Statistical Post-Processing for NAEFS

Bo Cui[1], Yuejian Zhu[2], Zoltan Toth[3],
Stéphane Beauregard[3]

[1]SAIC at Environmental Modeling Center, NCEP/NWS
[2]Environmental Modeling Center, NCEP/NWS
[3]Earth System Research Laboratory, NOAA
[4]Canadian Meteorological Centre, Meteorological Service of Canada

Acknowledgements

Richard Verret, Poulin Lewis CMC/MSC
Dingchen Hou EMC/NCEP/NWS/NOAA
David Michaud, Brent Gorden, Luke Lin NCO/NCEP/NWS/NOAA
Valery J. Dagostaro MDL/NWS/NOAA

May 17th 2010, 5th NAEFS workshop

Outline

- Background
  - Current NCEP/EMC post-processing system for NAEFS
- Statistical Post-Processing Techniques in NAEFS
  - Effect of bias-correction
  - Impact of combined ensemble (NAEFS)
  - Statistical downscaling for CONUS
  - Statistical downscaling techniques for Alaska
- Summary and future Plan

NCEP/EMC Statistical Post-Processing for NAEFS

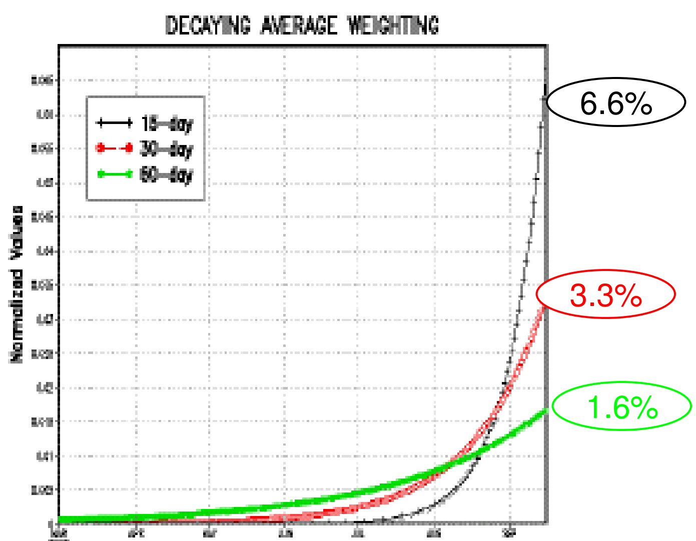
- Statistical Post-Processing Issues in NAEFS
  - GOAL
    - Improve reliability while maintaining resolution in NWP forecasts
      - Reduce systematic errors (improve reliability) while
      - Not increasing random errors (maintaining resolution)
        - Retain all useful information in NWP forecast
  - APPROACH – Computational efficiency
    - Bias Correction : remove lead-time dependent bias on <u>model grid</u>
      - Working on coarser model grid allows use of more complex methods
      - Feedback on systematic errors to model development
    - Downscaling: downscale bias-corrected forecast to <u>finer grid</u>
      - Further refinement/complexity added
        - No dependence on lead time
- Current NAEFS Statistical Post-Processing System
  - Bias corrected NCEP/CMC GEFS and GFS forecast (up to 180 hrs), same bias correction algorithm
    - Combine bias corrected GFS and NCEP GEFS ensemble forecasts
    - Dual resolution ensemble approach for short lead time
    - GFS has higher weights at short lead time
  - NAEFS products
    - Combine NCEP/GEFS (20m) and CMC/GEFS (20m), FNMOC ens. will be in soon
    - Produce Ensemble mean, spread, mode, 10% 50%(median) and 90% probability forecast at 1*1 degree resolution
    - Climate anomaly (percentile) forecasts also generated for ens. mean
  - Statistical downscaling
    - Use RTMA as reference - NDGD resolution (5km), CONUS only
    - Generate mean, mode, 10%, 50%(median) and 90% probability forecasts

Bias Correction Method & Application

- **Bias Correction Techniques** – array of methods
  - Estimate/correct bias moment by moment
    - Simple approach, implemented partially
    - May be less applicable for extreme cases
- **Moment-based method at NCEP**: apply adaptive (Kalman Filter type) algorithm

decaying averaging mean error = (1-*w*) * prior a.m.e + *w* * (f – a)

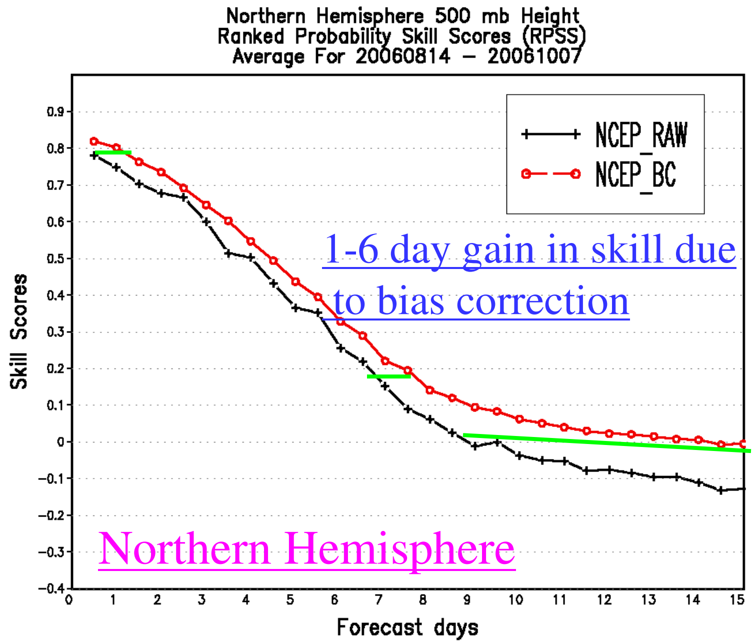
For separated cycles, each lead time and individual grid point, a.m.e = averaging mean error

