EMC Probabilistic Forecast Verification for Sub-season Scales

Yuejian Zhu
Environmental Modeling Center
NCEP/NWS/NOAA

Acknowledgement: Wei Li, Hong Guan and Eric Sinsky

Present for the DTC Test Plan and Metrics Workshop
July 30 – August 1st 2018
College Park, MD

Background

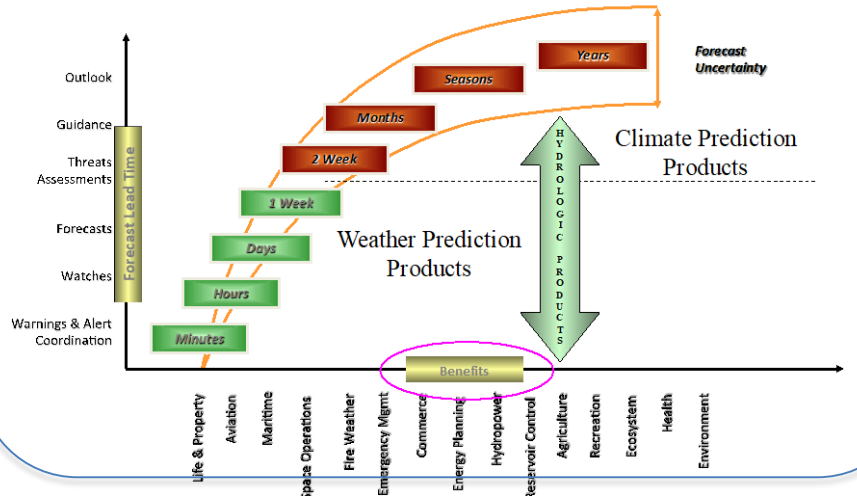
The development of probabilis3c forecast vea ific 3on across the scales

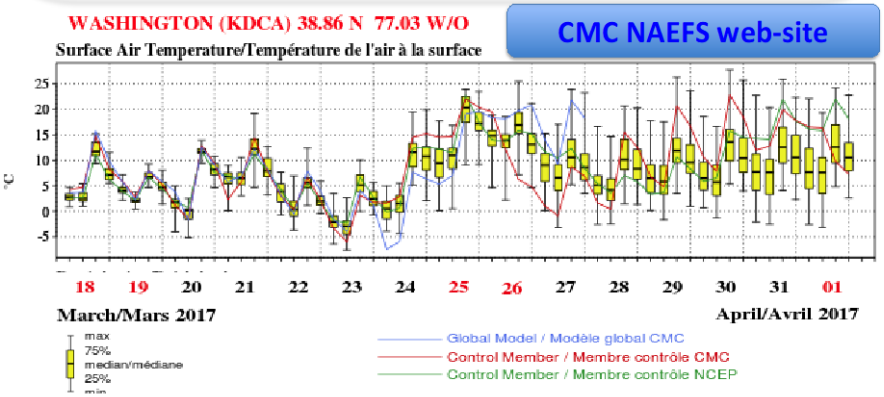
Yuejian Zhu
Chris Melhauser, Wei Li, Eric Sinsky and Yan Luo

Environmental Modeling Center
NCEP/NWS/NOAA

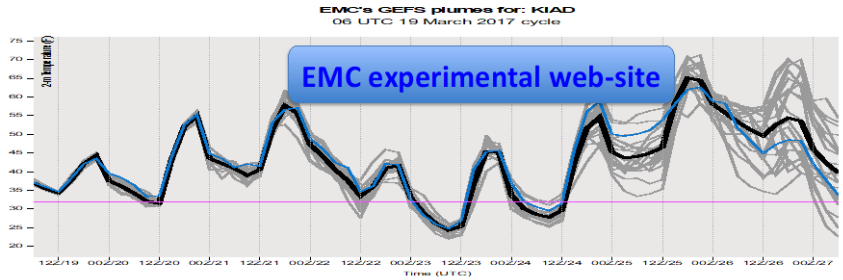
Present for 7th InternaF onal VerificaF on Method Workshop
May 8-11 2017
Berlin, Germany

Understand our forecasts across 3me scale



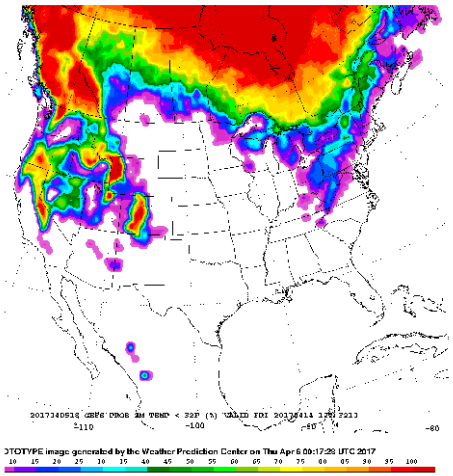


CMC NAEFS web-site

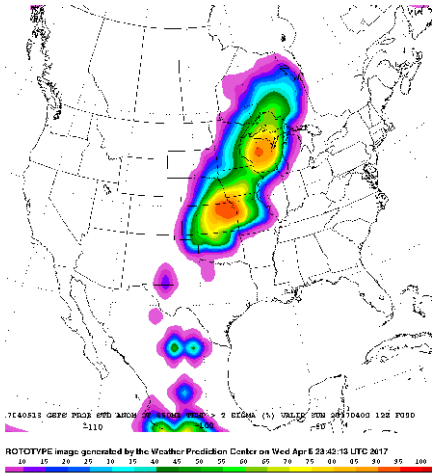


EMC experimental web-site

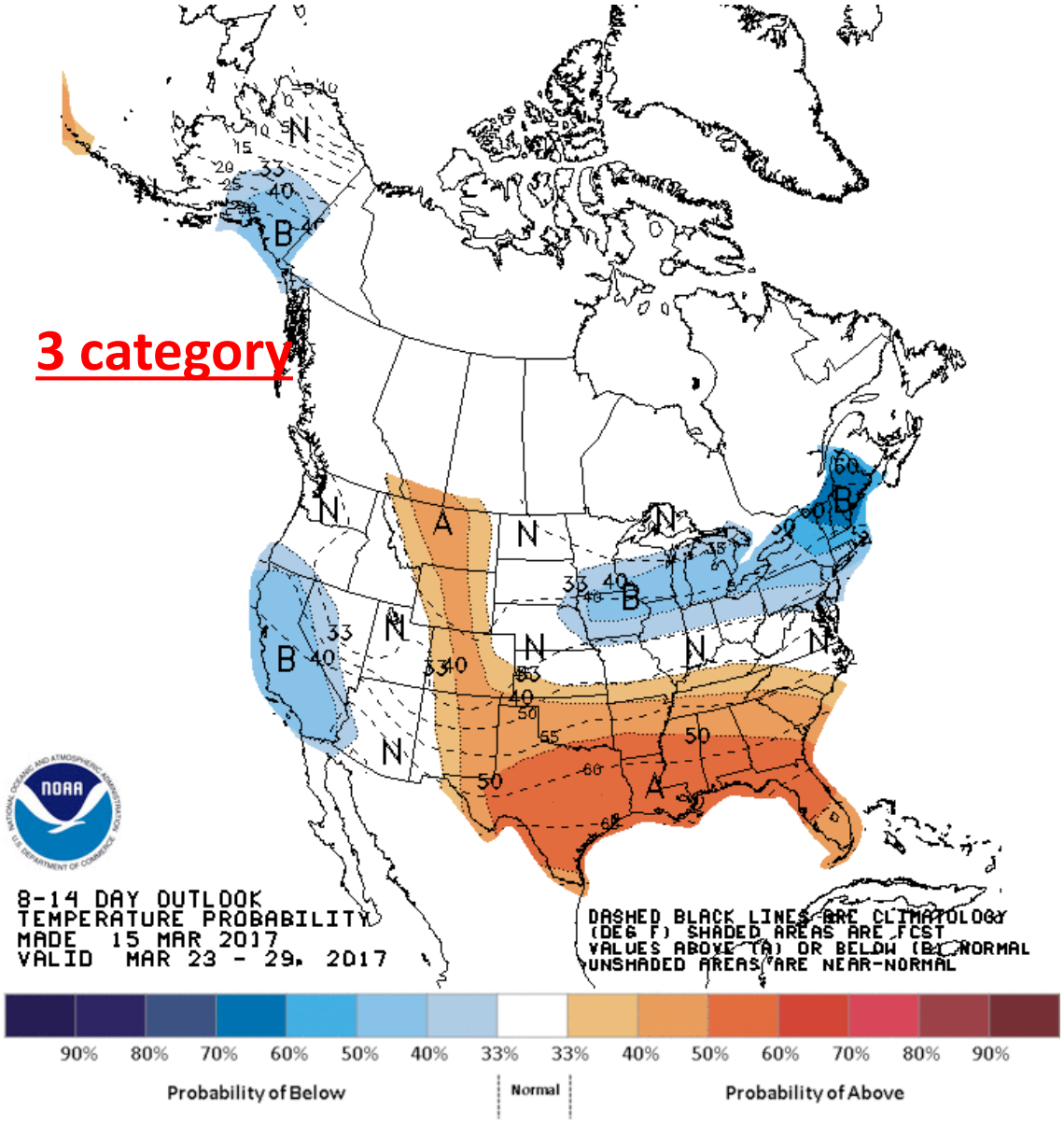
Probability (%) 2m temperature < 32° F for (f210)



Anomaly forecast of 850 temp. > 2 sigma (f90)

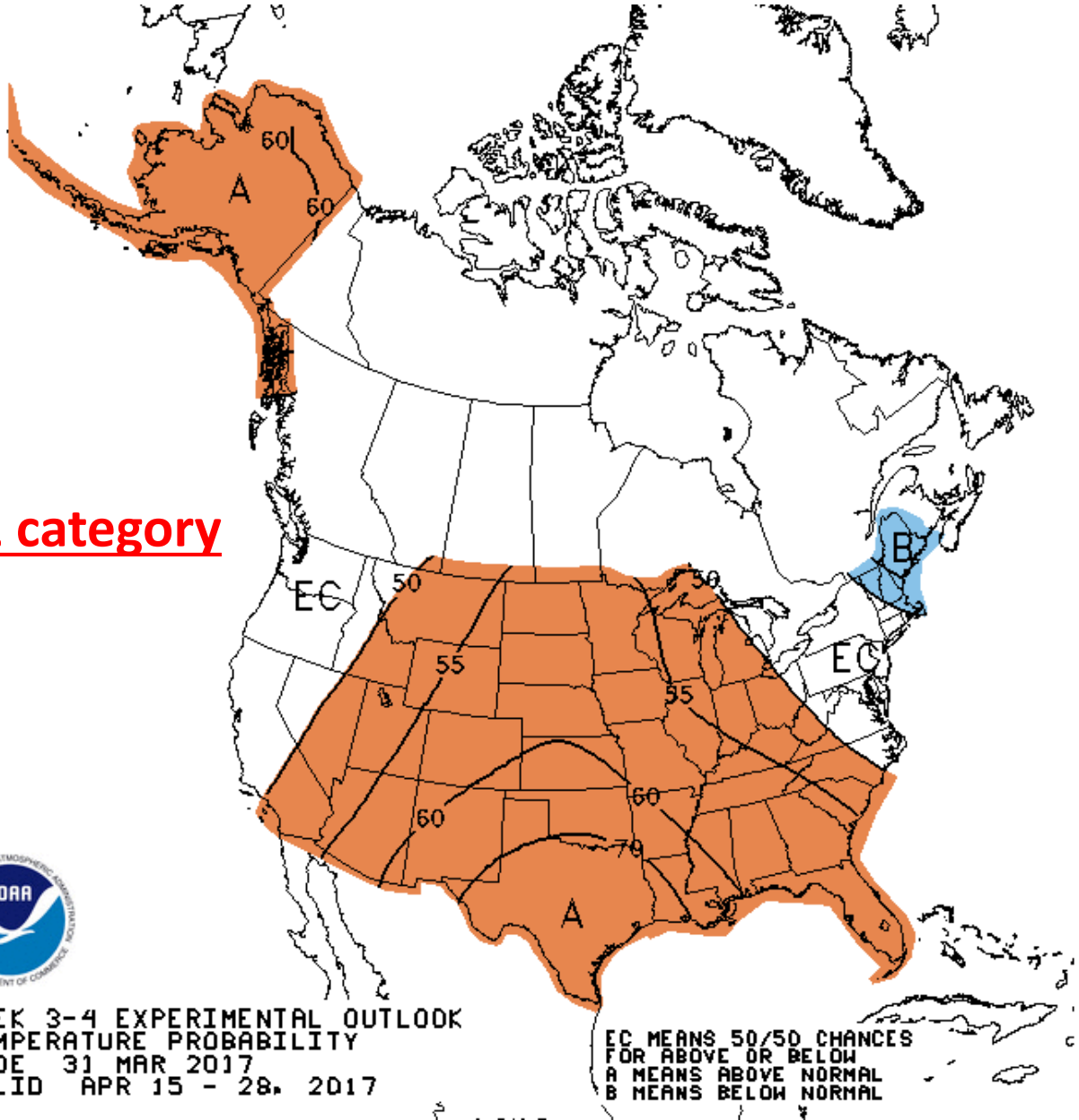


CPC's daily 8-14 day (week-2) outlook



The 8-14 day Outlook gives the confidence that a forecaster has, given as a probability, that the observed temperature, averaged over upcoming days 8, 9, 10, 11, 12, 13, and 14 will be in the range of one of three possible categories below (B), normal (N), or above (A). For any calendar 7-day period, these categories can be defined by separating the 30 years of the climatology period, 1981-2010 (30 years), into the coldest 10 years, the middle 10 years, and the warmest 10 years. Because each of these categories occurs 1/3 of the time (10 times) during 1981-2010, for any particular calendar 7-day period, the probability of any category being selected at random from the 1981-2010 set of 30 observations is one in three (1/3), or 33.33%. This is also called the climatological probability. The sum of the climatological probabilities of the three categories is 100%.

CPC's week 3&4 temperature forecast



2 category

These are experimental two category outlooks and differ from official operational current three category outlooks currently used for the monthly and seasonal forecasts.

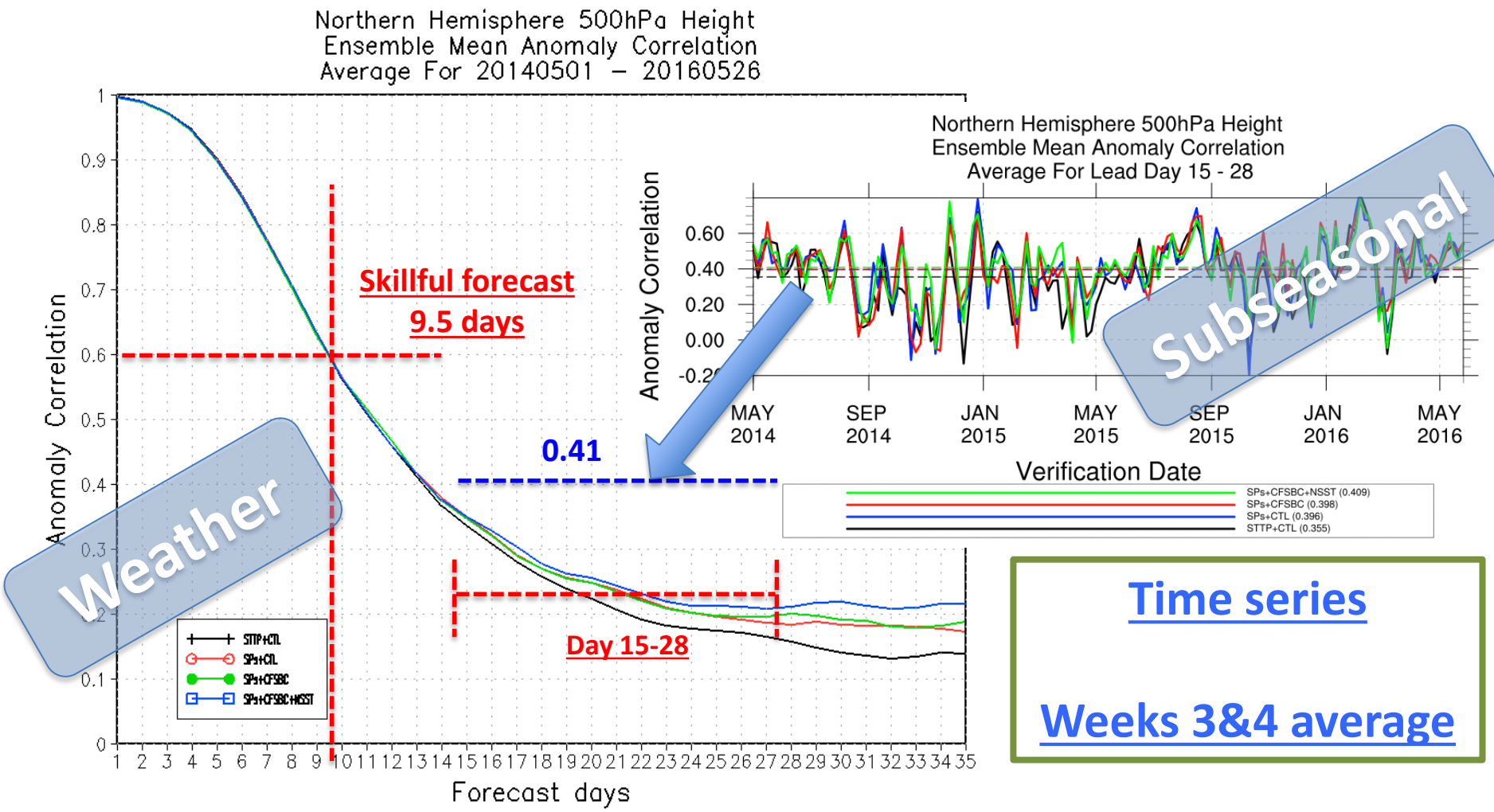
The shading on the temperature map depicts the most favored category, either above-normal (A) or below-normal (B) with the solid lines giving the probability (>50%) of this more likely category (above or below).

