# Bias Analysis and Application
## - SubX project

**Hong Guan**
Yuejian Zhu, Eric Sinsky, Wei Li, Xiaqiong Zhou,
Dingchen Hou, Christopher Melhauser, and Richard Wobus

Environmental Modeling Center
NCEP/NWS/NOAA
November 2017

Acknowledgment: Weizhong Zheng in EMC

Presentation for CWB
December 21 2017

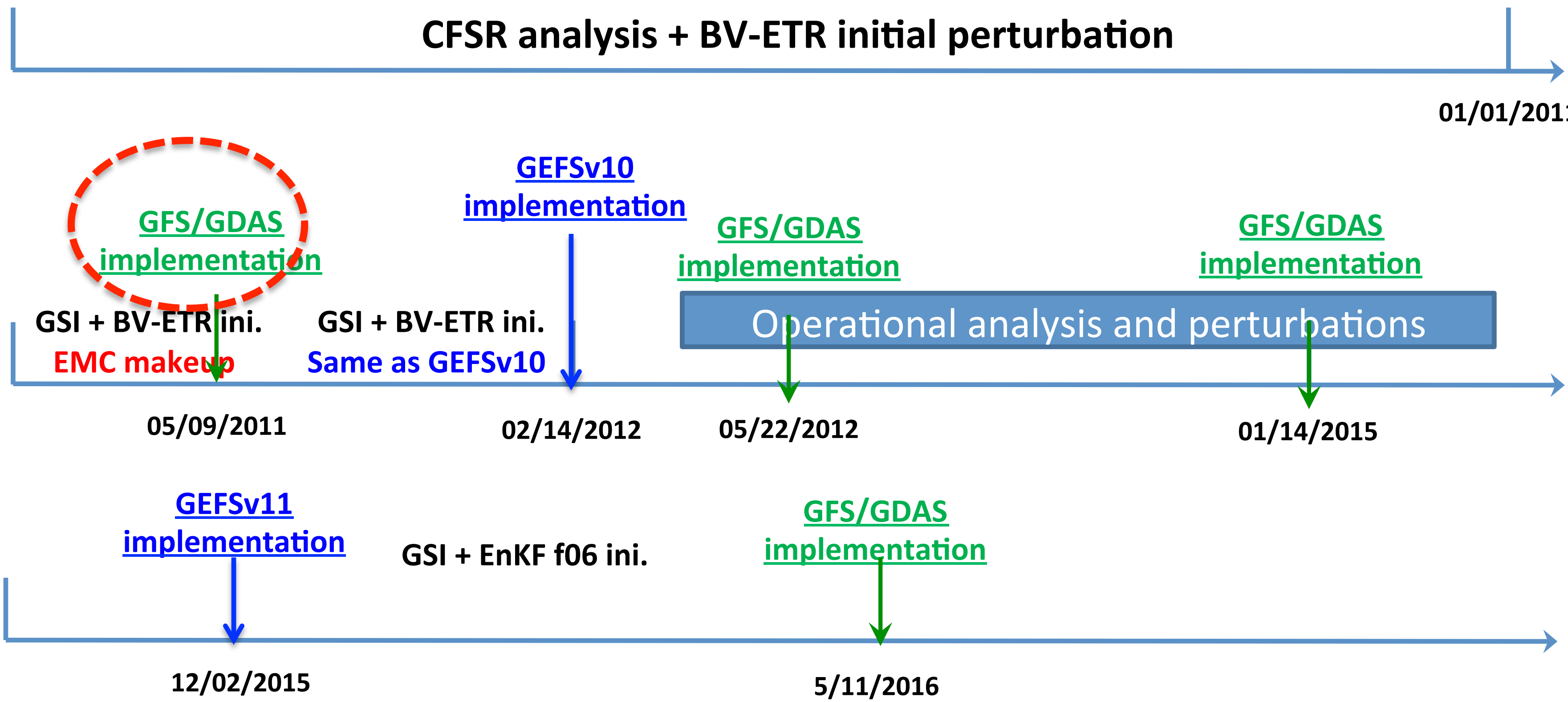
# Objectives

- Temporal and spatial distribution characteristics of T2m bias
- When does T2m error reaches to a saturated status?
  - ✓ The impact of inconsistent initial analysis on forecast
  - ✓ Only model physics drives bias
- Weeks 3&4 bias corrections
  - ✓ Bias correction with different sample-size (5yrs vs 17yrs) and analysis period (CFSR vs GDAS)
  - ✓ If the Week 2 2-m temperature bias can be used to calibrate week 3&4 forecasts
  - ✓ If analysis adjustment can improve bias corrections

# SubX Reforecast Configurations

- Model
  - GEFSv11 + SPs + SST_bc + SA_CV
- Resolutions
  - Tl574L64 (34km, 0-8 days); Tl382L64 (55km, 8-35 days)
- Memberships
  - 11 members
- Time period
  - 18 years (1999 – 2016), every Wednesday, 00 cycle only
- Initial analysis
  - CFSR (1/1/1999 – 12/31/2010)
  - GSI/GDAS (1/1/2011 – current)
  - Key GSI upgrade
    5/9/2011(roughness length), 5/22/2012, 1/4/2015, 5/11/2016
- Initial perturbations
  - BV-ETR (1/1/1999 – 12/2/2015)
  - EnKF f06 (12/2/2015 – current)

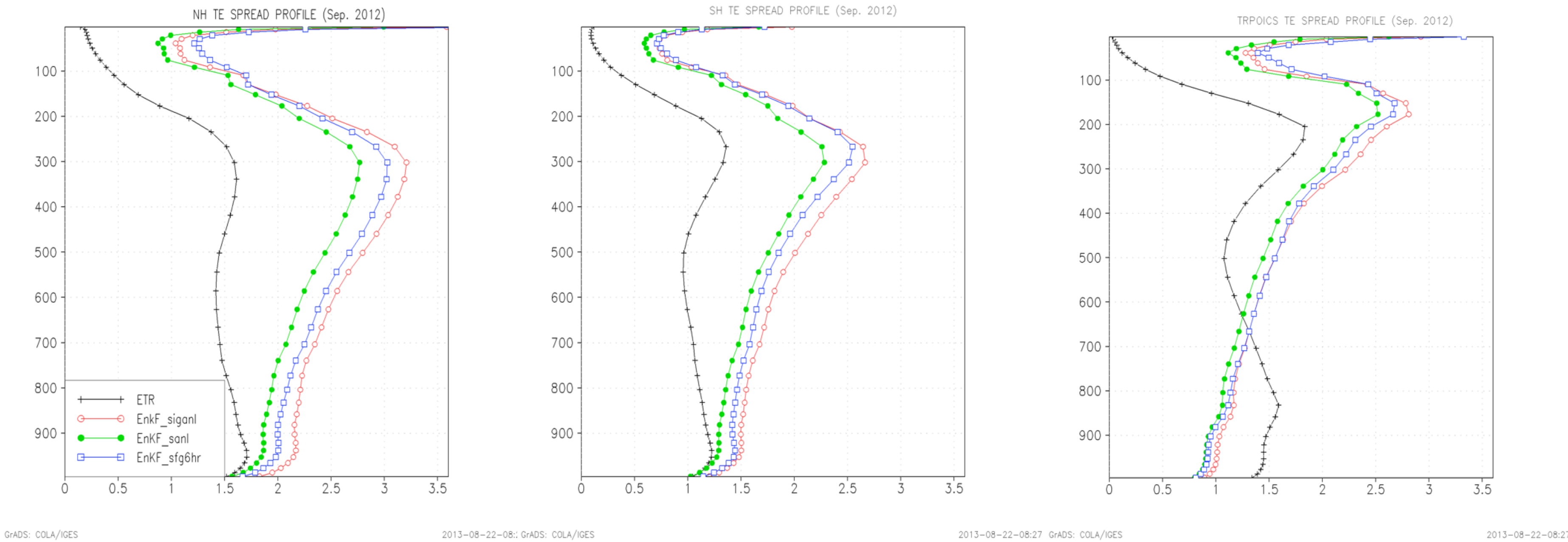
# Initial Analysis and Perturbations for 11-member v11+ (SubX) Reforecast




**18 years GEFS v11+ (SubX) reforecasts (Jan. 1999 – Dec. 2016)**

# Vertical structure of perturbation amplitude

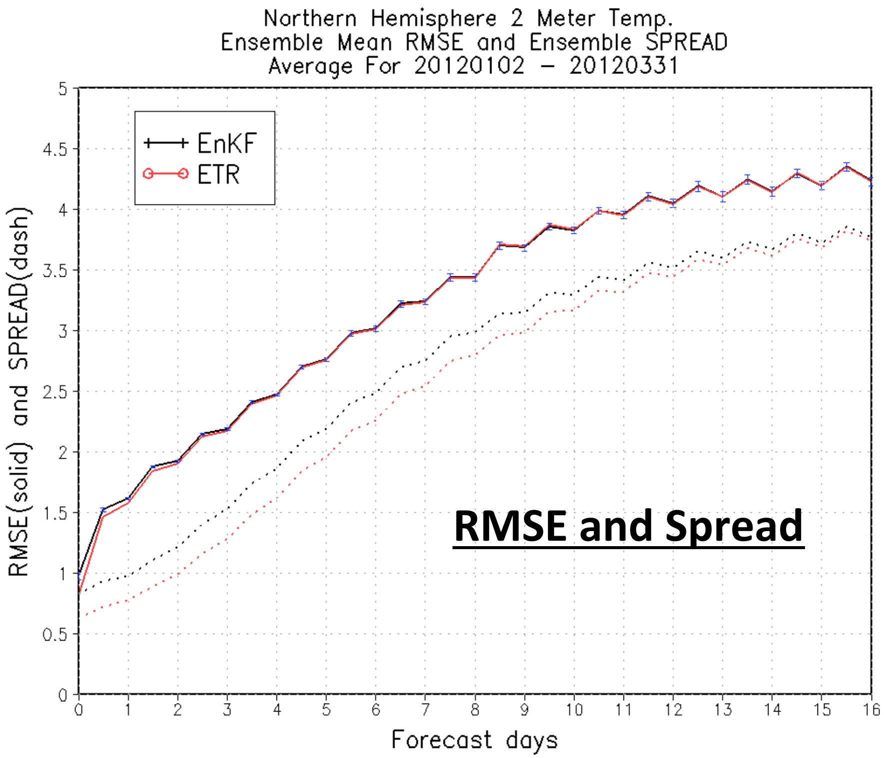
## Early study (2011-2012)

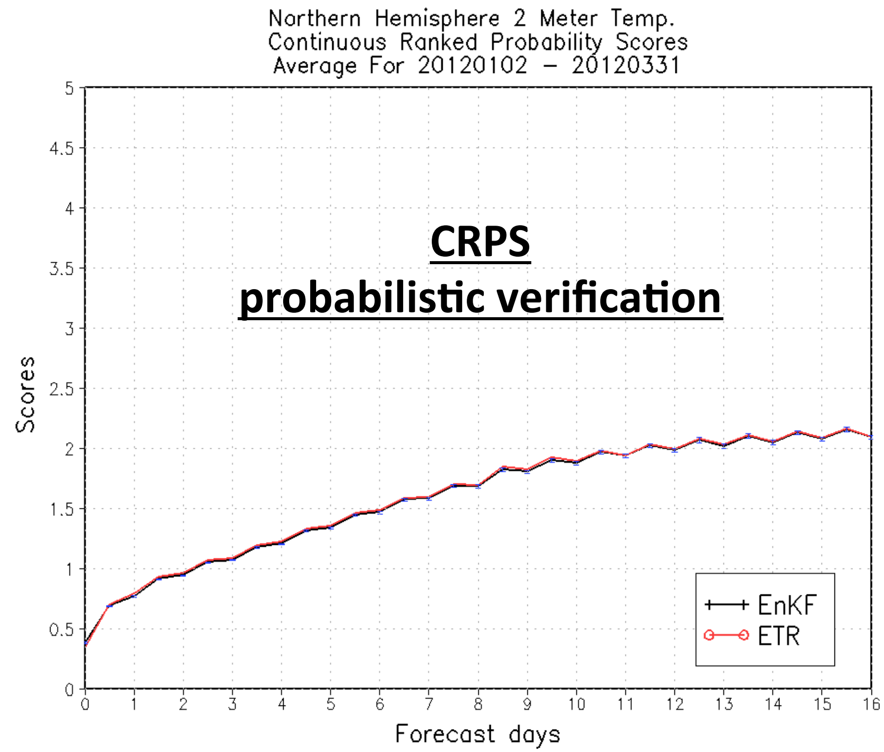



Black-BV-ETR; Green-EnKF analysis without additive inflation; Red-EnKF analysis; Blue-EnKF f06

Vertical profiles of initial perturbation spread in terms of total dry energy in the ETR and EnKF experiments over a) NH, b) SH and c) Tropics. Three EnKF profiles represent the spread of EnKF perturbations after multiple inflations (green curves), additive inflation (red) and 6-hr forecast (blue). The profiles are averaged from 1 July – 17 Oct. 2011.