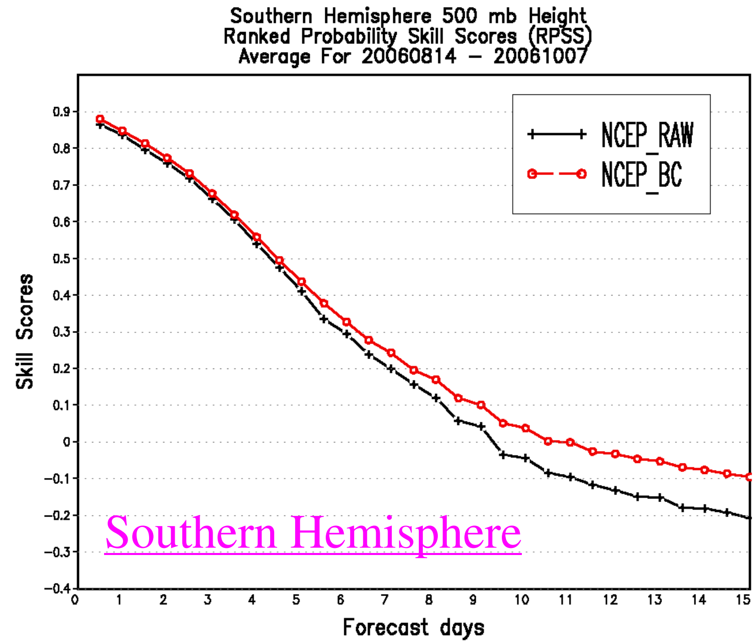


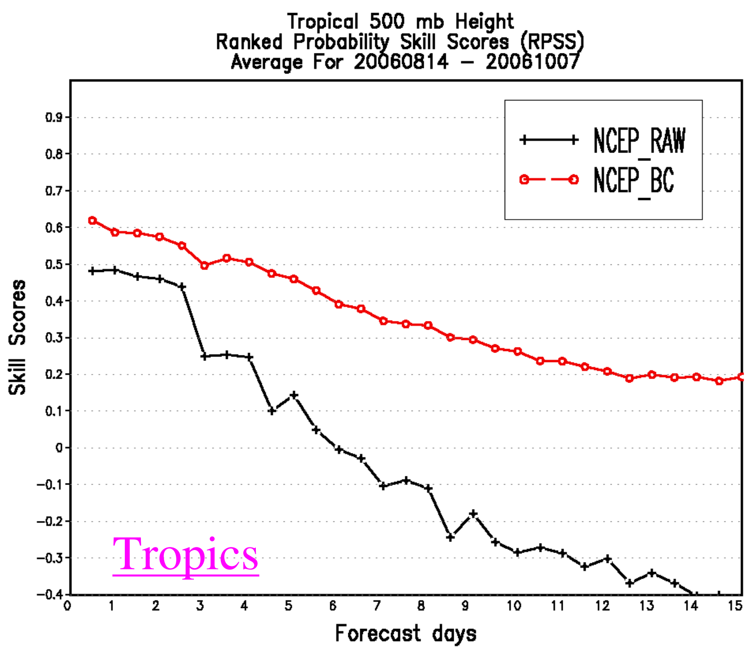
Toth, Z., and Y. Zhu, 2001

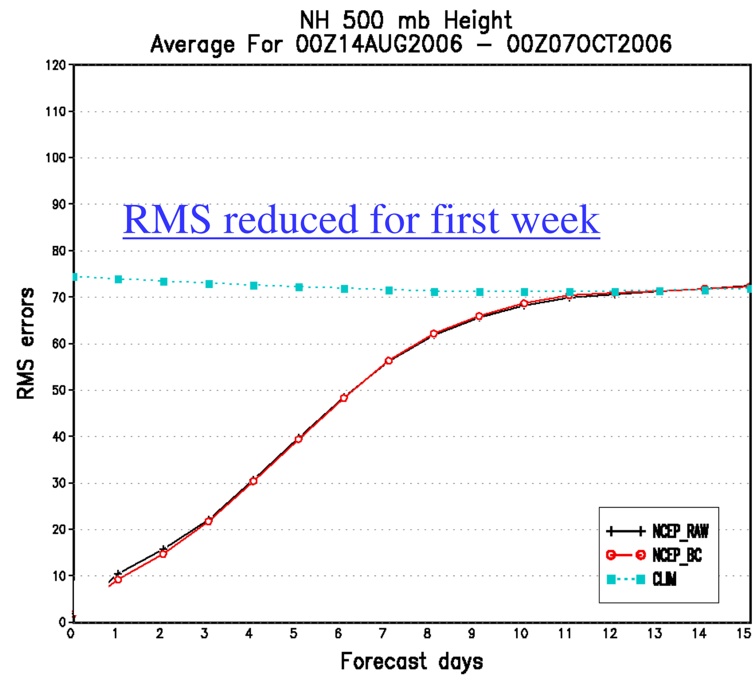
- Test different decaying weights. 0.25%, 0.5%, 1%, 2%, 5% and 10%, respectively
- Decide to use 2% (~ 50 days) decaying accumulation bias estimation

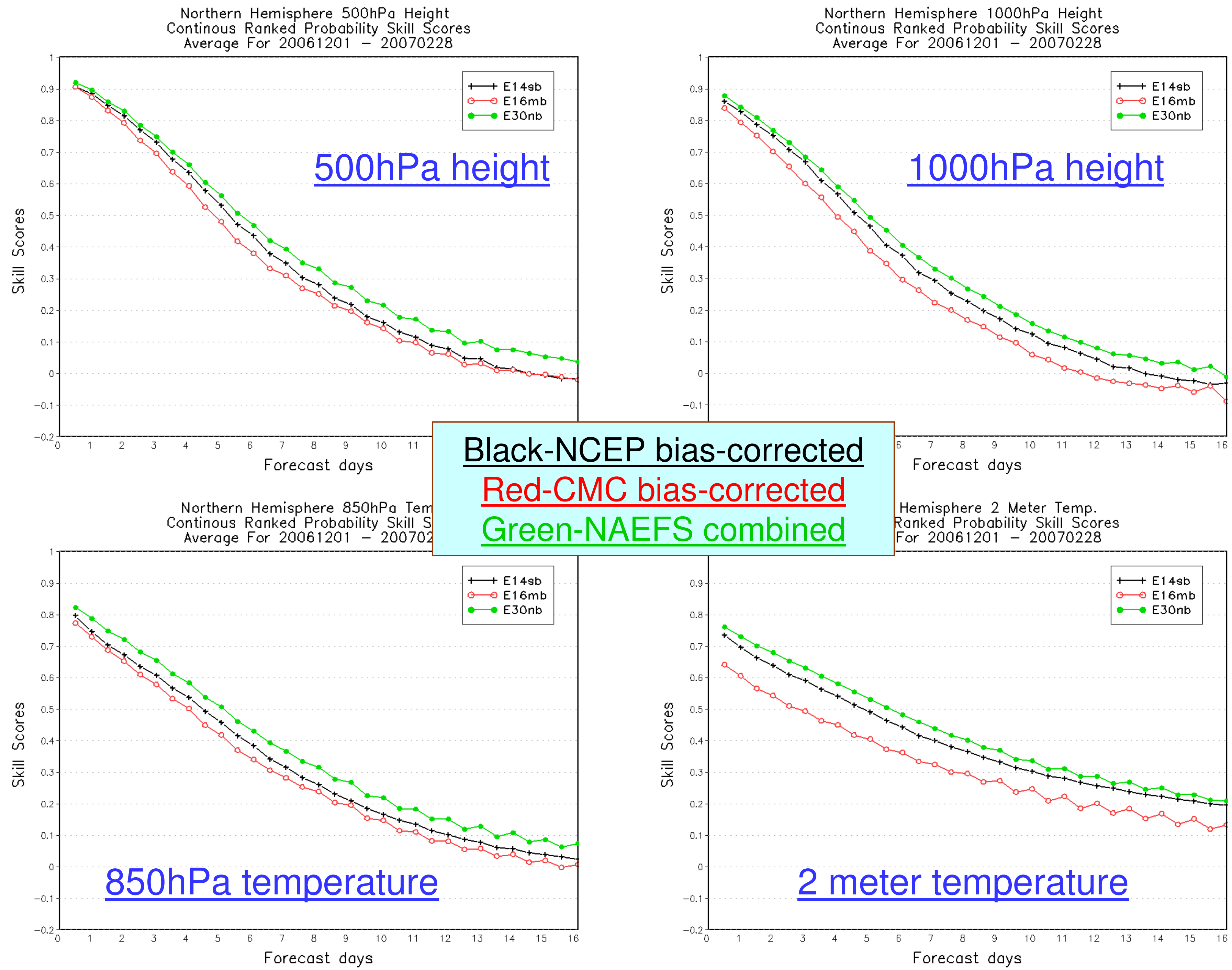
RPSS Before/After Bias Correction (NCEP 500 mb Height)





