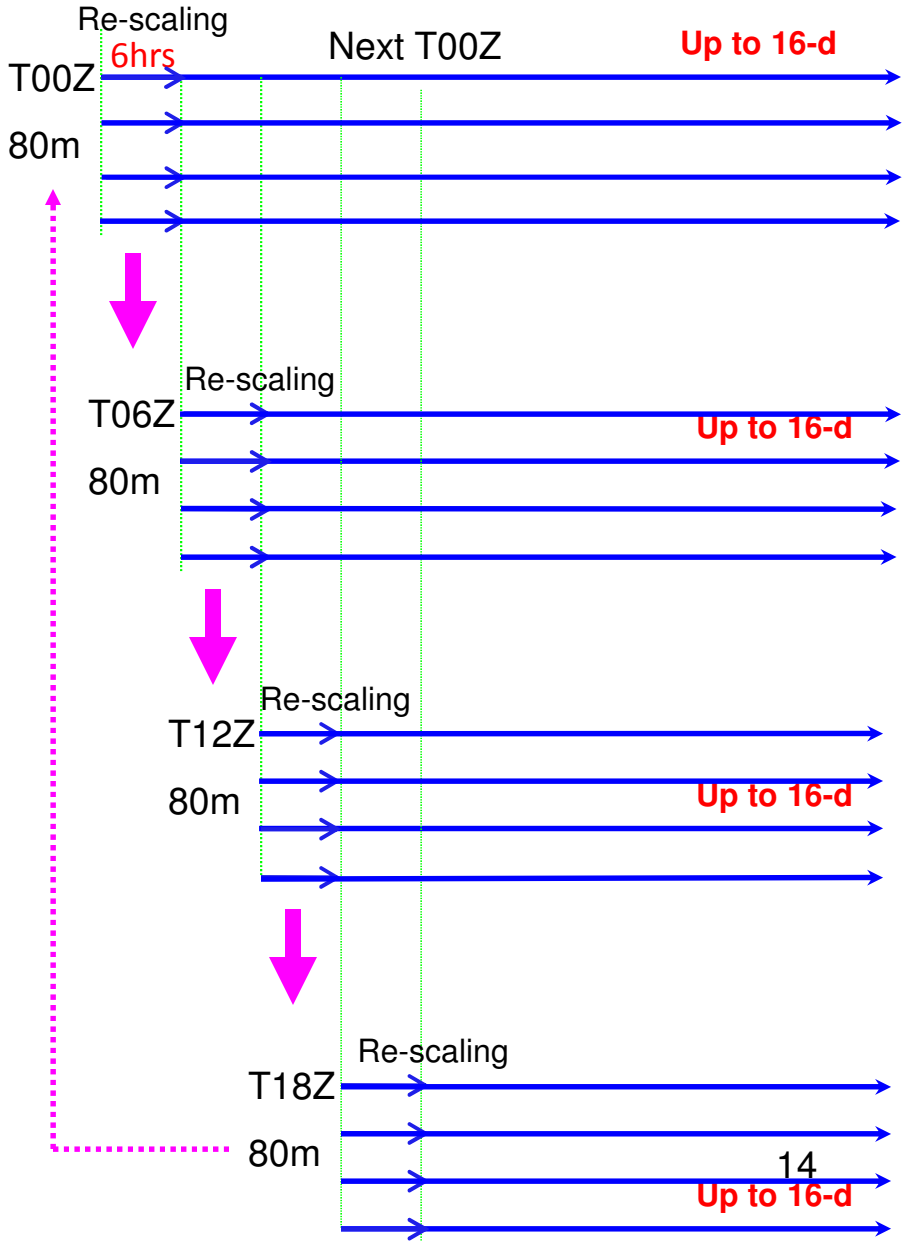
Variables: Z500, Z1000, T850, T2m, U10m, V10m

Verification: RMSE, PAC, CRPS, BSS and its decompositions of reliability and resolution

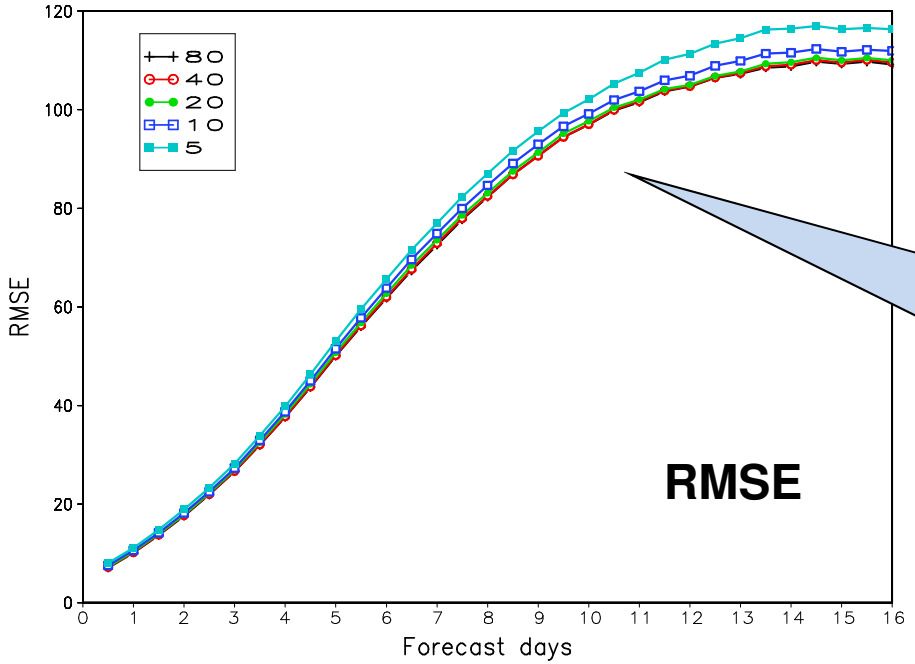
6-hour ETR cycle (operation)



6-hour ETR cycle (experiment)



NH 500hPa height



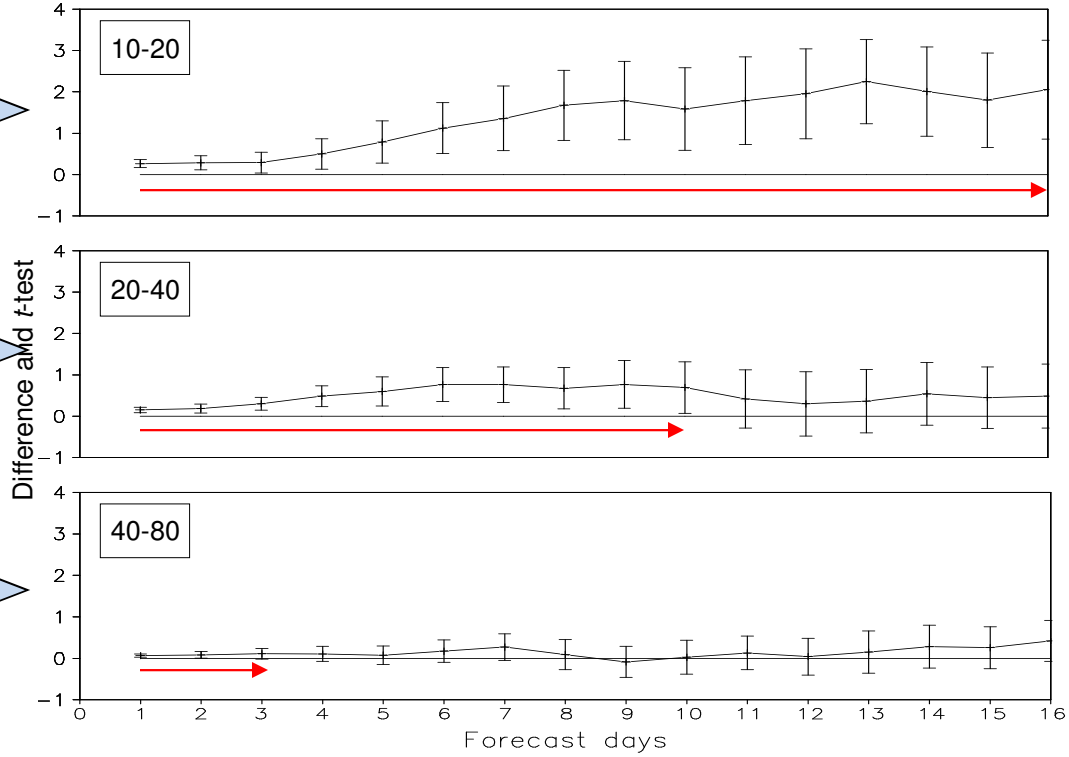
RMSE reductions are obvious with increasing ensemble sizes from 5 to 10, and from 10 to 20, and very small with further increasing.

Significant test

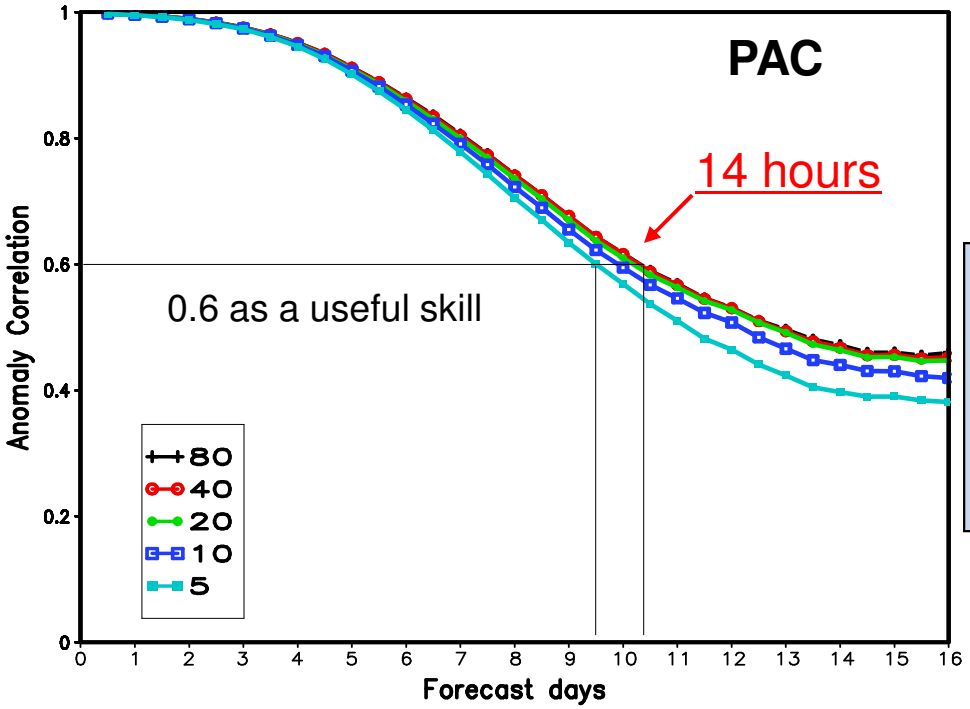
significant for all lead times

significant before 10 days

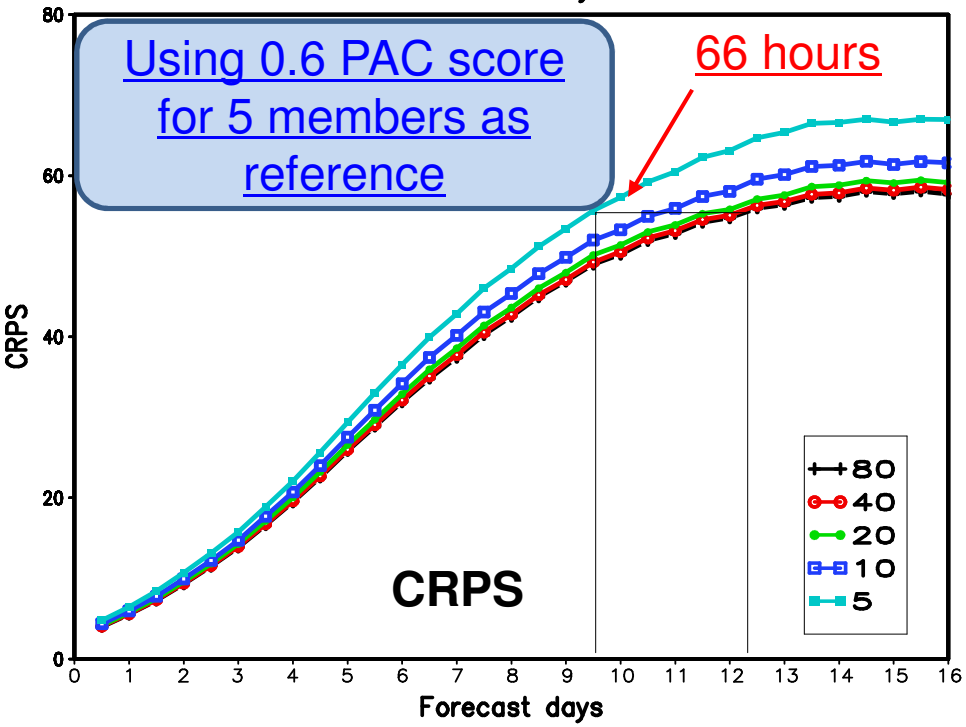
significant before 3 days



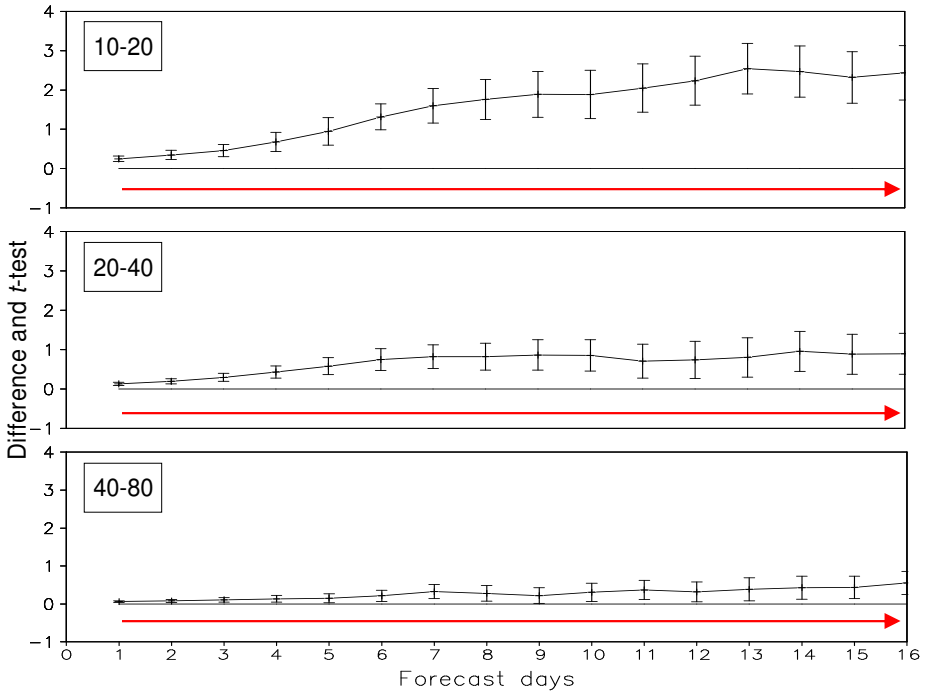
NH 500hPa height



The improvement of probabilistic forecast evaluation with increasing ensemble size is significantly larger than ensemble mean evaluation.