Probabilistic Verification Tools

- Ranking by forecasts
  - Histograms
  - Continues Ranked Probability Score (CRPS)
- Ranking by climatology
  - Brier scores – multi-categories and ranking
    - Decomposition: reliability and resolution
  - Ranked Probability Score (RPS)
  - Others
- Specified value (threshold)
  - Brier scores and decomposition
  - Others
- Two/three references
  - Proxy truth
    - Observation or best analysis?
      - Hard to have climatology at observation station
  - Climatology
    - Reanalysis
  - Specified value (or threshold) for extreme events
    - Hard to explore this for general ensemble application

Example one: daily score .vs 2-week average score

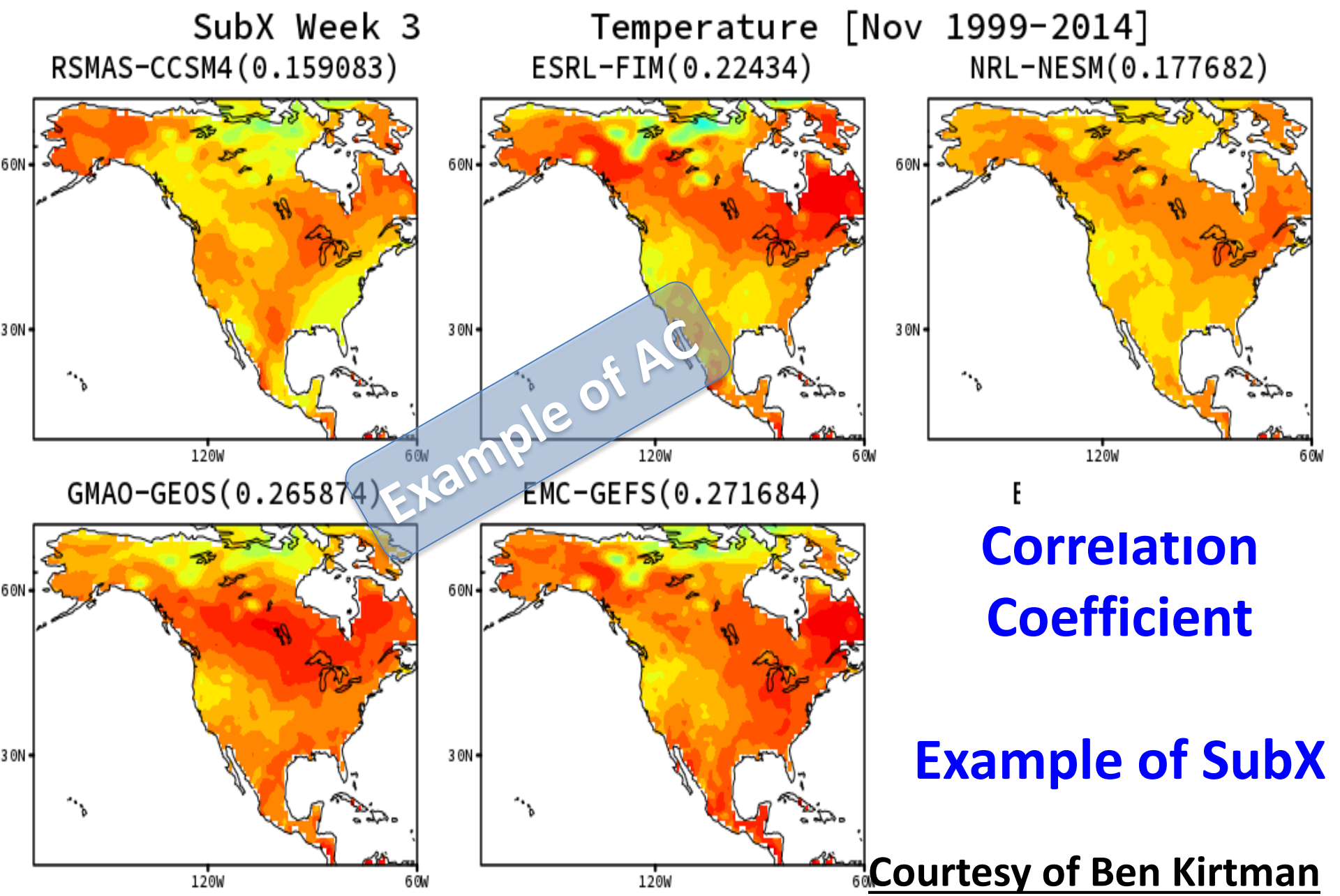
(deterministic – PAC or RMS error)



Time series

Weeks 3&4 average

Needs to convert daily forecast to period average

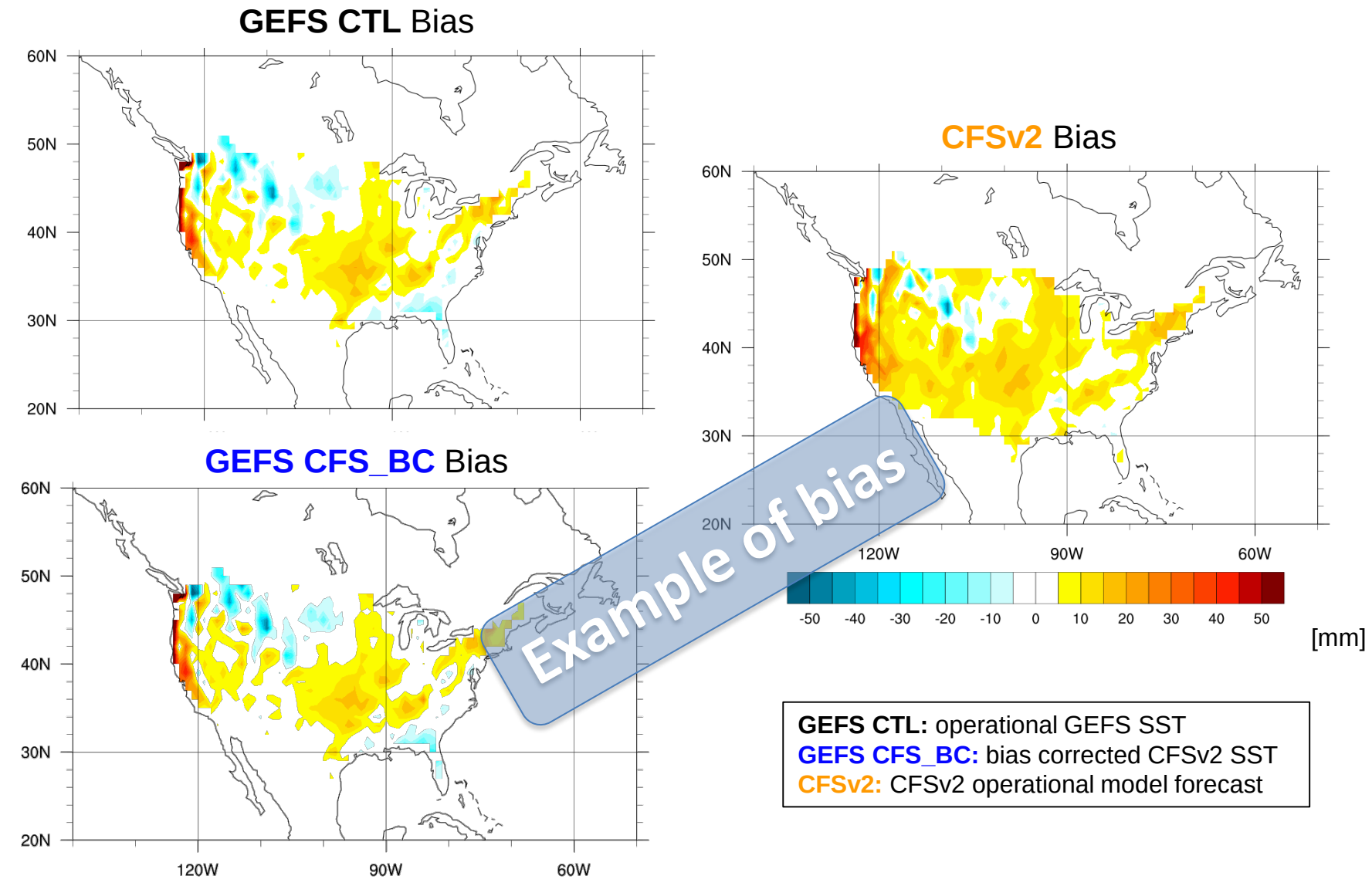


Correlation Coefficient

Example of SubX

Courtesy of Ben Kirtman

Week 3&4 Accumulated Precipitation



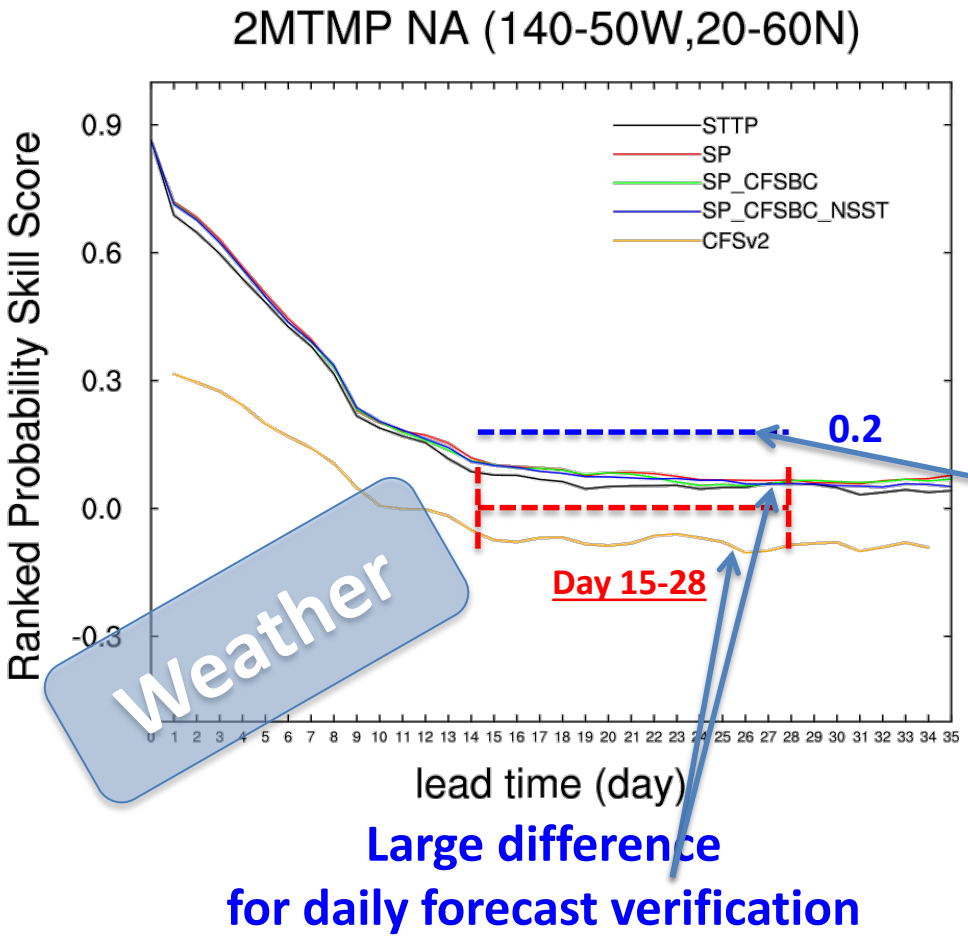
GEFS CTL: operational GEFS SST
GEFS CFS_BC: bias corrected CFSv2 SST
CFSv2: CFSv2 operational model forecast

Spatial (geographical) distribution is used for evaluation

Example two: daily score .vs 2-week average score

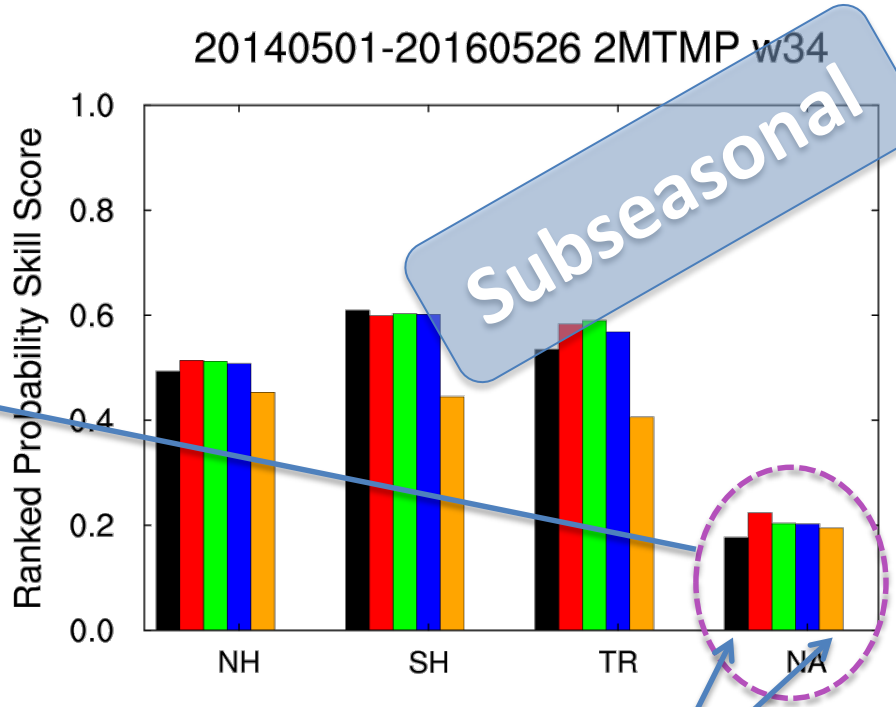
(Probabilistic – RPSS or CRPS)

Domain: Northern American, land only



Weather

Large difference
for daily forecast verification



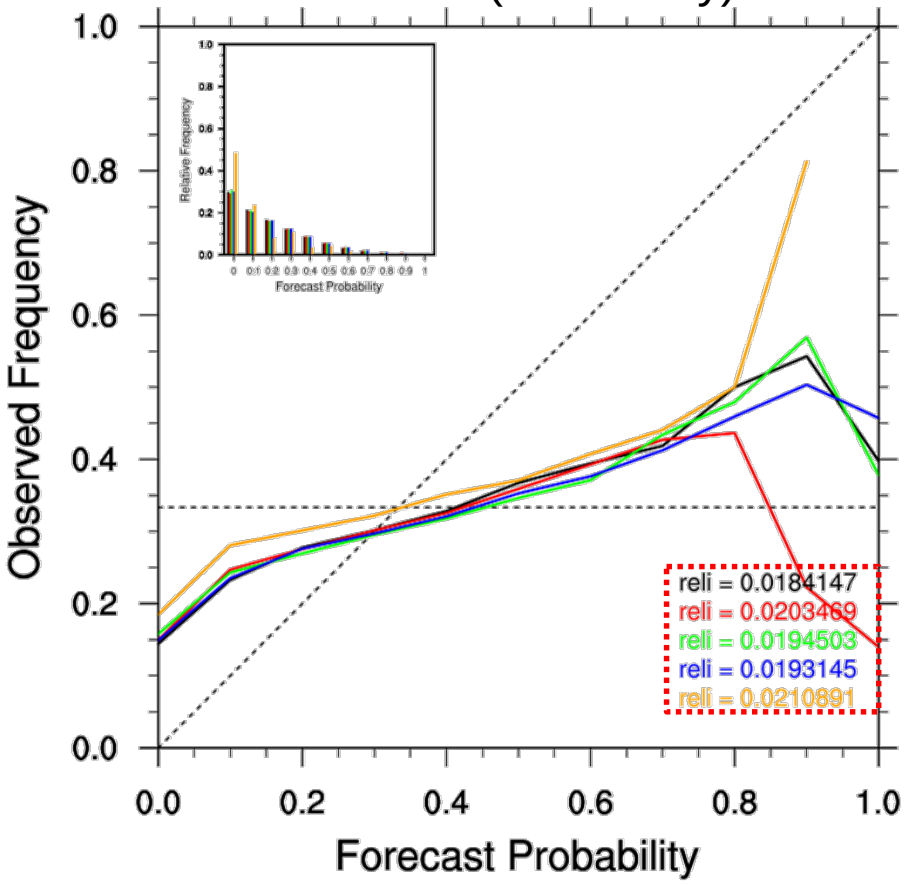
Subseasonal

Small difference
for 2-week average forecast verification

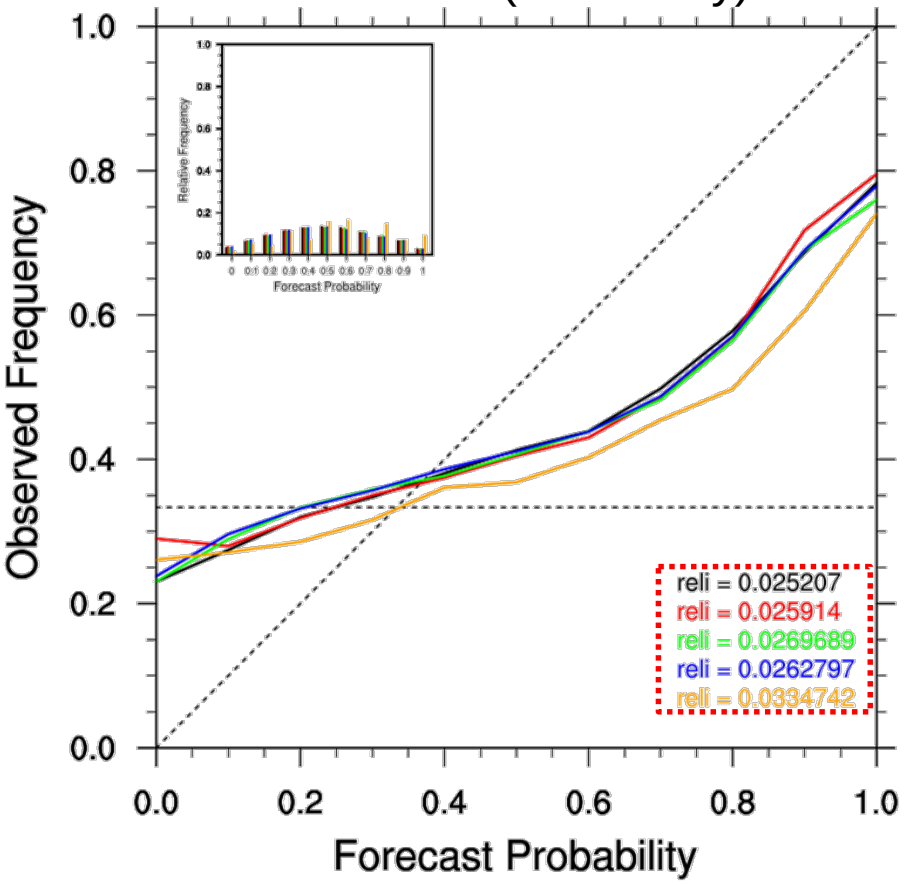
Needs to convert daily forecast to period average