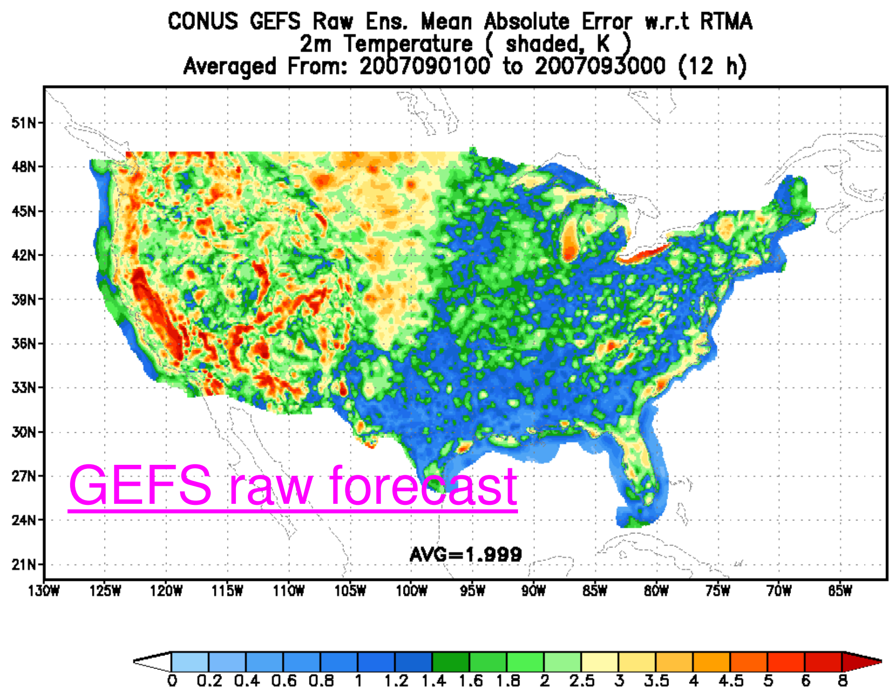


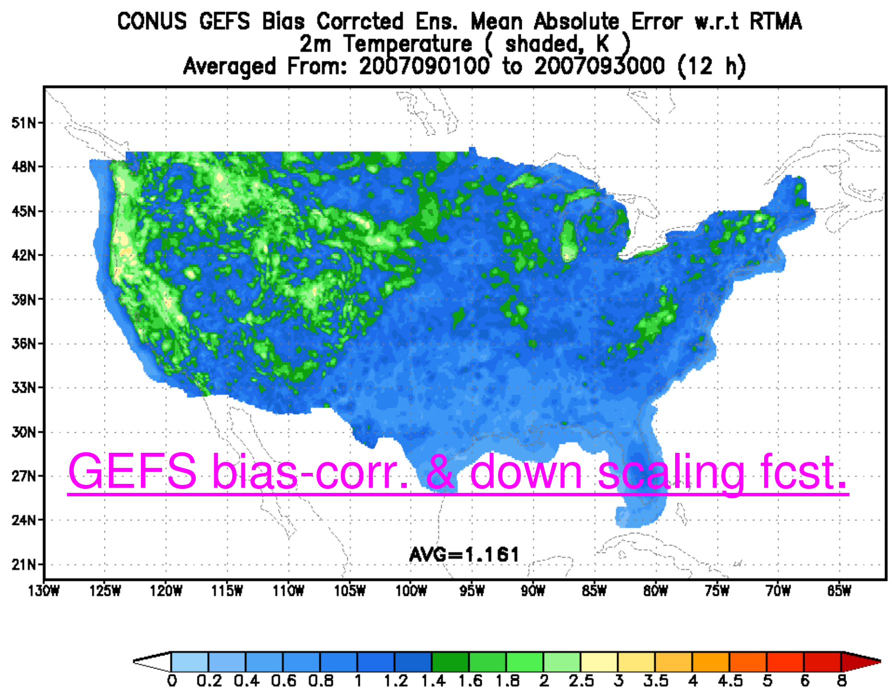


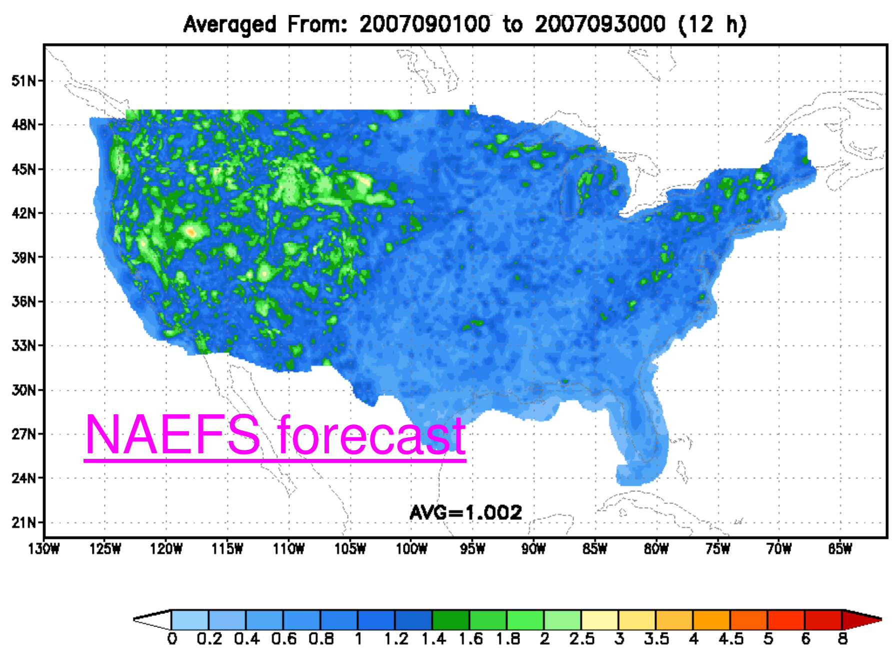

Northern Hemisphere 500hPa Height
Continous Ranked Probability Skill Scores
Average For 20061201 – 20070228
E14sb
E16mb
E30nb
500hPa height
Skill Scores
Forecast days
Northern Hemisphere 1000hPa Height
Continous Ranked Probability Skill Scores
Average For 20061201 – 20070228
E14sb
E16mb
E30nb
1000hPa height
Skill Scores
Forecast days
Black-NCEP bias-corrected
Red-CMC bias-corrected
Green-NAEFS combined
Northern Hemisphere 850hPa Tem
Continous Ranked Probability Skill S
Average For 20061201 – 200702
E14sb
E16mb
E30nb
850hPa temperature
Skill Scores
Forecast days
Hemisphere 2 Meter Temp.
Ranked Probability Skill Scores
For 20061201 – 20070228
E14sb
E16mb
E30nb
2 meter temperature
Skill Scores
Forecast days

Statistical downscaling for NAEFS forecast

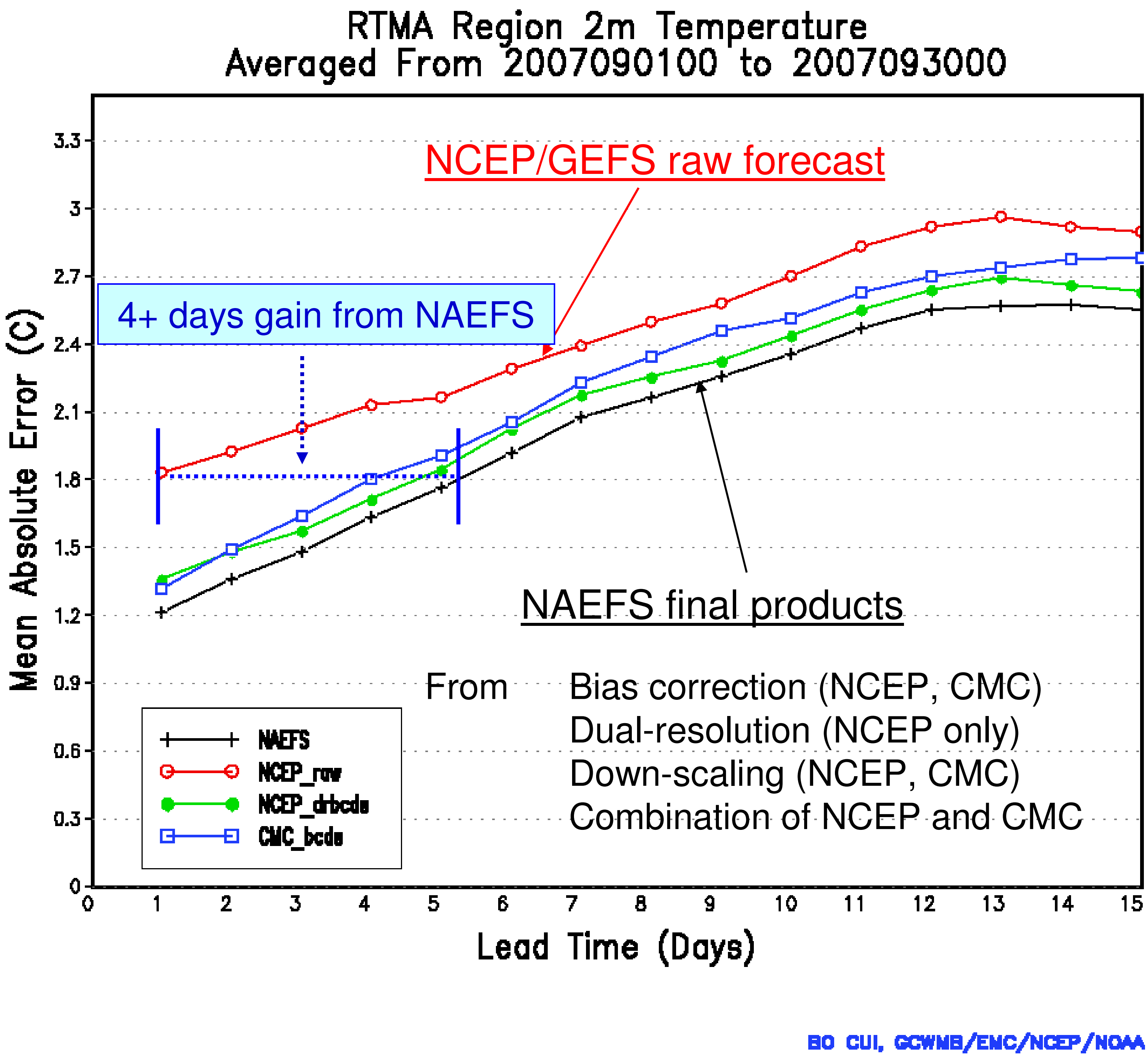
- Proxy for truth
  - *RTMA* at 5km resolution
  - Variables (surface pressure, 2-m temperature, and 10-meter wind)
- Downscaling vector
  - Interpolate GDAS analysis to 5km resolution
  - Compare difference between interpolated GDAS and RTMA
  - Apply *decaying weight* to accumulate this difference – *downscaling vector*
- Downscaled forecast
  - Interpolate bias corrected 1*1 degree NAEFS to 5km resolution
  - Add the downscaling vector to interpolated NAEFS forecast
- Application
  - Ensemble mean, mode, 10%, 50%(median) and 90% forecasts