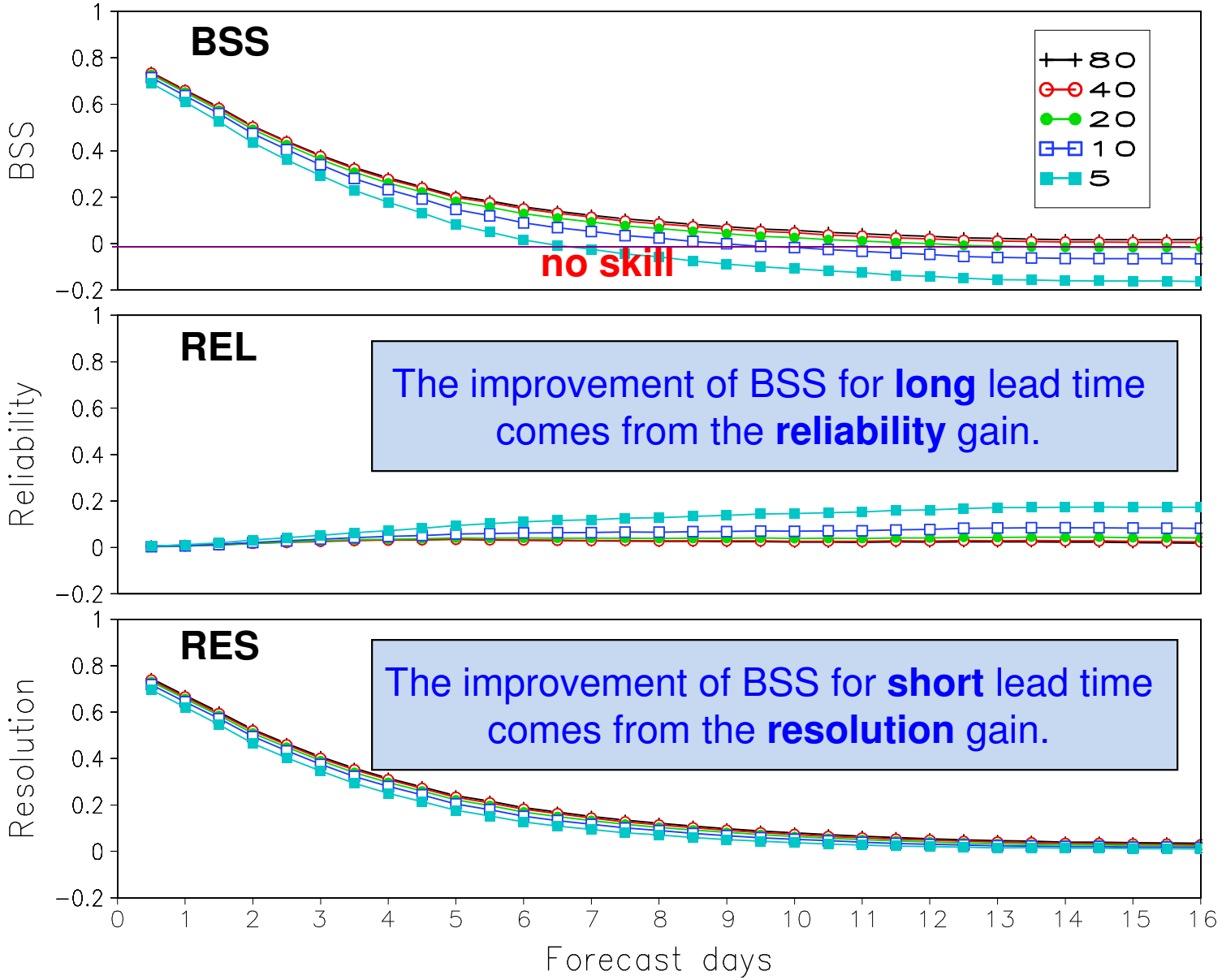


Significant test

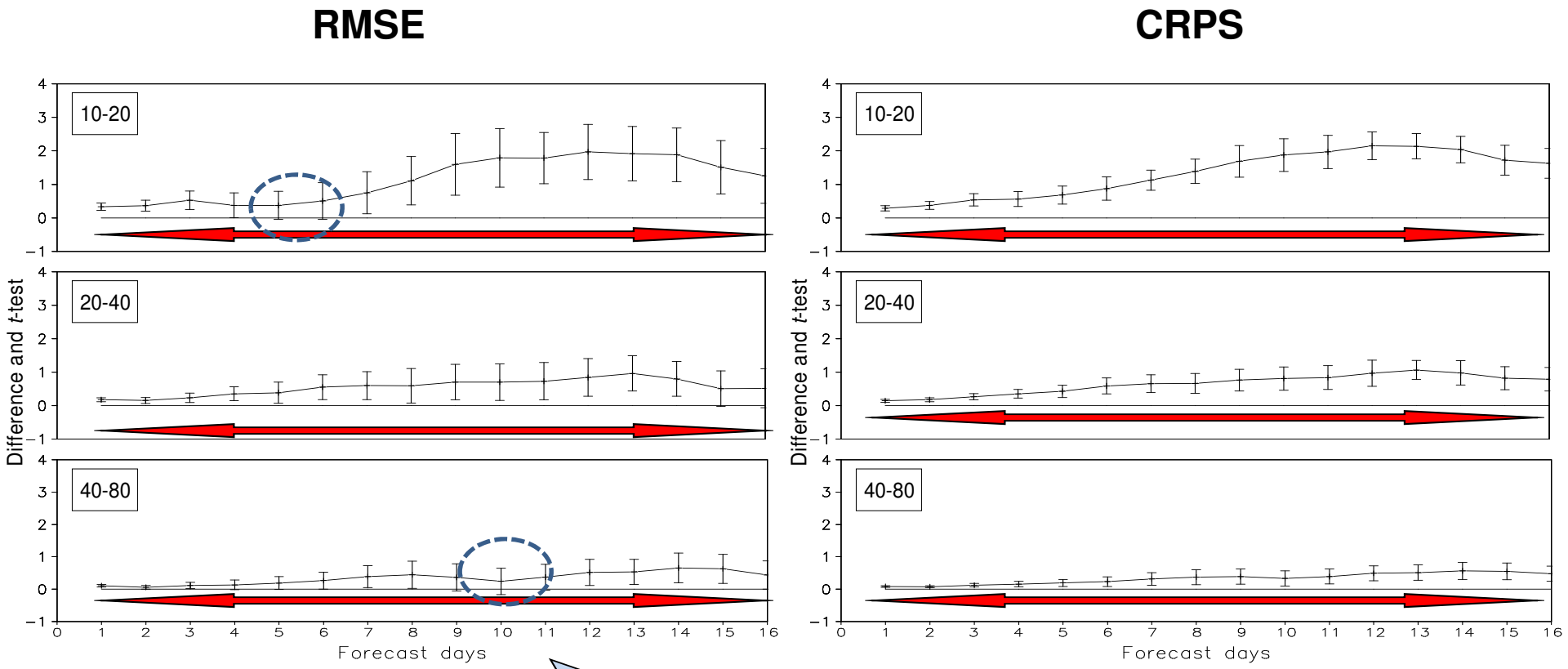


NH 500hPa height

BSS → Reliability + Resolution

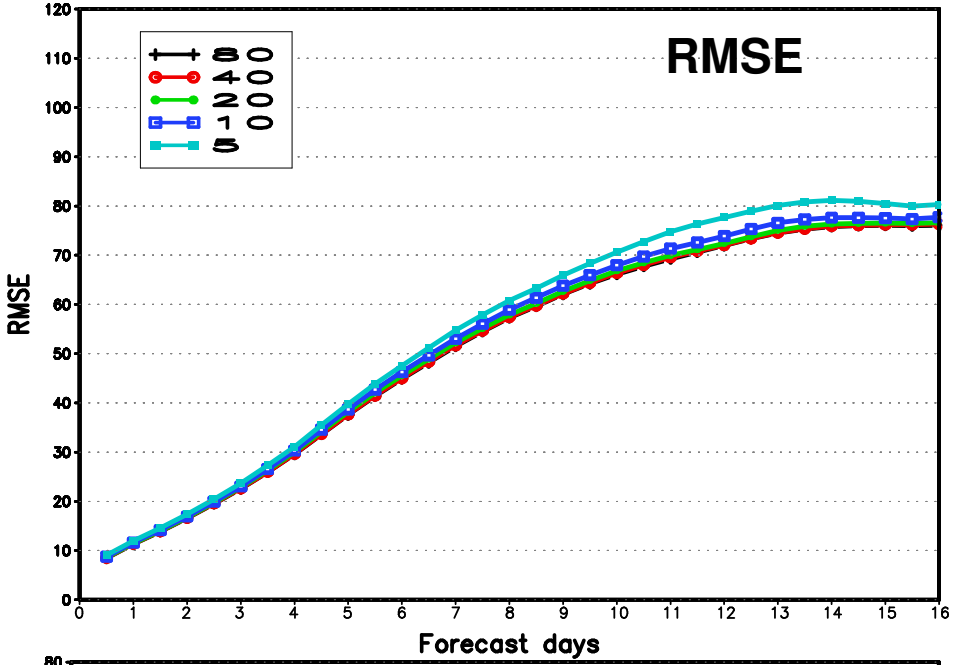


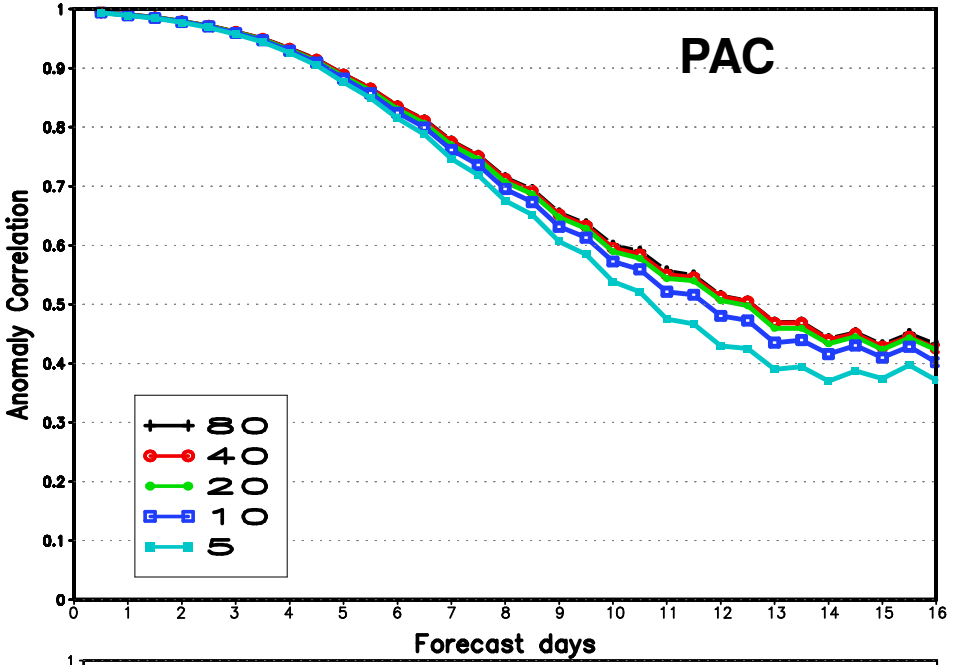
SH 500hPa height

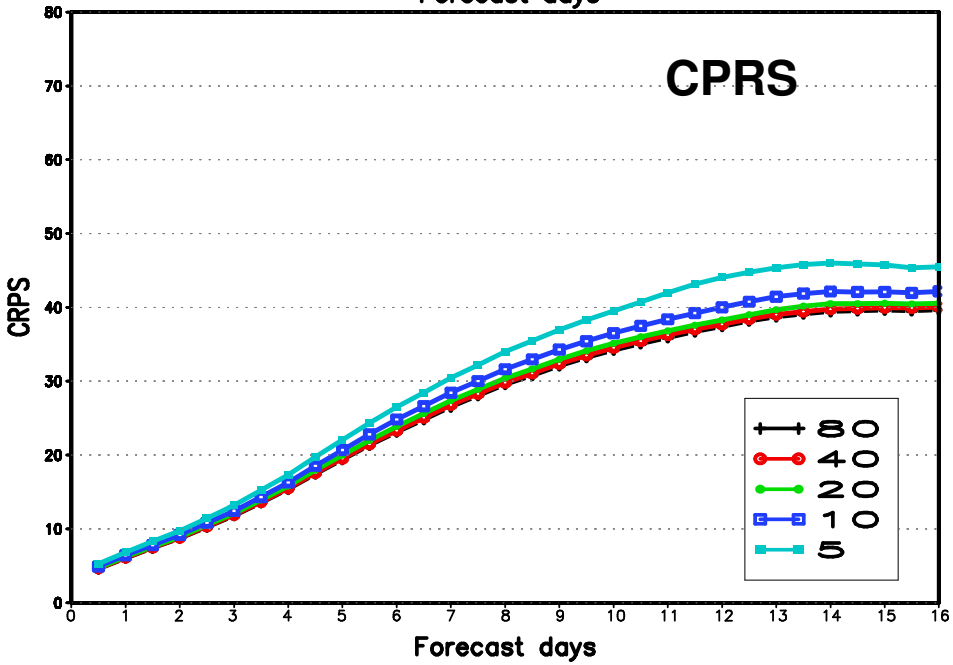


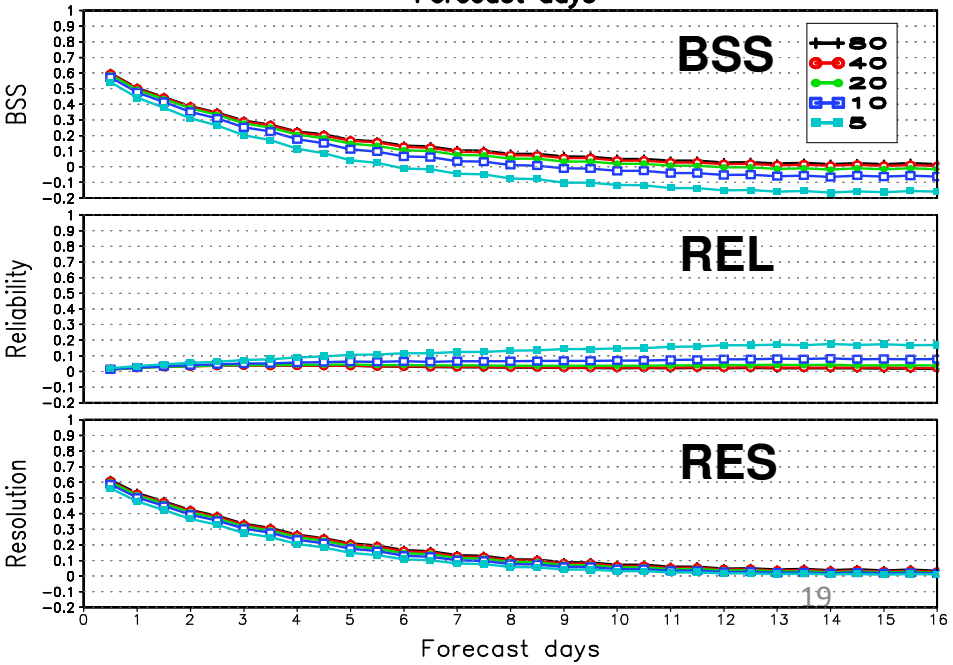
The impact of ensemble size for ensemble mean forecast over SH is greater than over NH.