# Comparison of BV-ETR and EnKF F06



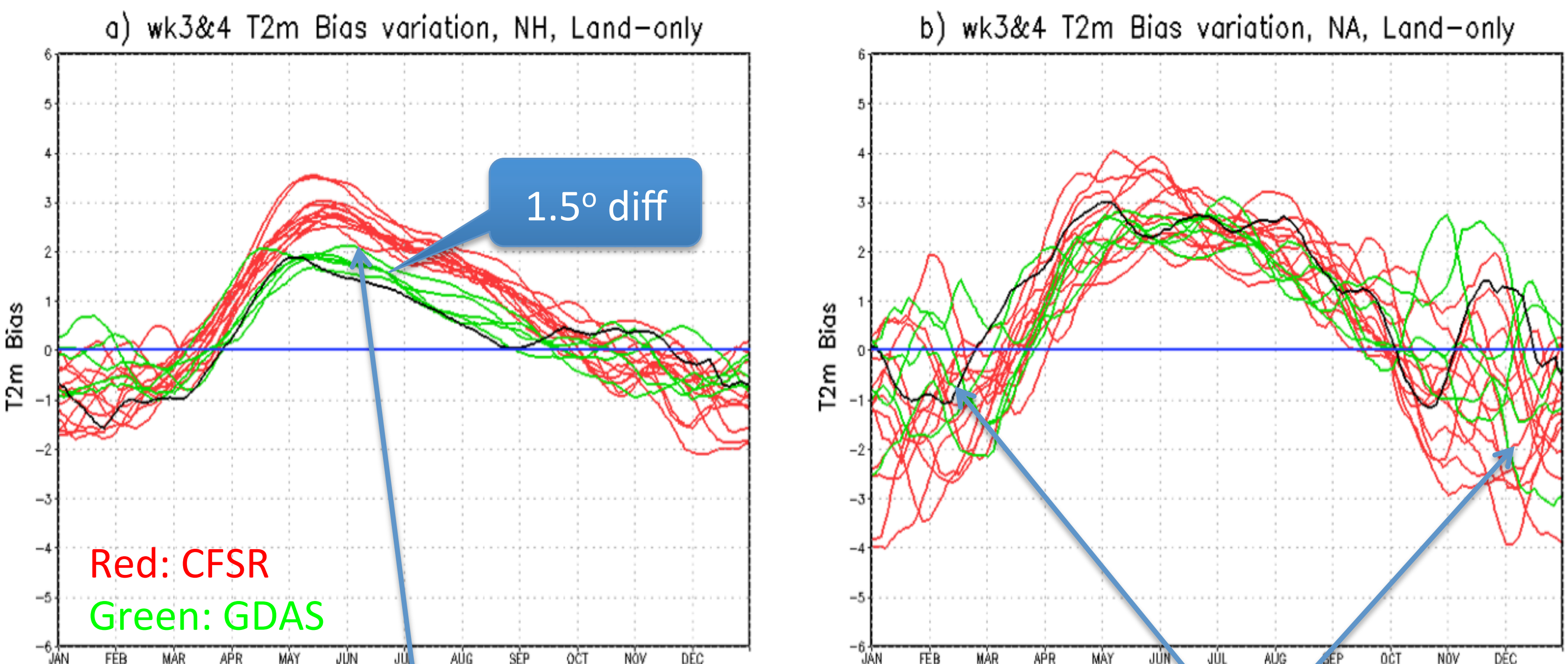




Zhou, X., Y. Zhu, D. Hou, and D. Kleist 2016,*Comparison of the ensemble transfoem and ensemble filter in the NCEP Global Ensemble Forecast System.* Wea. and Forecasting, Vol. 31, 2058-2074.

The difference of initial perturbations has impacted ensemble spread of T2m, but not for RMSE and skills!!

# Temporal and spatial distribution characteristics of T2m error

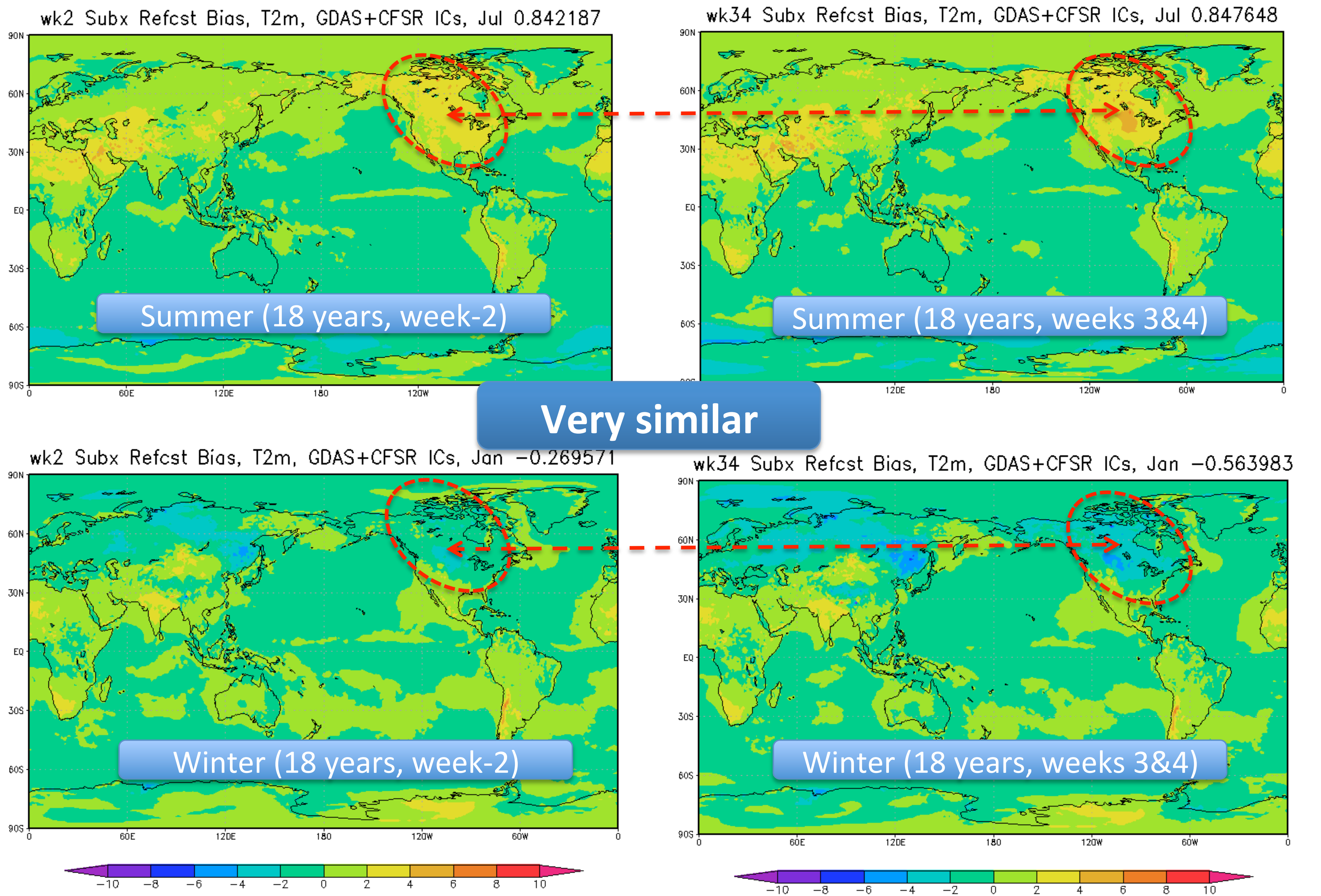
# T2m bias (weeks 3&4) time series for each year (18 years)