Week 3&4 Accumulated Precipitation Reliability

Lower Tercile
CONUS (Land Only)



Upper Tercile
CONUS (Land Only)

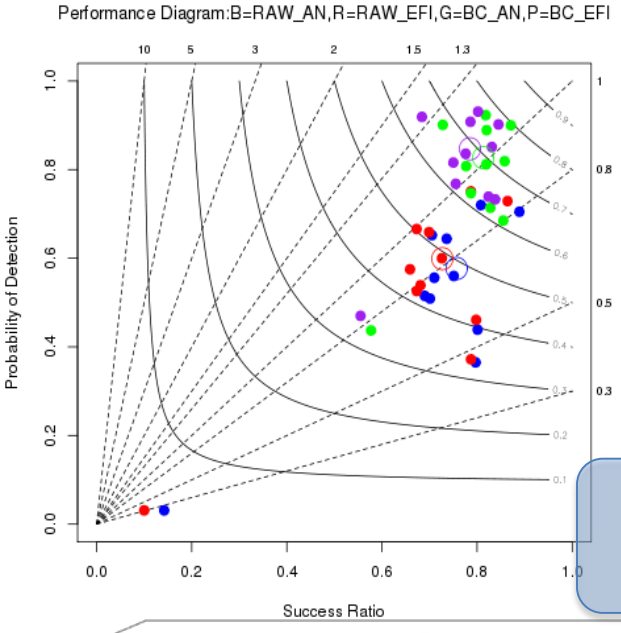


GEFS CTL: operational GEFS SST
GEFS RTG: analysis "optimal" SST
GEFS CFS: raw CFSv2 SST
GEFS CFS_BC: bias corrected CFSv2 SST
CFSv2: CFSv2 operational model forecast

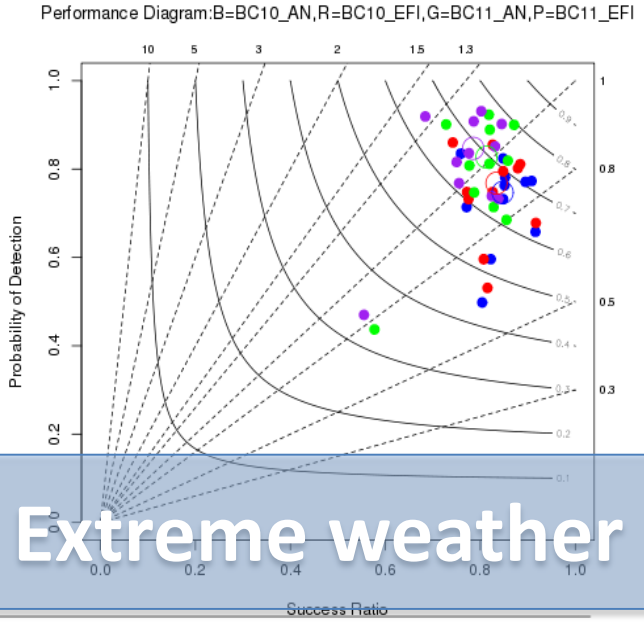
We need to look at reliability, too

Performance Diagram for Extreme Cold Events

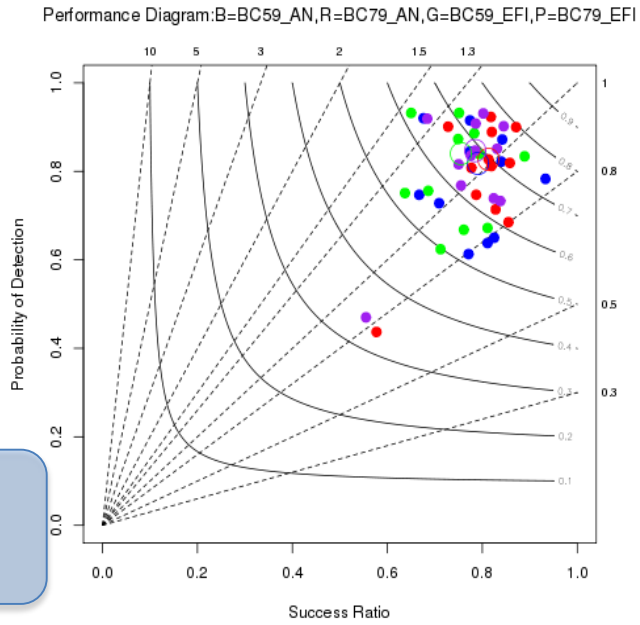
Raw vs. bias-corrected forecasts



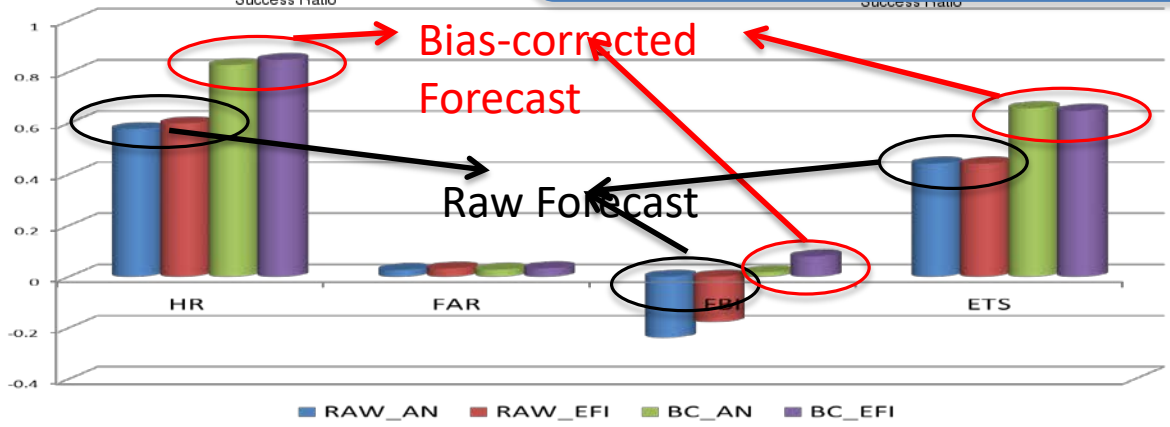
v10 vs. v11 forecasts



Reanalysis vs. CFSR



Extreme weather



Reference - Guan and Zhu, 2017: *"Development of Verification Methodology for Extreme Weather Forecasts"*, Wea. and Forecasting.

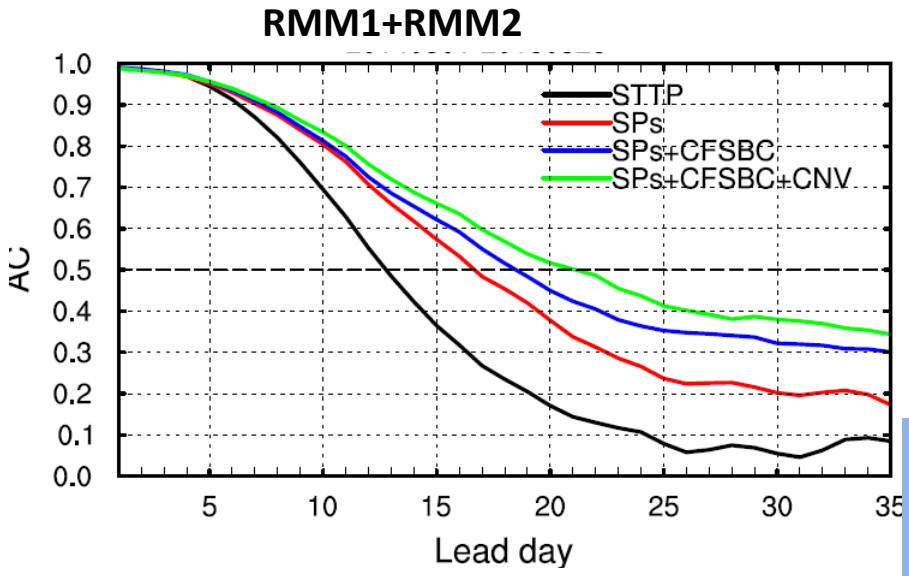
This method could be extended to subseasonal fcst

Statistics for extreme cold weather event (11 cases) for 13-14 winter (Raw and bias-corrected forecast (V11))

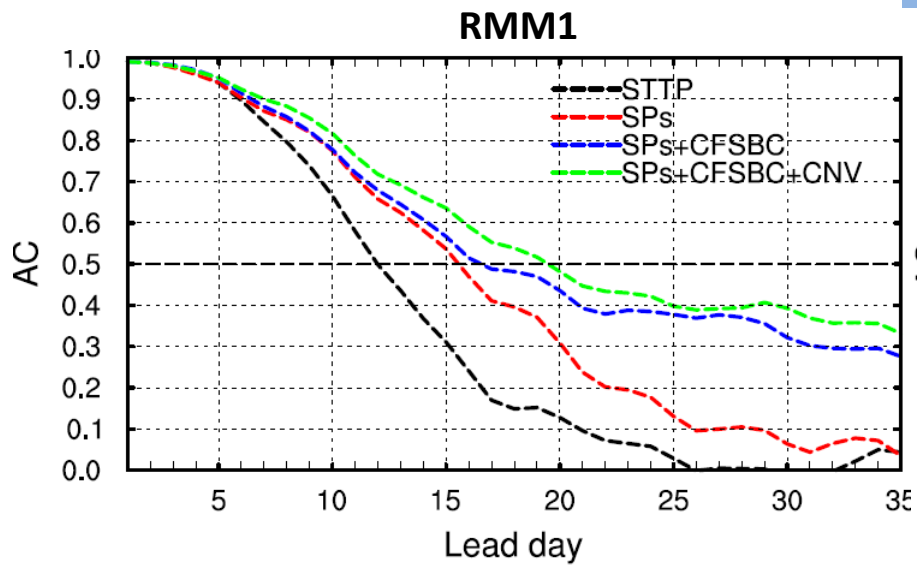
Tropical evaluation and diagnostics

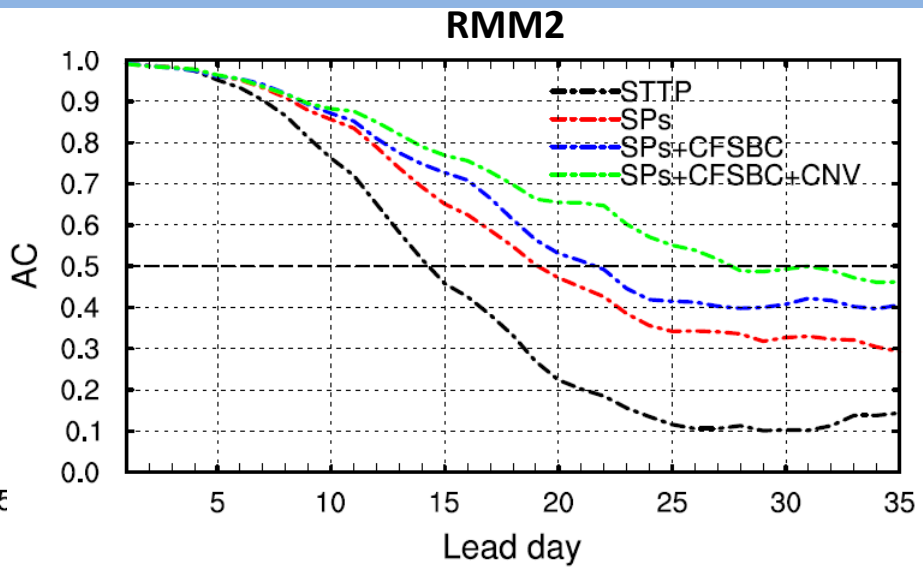
Wei Li

WH-MJO Forecast Skills

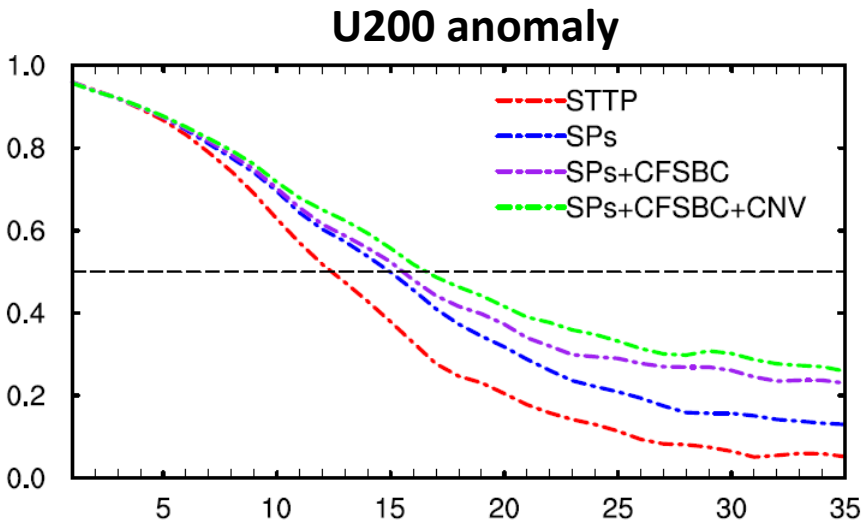


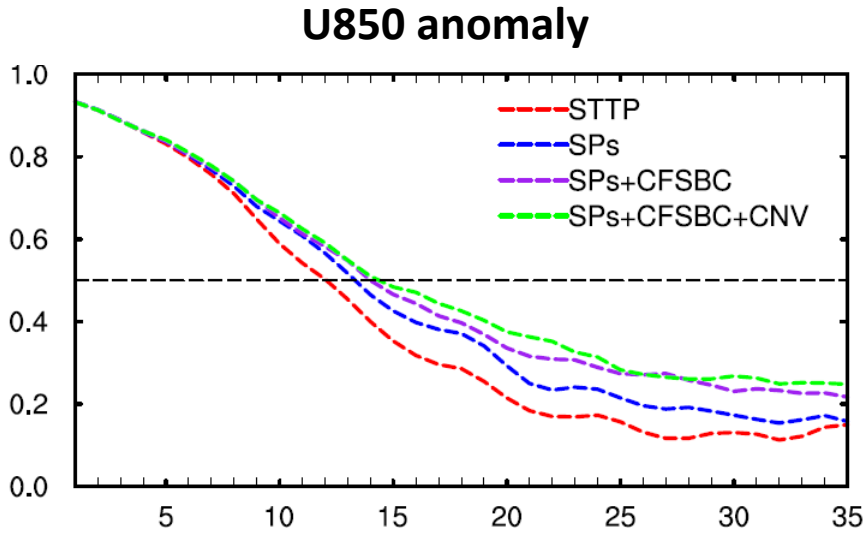
Wheeler Hendon (WH) MJO skill, which is defined as the bivariate anomaly correlation between the analysis and forecast of two principal component time series (Real-time Multivariate MJO - RMM1 and RMM2) from combined empirical orthogonal functions (EOF) of the MJO components using outgoing long wave radiation (OLR) and zonal wind at 200-hPa and 850-hPa respectively, i.e.