NH 1000hPa height

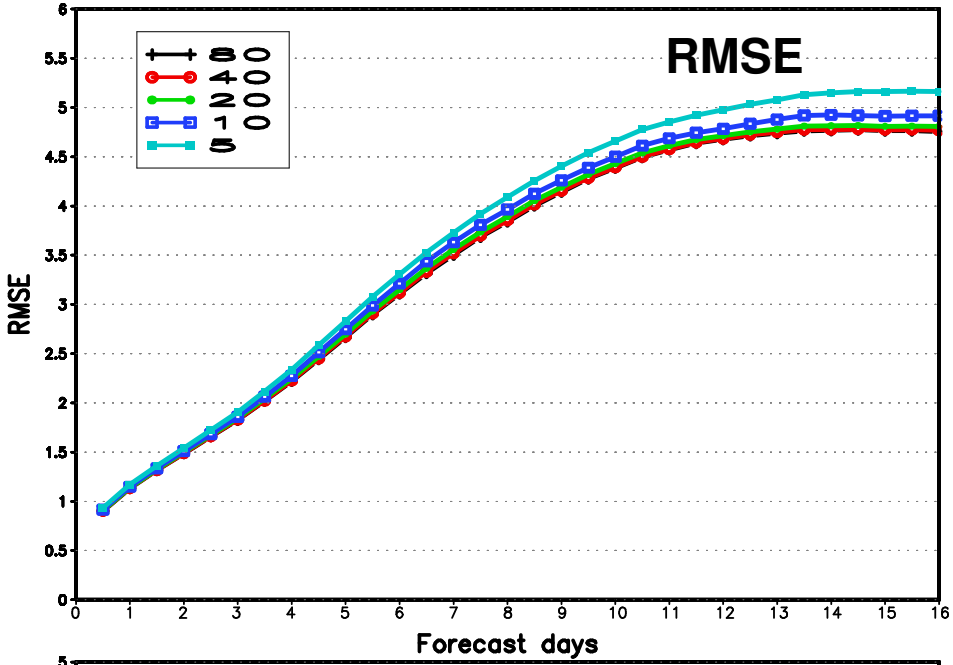


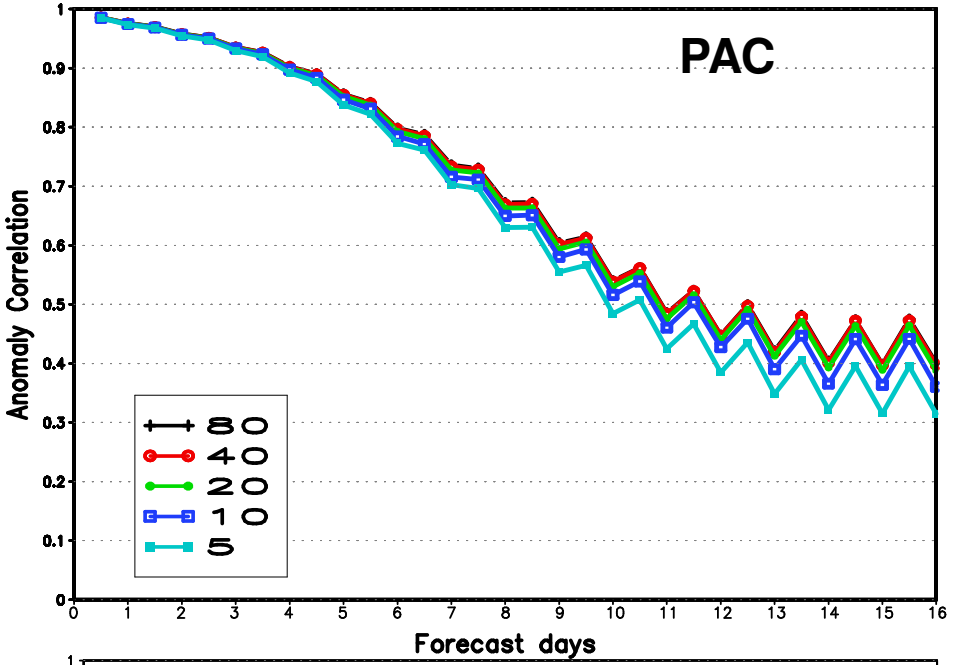


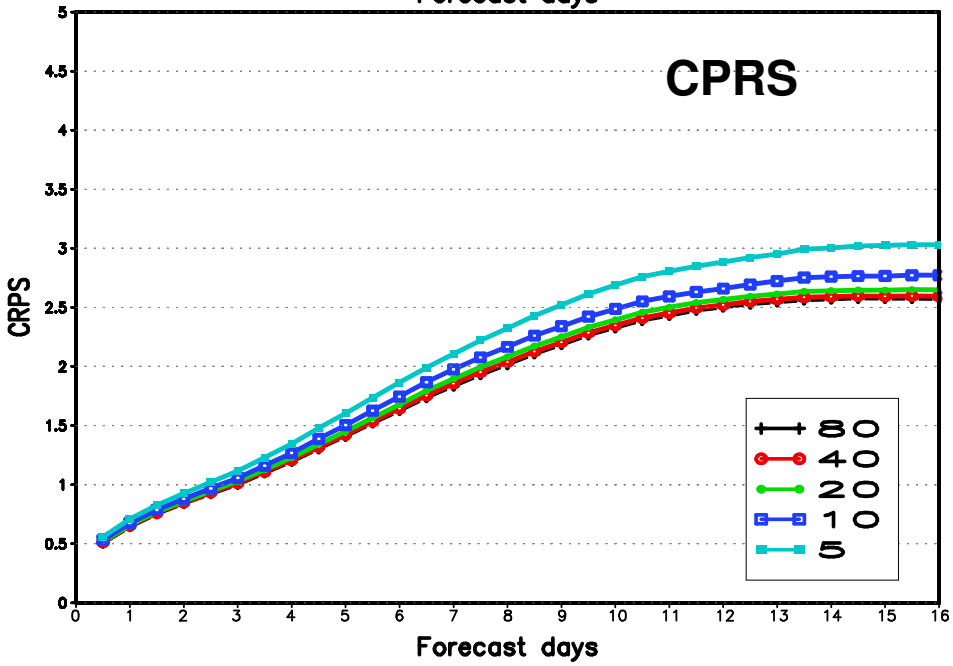


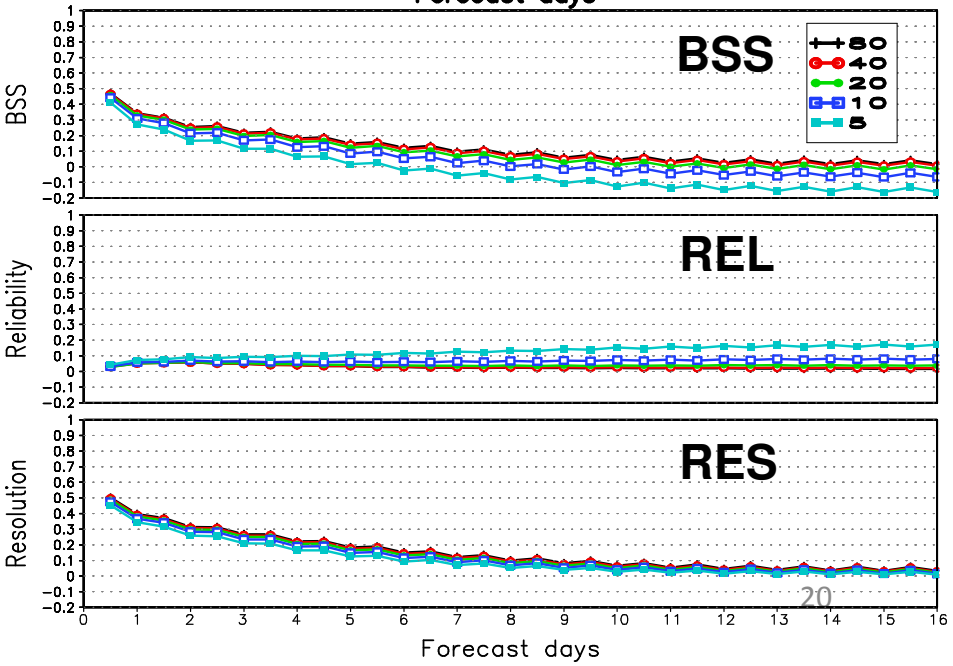


NH 850hPa Temperature

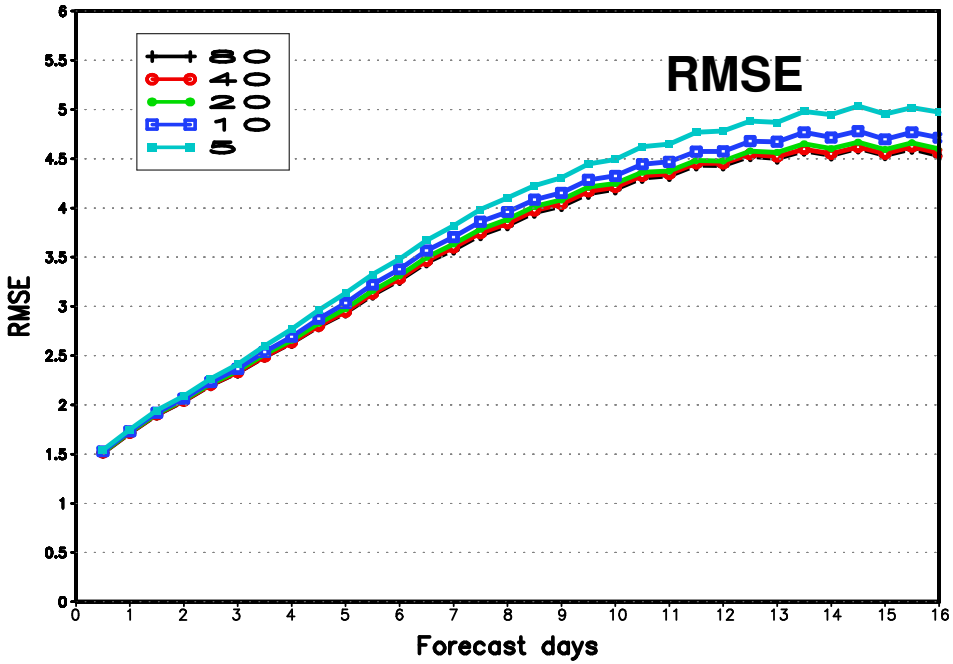


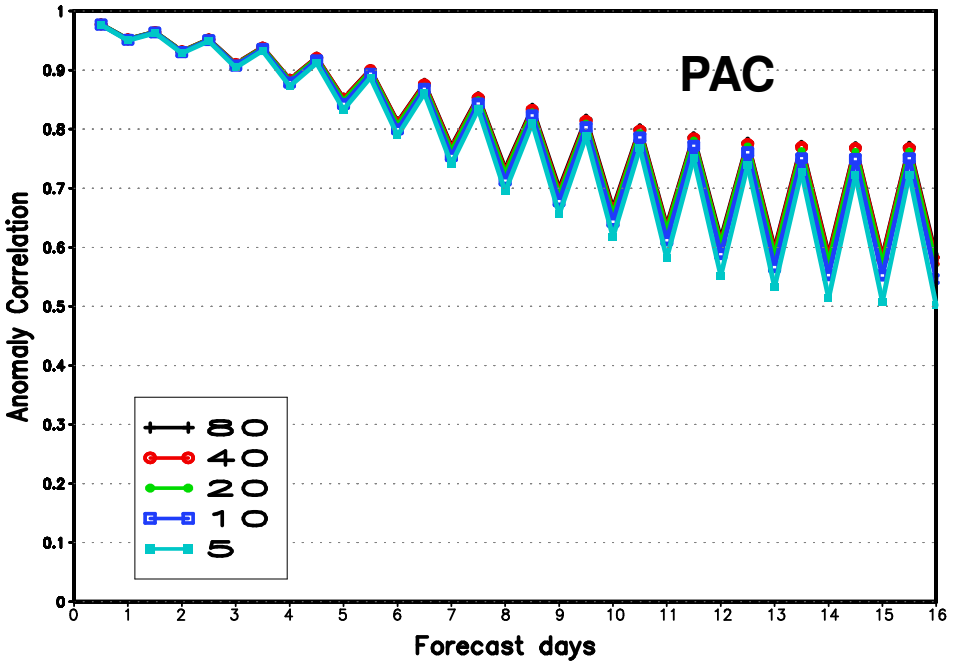


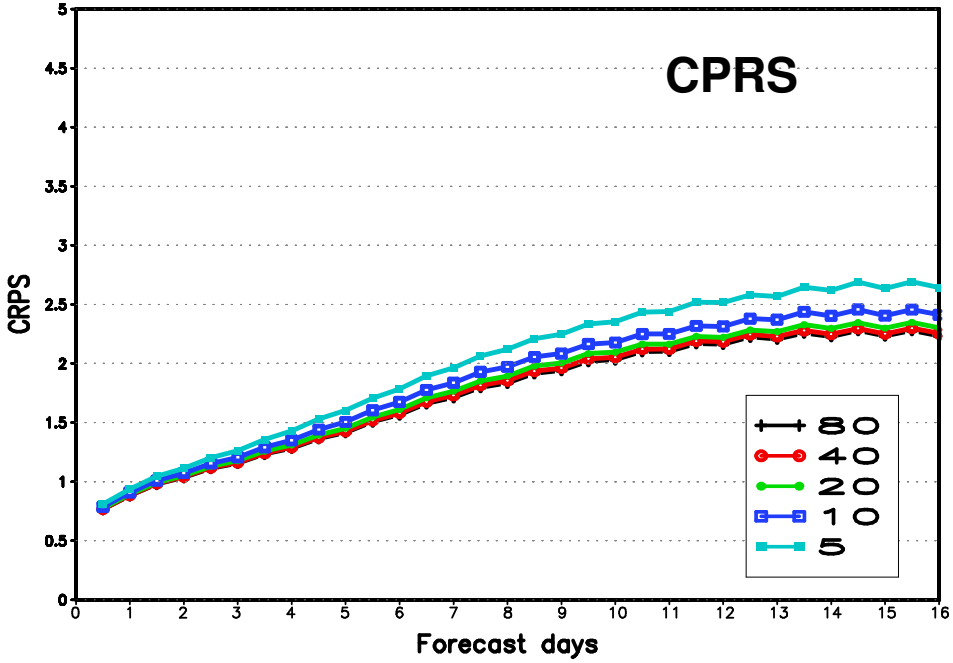


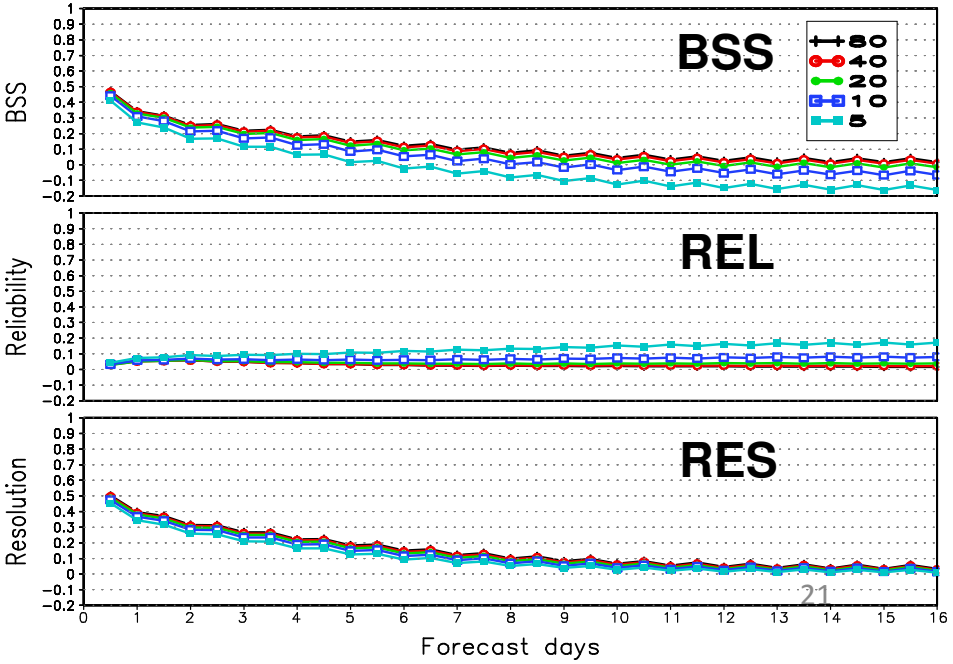


NH 2m Temperature