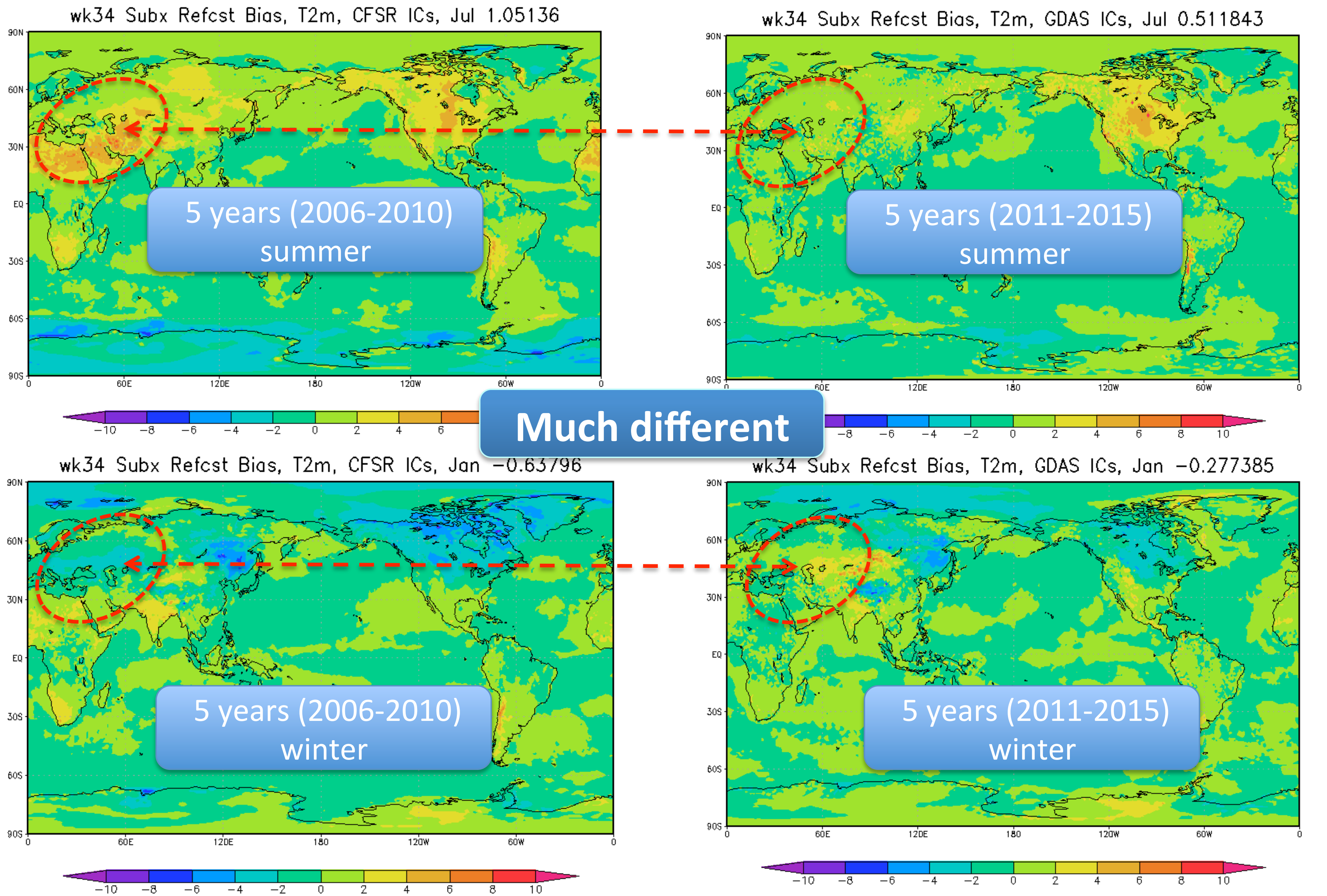

Impact mainly for arid area or dessert area

NA has a Larger bias variation

# 18-year T2m Bias for week-2, weeks 3&4

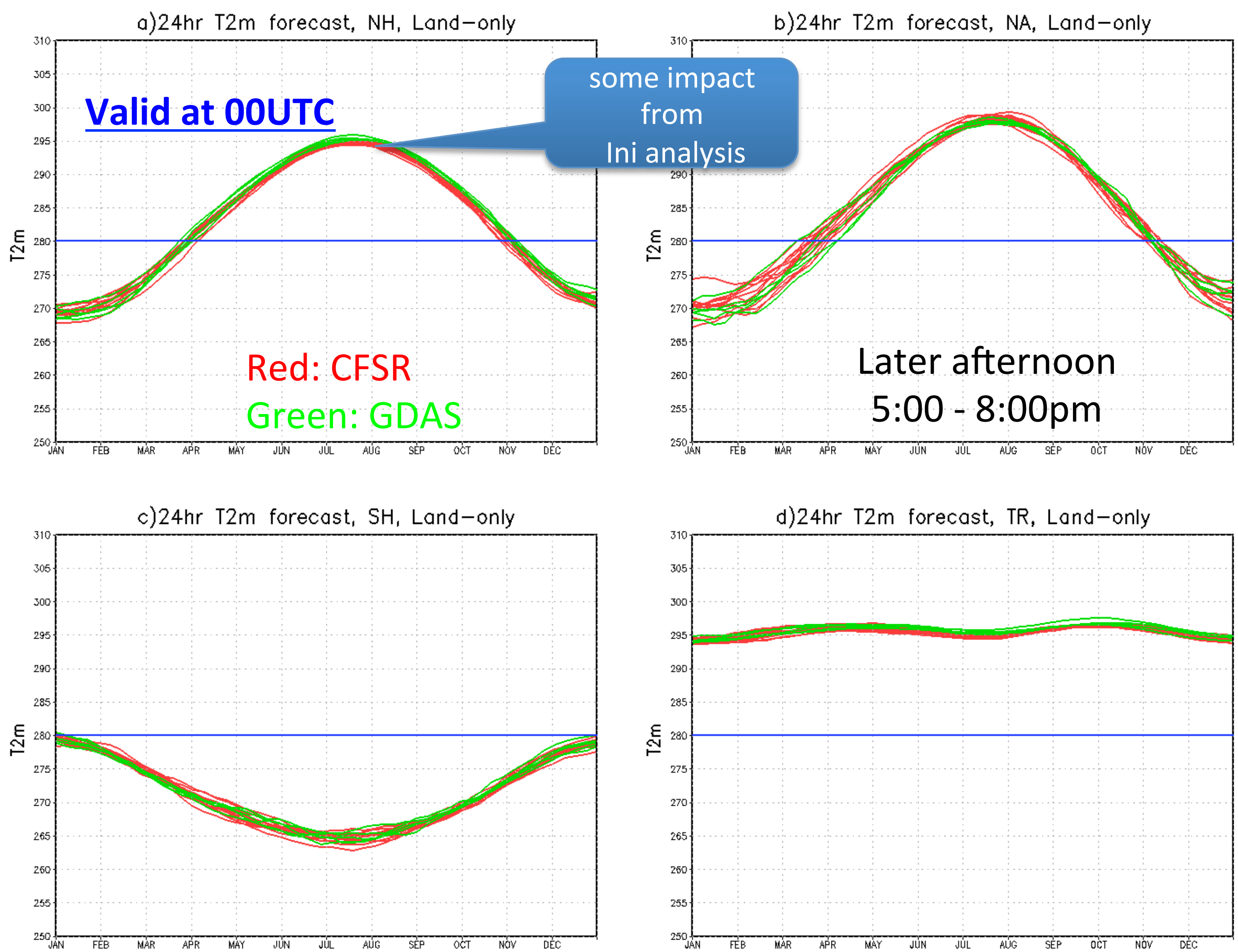


# 5-year average Bias of T2m for weeks 3&4

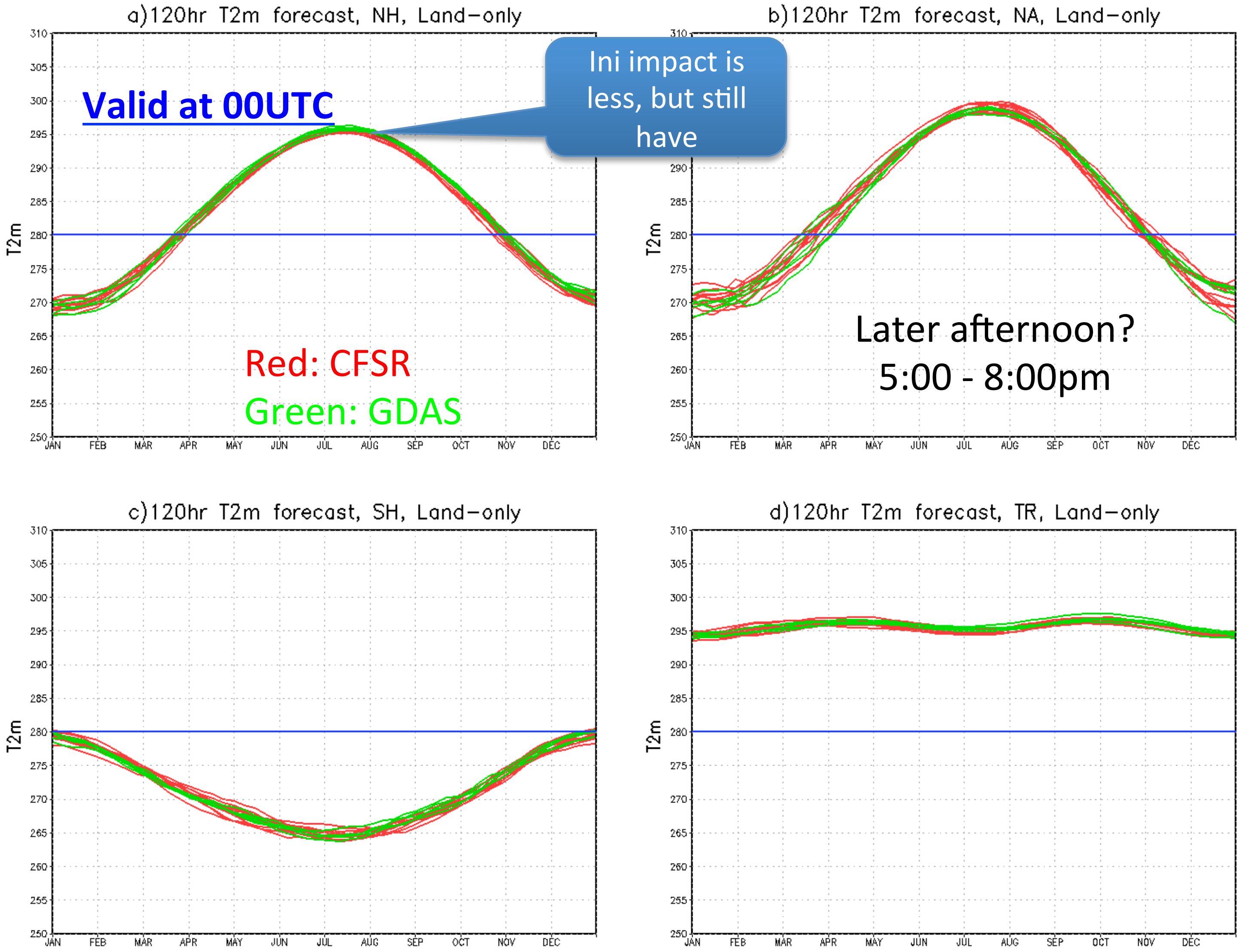


When does 2-m temperature error reach to a saturated status?

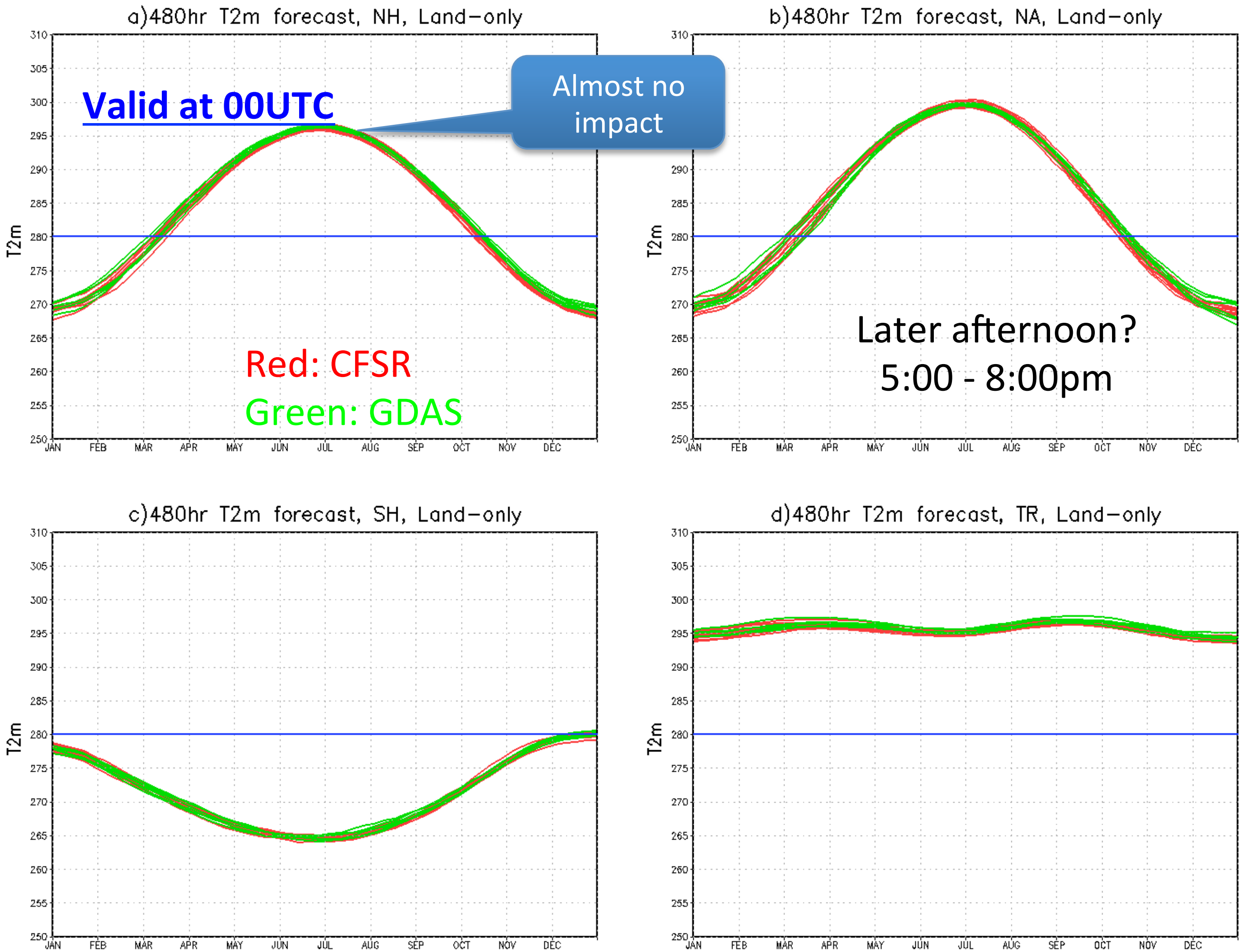
# Forecasts (24 hours) difference between CFSR (ini) and GDAS (ini)