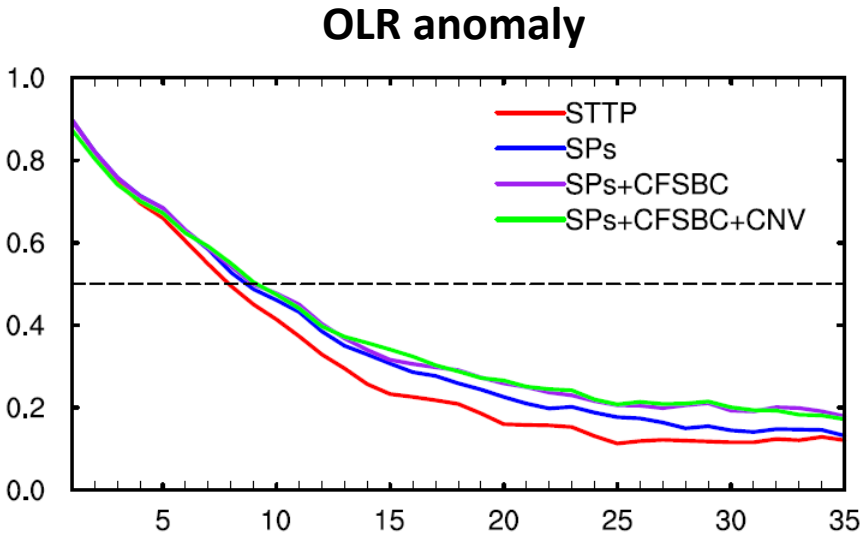




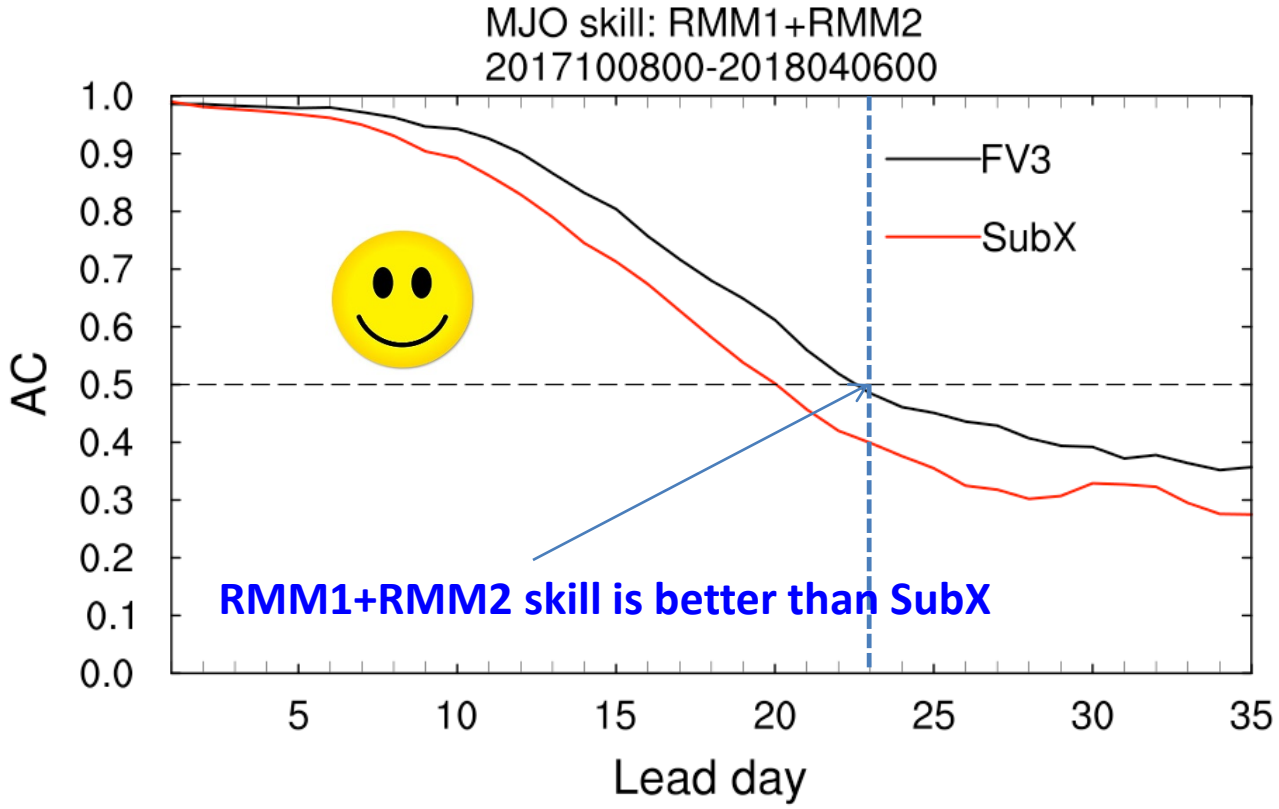
Forecast skill of the key Variables

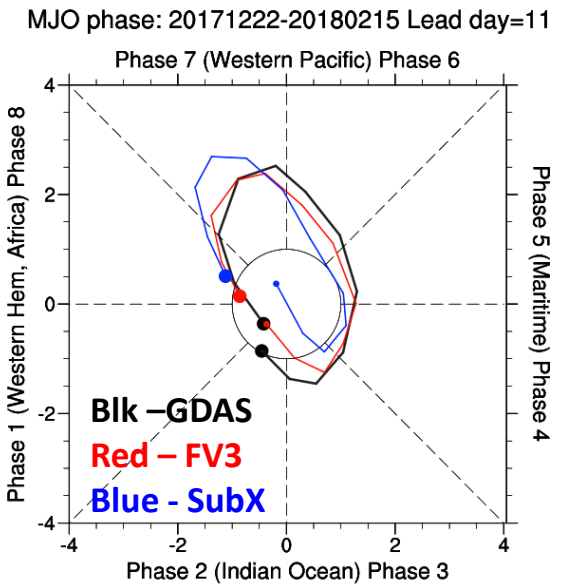




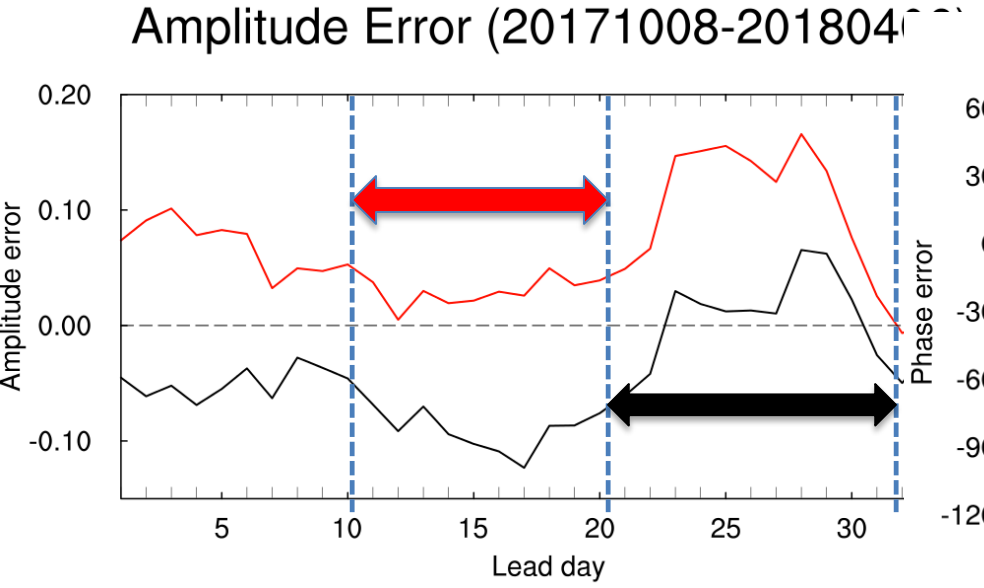


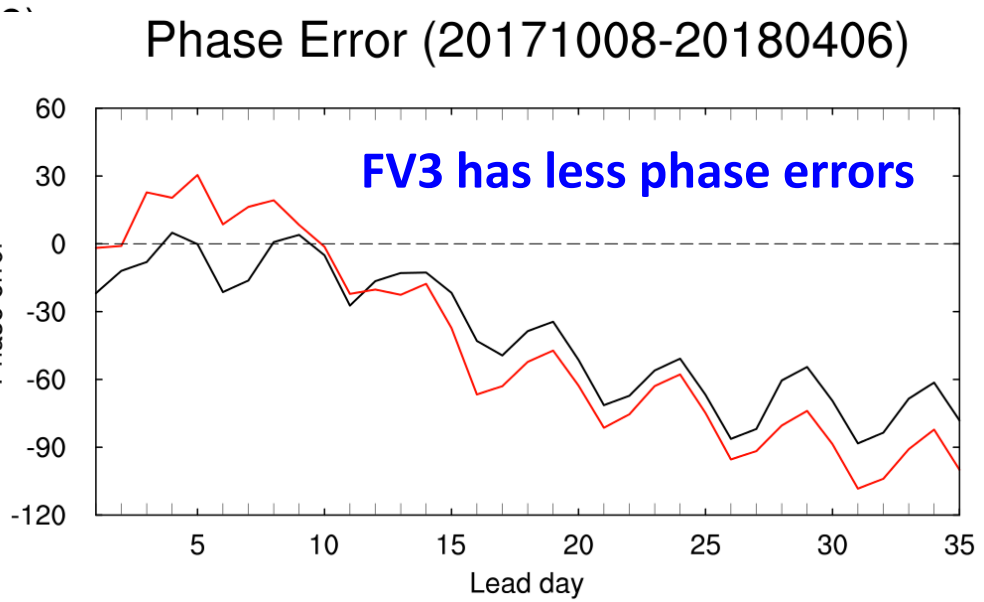
Correlation of time series of forecast and analysis anomaly



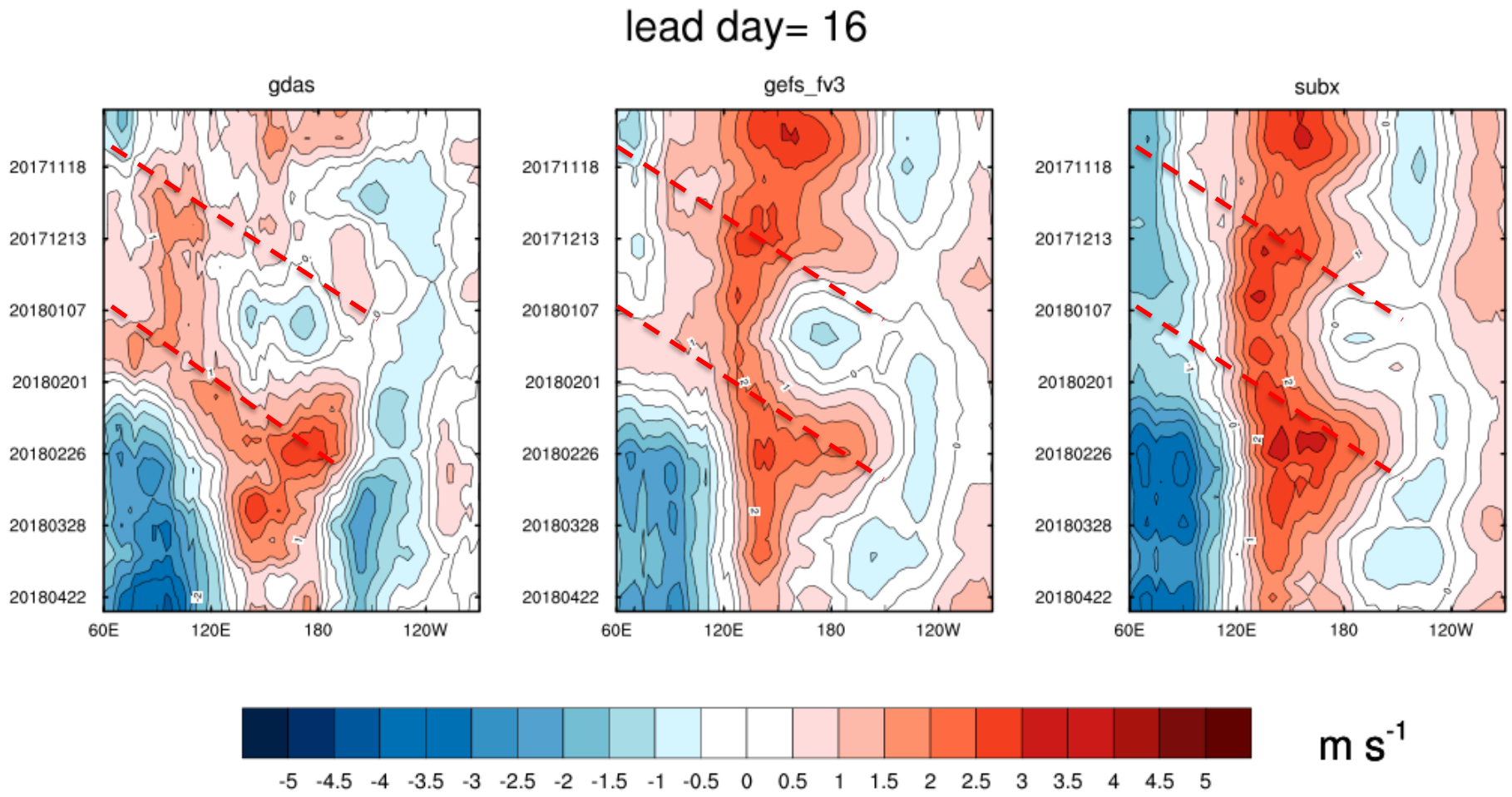


Example of one MJO phase (lead – 11 days)



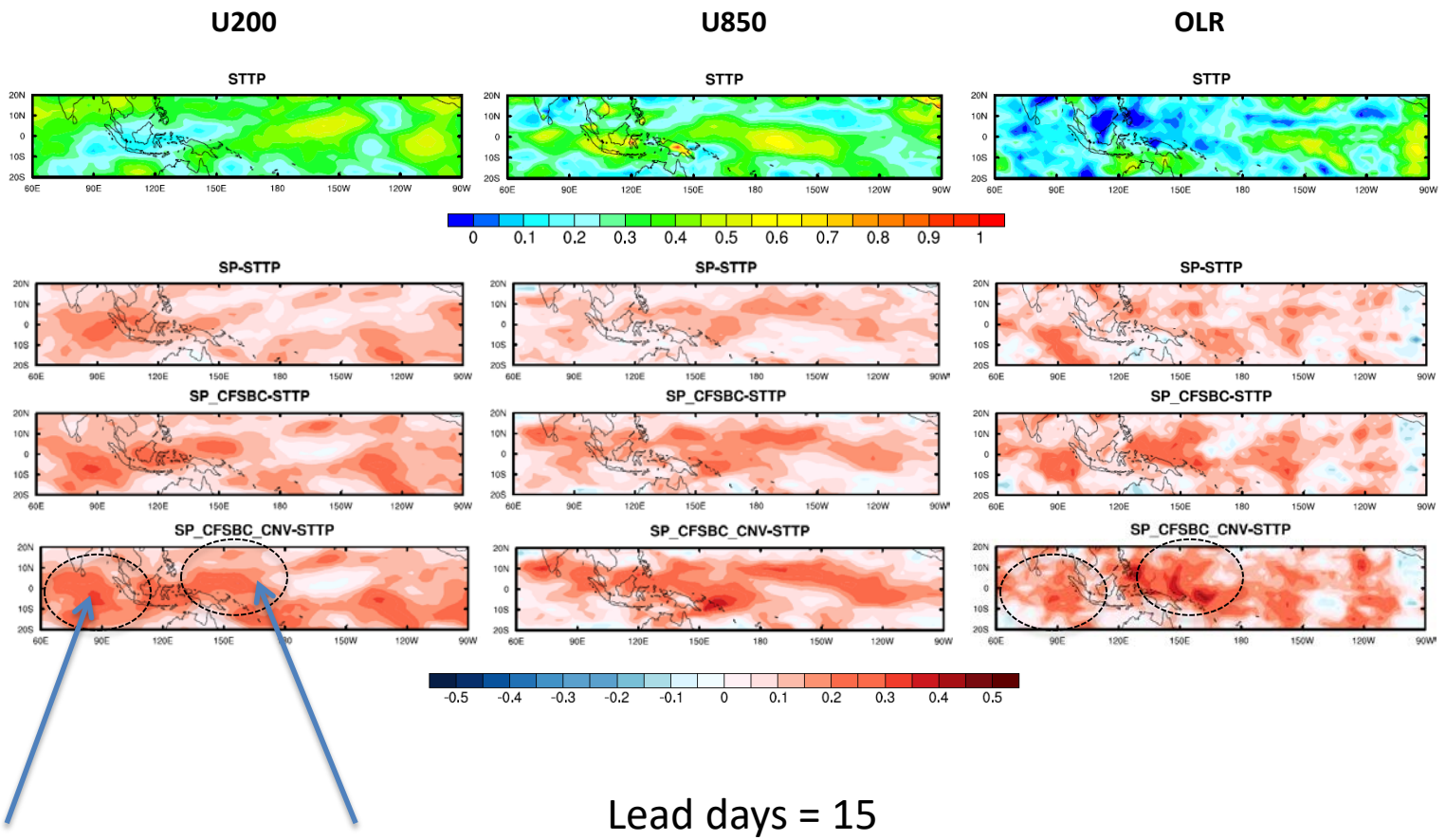


Example of 850hPa zonal wind anomaly (10°N–10°S)