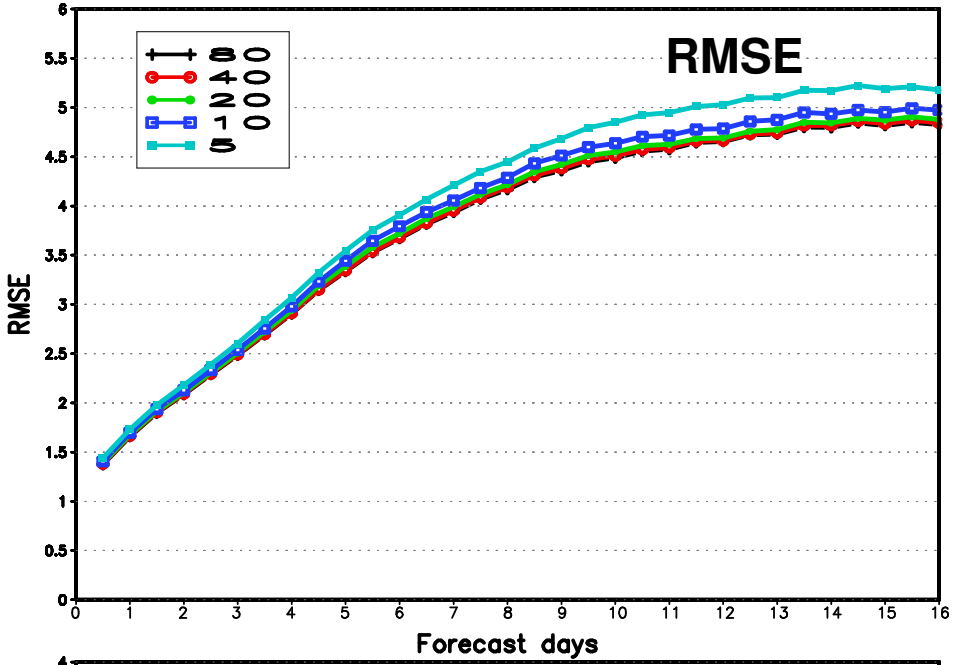


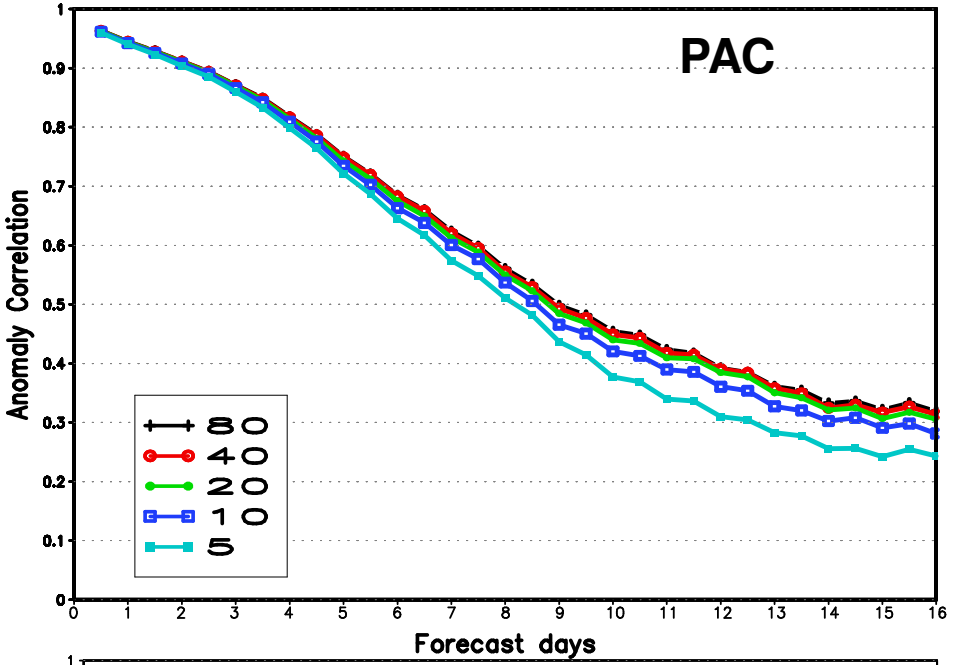


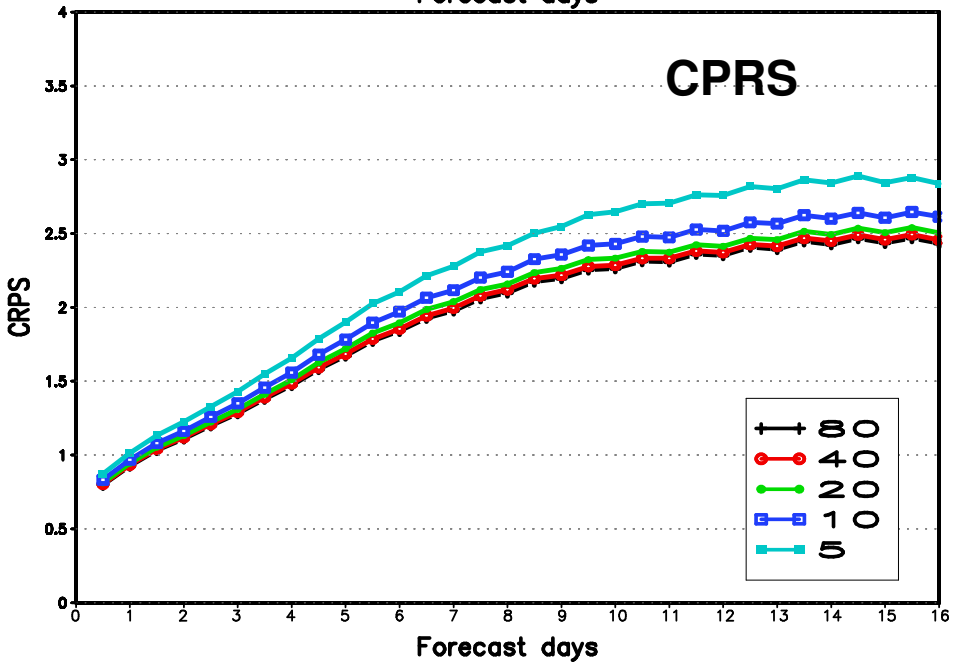


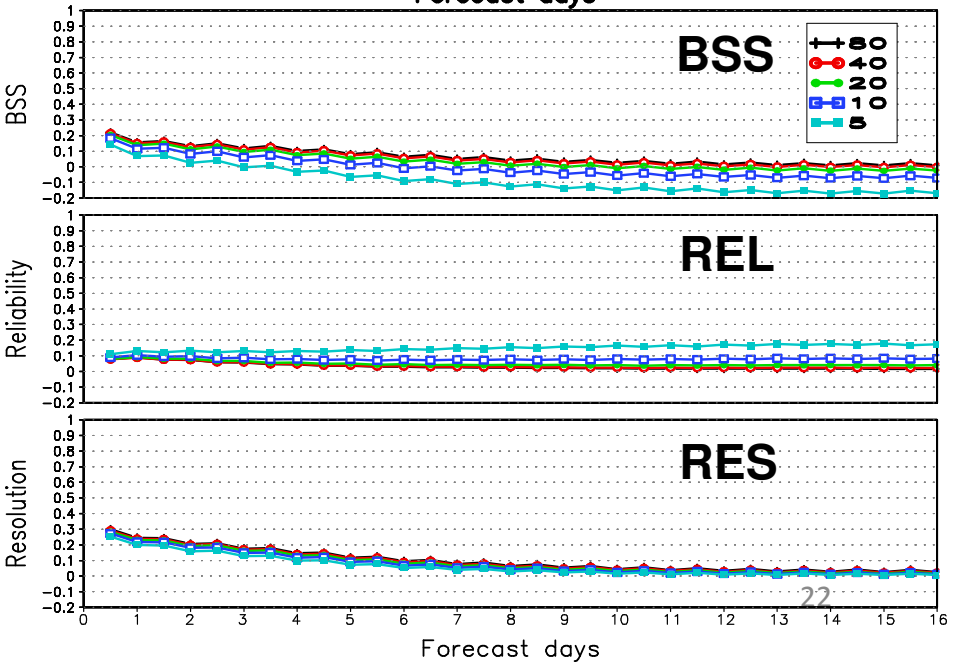


NH 10m Zonal Wind (U)

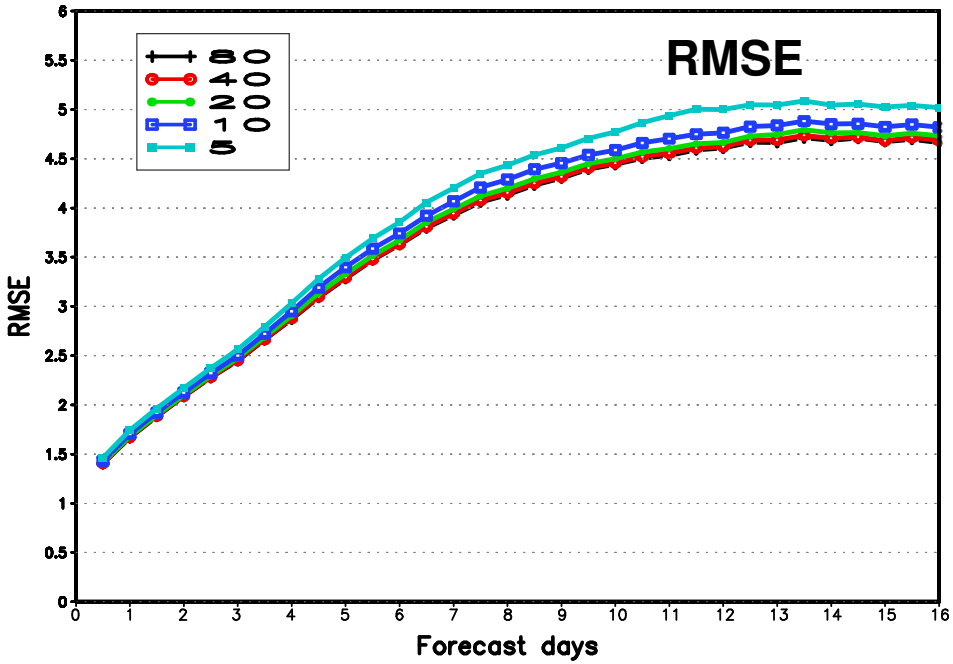


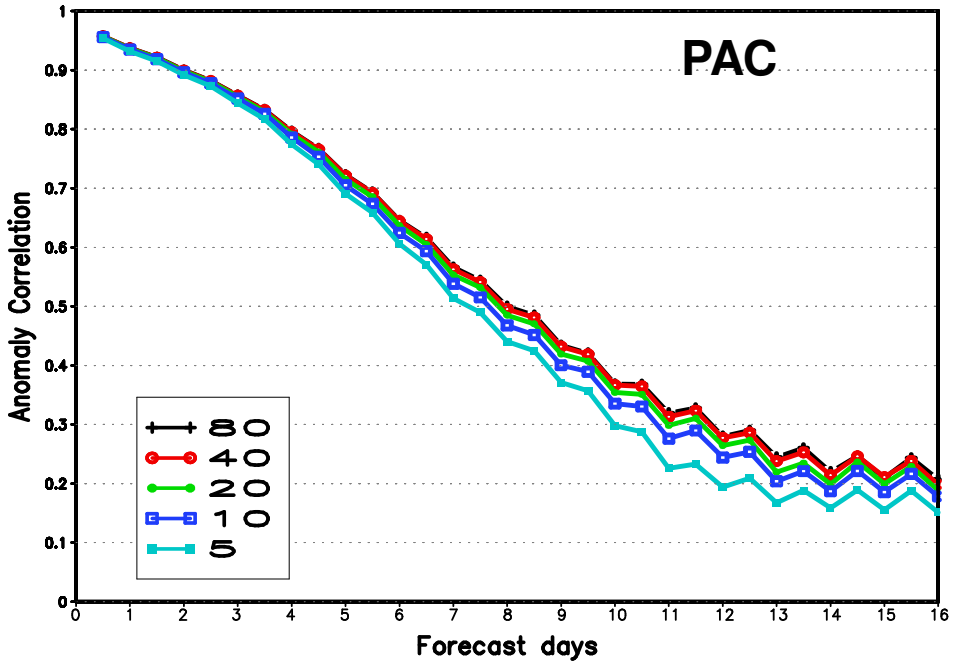


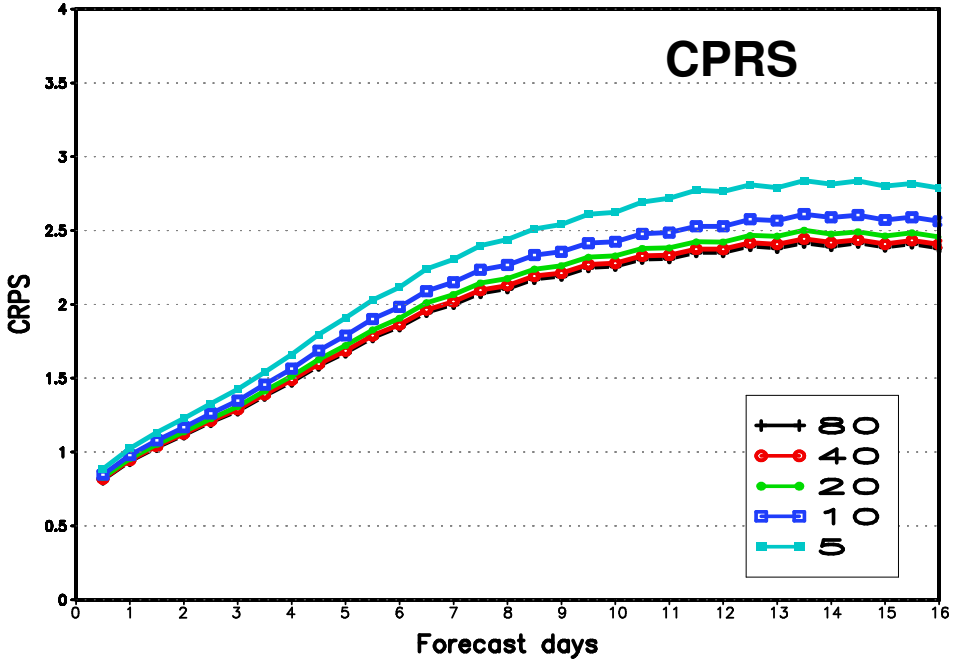


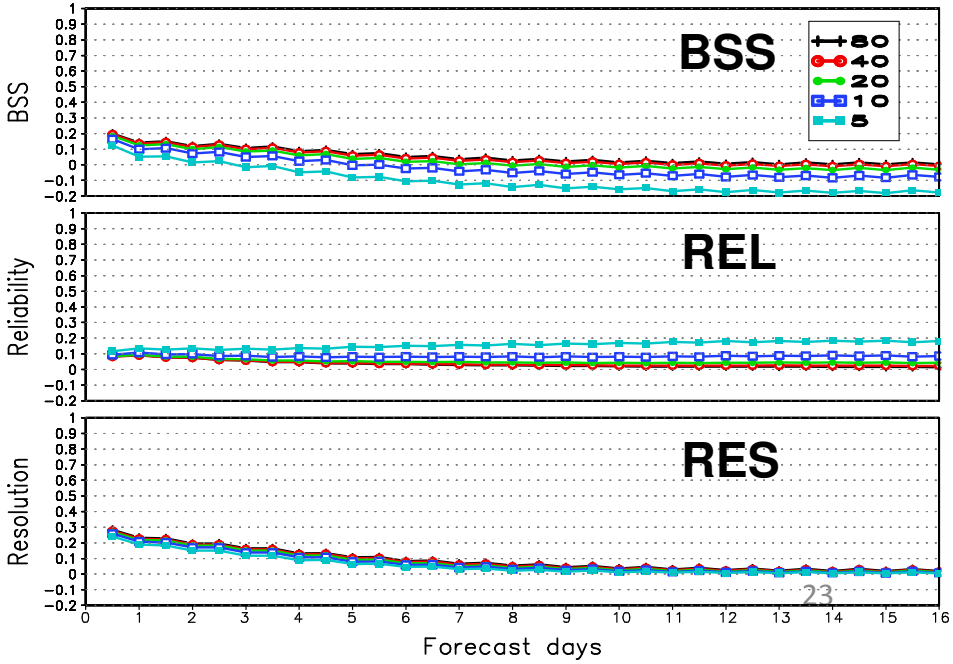


NH 10m Meridional Wind (V)