# Forecasts (120 hours) difference between CFSR (ini) and GDAS (ini)

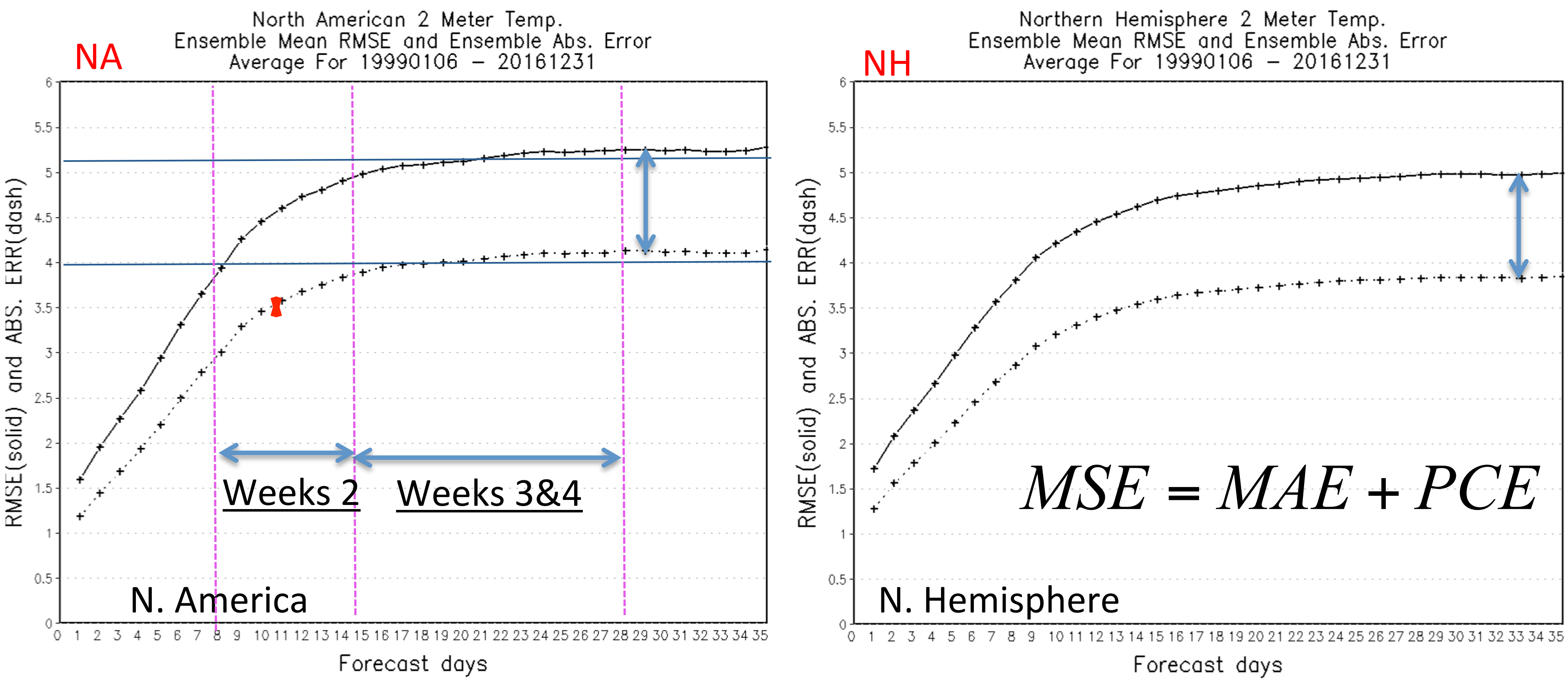


# Forecasts (480 hours) difference between CFSR (ini) and GDAS (ini)



# RMSE and MAE of T2m (18 years average)

## Day to day leads to 35 days, Land-only

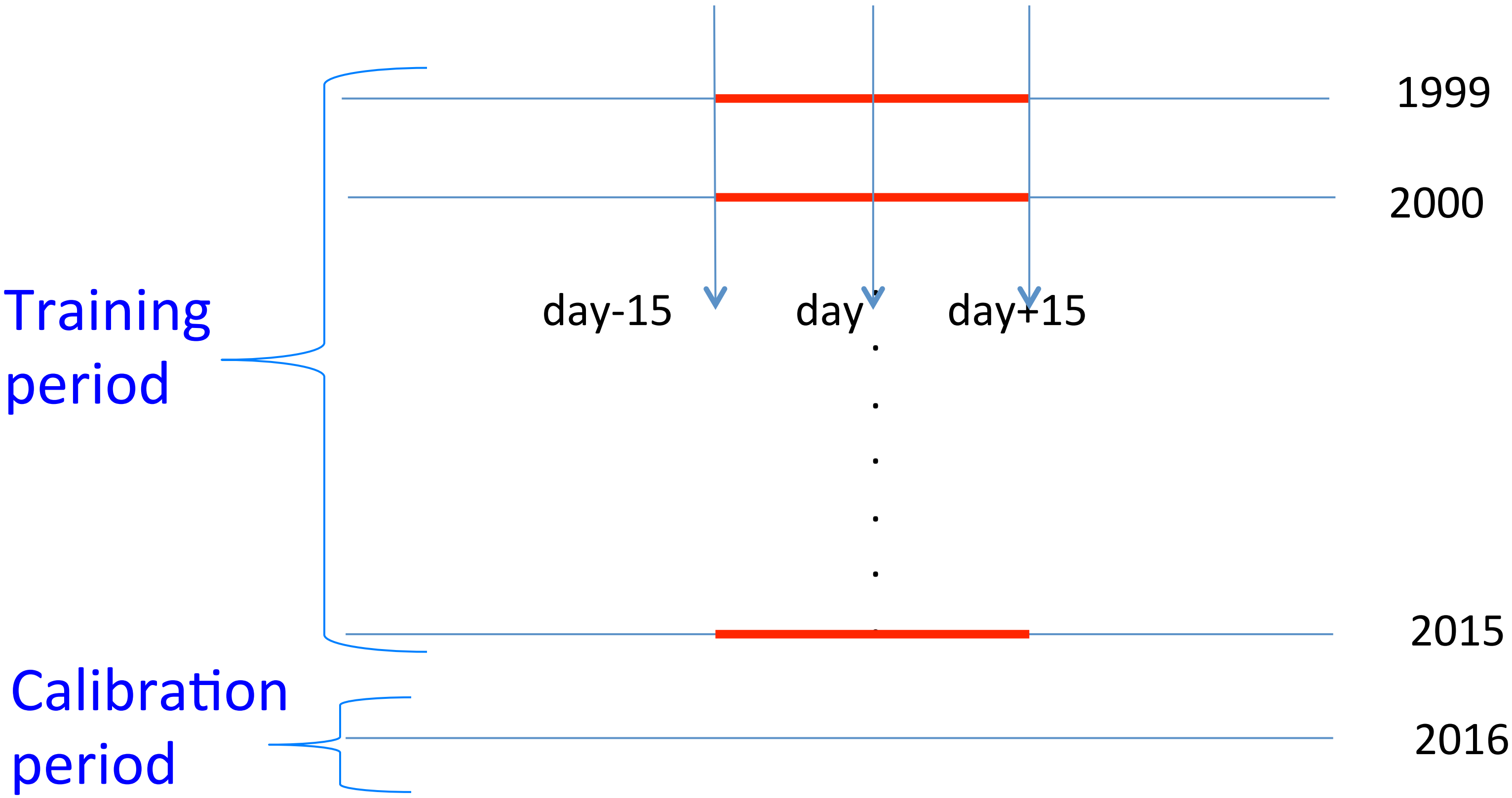



- MAE are about 75% - 80% of total error (RMSE) for NA and NH if they are in a saturated level.
- Error in NA is slightly larger than NH.
- Error of day 11 (week2) for NA is about 88% of its saturation value.

# Weeks 3&4 bias corrections

# Bias correction methodology (1)

## (Using 31day window and past 17yrs bias to calibrate 2016 forecasts)




Sample size is about 68 – 85 for each lead time and each grid point.

# Bias correction methodology (2)

Model bias in a time period

$$b_{i,j}(t) = \frac{1}{N}\sum_{k=1}^{N} \left(f_{i,j,k}(t) - a_{i,j,k}(t = t_0)\right)$$

t is valid (lead) forecast time, sample size N is about 68–85 for each lead time and each grid point.

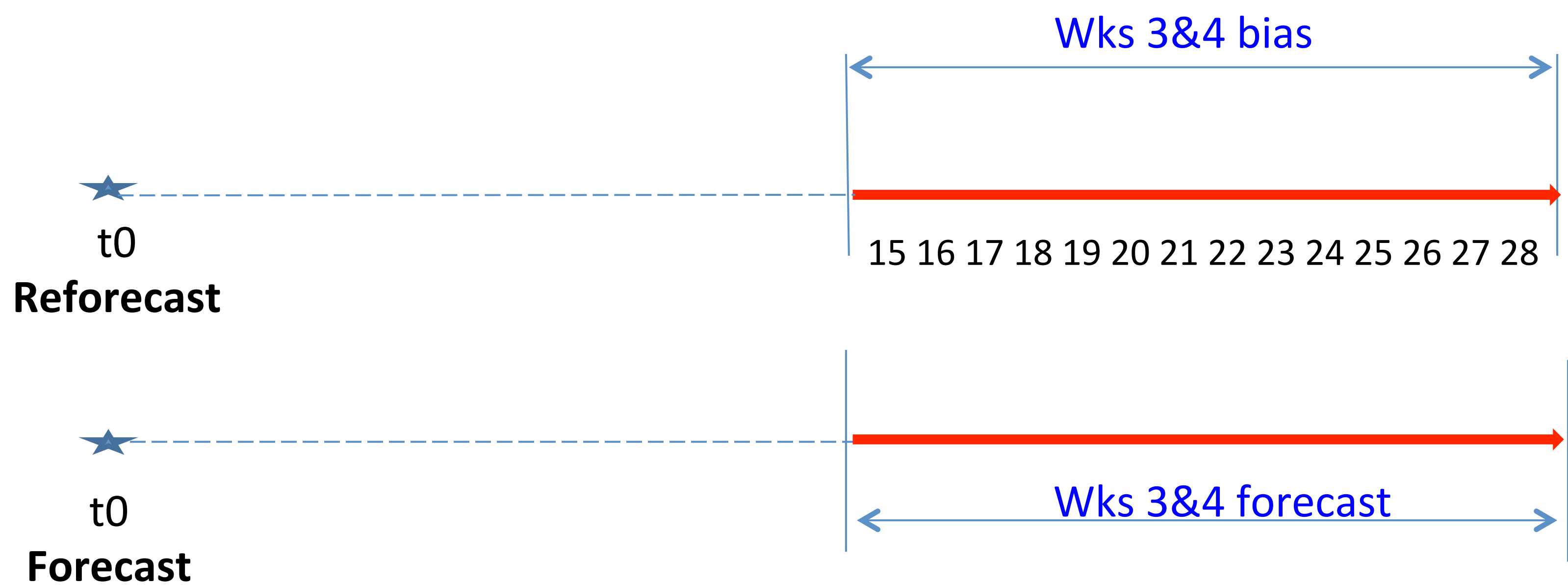
$$f_{i,j}^{bc}(t_{w34}) = f_{i,j}(t_{w34}) - b_{i,j}(t_{w34})$$

Or using week-2's bias to correct weeks 3&4 forecast

$$f_{i,j}^{bc}(t_{w34}) = f_{i,j}(t_{w34}) - b_{i,j}(t_{w2})$$

Or using adjusted bias (the same formula)

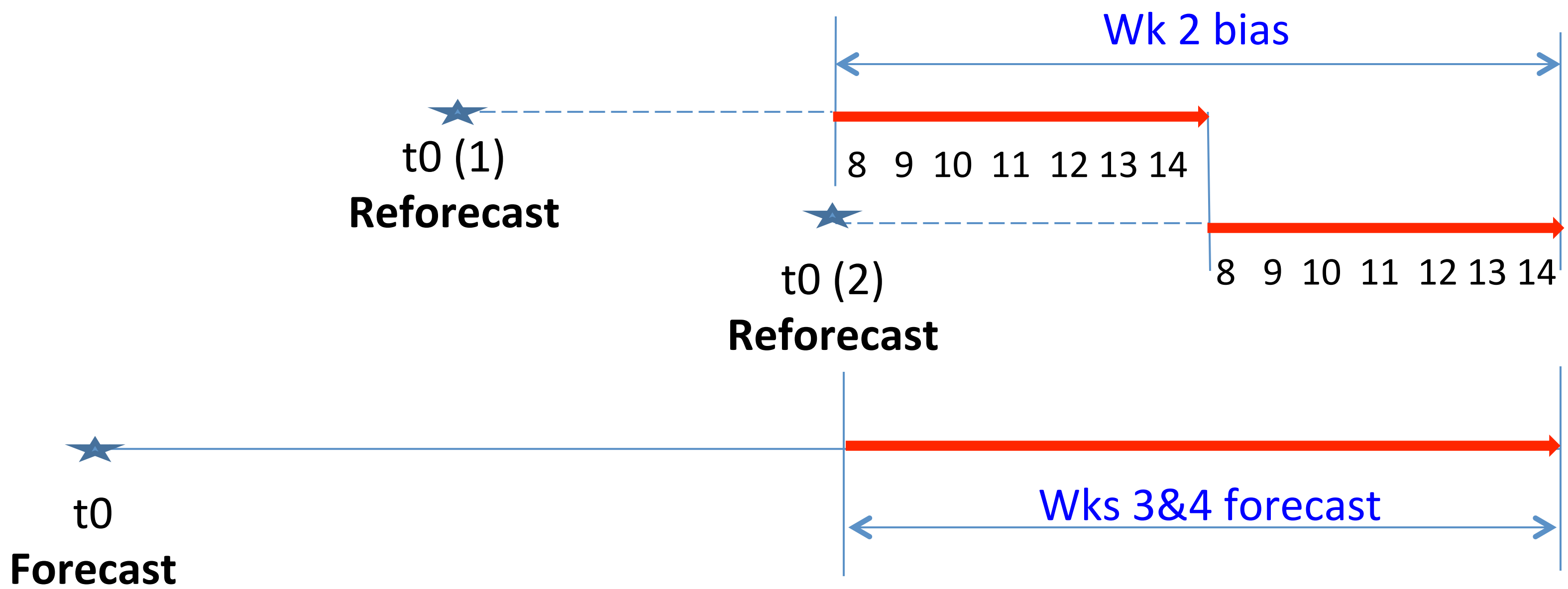
# Bias correction methodology (3)
## (Using weeks 3&4 bias to calibrate weeks 3&4 forecast)




Sample size is about 952 (68x14) to 1190 (85x14) for weeks 3&4 and each grid point.

# Bias correction methodology (4)

## (Using week 2 bias to calibrate weeks 3&4 forecast)




Sample size is about 952 (68x14) to 1190 (85x14) for week 2 and each grid point.

# Verification Metrics (1)

Mean Absolute Error (MAE)

$$MAE = \frac{1}{M * N} \sum_{i,j=1}^{i,j=M,N} | \overline{f} - a |$$

Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{M * N} \sum_{i,j=1}^{i,j=M,N} (\overline{f} - a)^2}$$

# Verification Metrics (2)

## Rank Probability Skill Score (RPSS)

$$RPSS = 1 - RPS_f / RPS_c$$

Where:

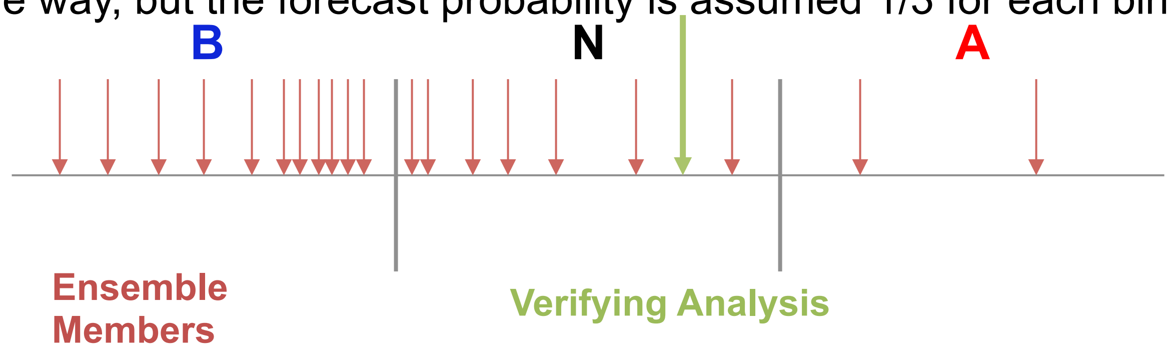
$RPS\downarrow f$ is the forecast rank probability score
$RPS\downarrow c$ is the climatological rank probability score

For each forecast-observation pair (*n*),

$$RPS_f = \frac{1}{N} \sum_{n=0}^{N} \left( (probB_n - obsB_n)^2 + (probN_n - obsN_n)^2 + (probA_n - obsA_n)^2 \right)$$

where $probX\downarrow n$ and $obsX\downarrow n$ are the ranked cumulative forecast probability and observation probability for each bin (B, N, and A). The ranked forecast probability for each bin are the cumulative number of ensemble members divided by the total number of ensemble members. The cumulative observation probability is either 1 or 0. $RPS\downarrow c$ is calculated the same way, but the forecast probability is assumed 1/3 for each bin.