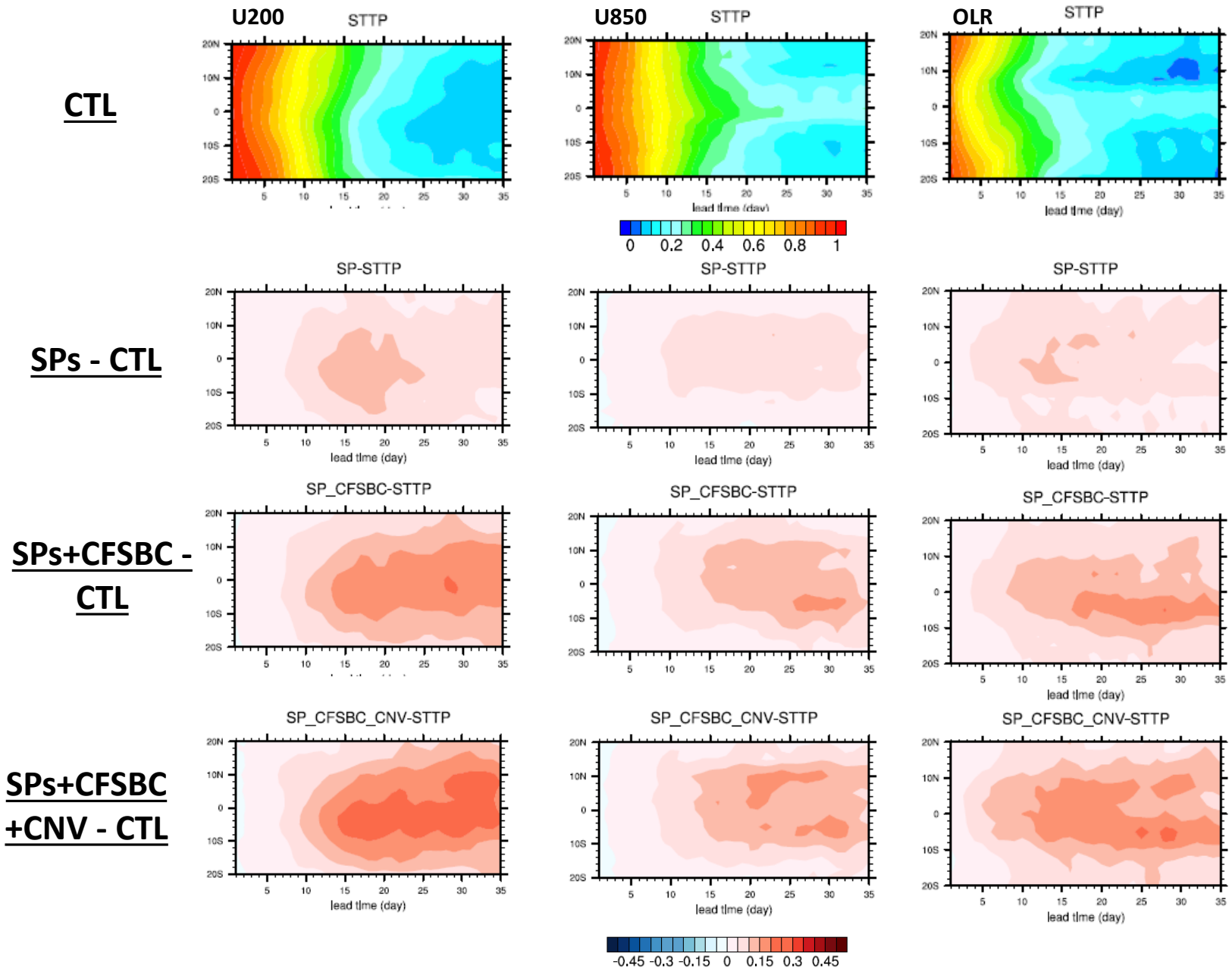
Hovmöller diagrams – for propagation

Correlation map of the key Variables

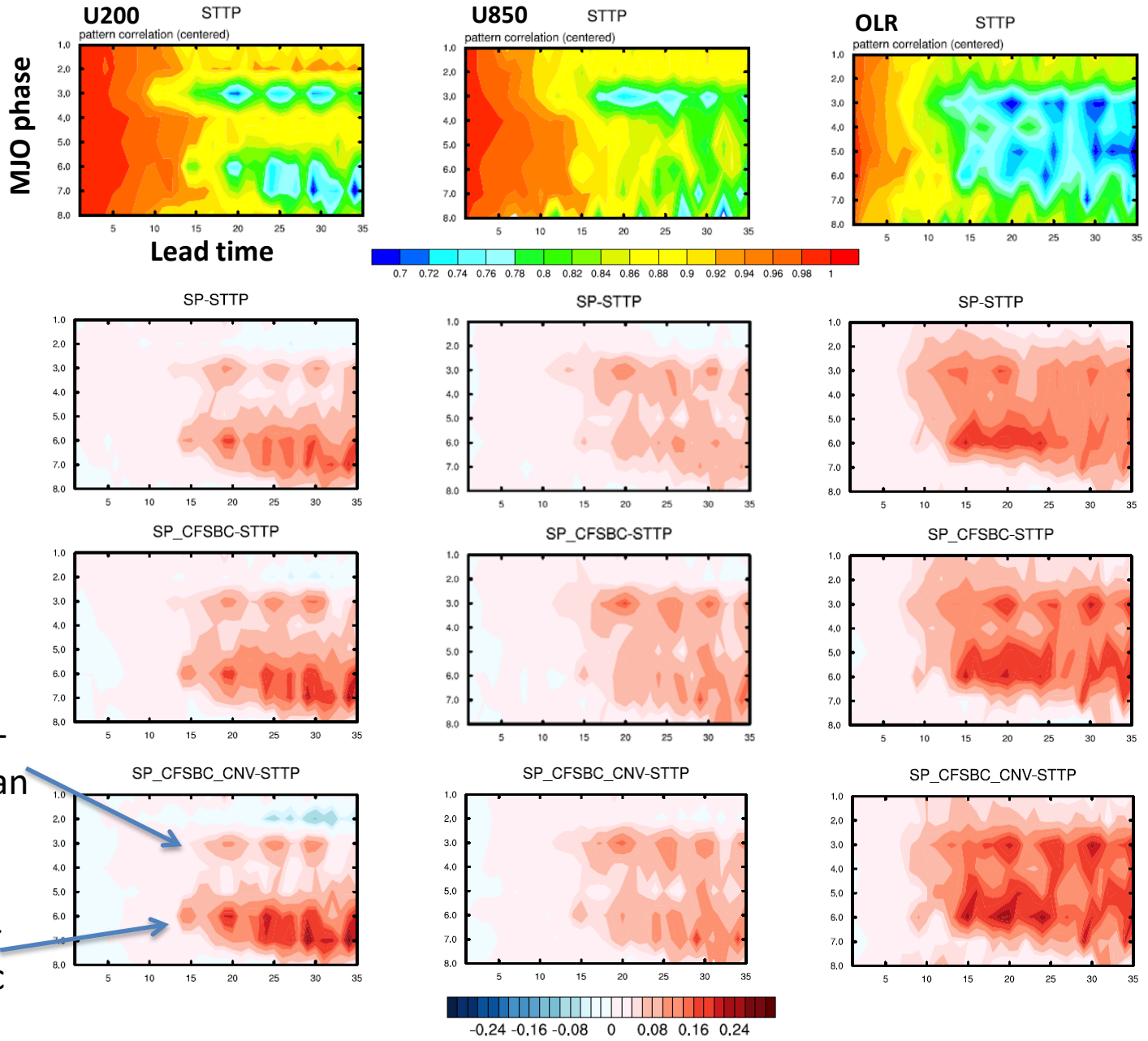


Lead days = 15

Correlation as a function of lead time



Pattern correlation of the composite variables in MJO phases



Phase3 – India Ocean

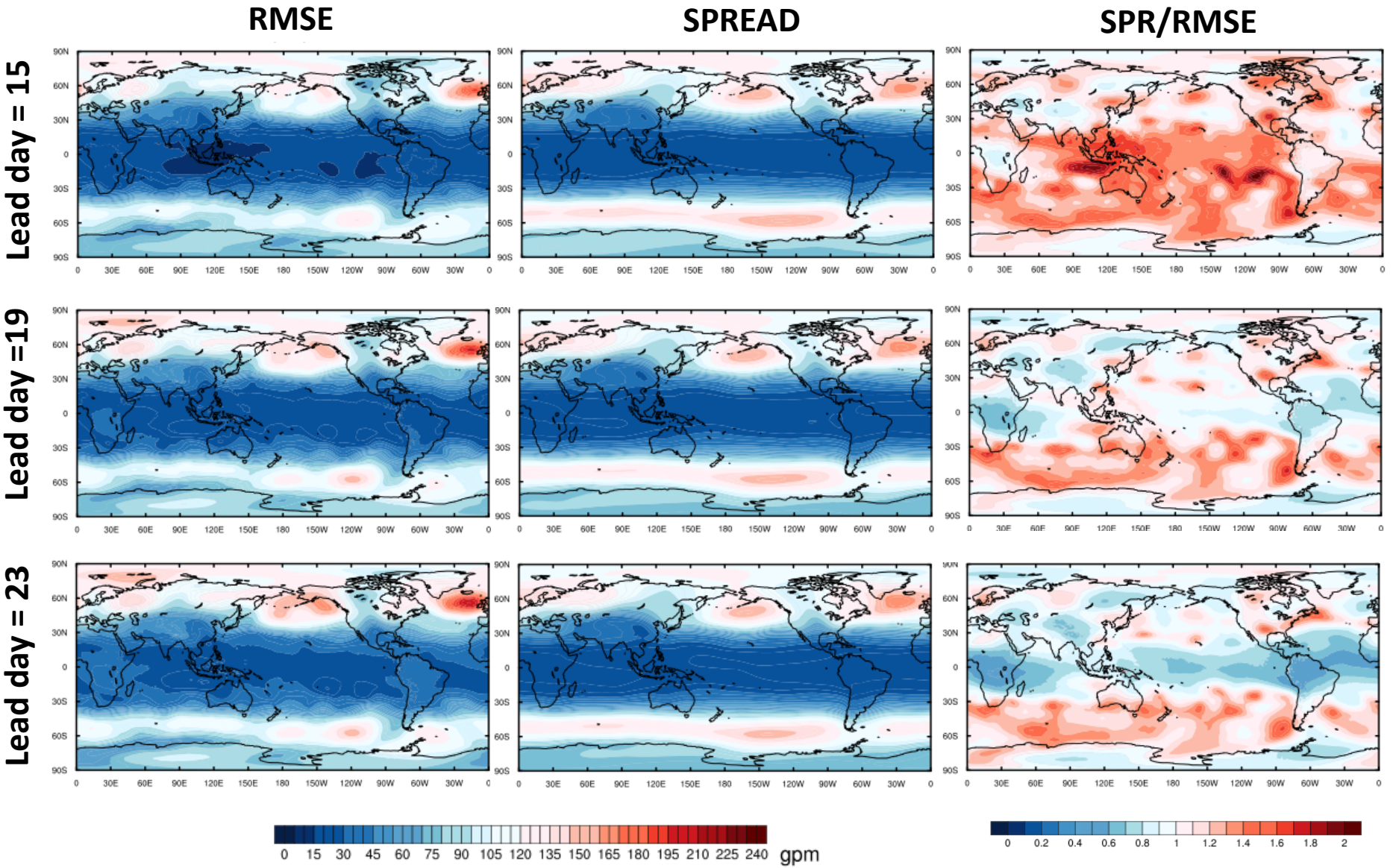
Phase6 – W. Pacific

Other Diagnostics/Evaluations

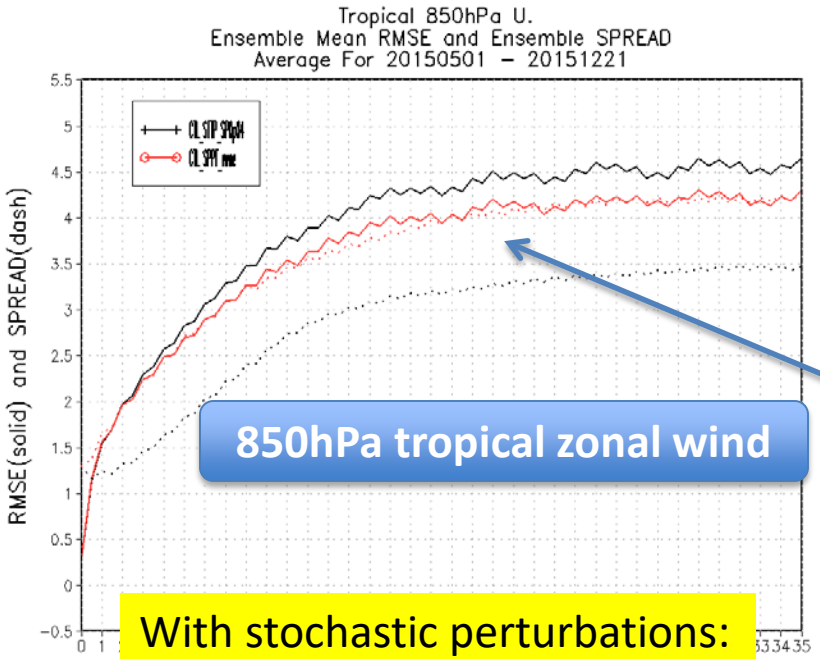
- Power (energy) spectrum
  - Could apply to many interesting variables for tropical area.
- Blocking
  - NH blocking index (500hPa height)
- Teleconnections
  - AO, NAO, PNA, NAO, AAO index
- Storm tracks
  - Globally, tropical and extra-tropical storms …
- …….

- All these are for deterministic forecast or ensemble mean (?)

Ensemble Mean VS. Spread: H500

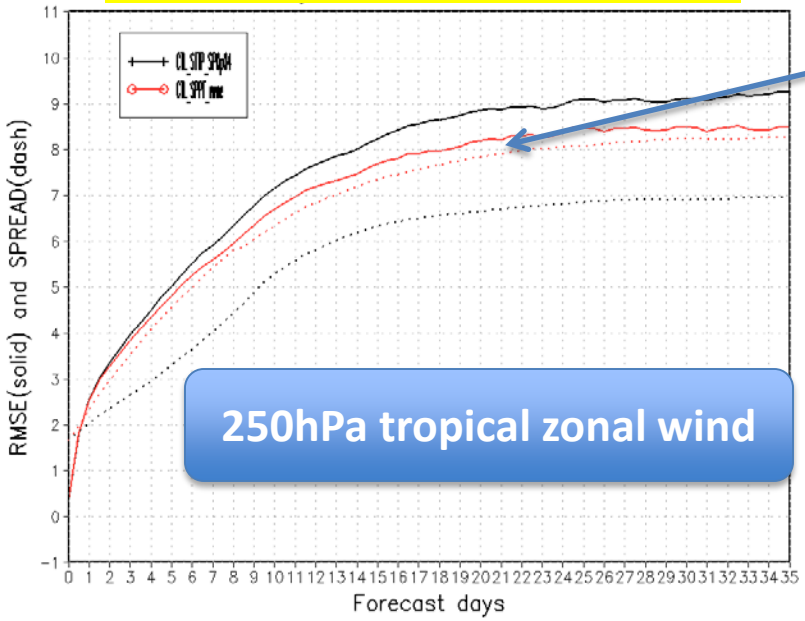


Zonal wind speed (f144 hours – 6 days)