Forecast Lead Times (days) of Statistically Significant Better

Variable	Ensemble Size	NH		SH	
		RMSE	CRPS	RMSE	CRPS
Z500	10-20	1-16	1-16	1-16	1-16
	20-40	1-10	1-16	1-15	1-16
	40-80	1-3	1-16	1-8&12-16	1-16
Z1000	10-20	1-16	1-16	1-16	1-16
	20-40	1-10	1-16	1-16	1-16
	40-80	1-3	1-16	1-15	1-16
T850	10-20	1-16	1-16	1-16	1-16
	20-40	1-16	1-16	1-16	1-16
	40-80	1-4	1-16	1-16	1-16
T2m	10-20	1-16	1-16	1-16	1-16
	20-40	1-16	1-16	1-16	1-16
	40-80	1-13	1-16	1-16	1-16
U10m	10-20	1-16	1-16	1-16	1-16
	20-40	1-16	1-16	1-16	1-16
	40-80	1-11	1-16	1-16	1-16
V10m	10-20	1-16	1-16	1-16	1-16
	20-40	1-16	1-16	1-16	1-16
	40-80	1-5	1-16	1-16	1-16

4. Ensemble size vs. horizontal resolution

- Design

Configuration: 20T190 vs. 70T126 (equivalent computational costs when they run at CCS – IBMP6)

Model: GFS v8.00

Analysis: GDAS/GSI

Initial perturbation : ETR based perturbation

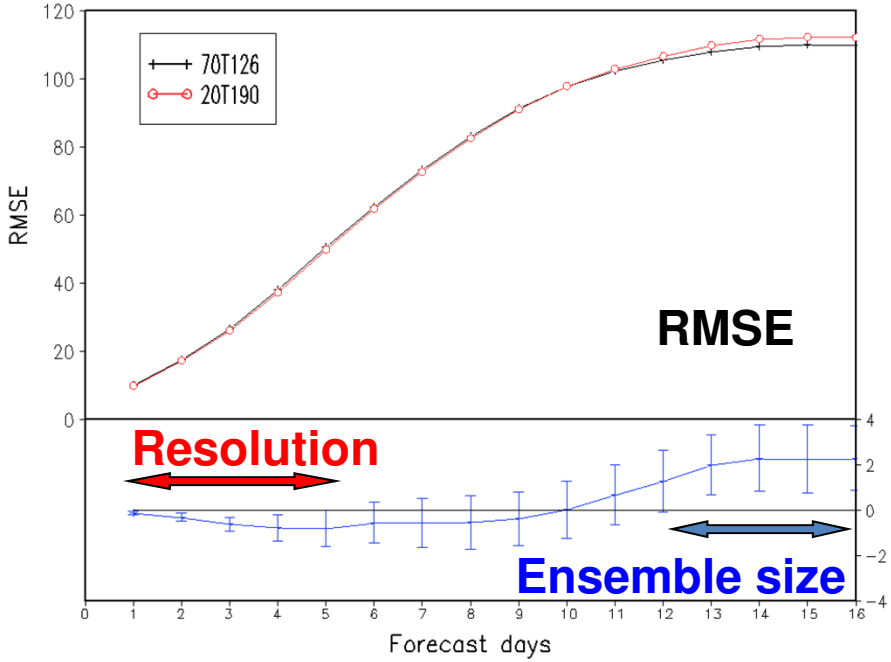
Forecast length: Out to 384 hours from each 00UTC

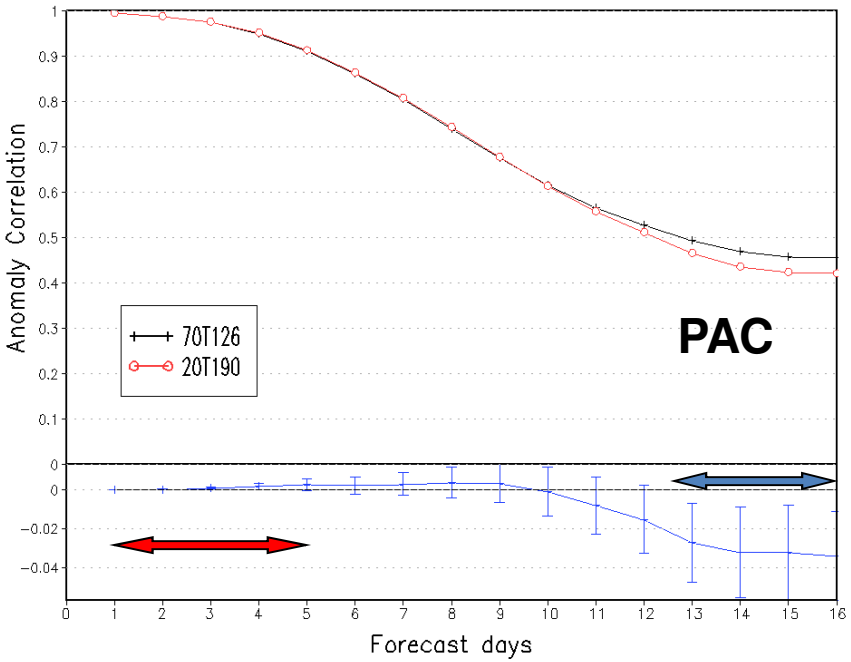
Period: December 1st, 2009 - January 31st, 2010

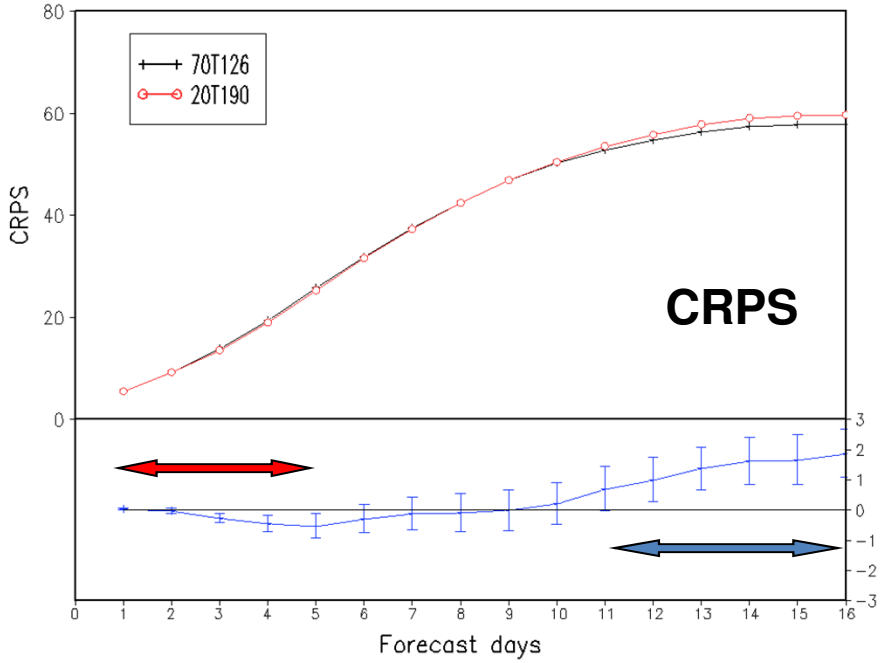
Variable: Z500, Z1000, T850, T2m, U10m, V10m