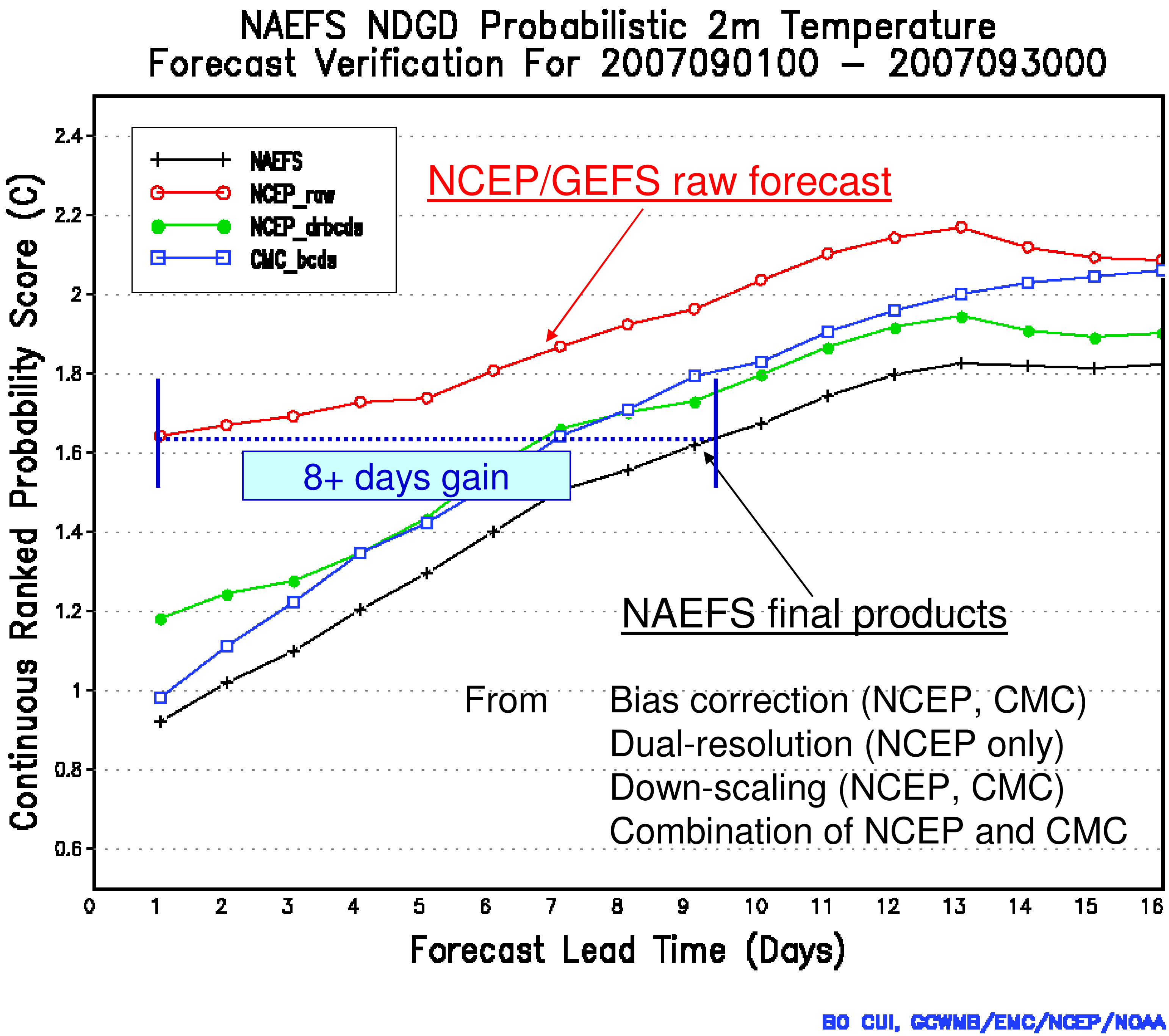






12hr 2m Temperature Forecast Mean Absolute Error w.r.t RTMA for CONUS Average for September, 2007


RTMA Region 2m Temperature
Averaged From 2007090100 to 2007093000
NCEP/GEFS raw forecast
4+ days gain from NAEFS
NAEFS final products
From
Bias correction (NCEP, CMC)
Dual-resolution (NCEP only)
Down-scaling (NCEP, CMC)
Combination of NCEP and CMC
NAEFS
NCEP_raw
NCEP_drbcds
CMC_bcds
Mean Absolute Error (C)
Lead Time (Days)
0
0.3
0.6
0.9
1.2
1.5
1.8
2.1
2.4
2.7
3
3.3
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
BO CUI, GCWMB/EMC/NCEP/NOAA



NAEFS NDGD Probabilistic 2m Temperature
Forecast Verification For 2007090100 – 2007093000
NAEFS
NCEP_raw
NCEP_drbcds
CMC_bcds
NCEP/GEFS raw forecast
8+ days gain
NAEFS final products
From
Bias correction (NCEP, CMC)
Dual-resolution (NCEP only)
Down-scaling (NCEP, CMC)
Combination of NCEP and CMC
Continuous Ranked Probability Score (C)
2.4
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Forecast Lead Time (Days)
BO CUI, GCWMB/EMC/NCEP/NOAA