850hPa tropical zonal wind

With stochastic perturbations:
Error is reduced
Spread is increased



250hPa tropical zonal wind

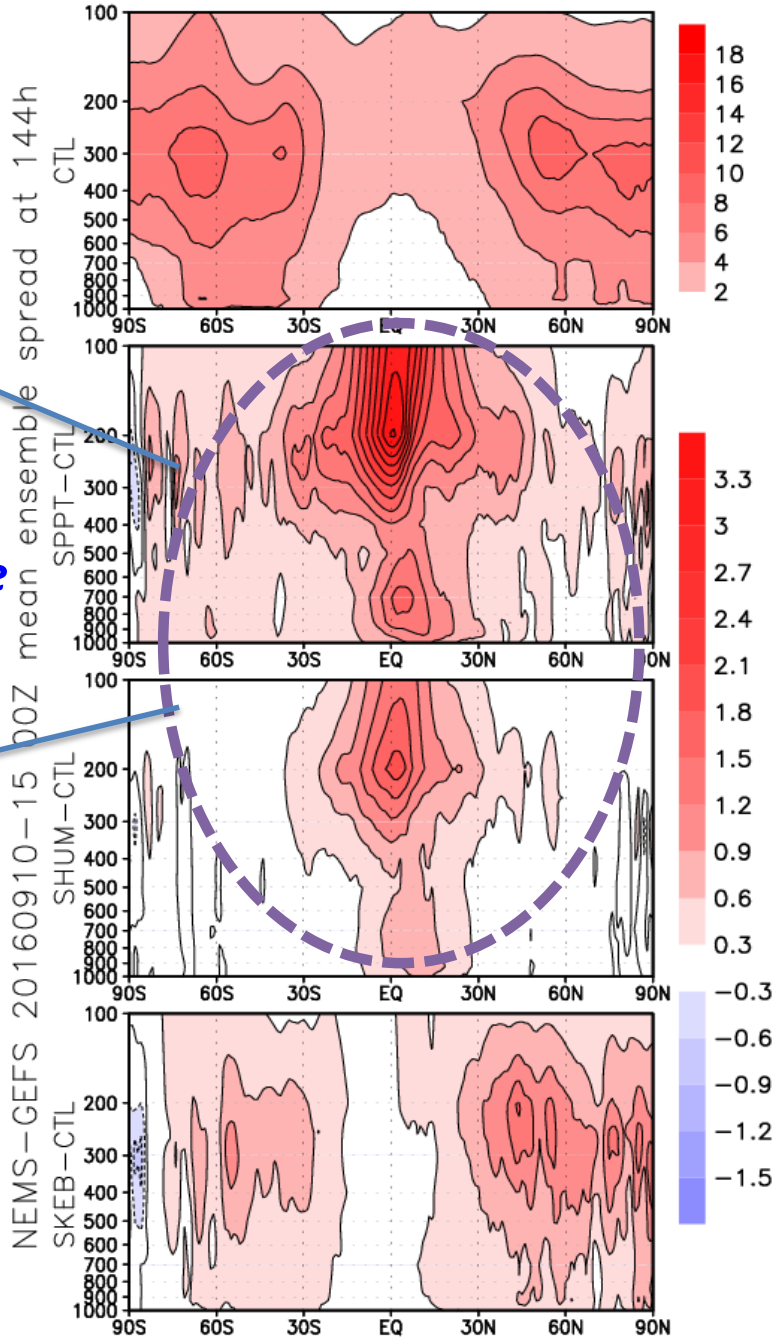
CTL

SPPT

5-scale

SHUM

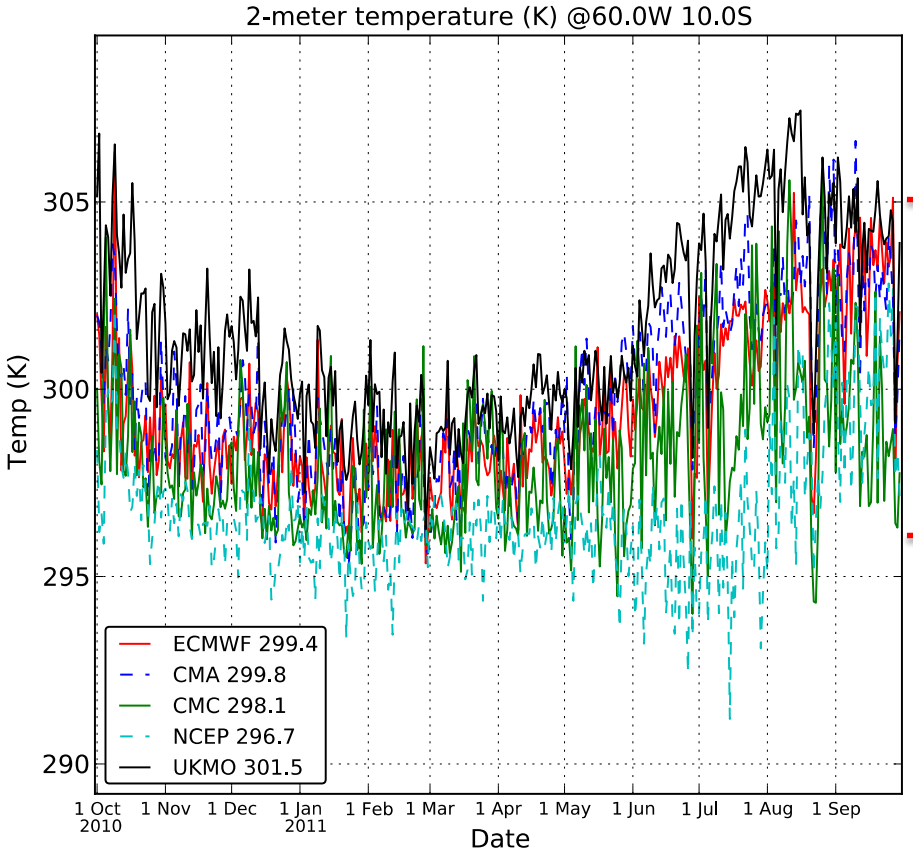
SKEB



Discussion

- What do we learn?
  - Anomaly forecast for week-2 and beyond.
  - Raw forecast skill is very low or negative.
  - Forecast is biased, need to be calibrated.
- What do we need?
  - What do we have from exist utilities?
    - NCL? GrADs?
  - Diagnostic tools for probabilistic forecast
    - For model development
    - For user evaluation/validation/verification
- Challenges
  - **Proxy truth for verification** – observation? Best analysis?
  - Reference for skill – climatology? On observation stations?
  - Limited sample size?
  - No calibration, hindcast is not available mostly.

Time series of analyses, Amazon basin, South America



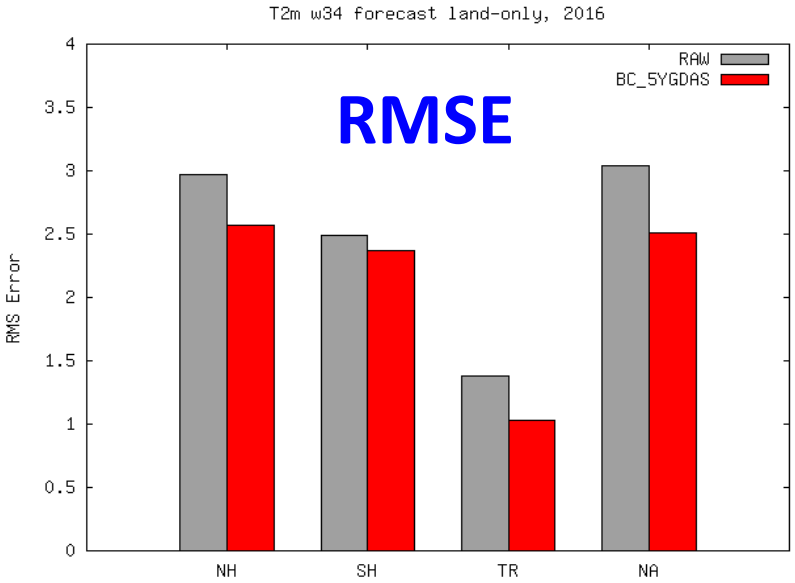
Courtesy of Dr. Tom Hamill

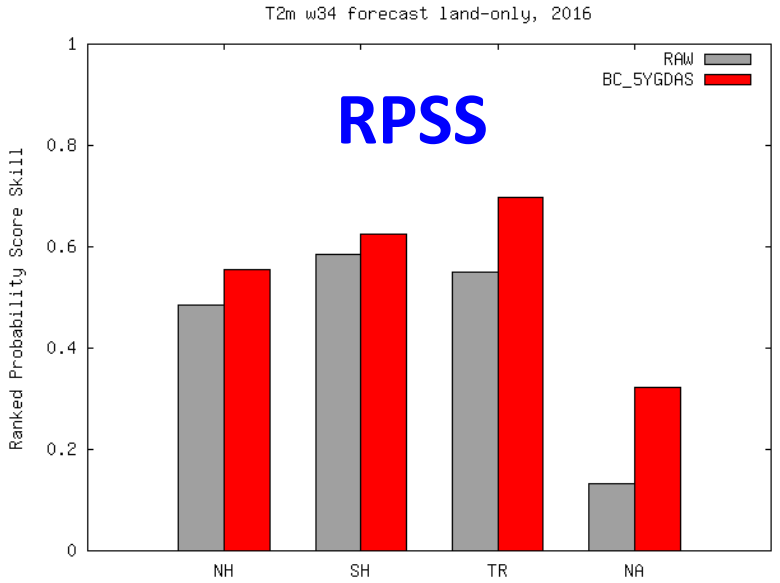
At this point in the Amazon basin, the analyses are quite different from each other, even in **yearly means**, which range from 296.7K to 301.5K. NCEP is rather consistently the coldest, UKMO the warmest. This suggests that a primary difference between analyses, especially in relatively data-sparse regions, may be a systematic bias in each analysis system.

Approximately 10 K difference from these five analysis in summer

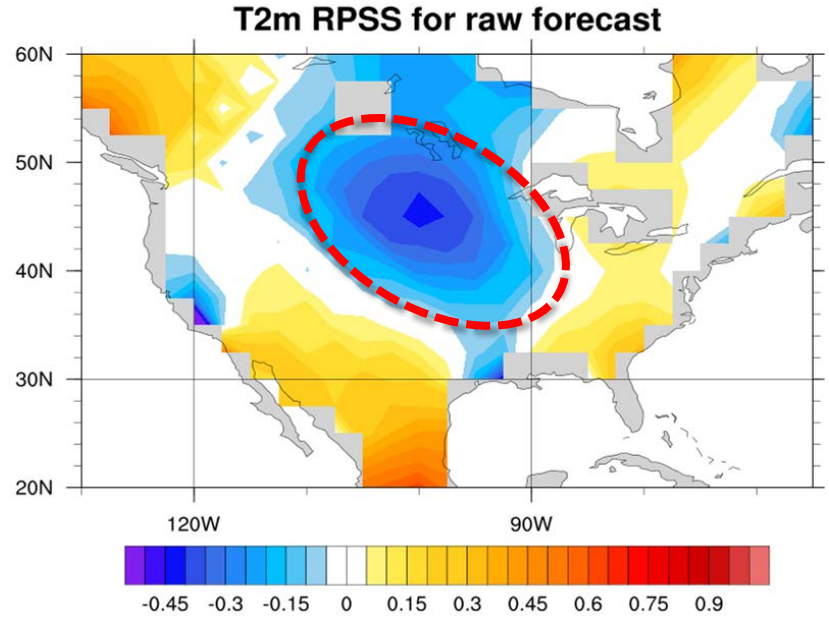
Extra slides!!!

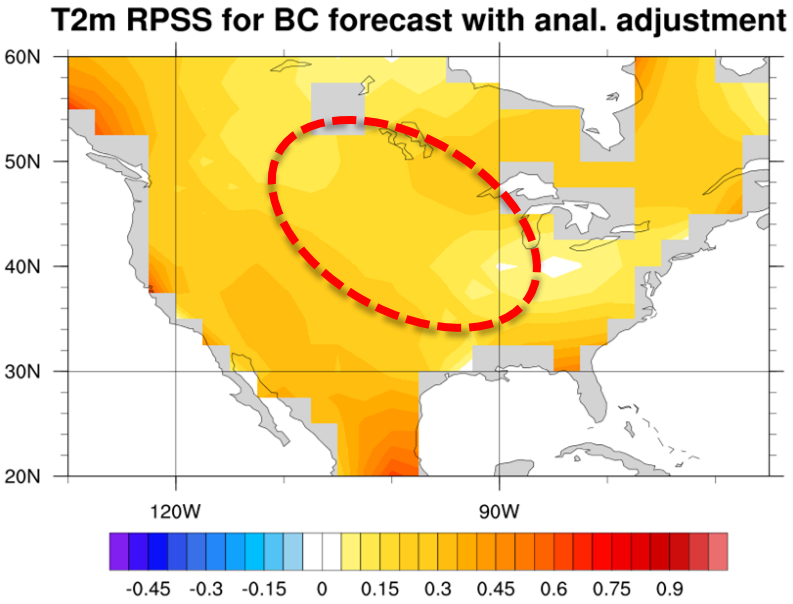
Bias correction for T2m (weeks 3&4)





Land only





Ensemble Mean PNA and NAO Correlation

NCEP GEFS has best score of PNA and NAO (green) based on 16 years hindcast

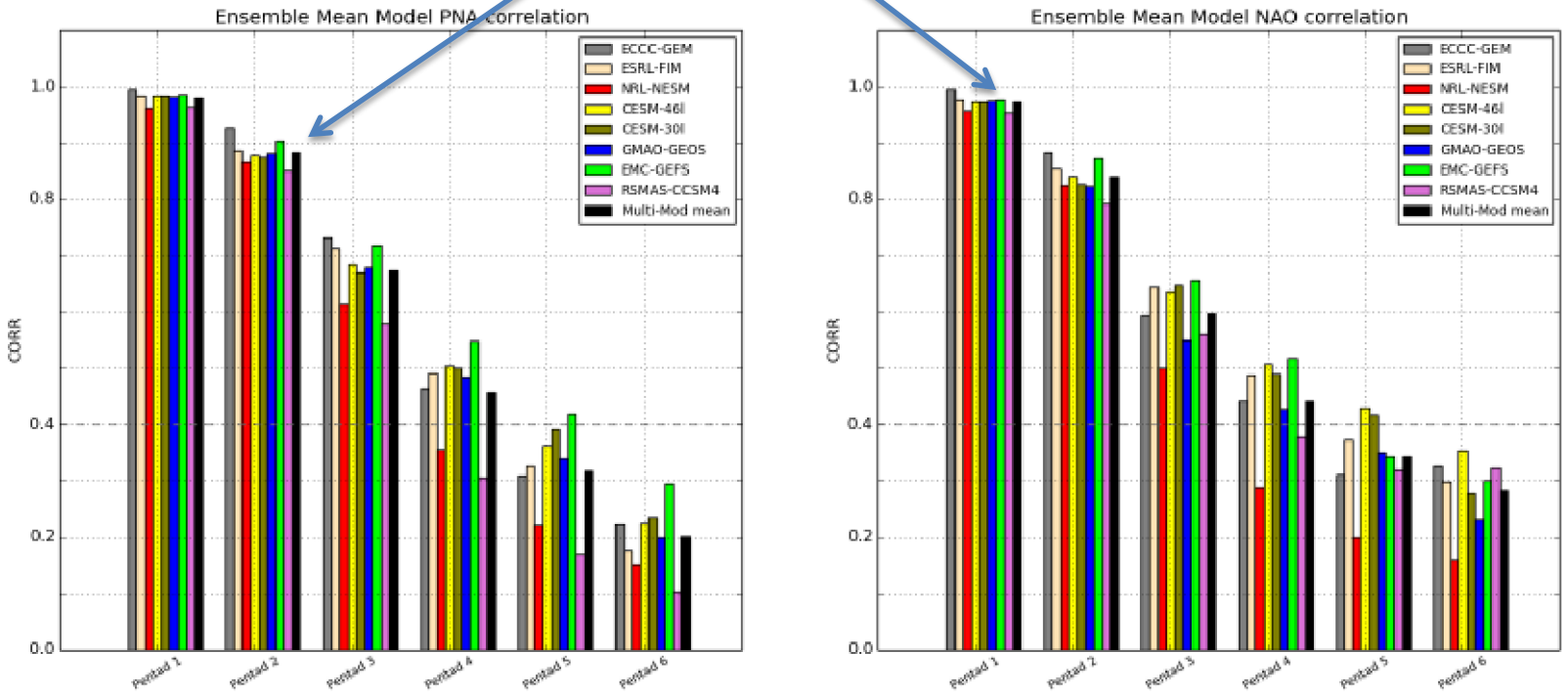
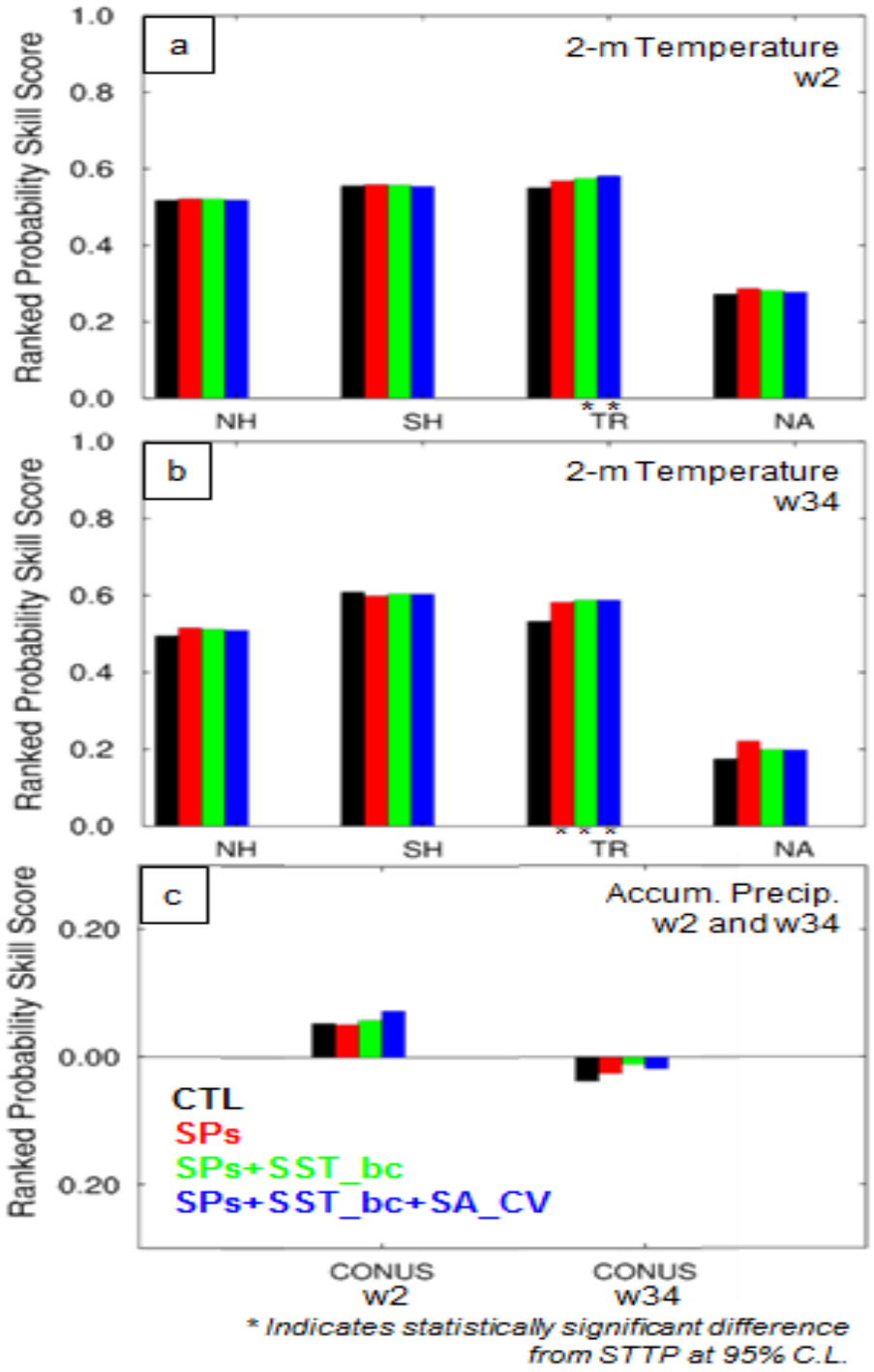


Figure 4 : Ensemble mean PNA and NAO Correlation SKILL over 6 pentads forecast range SubX datasets. The hindcast period spans from 1999 to 2014 and over the extended winter time (November to March). Roughly 350 forecast sample was used for each model