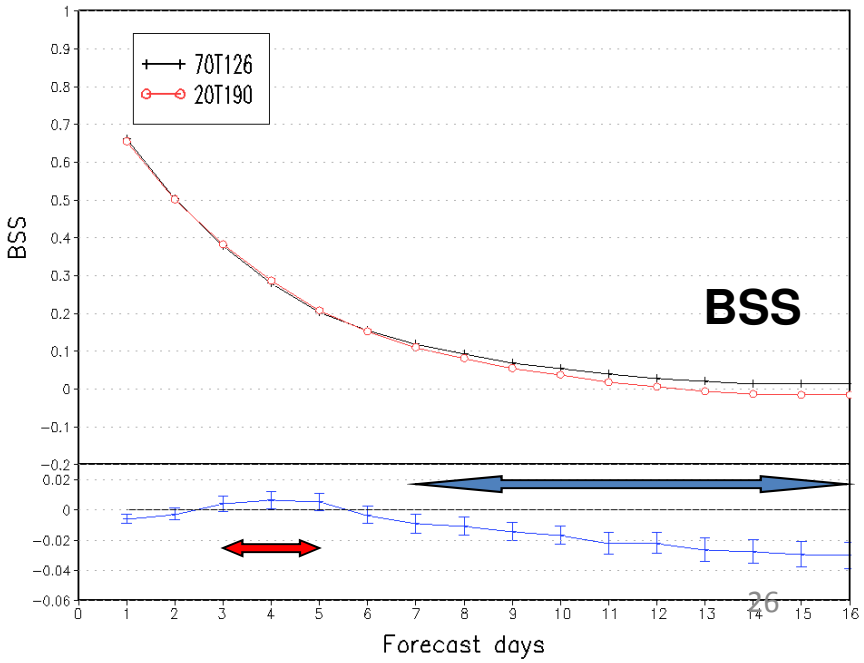
Verification: RMSE, PAC, CRPS and BSS

NH 500hPa height

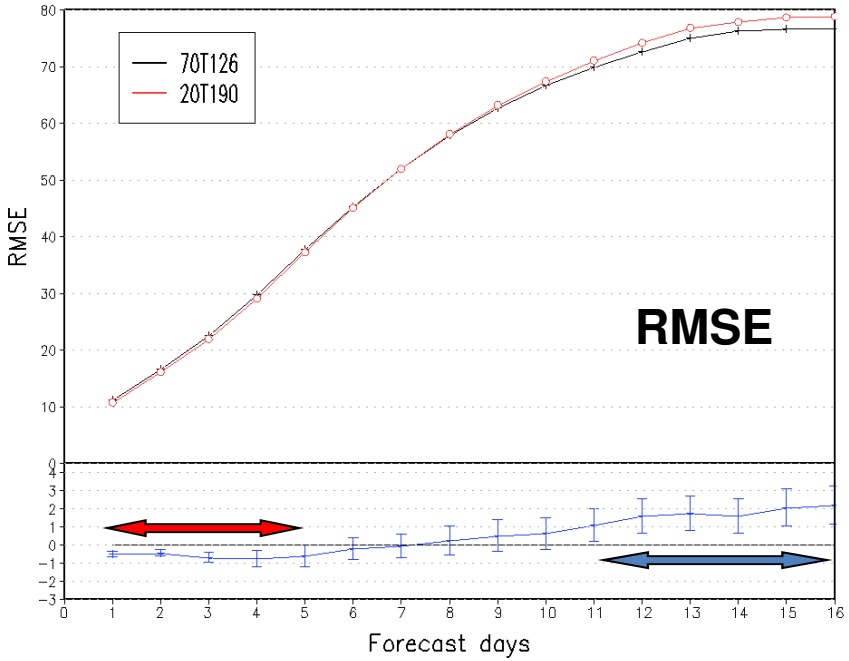


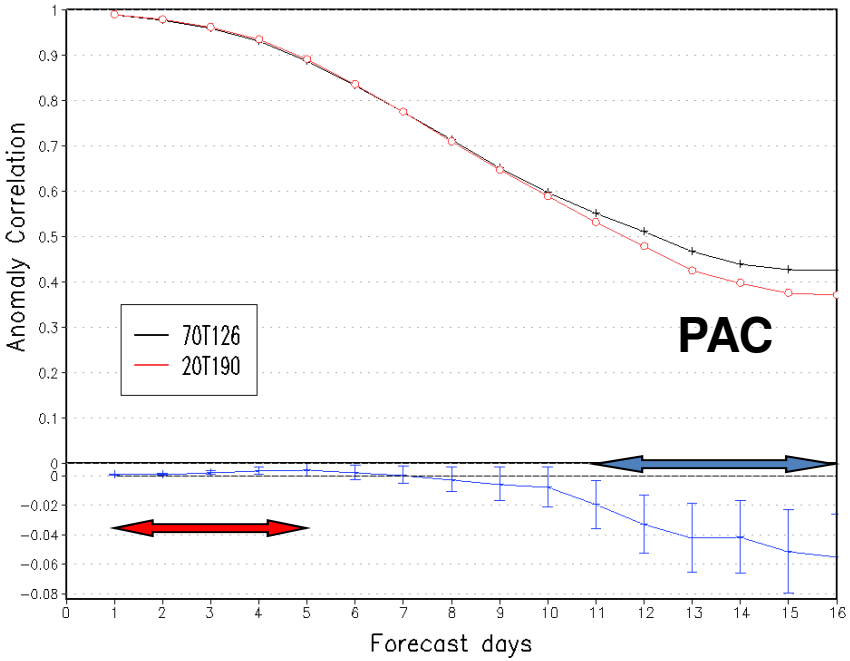


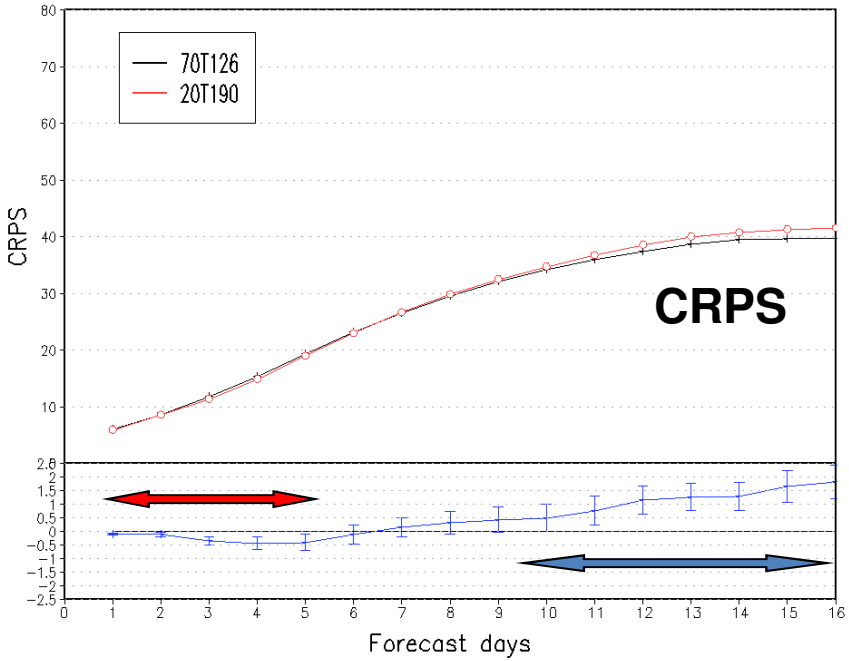


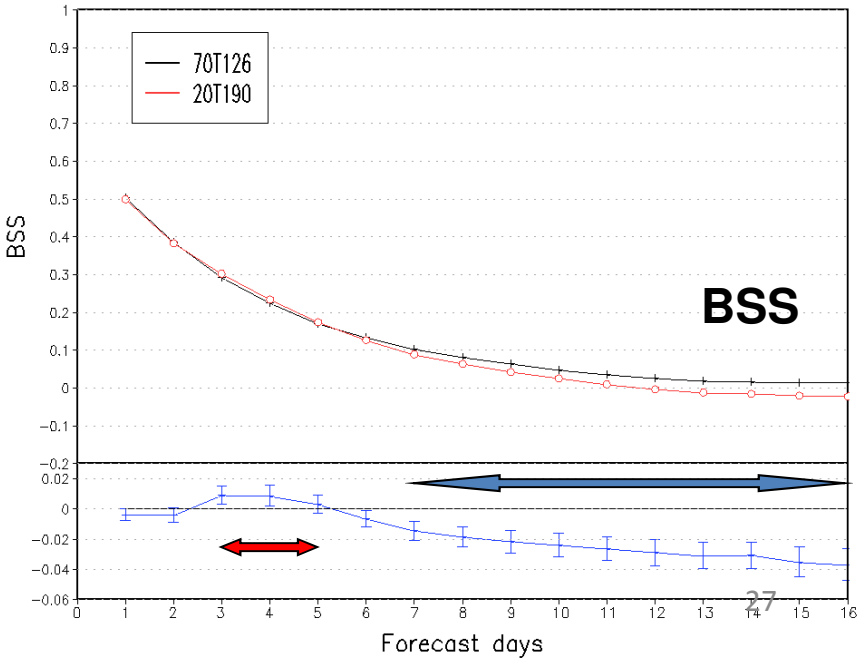


NH 1000hPa height

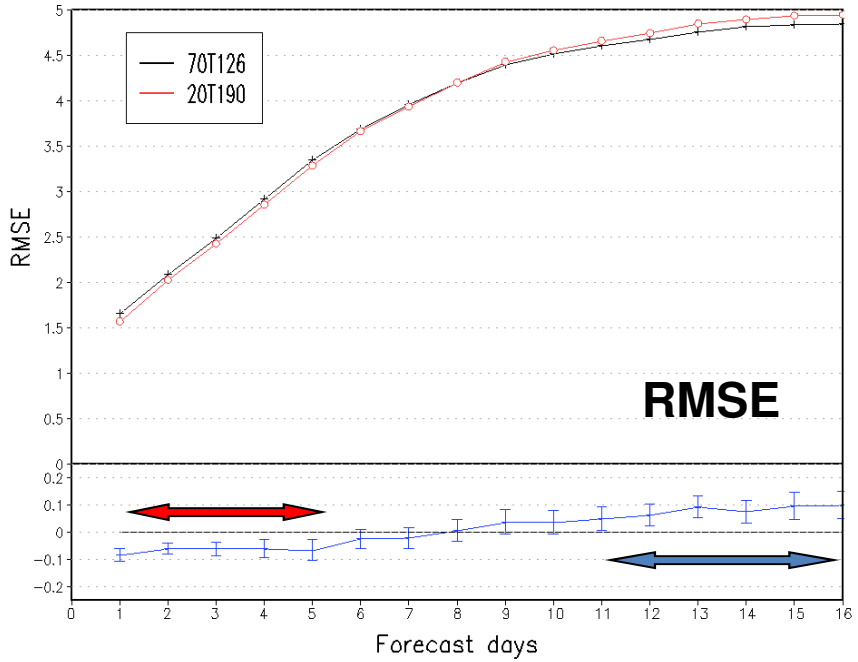


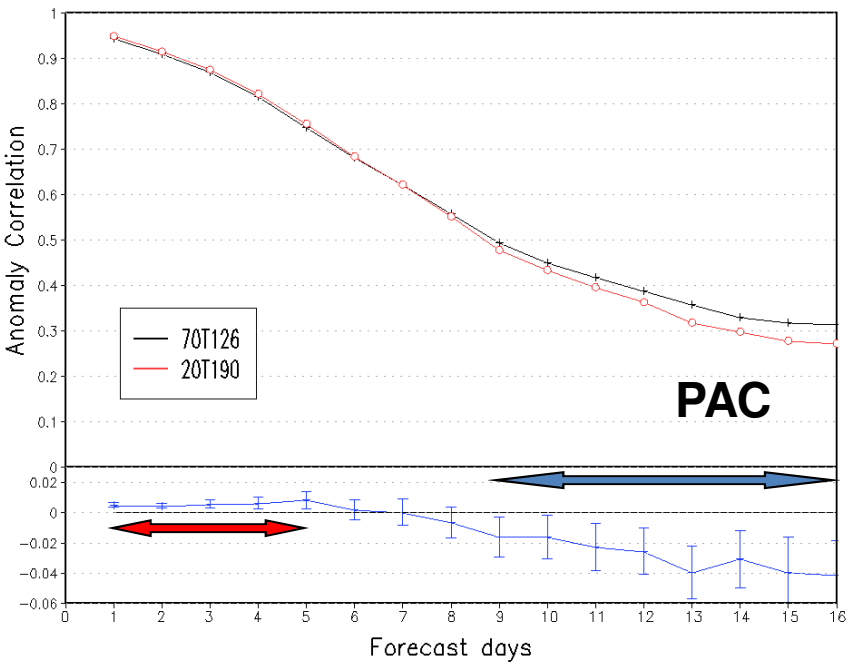


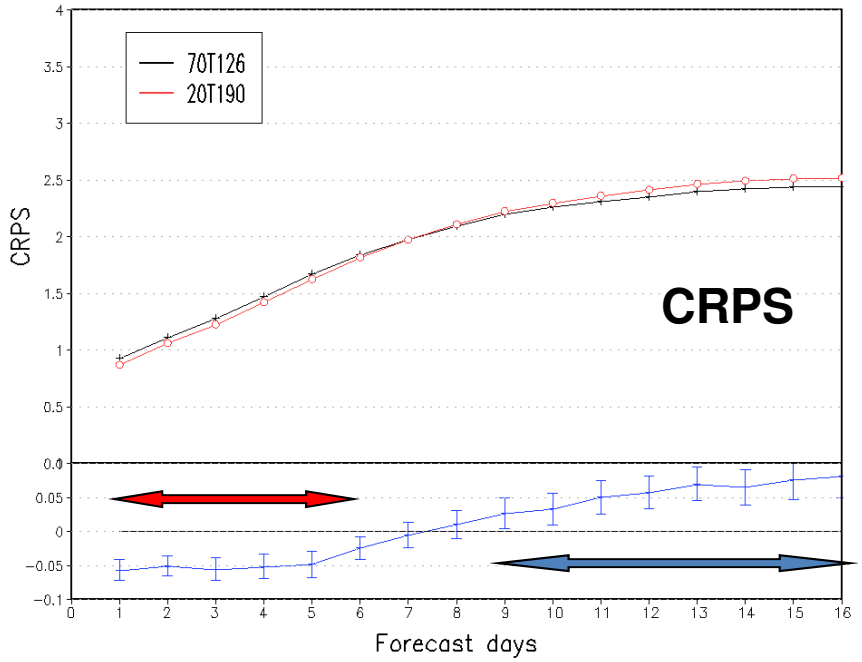


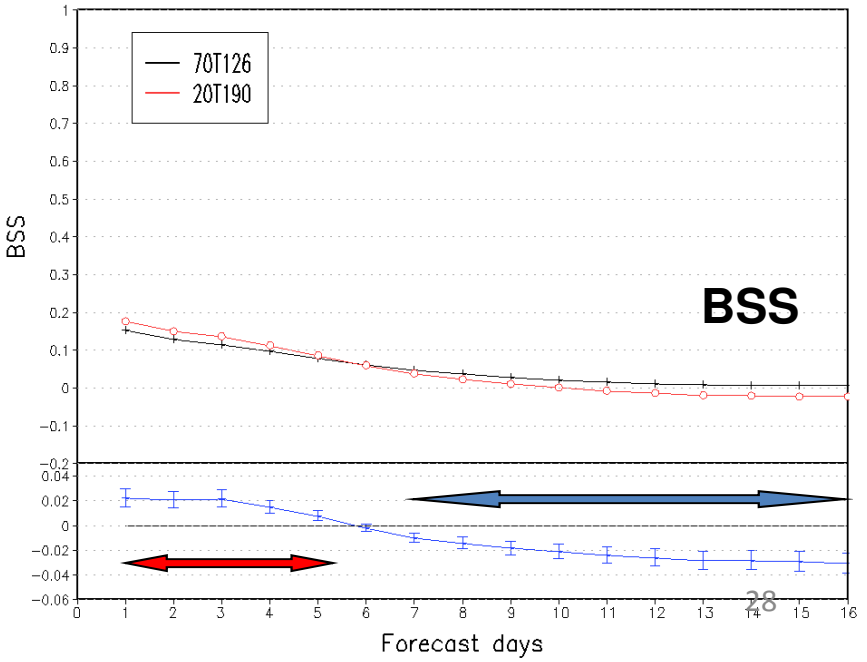


NH 10m Zonal Wind (U)

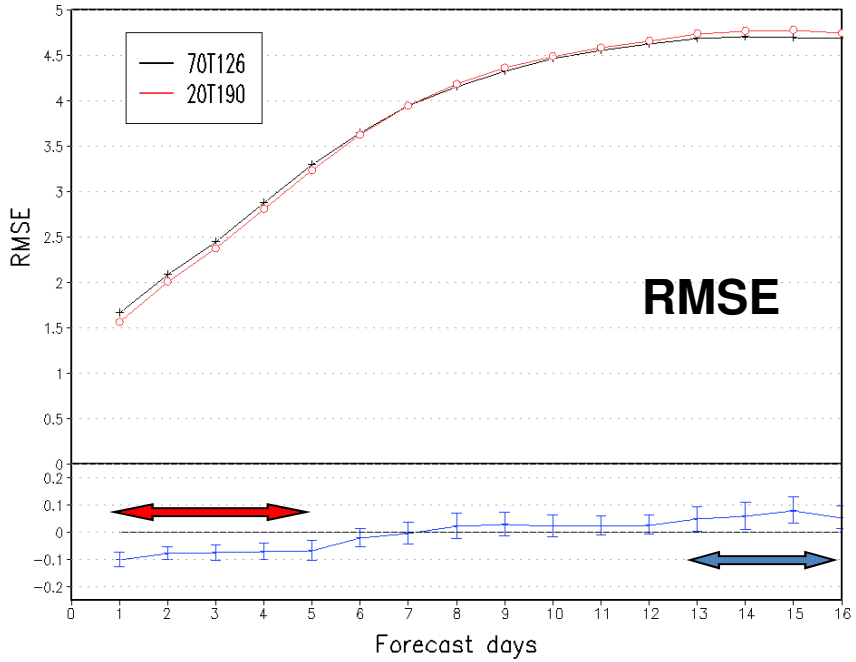


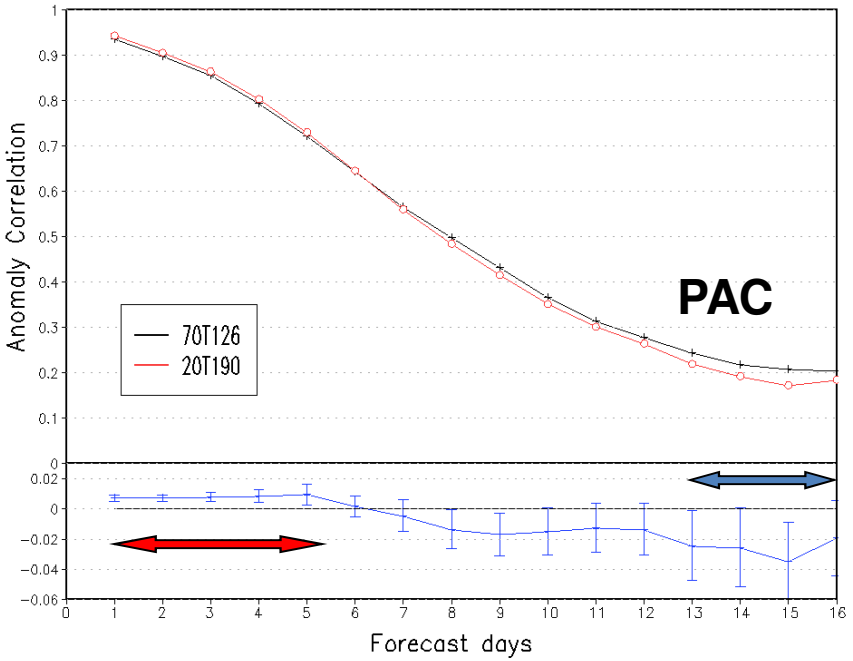


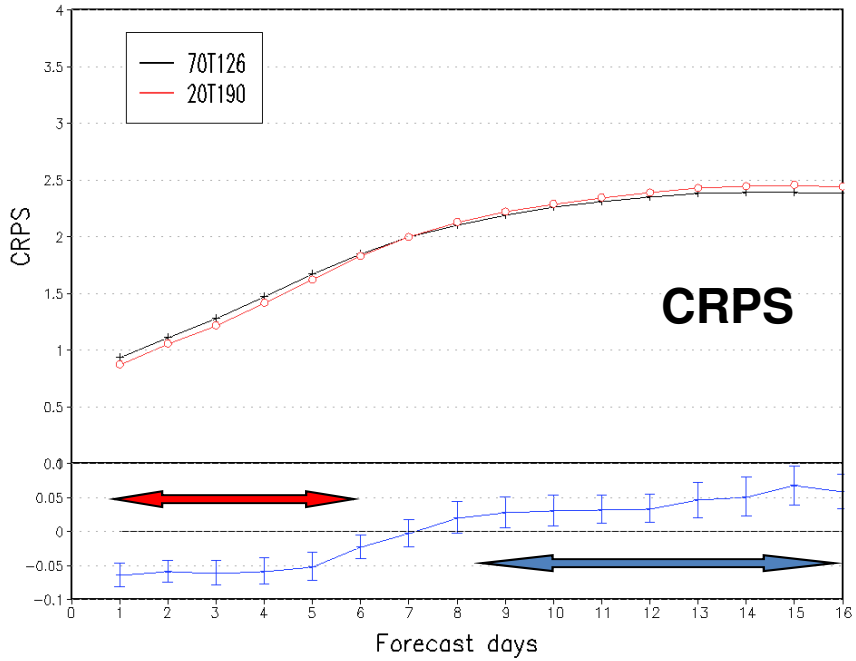


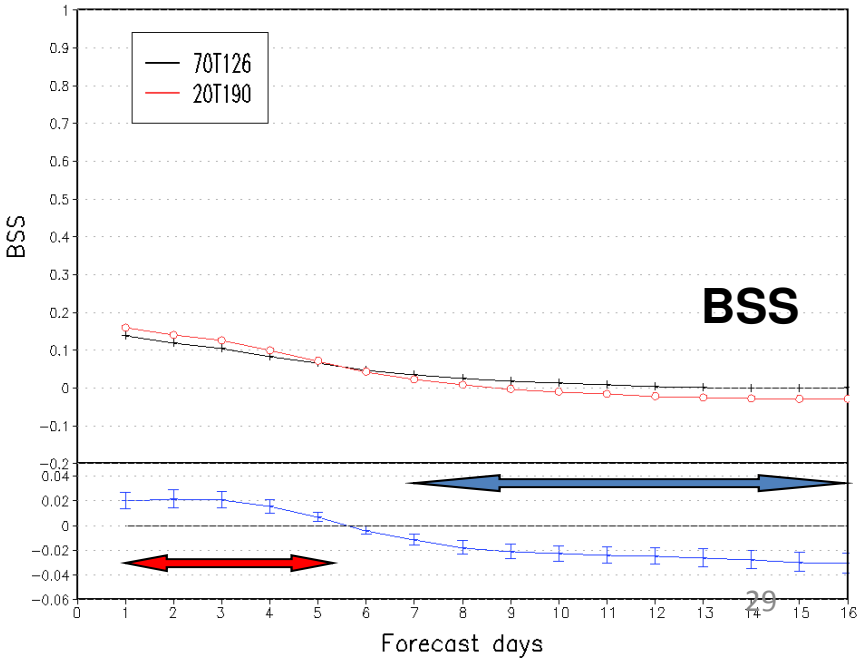


NH 10m Meridional Wind (V)