**For example, for a single forecast observation pair:**




$probB\uparrow f = 11/20$ $probB\uparrow c = 1/3$

$obsB = 0$

$probB\uparrow f = 18/20$ $probB\uparrow c = 2/3$

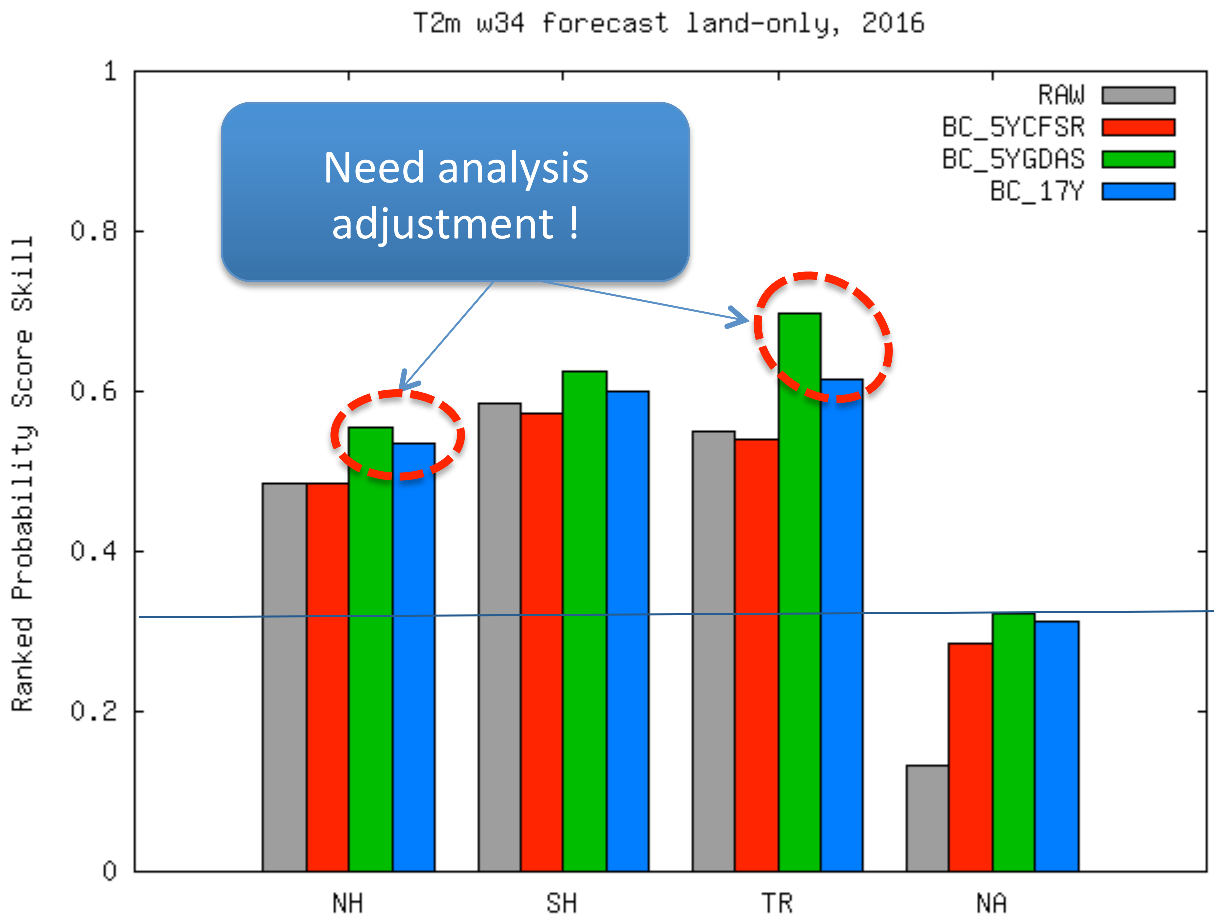
$obsB = 1$

$probA\uparrow f = 20/20$ $probA\uparrow c = 3/3$

$obsB = 1$

# What do we get for domain average scores (RPSS) before analysis adjustment?

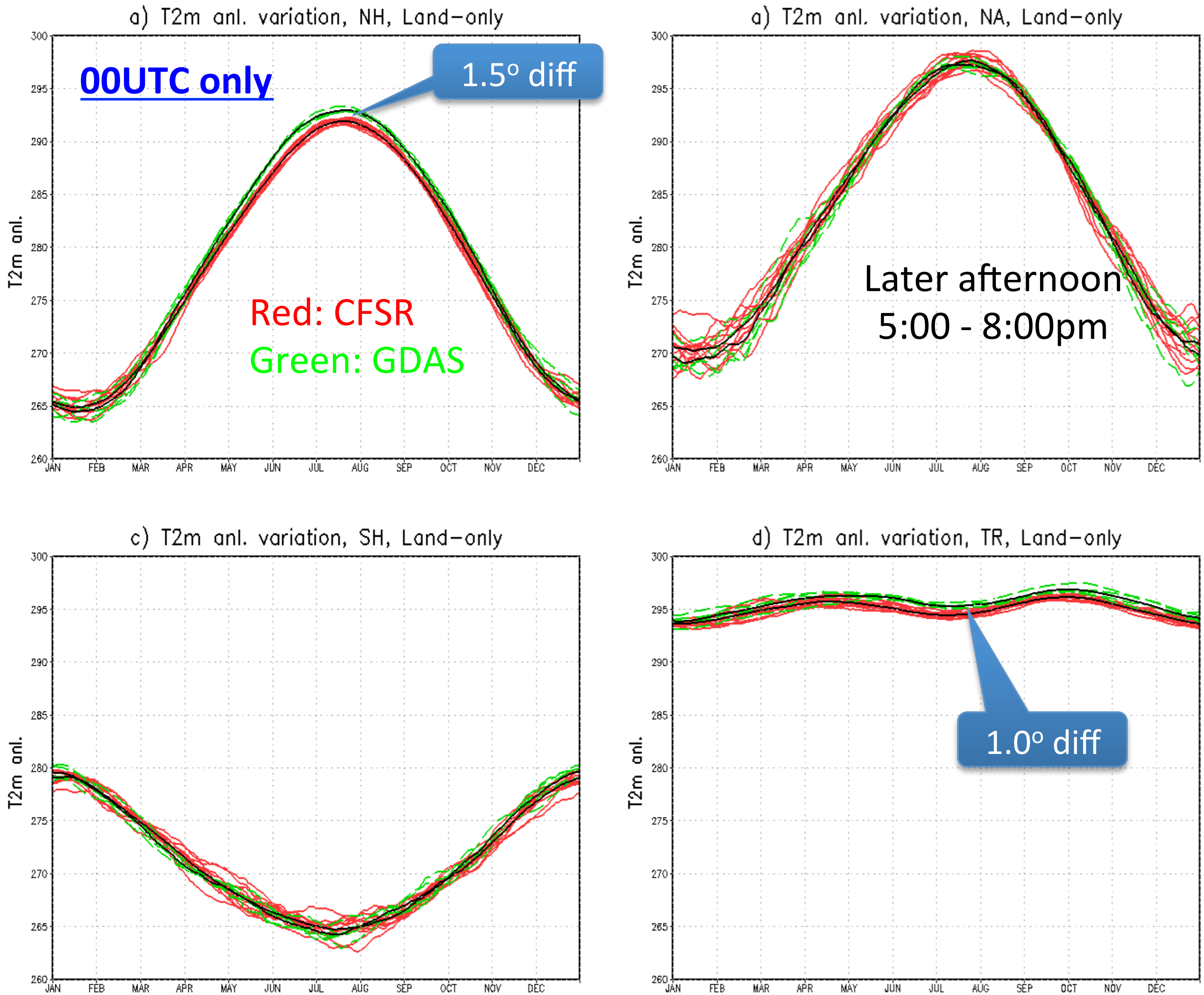
Bias correction comparison of later 5 years (2011-2015) .vs prior 5 years (2006-2010) .vs 17 years (1999 – 2015).




Clearly – bias correction has less skills from **17 years bias** comparing to later **5 years**

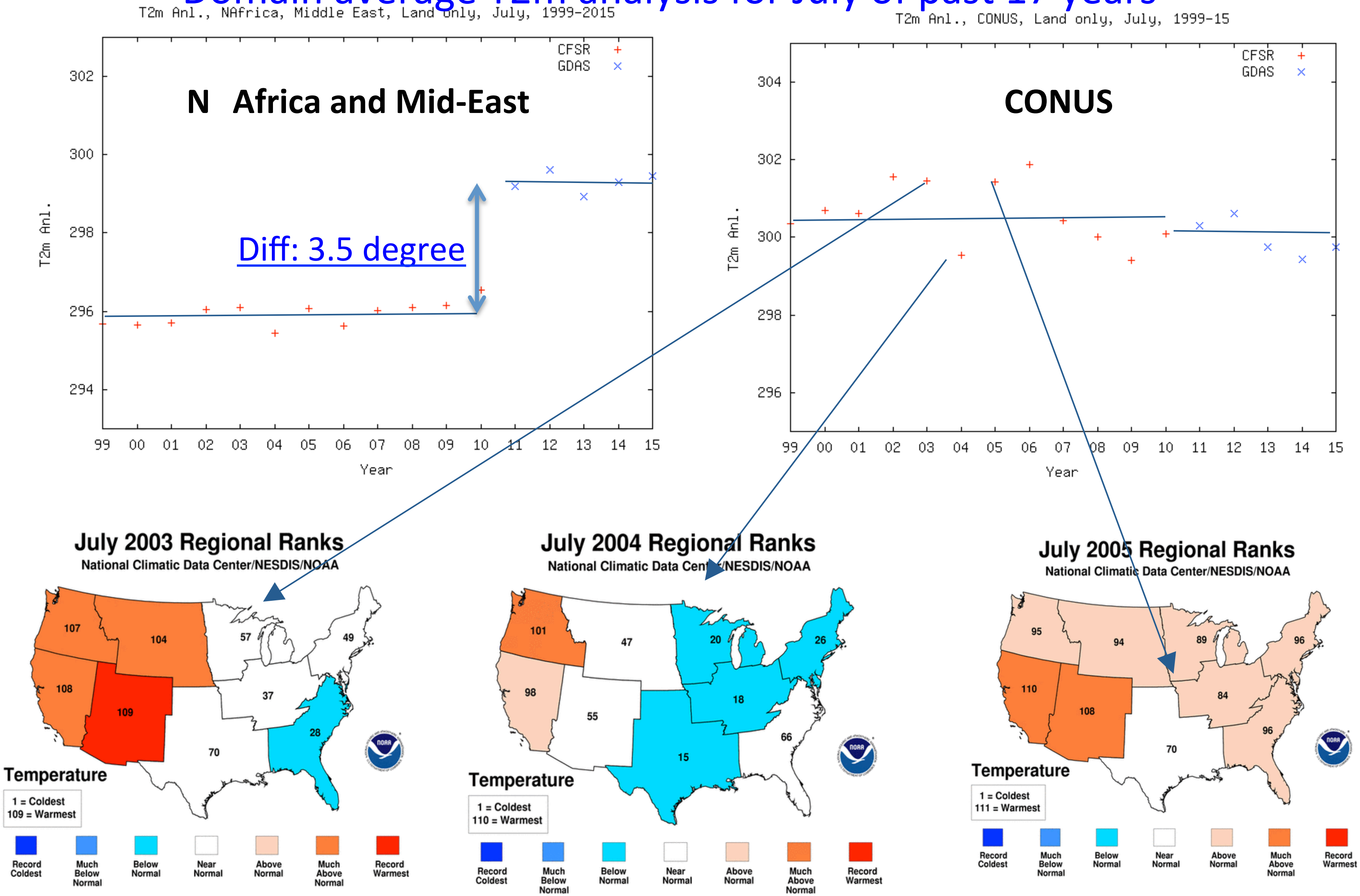
# Analysis adjustment (1)

## Domain average T2m (analysis) time series for each year (17 years)



# Analysis adjustment (2)

## Domain average T2m analysis for July of past 17 years




https://www.ncdc.noaa.gov/temp-and-precip/

# Analysis adjustment (3)

Analysis of T2m is much different from early DA to later DA due to model upgrade in May 2011, to change DA reference (background). Therefore, it is necessary to make some adjustment for early analysis