Courtesy of E. Poan and H. Lin



Evaluation of Surface Elements

RPS forecast skills

Surface temperature

Raw forecast

Land only

Week 2 averages

Weeks 3&4 average

Significant test

Precipitation

Raw forecast

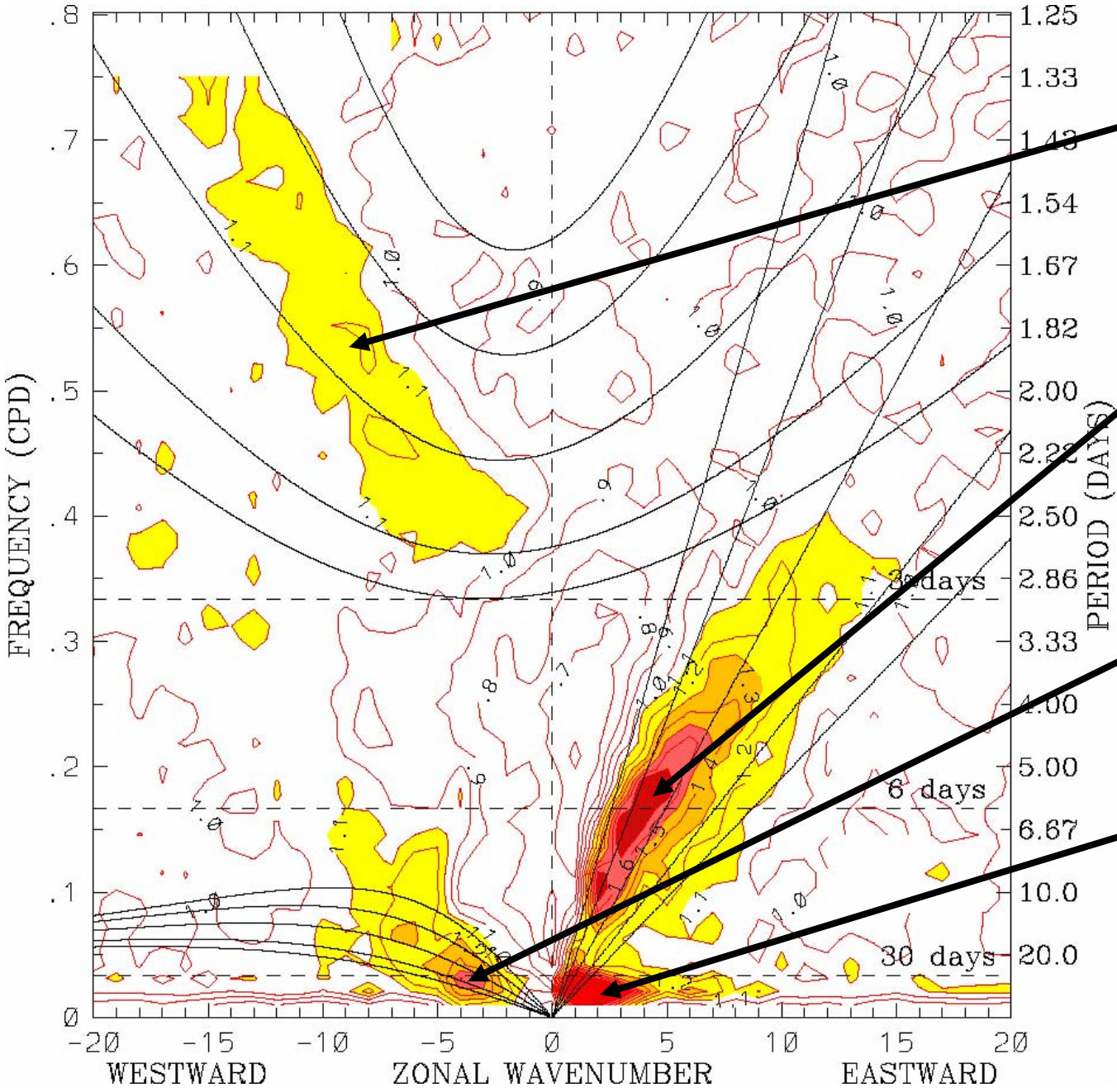
CONUS only

Week 2 accumulation

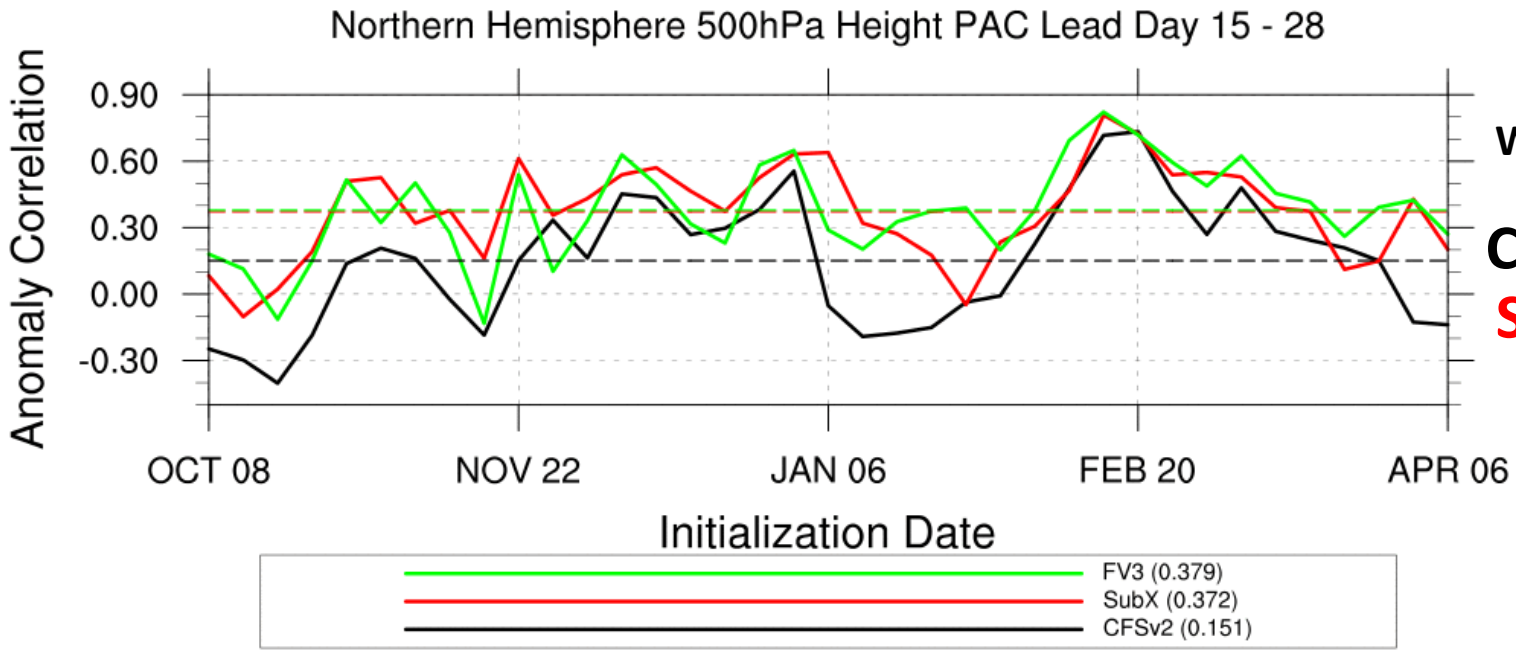
Weeks 3&4 accum.

Significant test

OLR power spectrum, 1979–2001 (Symmetric)



from
Wheeler and Kiladis, 1999

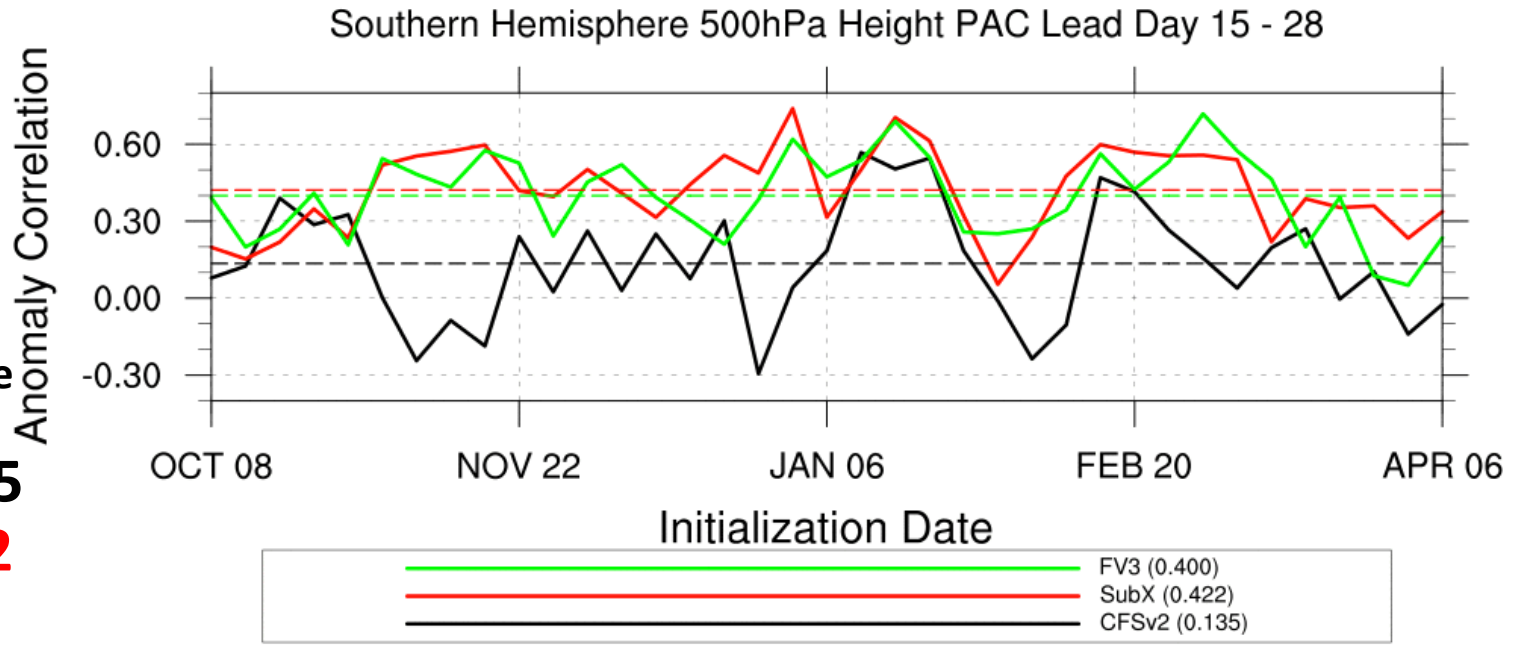


Weeks 3&4 average

CFSv2 --- 0.151

SubX --- 0.372

FV3 --- 0.379



Weeks 3&4 average

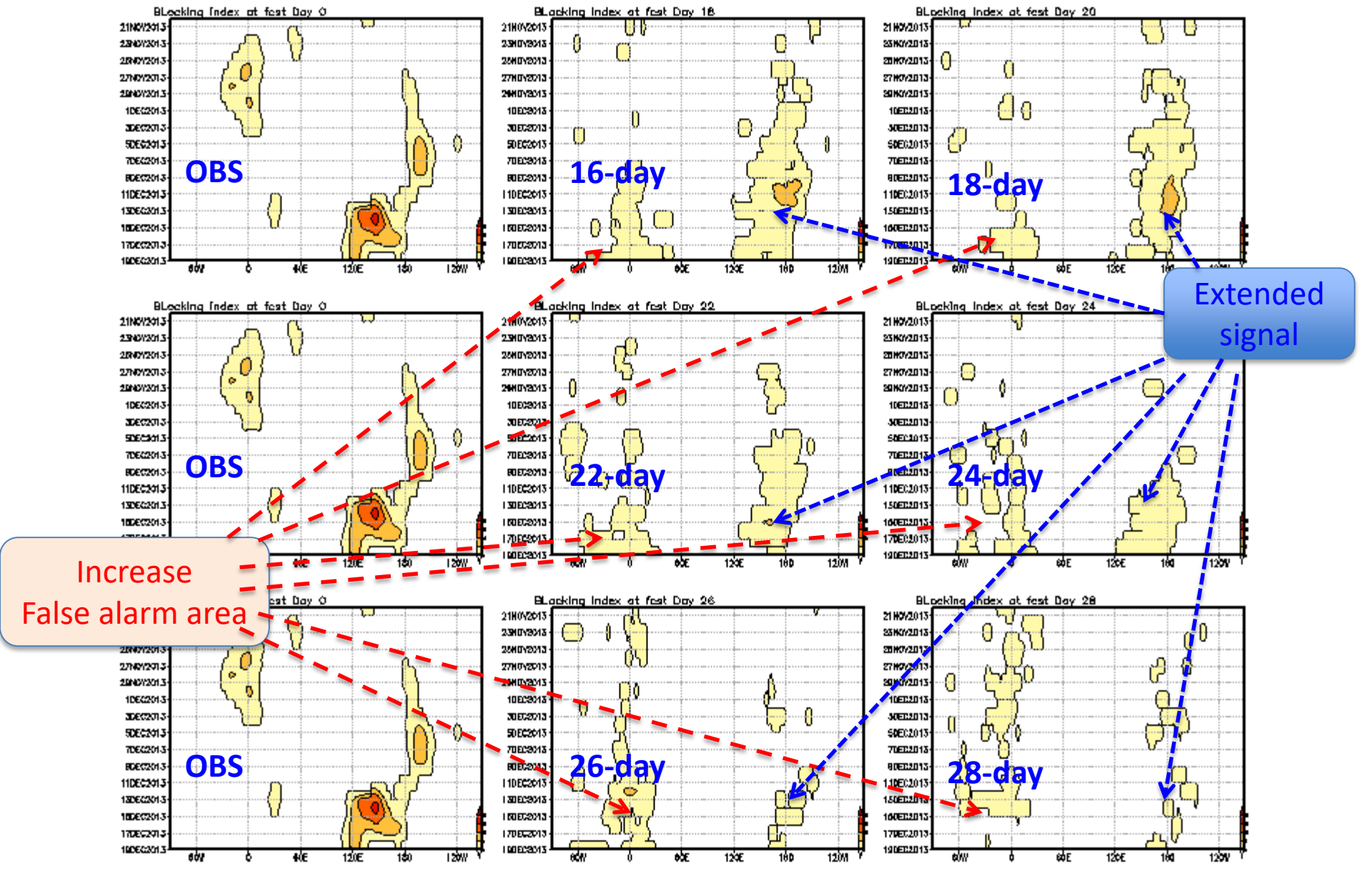
CFSv2 --- 0.135

SubX --- 0.422

FV3 --- 0.400

Northern Hemisphere Blocking Indices for Winter (2013-2014)

GEFS probabilistic forecast (10%) – parallel (control) extended fcst



Blocking Index Definition

Northern Hemisphere

For each longitude, the southern 500 hPa geopotential height gradient ($GHGS$) and the northern 500 hPa geopotential height gradient ($GHGN$) are computed as follows:

$$GHGS = \left[\frac{Z(\phi_o) - Z(\phi_s)}{\phi_o - \phi_s}\right]$$

$$GHGN = \left[\frac{Z(\phi_n) - Z(\phi_o)}{\phi_n - \phi_s}\right]$$

where

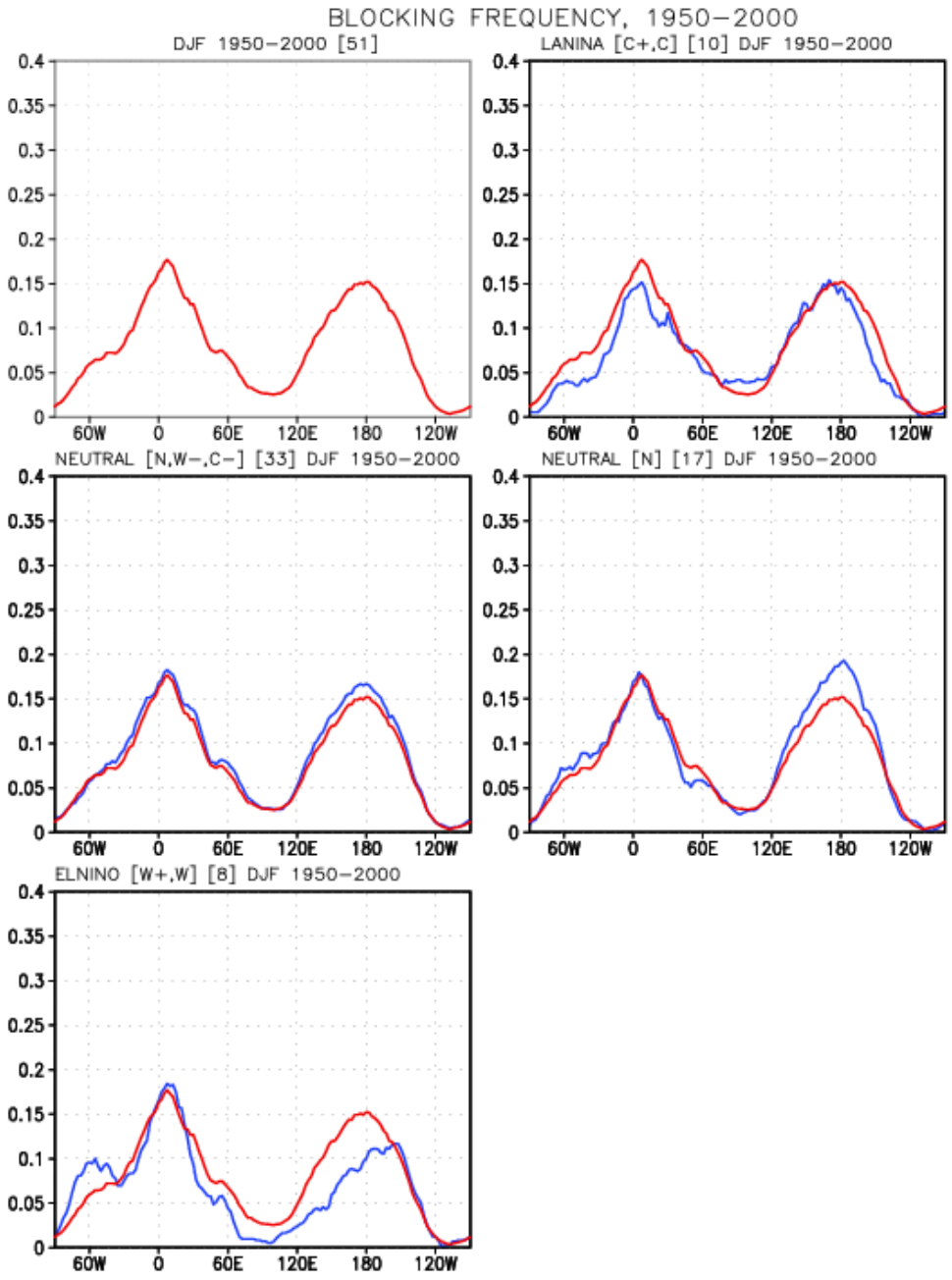
$\phi_n = 80^o N + \delta,\ \phi_o = 60^o N + \delta,\ \phi_s = 40^o N + \delta,\ \delta = -5^o, 0^o, 5^o.$

A given longitude is said to be blocked at a given time if the following conditions are satisfied for at least one value of δ:

(1) $GHGS > 0$

(2) $GHGN$ < -10 m/deg latitude.