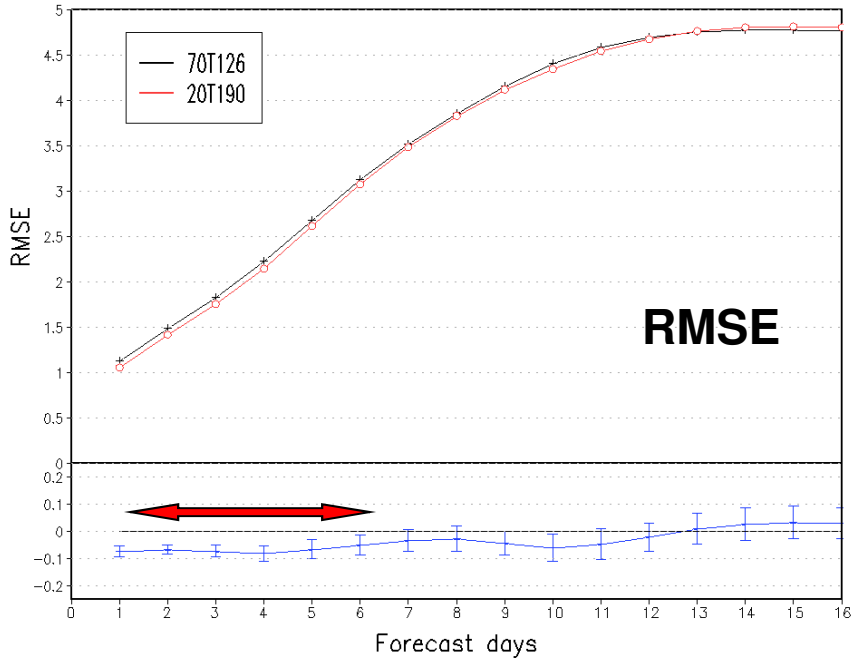


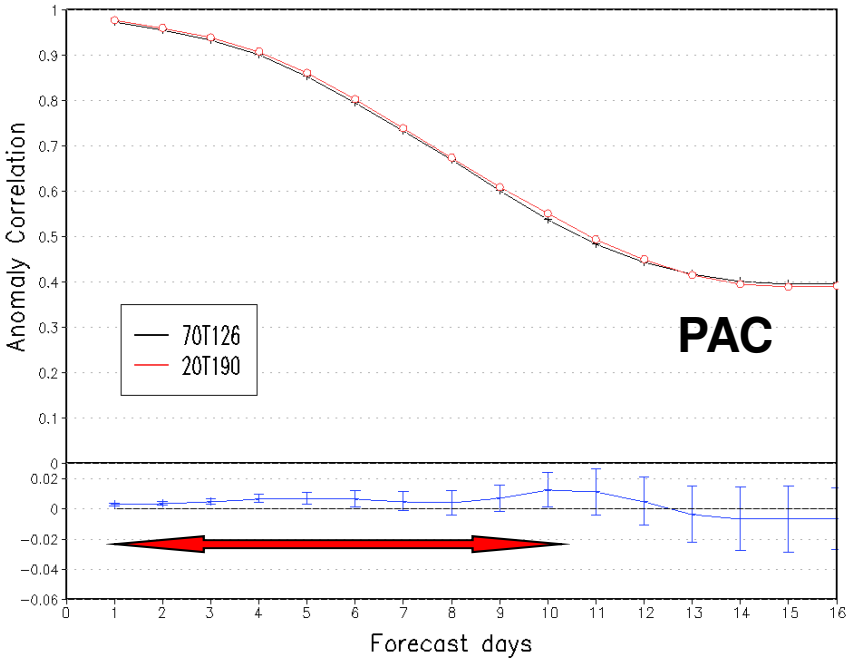


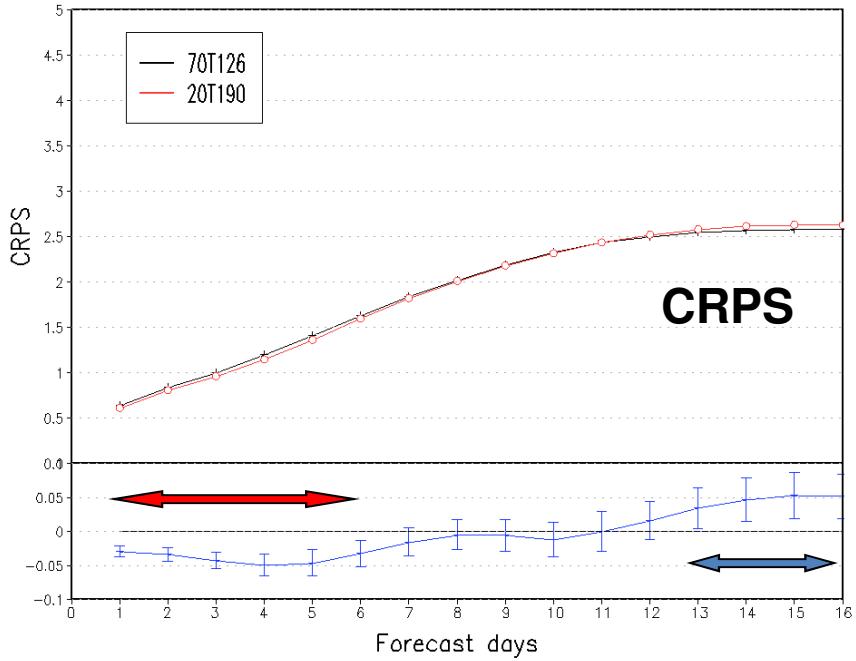


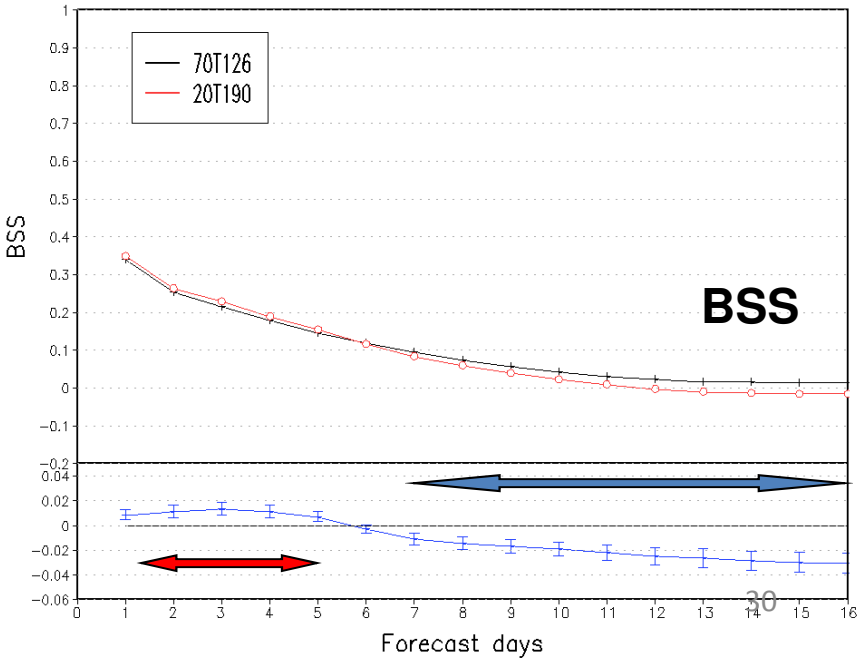


NH 850hPa Temperature

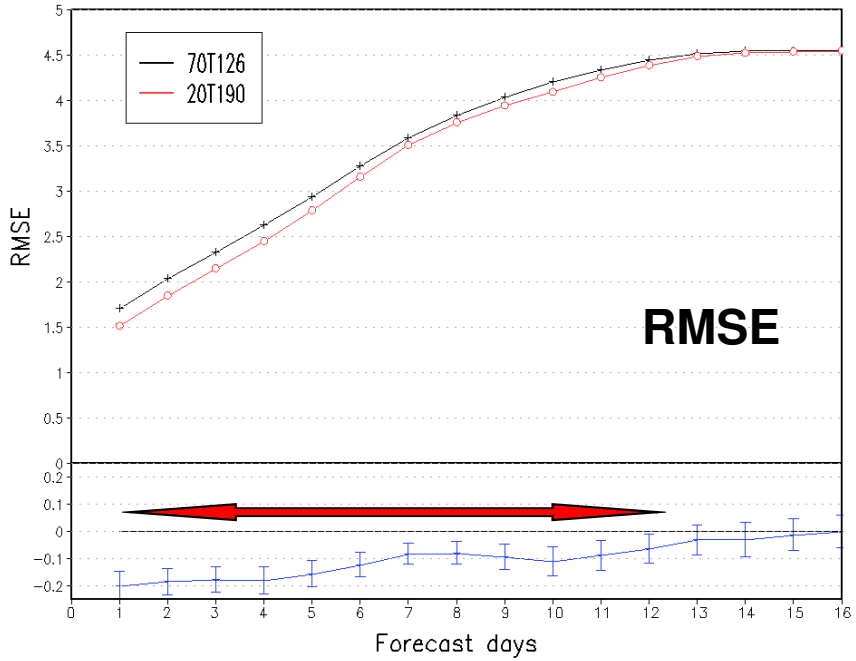


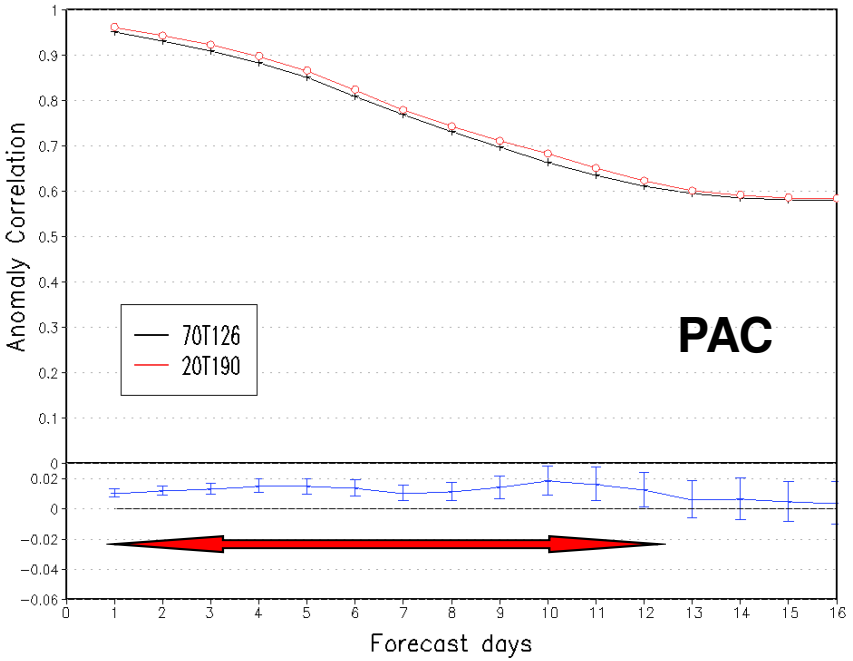


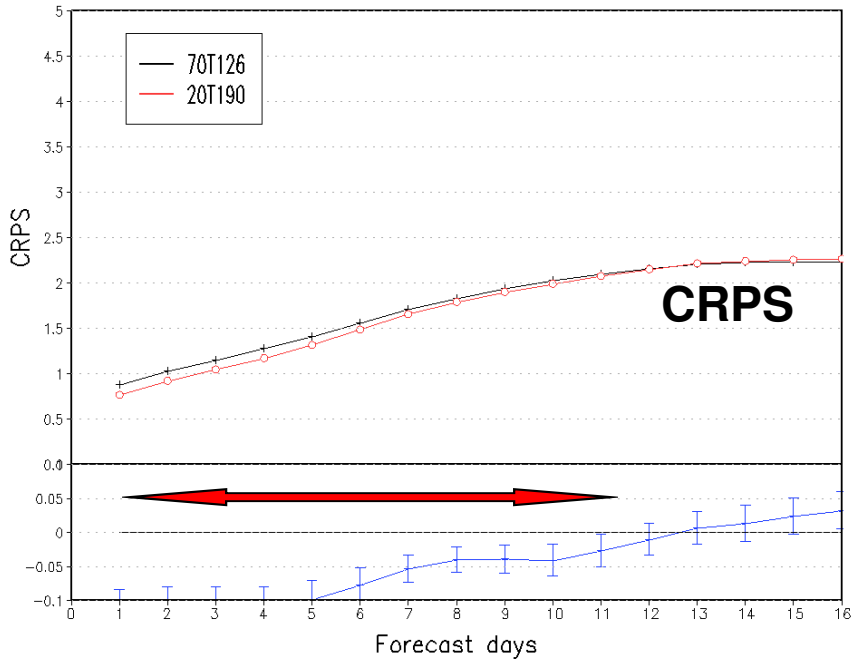


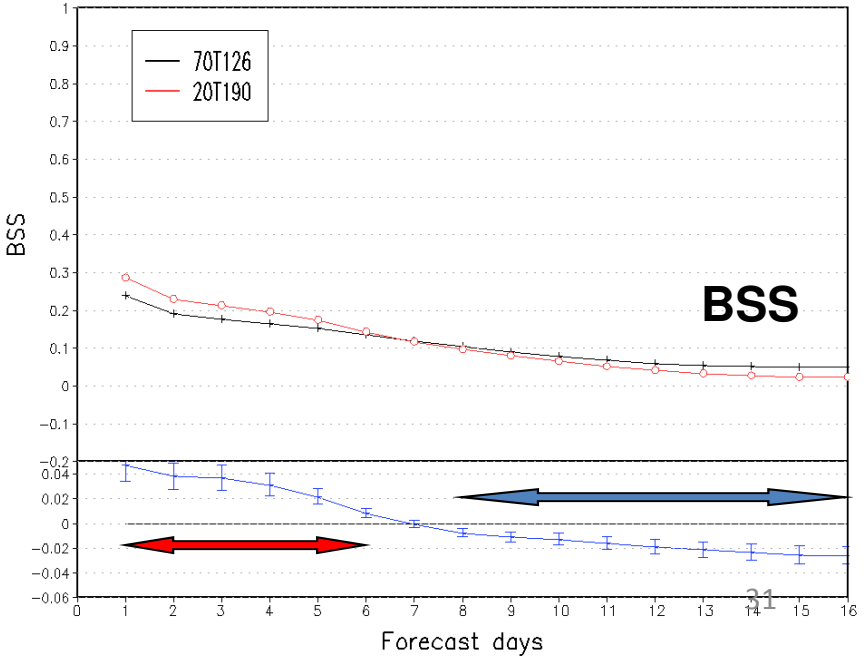


NH 2m Temperature









Forecast Lead Times (days) of Statistically Significant Better

Variable	Configuration	NH		SH	
		RMSE	CRPS	RMSE	CRPS
Z500	20T190	1-4	2-5	1-3	-
	70T126	13-16	12-16	15-16	14-16
Z1000	20T190	1-5	1-5	1-5	1-4
	70T126	11-16	10-16	14-16	12-16
T850	20T190	1-6	1-6	1-6	1-4
	70T126	-	13-16	-	-
T2m	20T190	1-12	1-11	1-16	1-11
	70T126	-	-	-	-
U10m	20T190	1-5	1-6	1-3	1-3
	70T126	11-16	9-16	12-16	6-16
V10m	20T190	1-5	1-6	1-3	1-3
	70T126	13-16	9-16	13-16	6-16

5. Summary and Reference

- Increasing ensemble size is beneficial to improve skill of ensemble mean for small ensemble members (especially less than 40-member). The skill of probabilistic forecast will be significantly improved with further increasing ensemble members.