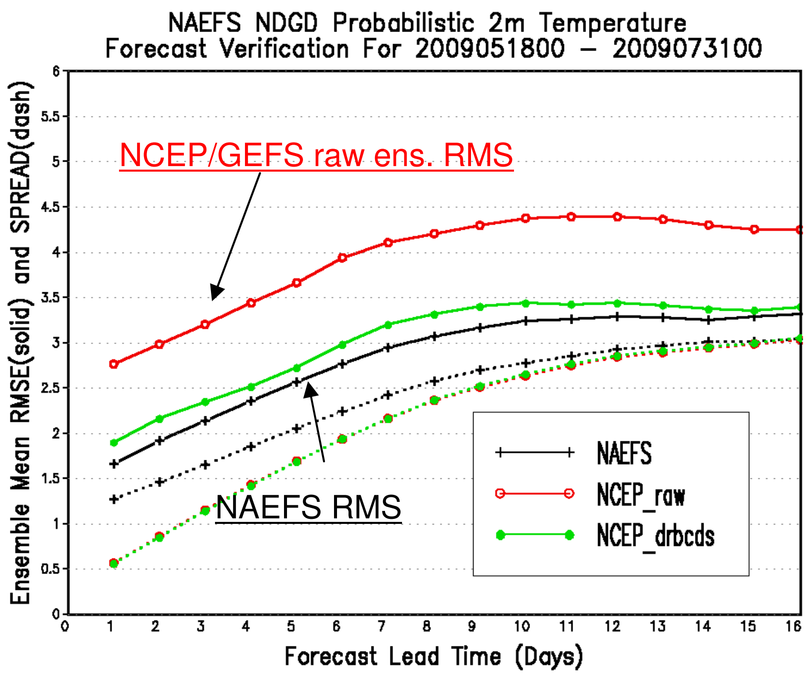
Statistical Down-Scaling for Alaska

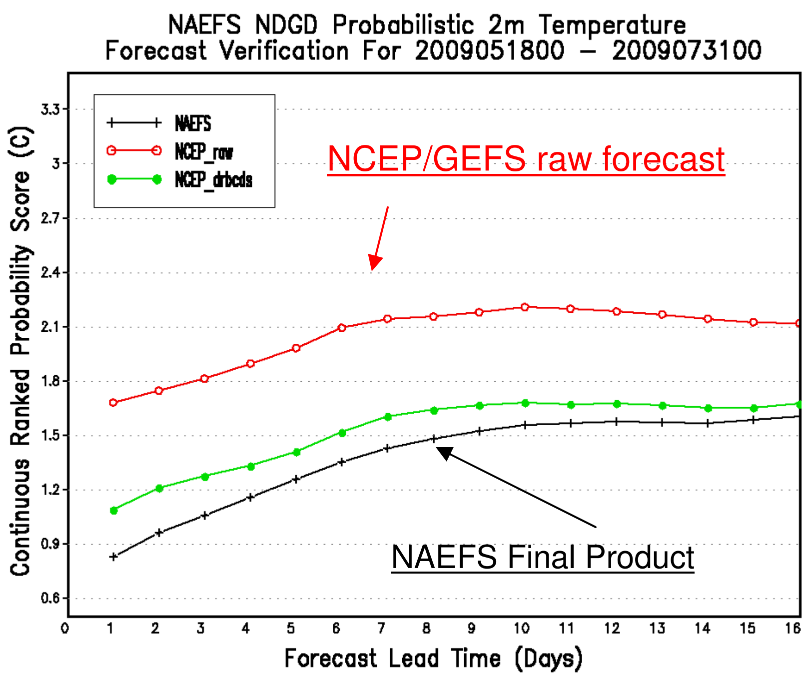
- Downscaling NAEFS products has been implemented in NCEP by
  - December 4th 2007 1200UTC
  - CONUS, 4 variables (surface pressure, 2m temperature and 10m u and v)

- Apply statistical downscaling method to Alaska region
  - Add new variables, wind speed/direction, maximum/minimum temperature

- Statistical Down-Scaling Techniques for Alaska
  - Variable: surface pressure, 2-m temperature, 10-meter wind component
    - work well using current operational technique for CONUS
  - Variable: Tmax and Tmin
    - Choose proper period definition in code/scripts
    - Modification for definition changed July, 2009
  - Variable: wind speed and direction
    - Problem exist in utility "copygb" for wind direction
    - Solution to avoid interpolation of wind speed
    - Not bad, difficult for wind direction improvement
  - Variable: 2m dew point temp and 2m relative humility
    - How to improve methods, future inclusion

- Alaska Verification
  - Next images show some verification for Tmax, Tmin, wind direction/speed





2m Temperature verification
Average for 2 and half months

NAEFS Final Product

From
- Bias correction (NCEP, CMC)
- Dual-resolution (NCEP only)
- Down-scaling (NCEP, CMC)
- Combination of NCEP and CMC

Process to Downscale Tmax & Tmin for Alaska

- Based on 1°×1° 6-hr bias corrected Tmax/Tmin and down-scaling vectors (DV) for T2m at each 6-hr cycle
  - Definition of Tmax/Tmin for Alaska region
    - Tmax period: 13UTC (5am-local) – 04UTC (8pm-local) – local daylight time
    - Tmin period: 01UTC (5pm-local) – 19UTC (11am-local) – local daylight time
  - Definition of approximated period for Tmax/Tmin for giving initial cycle
  - Mean DV of T2m for 6-hr period: weighted average of two instantaneous DVs
  - Interpolating bias corr. 6-hr Tmax/Tmin (1°×1°) to 6km NDGD grid for Alaska
- Downscaling detailed process
  - Apply mean DV to each grid point, each ens. member, and each 6-hr lead-time period, to produce down-scaled Tmax and Tmin for each 6-hr lead-time period
  - Find out highest Tmax and lowest Tmin for approximated period
  - For different grid points, different ens. members, highest Tmax could be in different 6-hr period, the same for lowest Tmin
  - Only one down-scaled Tmax and Tmin for every 24-hr. fcst, up to 384 hours
- Calculate the Tmax/Tmin statistical outputs: mean, spread, mode, 10%, 50% and 90% based on above step

EFFECTIVE 1200UTC JULY 28 2009, DEFINITIONS OF MAX T AND MIN T FOR ALASKA REGION
CHANGE FROM: **MAX T 7:00 AM - 7:00 PM LST, MIN T 7:00 PM - 8:00 AM LST**
TO: **MAX T 5:00 AM - 8:00 PM LST, MIN T 5:00 PM - 11:00 AM LST**

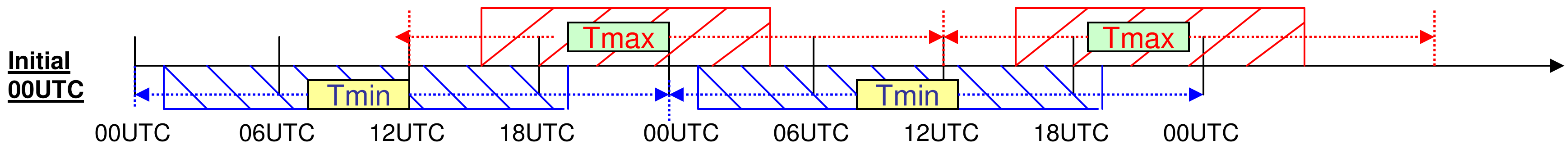
Tmax and Tmin calculations for Alaska region (2009)

Alaska Daylight Time : Tmax period: 13UTC (5am-local) – 04UTC (8pm-local) – local daylight time

Tmin period: 01UTC (5pm-local) –19UTC (11am-local) – local daylight time

OUTPUT: Tmax: f18(no), f24(no), f30(no), f36(yes:12-36hrs), f42(no),f48(no), f54(no), f60(yes:36-60hrs), ……

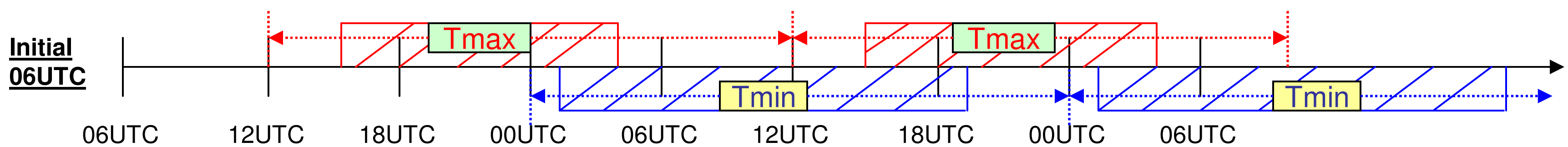
(00UTC) Tmin: f06(no), f12(no), f18(no), f24(yes:0-24hrs), f30(no), f36(no), f42(no), f48(yes:24-48hrs), ……



00UTC temperature forecast at valid time

OUTPUT: Tmax: f12(no), f18(no), f24(no), f30(yes:06-30hrs), f36(no), f42(no), f48(no), f54(yes:30-54hrs), ……