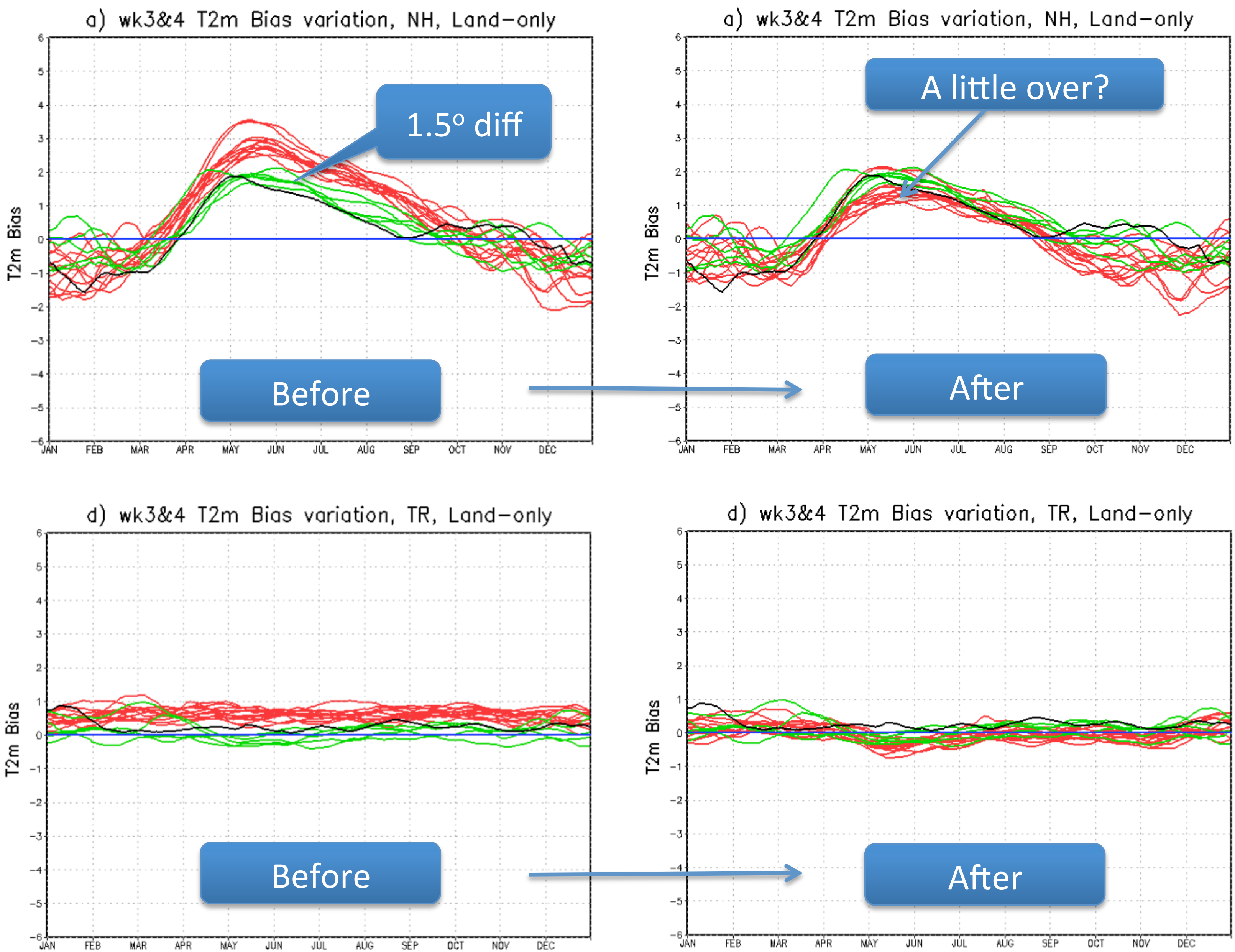
When we have 12-year (1999-2010) average and 5-year (2011-2015) T2m

Difference could be $$a'_{i,j} = a^{12y}_{i,j} - a^{5y}_{i,j}$$

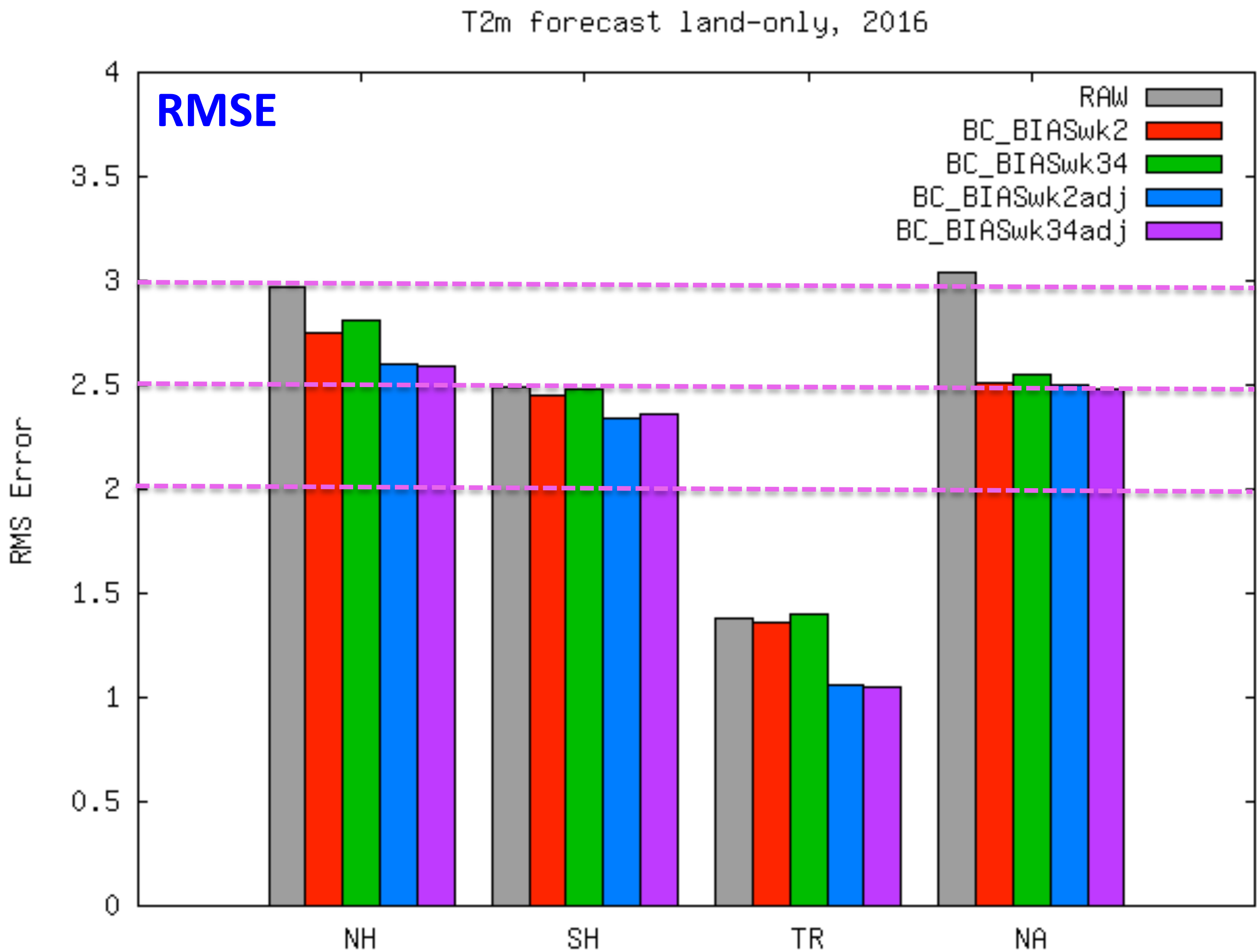
Then, apply this difference to first 12 years analysis

$$a^{adj}_{i,j} = a_{i,j} - a'_{i,j}$$

# Analysis adjustment (5)

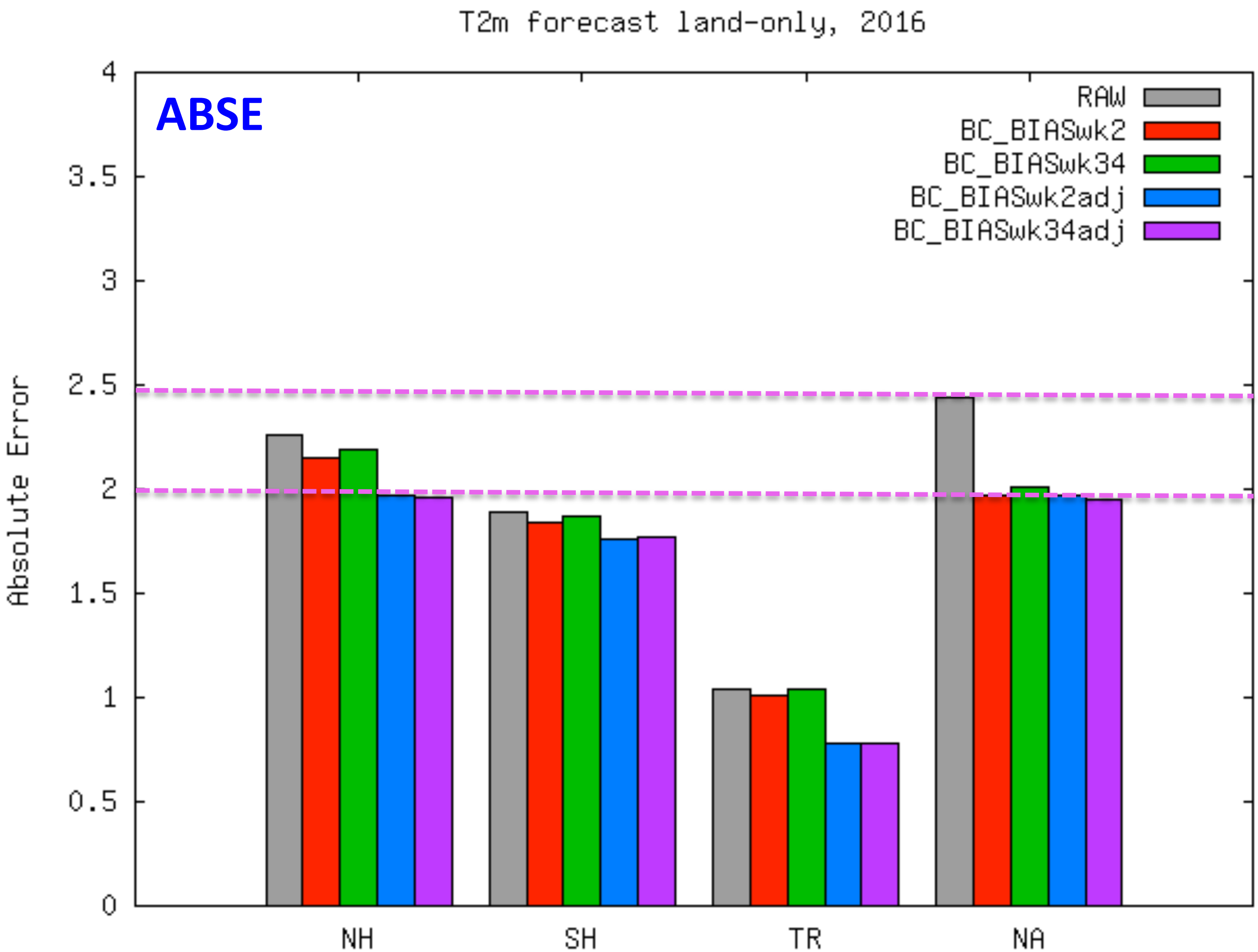


# Verification of Weeks 3&4 bias correction (1)



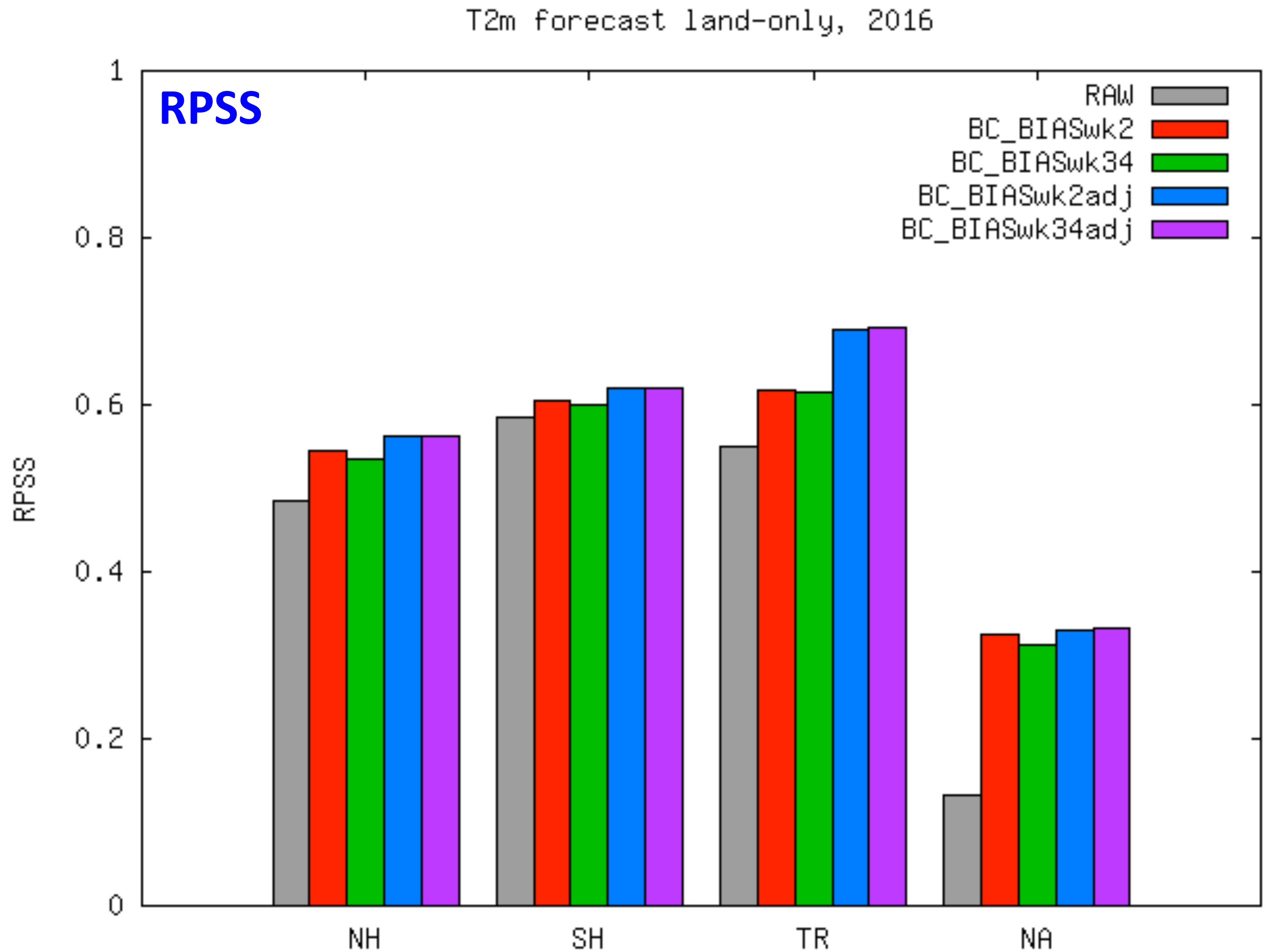

RMS errors are reduced after bias correction. Analysis adjustment is excellent for NH and tropical, but not for NA. Nearly 20% errors are removed for NA through various process