A 5 day running mean can be applied to the 500 hPa height field prior to calculating both $GHGS$ and $GHGN$ to isolate potential blocking episodes of sufficient duration.



EMC Ensemble Probabilistic Verification

Yuejian Zhu

Environmental Modeling Center

NCEP/NWS/NOAA

Metrics

- Upper atmosphere and/or continuous
  - Variables
    - 500hPa, 1000hPa geopotential height
    - 850hPa 2-meter temperature
    - 850hPa, 250hPa, 10 meter winds
  - Proxy truth – best analysis
  - Climatology
    - NCEP/NCAR 40-y reanalysis and CFSRR (30 years)
    - PDF – 10 equal-likely-bin
- Special variables
  - Precipitation for CONUS
    - Proxy truth – CCPA and rain gauge
    - Climatology – CCPA since 2002-current
  - Tropical storm
    - Proxy truth – best track (observation)
- Ranking by forecasts
  - Histograms
  - Continues Ranked Probability Score (CRPS)
- Ranking by climatology (10 equal-likely-bin)
  - Brier scores – multi-categories and ranking
    - Decomposition: reliability and resolution
  - Ranked Probability Score (RPS)
  - ROC and Economic Values
  - Many others
- Specified value (threshold)
  - Brier scores and decomposition
  - ROC and Economic Values
  - Many others
- Ensemble mean related
  - Pattern Anomaly Correlation (PAC)
  - RMS error
  - Spread
- Challenges?
  - Proxy truth
    - Observation or best analysis?
      - Hard to have climatology at observation station
  - Specified value (or threshold) for extreme events
    - Hard to explore this for general ensemble application

EMC ensemble verification web-page

Since 2000 (quarterly average scores)

File Edit View History Bookmarks Tools Help

NOAA/NWS/EMC/GMB Ensem

www.emc.ncep.noaa.gov/gmb/yzhu/html/opr/naefs.htm

Most Visited Internet Lookup New&Cool

Period	NCEP .vs NCEPb	CMC .vs CMCb	NCEP .vs CMC	NCEPb .vs CMCb	NAEFS	NCEP .vs CMC .vs ECMWF	FN .vs FNb	NAEFS .vs NUOPC
win1718	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
fal2017	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
sum2017	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
spr2017	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
win1617	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
fal2016	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
sum2016	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
spr2016	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
win1516	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
fal2015	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
sum2015	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
spr2015	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
win1415	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
fal2014	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
sum2014	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
spr2014	Yes	Yes	Yes	Yes	Yes	Yes		
win1314	Yes	Yes	Yes	Yes	Yes	Yes		
fal2013	Yes	Yes	Yes	Yes	Yes	Yes		
sum2013	Yes	Yes	Yes	Yes	Yes	Yes		
spr2013	Yes	Yes	Yes	Yes	Yes	Yes		
win1213	Yes	Yes	Yes	Yes	Yes	Yes		
fal2012	Yes	Yes	Yes	Yes	Yes	Yes		
sum2012	Yes	Yes	Yes	Yes	Yes	Yes		
spr2012	Yes	Yes	Yes	Yes	Yes	Yes		
win1112	Yes	Yes	Yes	Yes	Yes	Yes		
fal2011	Yes	Yes	Yes	Yes	Yes	No		
sum2011	Yes	Yes	Yes	Yes	Yes	No		
spr2011	Yes	Yes	Yes	Yes	Yes	No		
win1011	Yes	Yes	Yes	Yes	Yes	No		
fal2010	Yes	Yes	Yes	Yes	Yes	No		

Scroll down for more

File Edit View History Bookmarks Tools Help

NCEP .vs NCEPb

www.emc.ncep.noaa.gov/gmb/yluo/html/opr/naefs/NCE

Most Visited Internet Lookup New&Cool

Global Ensemble model evaluation: (NCEP against NCEPb)

500 hPa Height Scores NCEP .vs NCEPb										
NH	ROC	EV	RPSS	BSS	CRP	CRPS	RMS/SPRD	ERR/ABSE	PAC	HISTOGRAM
SH	ROC	EV	RPSS	BSS	CRP	CRPS	RMS/SPRD	ERR/ABSE	PAC	HISTOGRAM
TROP	ROC	EV	RPSS	BSS	CRP	CRPS	RMS/SPRD	ERR/ABSE	PAC	HISTOGRAM

1000 hPa Height Scores (NCEP .vs NCEPb)										
NH	ROC	EV	RPSS	BSS	CRP	CRPS	RMS/SPRD	ERR/ABSE	PAC	HISTOGRAM
SH	ROC	EV	RPSS	BSS	CRP	CRPS	RMS/SPRD	ERR/ABSE	PAC	HISTOGRAM
TROP	ROC	EV	RPSS	BSS	CRP	CRPS	RMS/SPRD	ERR/ABSE	PAC	HISTOGRAM

850 hPa Temperature Scores(NCEP .vs NCEPb)										
NH	ROC	EV	RPSS	BSS	CRP	CRPS	RMS/SPRD	ERR/ABSE	PAC	HISTOGRAM
SH	ROC	EV	RPSS	BSS	CRP	CRPS	RMS/SPRD	ERR/ABSE	PAC	HISTOGRAM
TROP	ROC	EV	RPSS	BSS	CRP	CRPS	RMS/SPRD	ERR/ABSE	PAC	HISTOGRAM

EP .vs NCEPb)

ERR/ABSE PAC HISTOGRAM

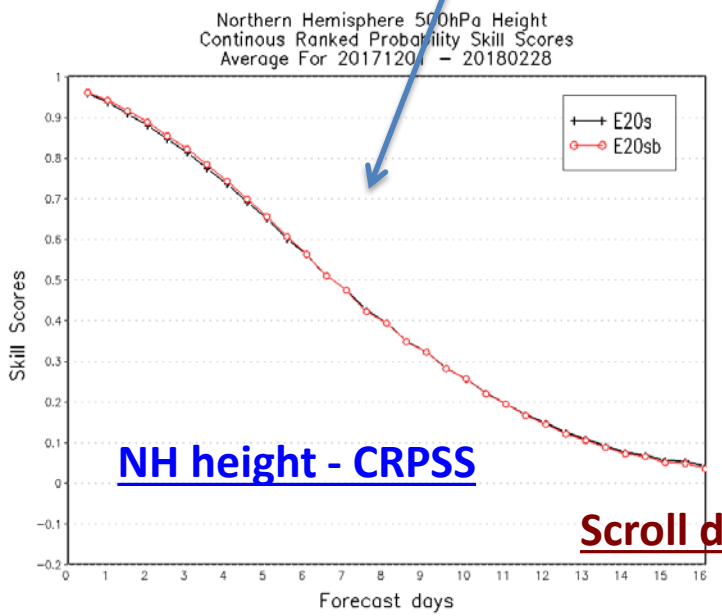
ERR/ABSE PAC HISTOGRAM

ERR/ABSE PAC HISTOGRAM

P .vs NCEPb)

ERR/ABSE PAC HISTOGRAM

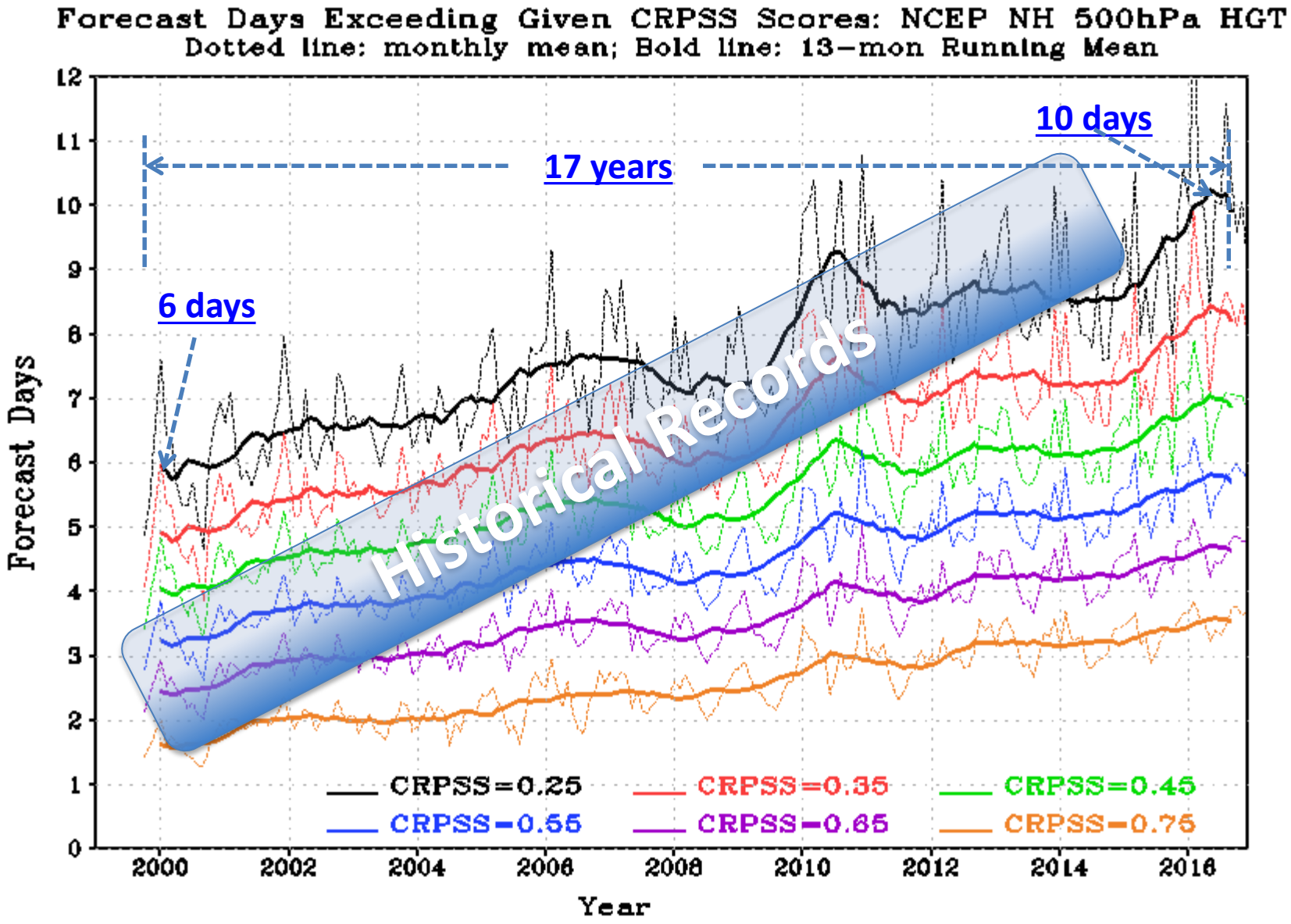
ERR/ABSE PAC HISTOGRAM



NH height - CRPSS

Scroll down for more

CRPSS for NH 500hPa geopotential height



Scores Card: GEFSv11 21m .vs 41m (August 1 – October 1 2013)

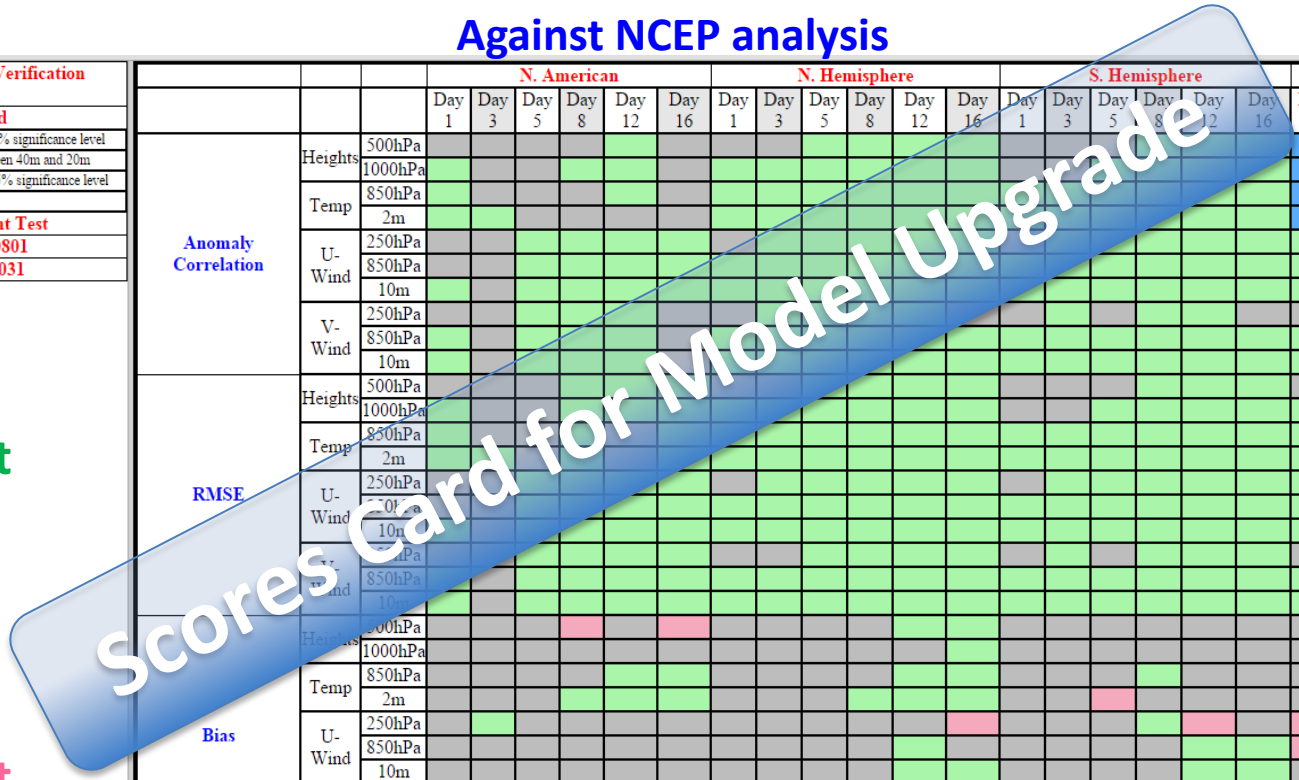
Against NCEP analysis

EMC Global Ensemble Verification Scorecard
Symbol Legend
40m is better than 20m at the two-tailed 95% significance level
No statistically significant difference between 40m and 20m
40m is worse than 20m at the two-tailed 95% significance level
Not statistically relevant
Bootstrap Significant Test
Start Date: 20130801
End Date: 20131031

Green: significant better (95%)

Pink: significant worse (95%)

Grey: In-significant or neutral



			N. American						N. Hemisphere						S. Hemisphere						Tropics					
			Day 1	Day 3	Day 5	Day 8	Day 12	Day 16	Day 1	Day 3	Day 5	Day 8	Day 12	Day 16	Day 1	Day 3	Day 5	Day 8	Day 12	Day 16	Day 1	Day 3	Day 5	Day 8	Day 12	Day 16
Anomaly Correlation	Heights	500hPa																								
		1000hPa																								
	Temp	850hPa																								
		2m																								
	U-Wind	250hPa																								
		850hPa																								
		10m																								
	V-Wind	250hPa																								
		850hPa																								
		10m																								
RMSE	Heights	500hPa																								
		1000hPa																								
	Temp	850hPa																								
		2m																								
	U-Wind	250hPa																								
		850hPa																								
		10m																								
	V-Wind	250hPa																								
		850hPa																								
		10m																								
Bias	Heights	500hPa																								
		1000hPa																								
	Temp	850hPa																								
		2m																								
	U-Wind	250hPa																								
		850hPa																								
		10m																								
	V-Wind	250hPa																								
		850hPa																								
		10m																								
CRPSS	Heights	500hPa																								
		1000hPa																								
	Temp	850hPa																								
		2m																								
	U-Wind	250hPa																								
		850hPa																								
		10m																								
	V-Wind	250hPa																								
		850hPa																								
		10m																								
BSS	Heights	500hPa																								
		1000hPa																								
	Temp	850hPa																								
		2m																								
	U-Wind	250hPa																								
		850hPa																								
		10m																								
	V-Wind	250hPa																								
		850hPa																								
		10m																								