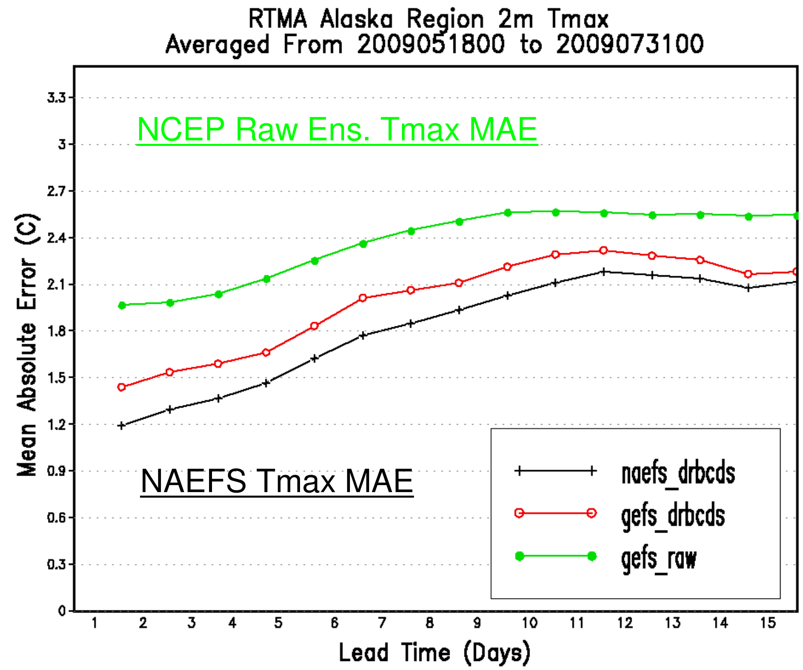
(06UTC) Tmin: f24(no), f30(no), f36(no), f42(yes:18-42hrs), f48(no), f54(no), f60(no), f66(yes:42-66hrs), ……

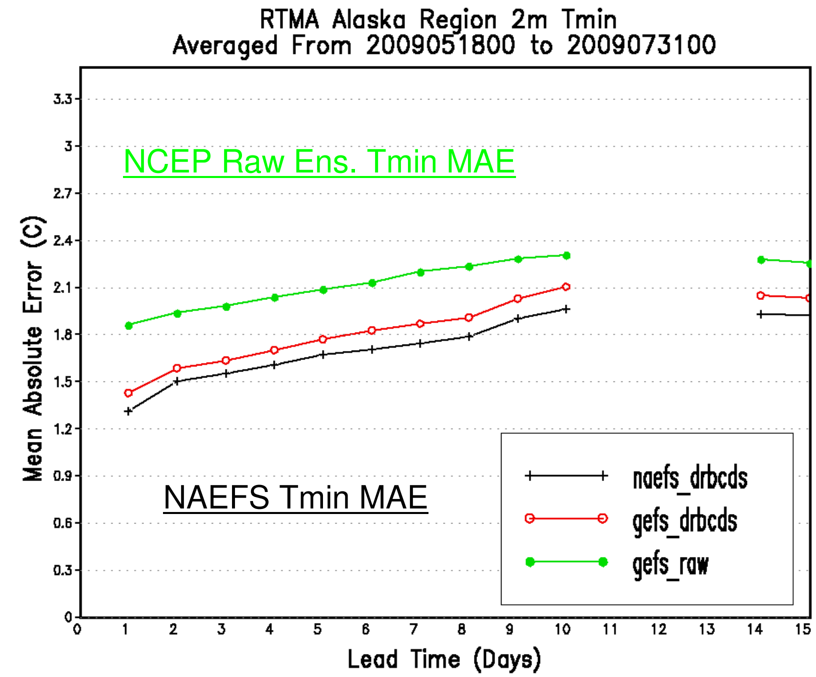


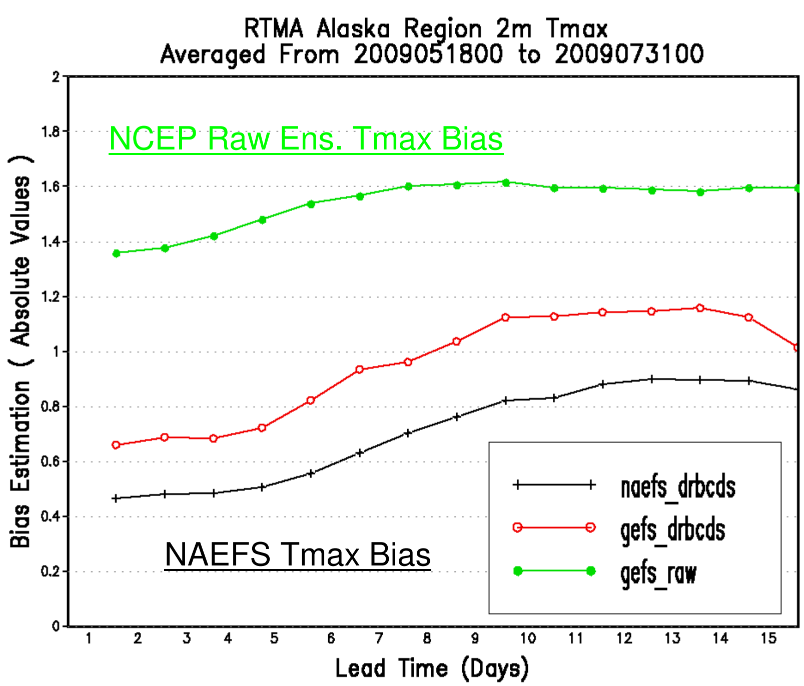
06UTC temperature forecast at valid time

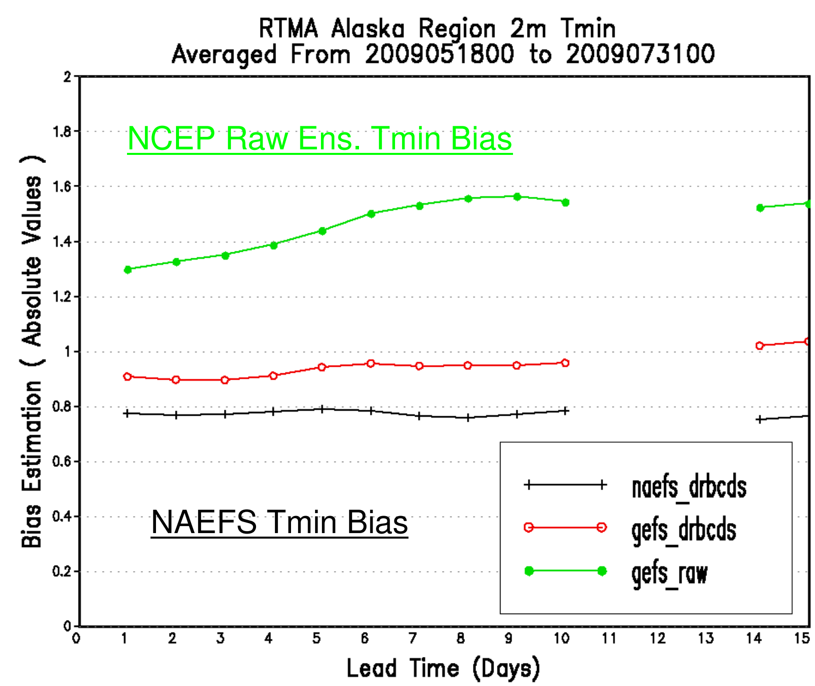
Alaska Standard Time : Tmax period: 14UTC (5am-local) – 05UTC (8pm-local)

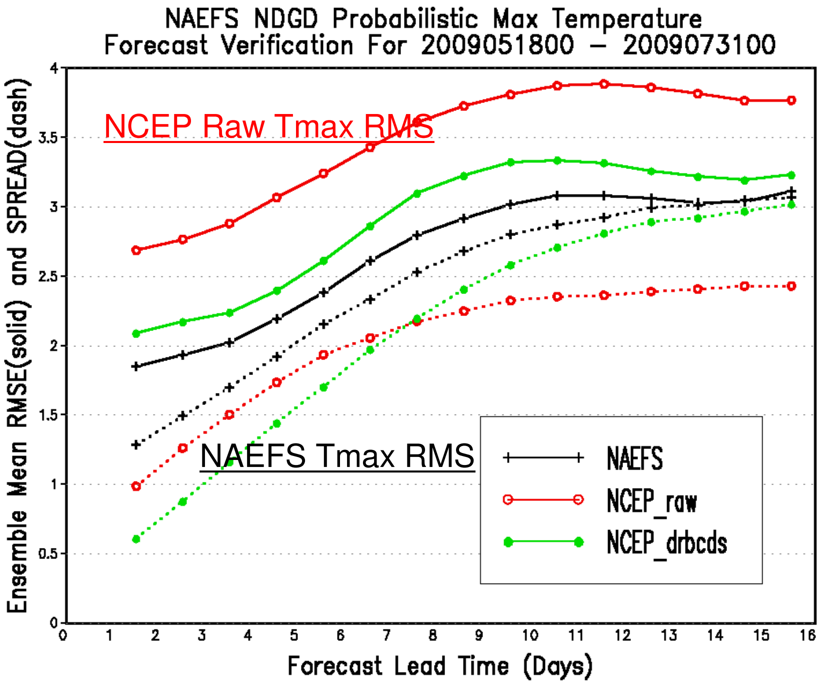
Tmin period: 02UTC (5pm-local) – 20UTC (11am-local)