# Verification of Weeks 3&4 bias correction (2)




Bias correction is important to all domains (land only) Week-2 bias could correct weeks 3&4 forecast After adjustment – NA has similar errors as NH

# Verification of Weeks 3&4 bias correction (3)

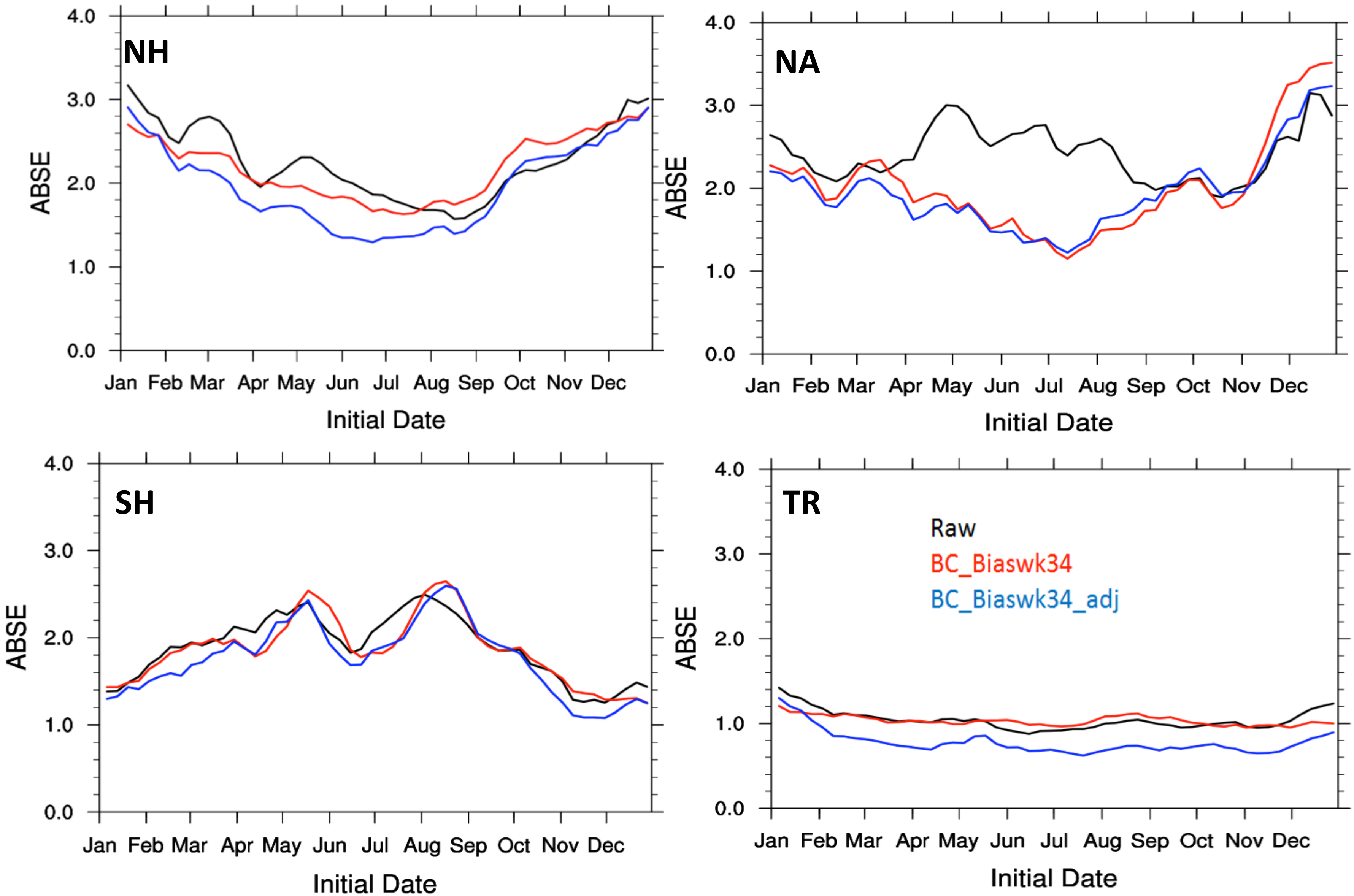



Bias correction improves weeks 3&4 forecast skills for all domains

Biggest improvement is for NA, but it is still lowest skill because NA's bias variance is much larger than other domain– **less predictability**

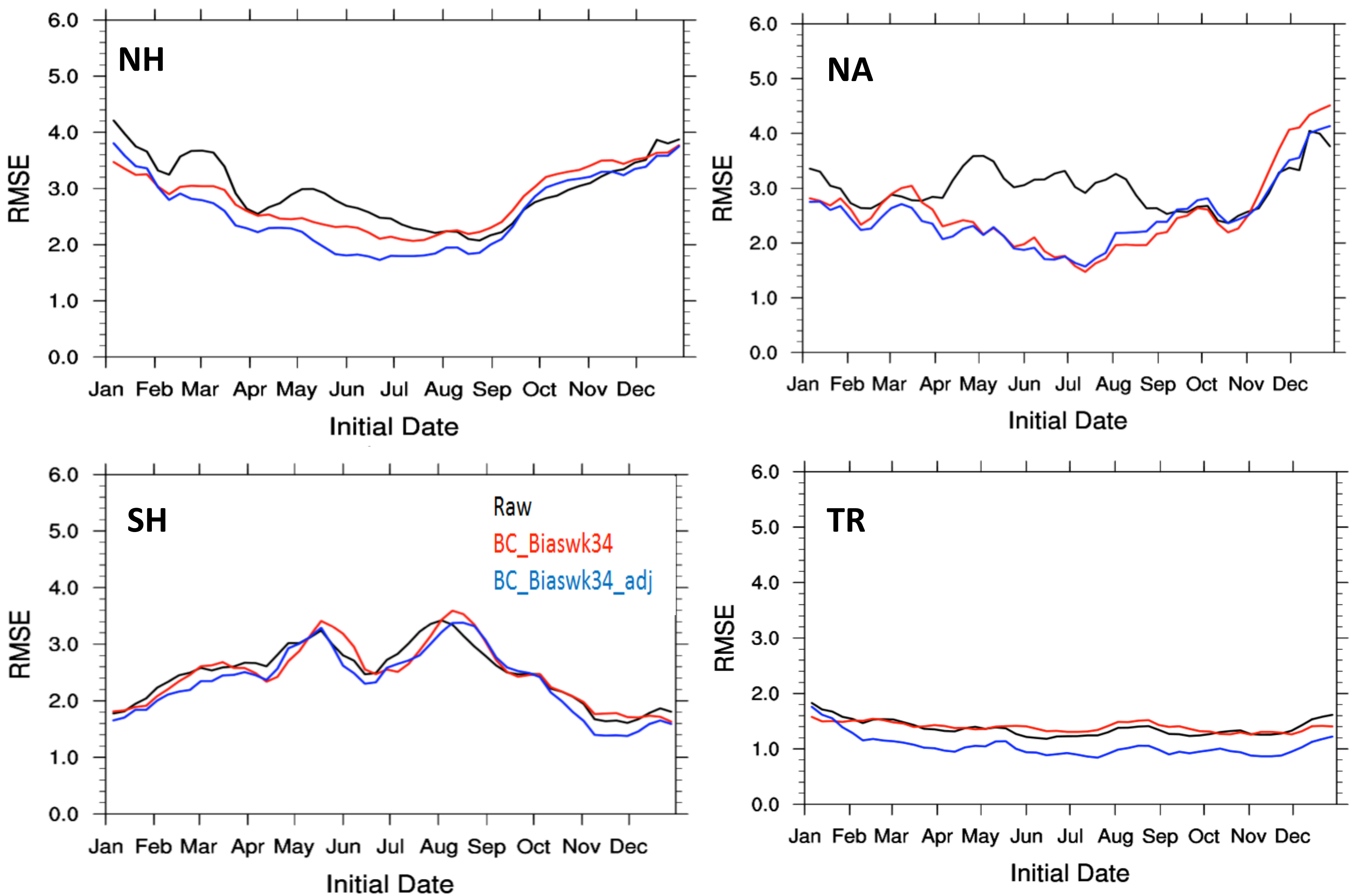
# Verification of Weeks 3&4 bias correction (4)

(ABSE, 2016)



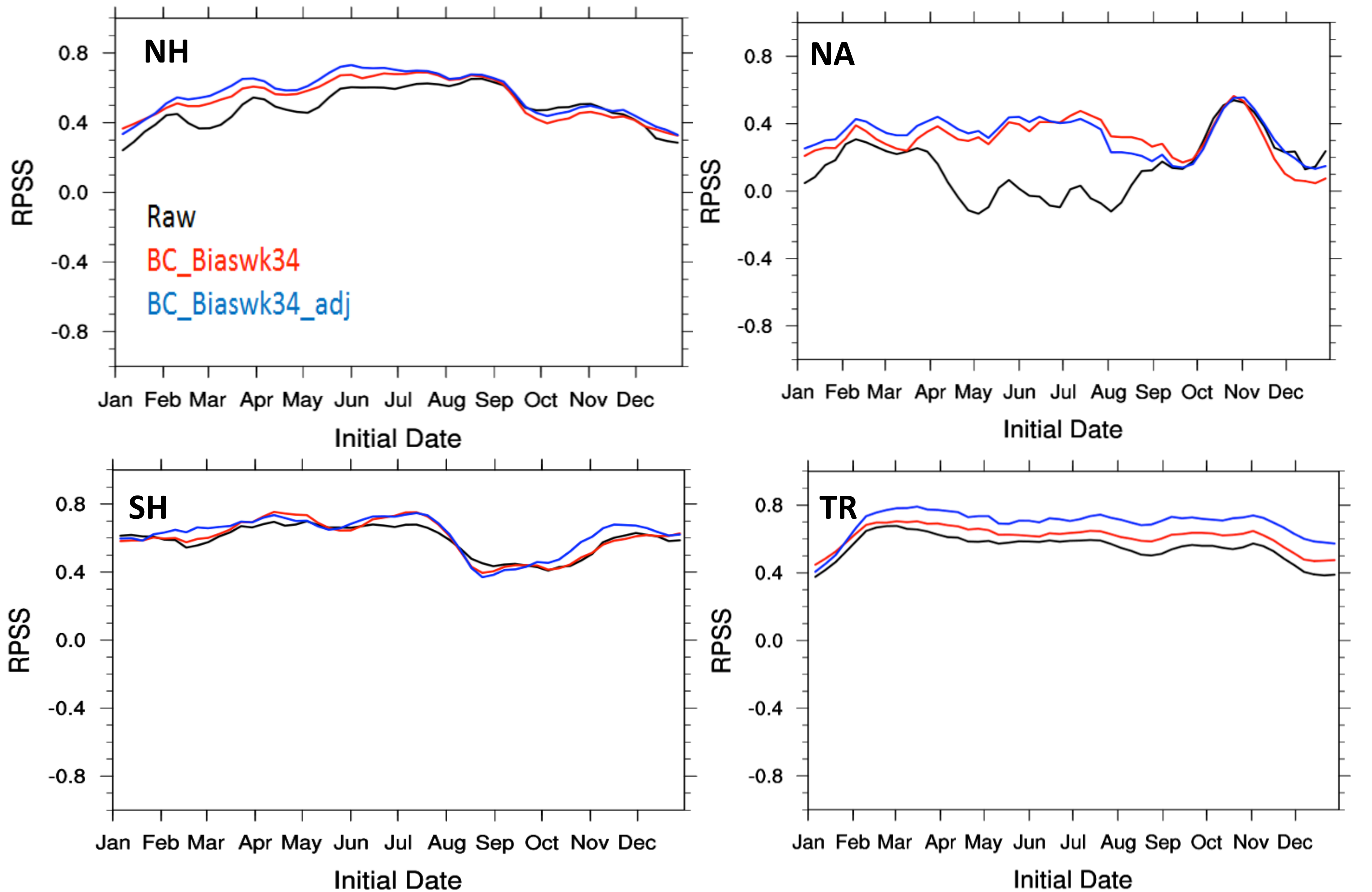
# Verification of Weeks 3&4 bias correction (5)

(RMSE, 2016)