- Increasing model resolution is more beneficial than increasing ensemble size for short lead times. Increasing ensemble size is more important than increasing model resolution for long lead times.

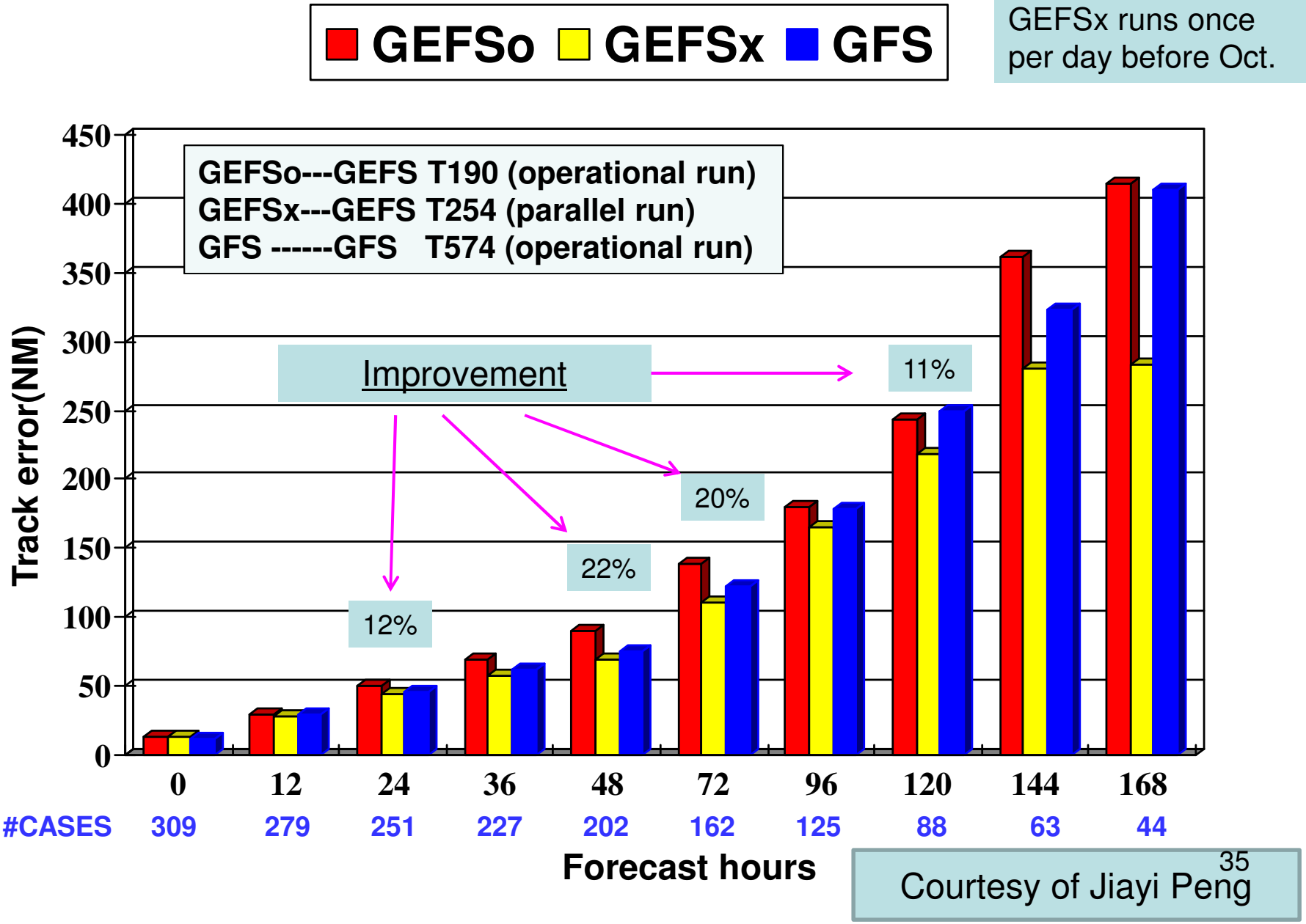
- Reference: Ma, J., Y. Zhu, R. Wobus and P. Wang. 2012: "*A Effective Configuration of Ensemble Size and Horizontal Resolution for NCEP GEFS*". *Adv. Atmos. Sci.,* 29(4), 782-794, doi: 10.1007/s00376-012-1249-y.

NCEP GEFS Configuration

- Latest upgrade - Feb. 14 2012

- Model and initialization
  - Using GFS V9.01 (current operational GFS) instead of GFS V8.00
  - Improved Ensemble Transform with Rescaling (ETR) initialization
  - Improved Stochastic Total Tendency Perturbation (STTP)
- Configurations
  - T254 (55km) horizontal resolution for 0-192 hours
  - T190 (70km) horizontal resolution for 192-384 hours
  - L42 vertical levels for 0-384 hours (from L28)
- Add Sunshine duration for TIGGE data exchange
- Part of products will be delayed by approximately 20 minutes
  - Due to limit CCS resources
  - 40-42 nodes for 70 minutes (start +4:35 end: +5:45)
- Unchanged:
  - 20+1 members per cycle, 4 cycles per day
  - pgrb file output at 1*1 degree every 6 hours
  - GEFS and NAEFS post process output data format
- What do we expect from this implementation?
  - Improve general probabilistic forecast skill overall
  - Significant improvement of tropical storm tracks (especially for Atlantic basin)

Atlantic, AL01~19 (06/01~12/31/2011)



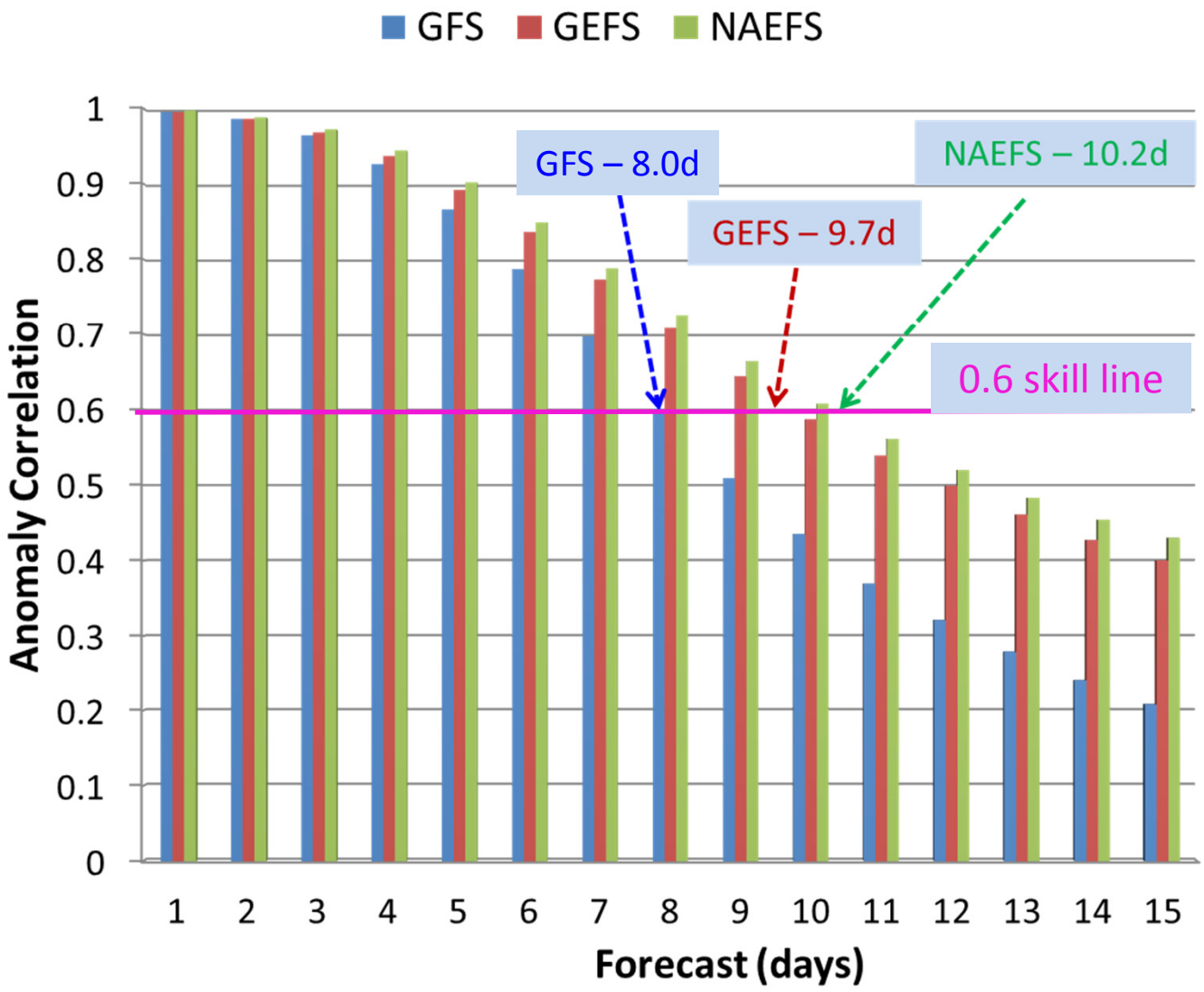
NAEFS/NUOPC Configuration

Updated: February 14 2011

	NCEP	CMC	FNMOC
Model	GFS	GEM	Global Spectrum
Initial uncertainty	ETR	EnKF	(9) Banded ET
Model uncertainty Stochastic	Yes (STTP)	Yes (multi-physics)	None
Tropical storm	Relocation	None	None
Daily frequency	00,06,12 and 18UTC	00 and 12UTC	00 and 12UTC
Resolution	T254L42 (d0-d8)~55km T190L42 (d8-16)~70km	L40 ~ 66km	T159L42 ~ 80km
Control	Yes	Yes	No
Ensemble members	20 for each cycle	20 for each cycle	20 for each cycle Totally 60 perturbed ensemble forecasts in NCEP operation
Forecast length	16 days (384 hours)	16 days (384 hours)	16 days (384 hours)
Post-process	Bias correction for ensemble mean	Bias correction for each member	Bias correction for member mean
Last implementation	February 14 2012	August 17th 2011	September 14 2011

NH Anomaly Correlation for 500hPa Height

Period: January 1st – December 31st 2010



Benefit for forecast:
1.Ensemble mean will extend 1.7 days forecast ability
2.NAEFS will add additional 0.5 day forecast skill
3.Post process will add another additional

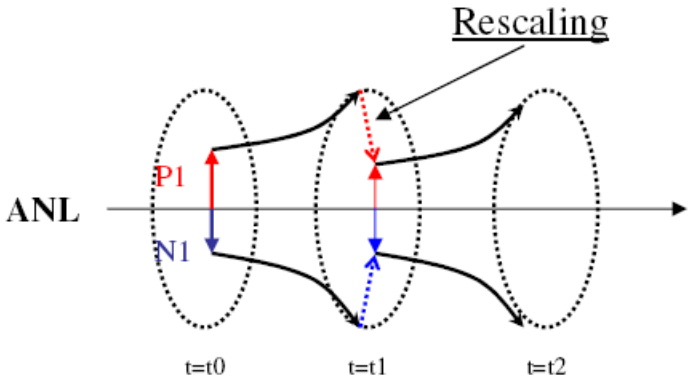
Thank you!

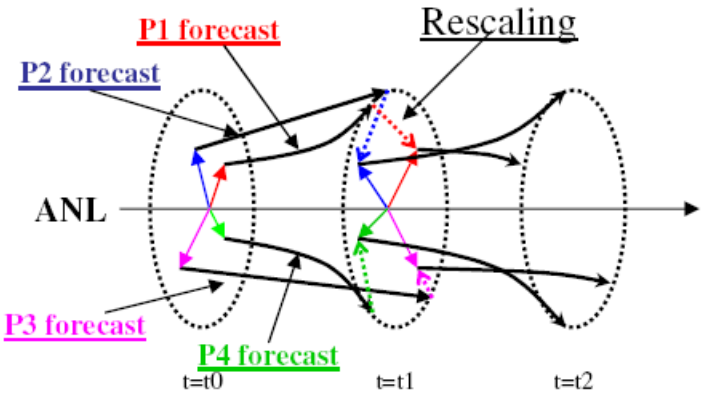
Background !!!

Bred Vector (←2006)

Ensemble Transform with Rescaling (2006 →)

2006

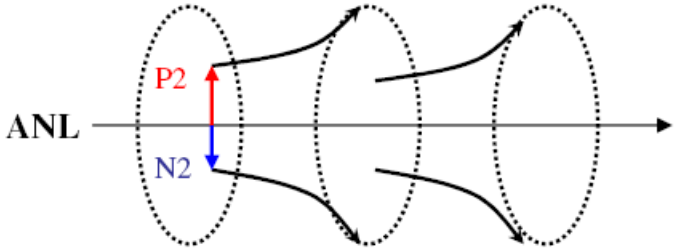




P#, N# are the pairs of positive and negative

P1 and P2 are independent vectors

Simple scaling down (no direction change)



P1, P2, P3, P4 are orthogonal vectors

No pairs any more

To centralize all perturbed vectors (sum of all vectors are equal to zero)

Scaling down by applying mask,

The direction of vectors will be tuned by ET.

References:

1. Wei and et al: 2006 Tellus
2. Wei and et al: 2008 Tellus

NH Anomaly Correlation for 500hPa Height

Period: September 1st – November 30th 2011