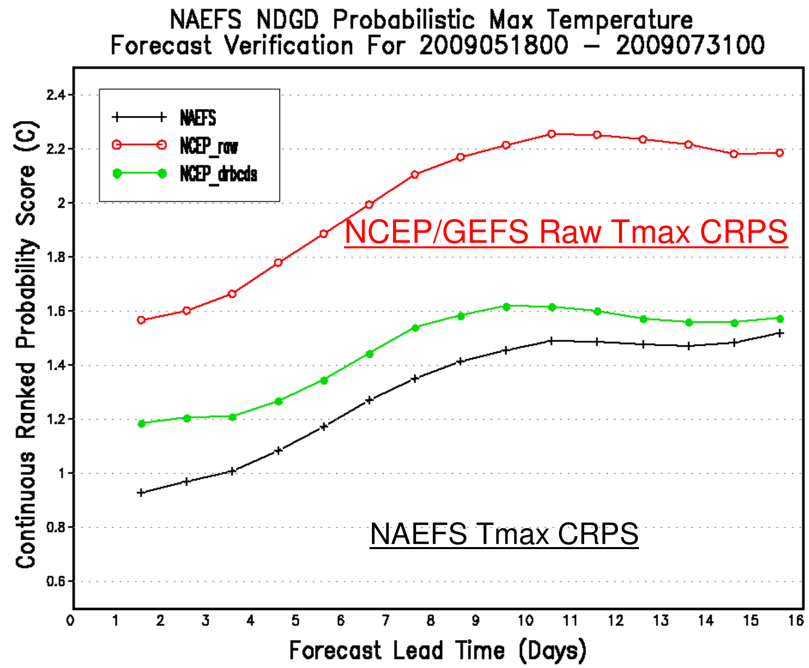

RTMA Alaska Region 2m Tmax
Averaged From 2009051800 to 2009073100
NCEP Raw Ens. Tmax MAE
NAEFS Tmax MAE
naefs_drbcds
gefs_drbcds
gefs_raw
Mean Absolute Error (C)
Lead Time (Days)

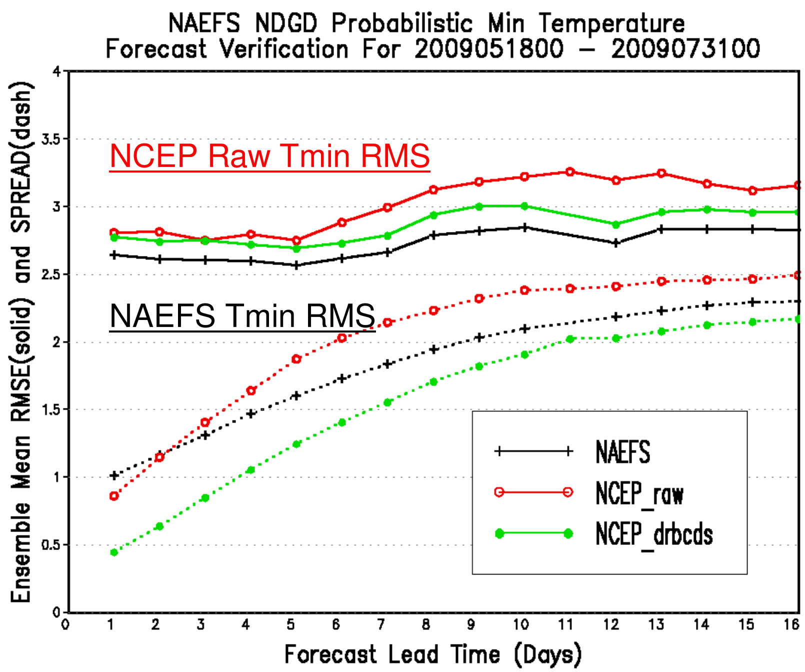

RTMA Alaska Region 2m Tmin
Averaged From 2009051800 to 2009073100
NCEP Raw Ens. Tmin MAE
NAEFS Tmin MAE
naefs_drbcds
gefs_drbcds
gefs_raw
Mean Absolute Error (C)
Lead Time (Days)

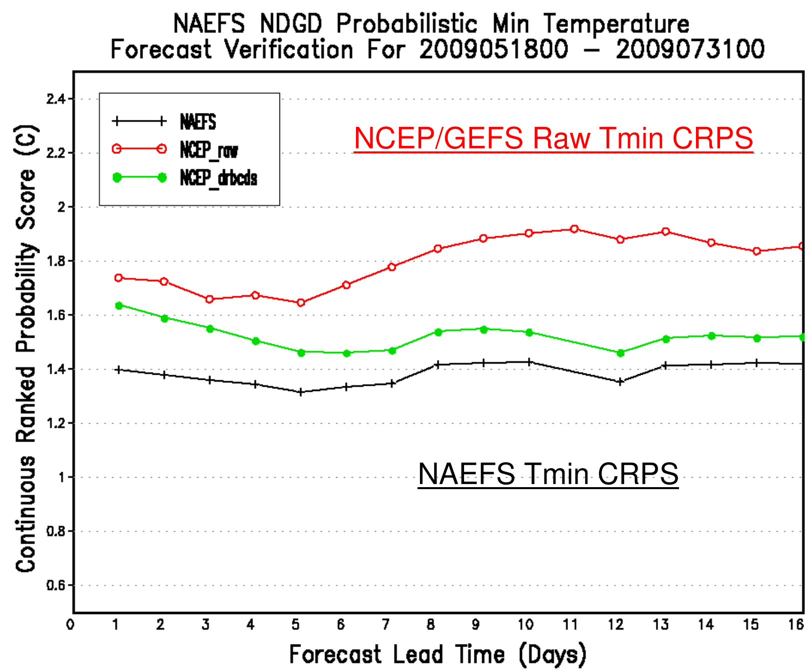

RTMA Alaska Region 2m Tmax
Averaged From 2009051800 to 2009073100
NCEP Raw Ens. Tmax Bias
NAEFS Tmax Bias
naefs_drbcds
gefs_drbcds
gefs_raw
Bias Estimation (Absolute Values)
Lead Time (Days)


RTMA Alaska Region 2m Tmin
Averaged From 2009051800 to 2009073100
NCEP Raw Ens. Tmin Bias
NAEFS Tmin Bias
naefs_drbcds
gefs_drbcds
gefs_raw
Bias Estimation (Absolute Values)
Lead Time (Days)


NAEFS NDGD Probabilistic Max Temperature
Forecast Verification For 2009051800 – 2009073100
NCEP Raw Tmax RMS
NAEFS Tmax RMS
NAEFS
NCEP_raw
NCEP_drbcds
Ensemble Mean RMSE(solid) and SPREAD(dash)
Forecast Lead Time (Days)


NAEFS NDGD Probabilistic Max Temperature
Forecast Verification For 2009051800 – 2009073100
NAEFS
NCEP_raw
NCEP_drbcds
NCEP/GEFS Raw Tmax CRPS
NAEFS Tmax CRPS
Continuous Ranked Probability Score (C)
Forecast Lead Time (Days)


NAEFS NDGD Probabilistic Min Temperature
Forecast Verification For 2009051800 – 2009073100
NCEP Raw Tmin RMS
NAEFS Tmin RMS
NAEFS
NCEP_raw
NCEP_drbcds
Ensemble Mean RMSE(solid) and SPREAD(dash)
Forecast Lead Time (Days)


NAEFS NDGD Probabilistic Min Temperature
Forecast Verification For 2009051800 – 2009073100
NAEFS
NCEP_raw
NCEP_drbcds
NCEP/GEFS Raw Tmin CRPS
NAEFS Tmin CRPS
Continuous Ranked Probability Score (C)
Forecast Lead Time (Days)