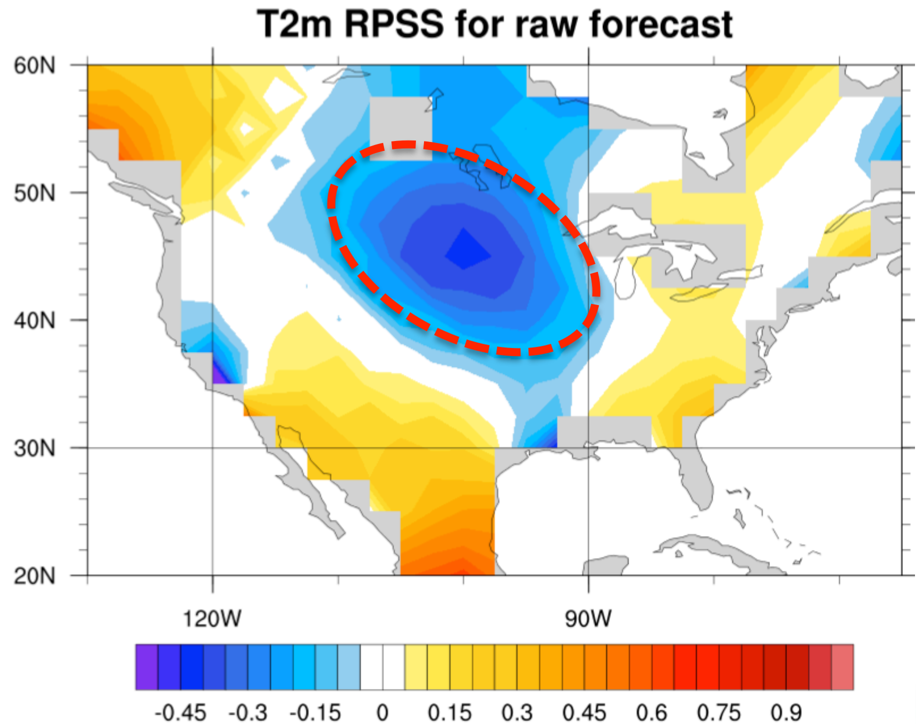


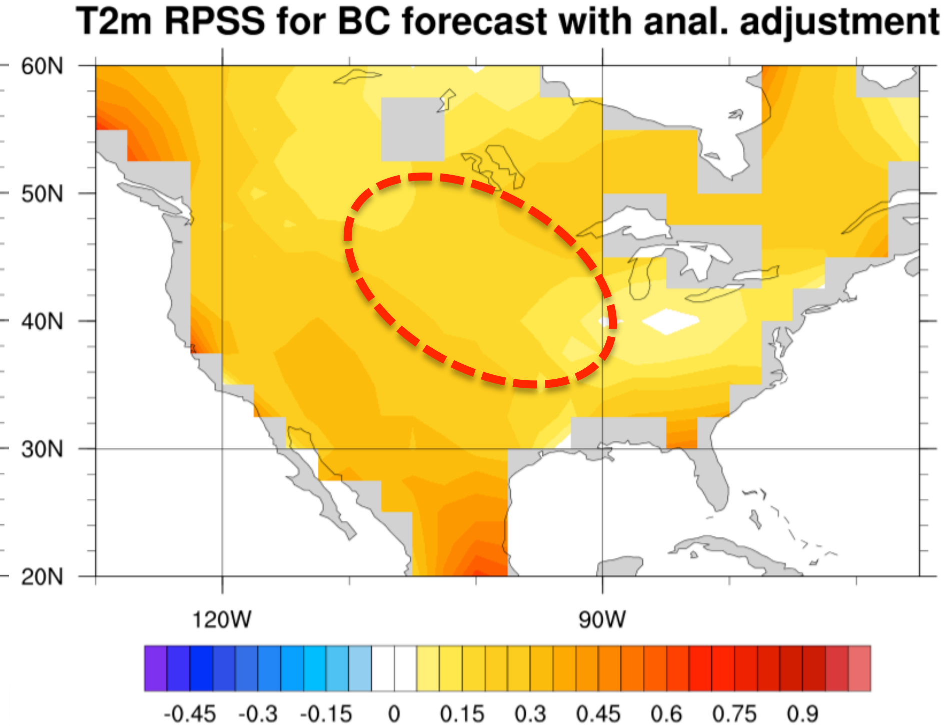
# Verification of Weeks 3&4 bias correction (6)

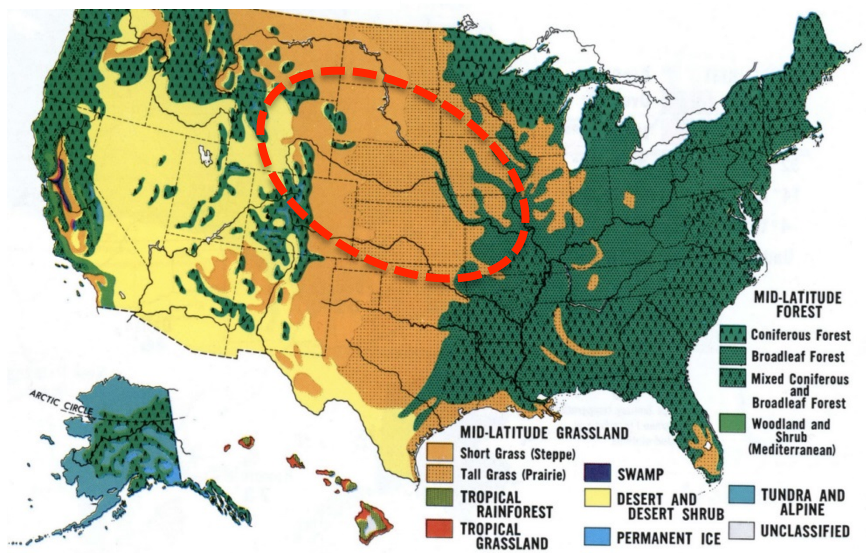
(RPSS, 2016)



# Bias correction improves Wks 3&4 forecast skill (2016)

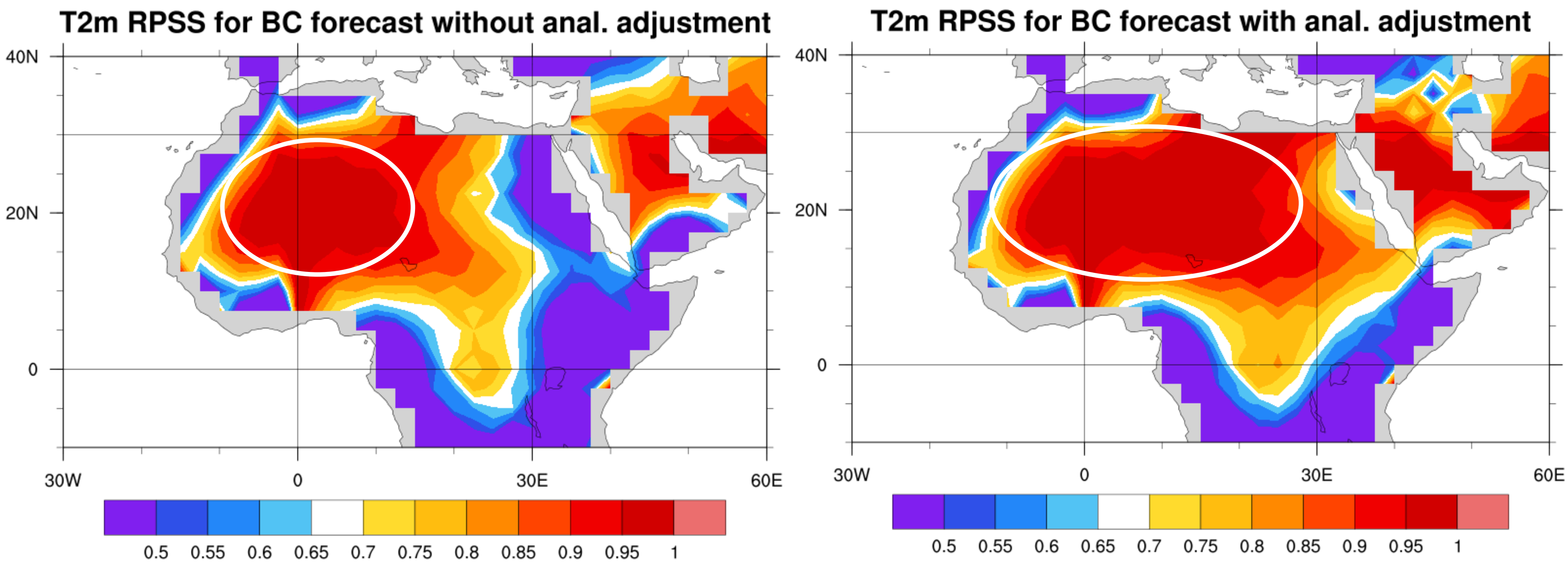


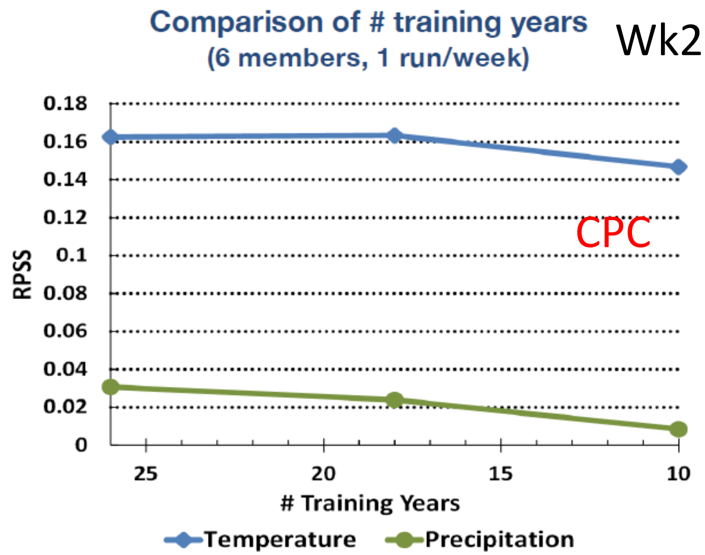



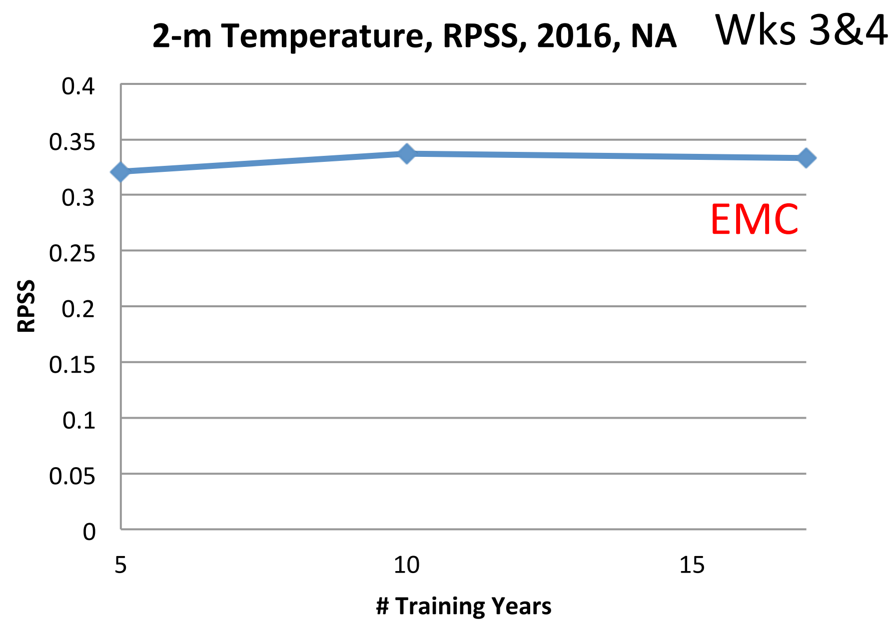
# Analysis adjust. improves Wks 3&4 forecast skill

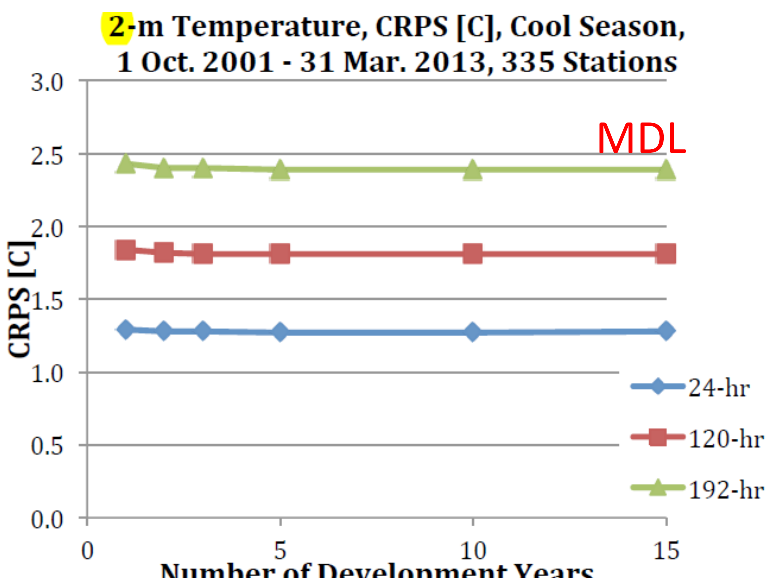
## (May-Sept, 2016)



# Sensitivity of T2m bias correction on the number of training years




Forecast skill (RPSS or CRPS) for 2-m temperature as function of the number of years of training data for the three studies.

Courtesy of Dr. Tom Hamill

# Summary

- Initial conditions
  - Analysis – consistent analysis is very important to generate reforecast and real-time forecast
    - Post-adjust analysis could help to improve the skills
  - Weeks 3&4 bias has less impact from initial condition

- Characters of model systematic errors
  - Forecast systematic errors could be impacted from initial condition
  - Large year-to-year variation of T2m bias for NA land
  - Year-to-year variation is lager for cold season than warm season for both NH and NA land
  - Large warm bias for warm season of NA

- For week 3&4 time range
  - Forecasts are strong biased, and bias is mostly saturated
  - Bias correction is very important, reduce errors and increase skills
  - Longer historical reforecast could improve scores if initial analyses are consistent.
  - Using week-2's bias could correct weeks 3&4's forecast
    - This could save huge computation resource to increase reforecast samples
    - (both frequency and ensemble size).