Process to Statistically Downscale Wind Speed and Direction

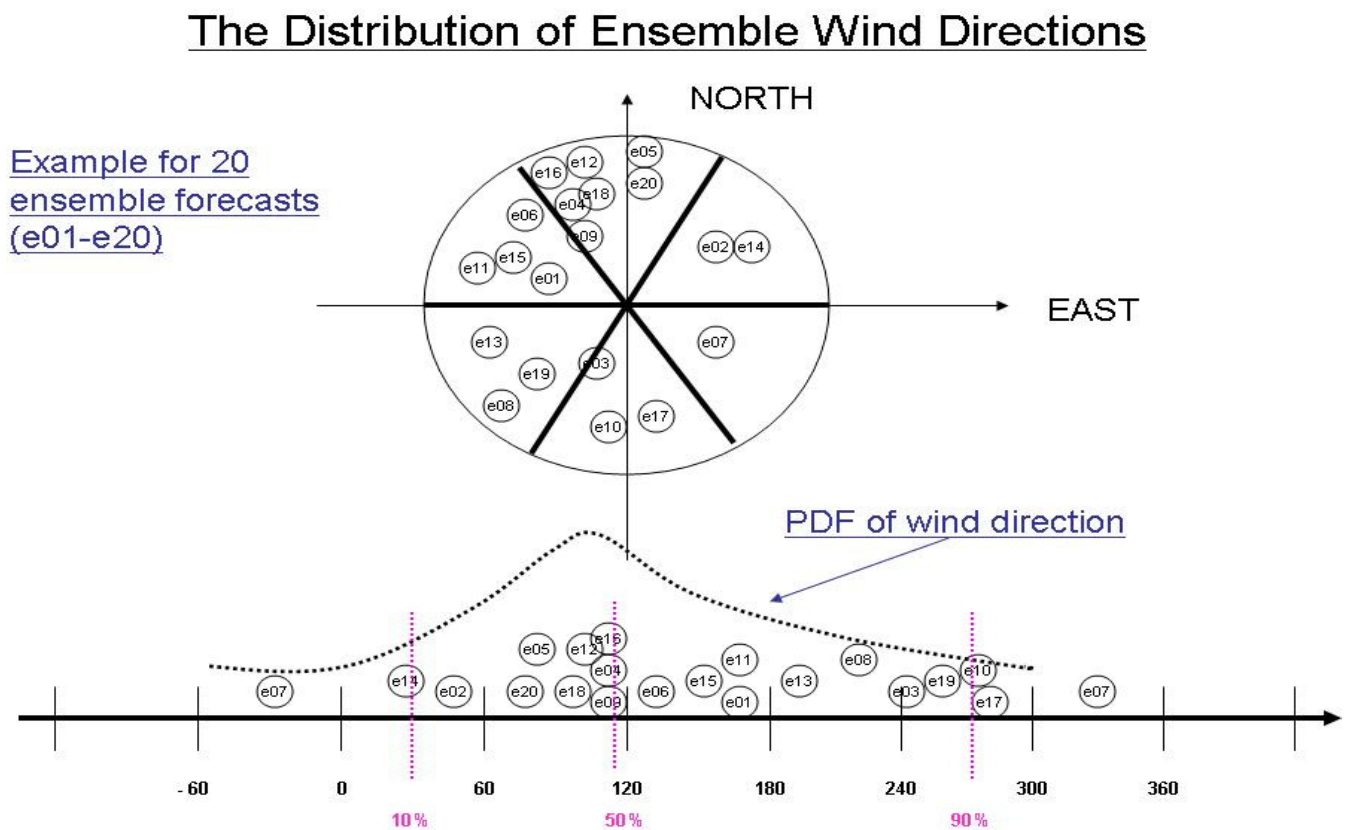
- Wind speed & direction calculation based on downscaled u10m & v10m

$$W_s = \sqrt{u^2 + v^2}$$

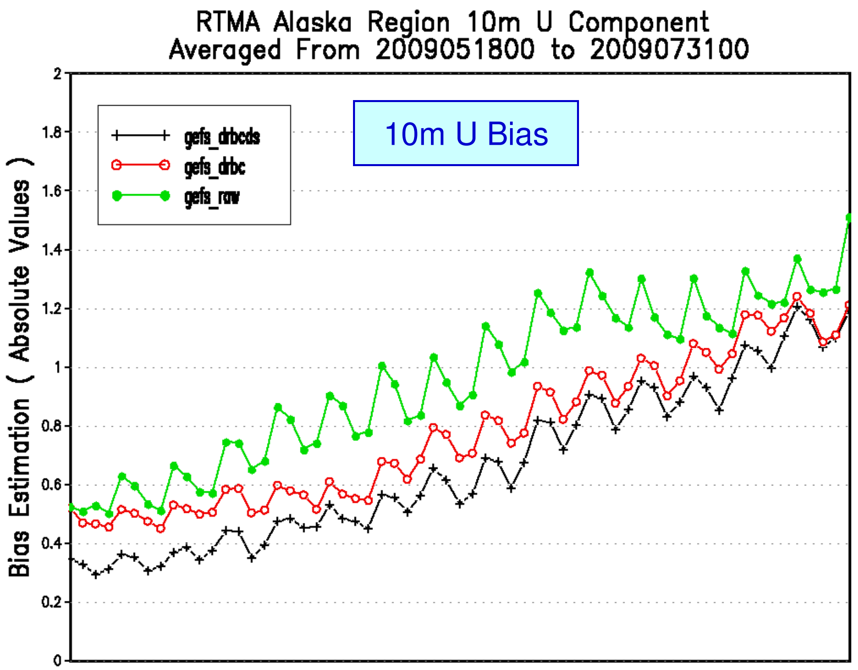
$$W_d = sign(u \cdot v) \cdot \arctan\left|\frac{u}{v}\right| + d_p \text{ where } d_p = \begin{cases} 0, if\_u \le 0, v < 0 \\ 180, if\_u < 0, v \ge 0 \\ 180, if\_u \ge 0, v > 0 \\ 360, if\_u > 0, v \le 0 \end{cases}$$

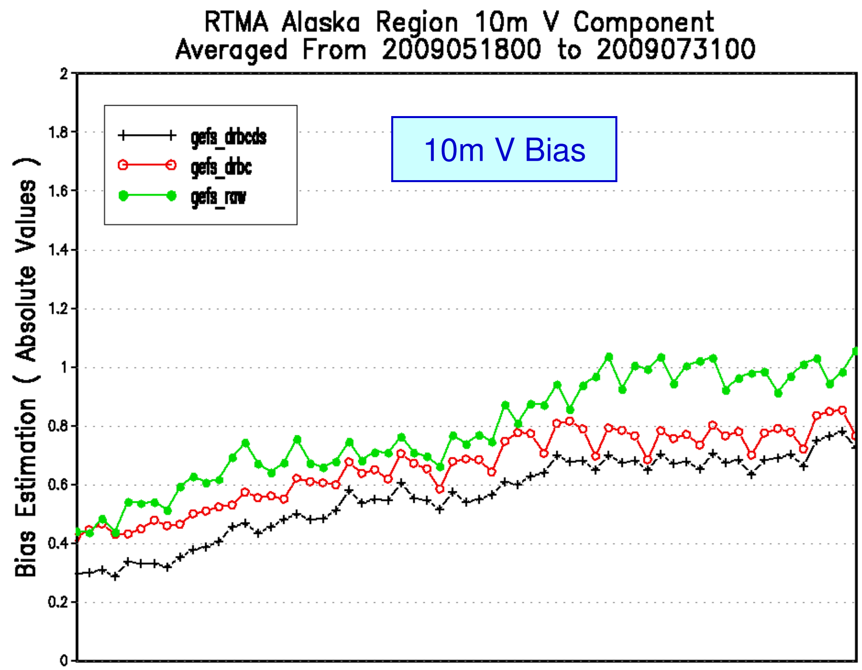
- Downscaling process
  - No action for data on 1°, whole process based and completed on 6km grid u10m & v10m
  - Downscaling process
    - Apply "copygb" utility to interpolate u10m & v10m for NCEP/CMC each ens. member
    - Apply DV to produce down-scaled u10m & v10m for each member and compute its wind speed & direction
    - Combine NCEP/CMC 10m wind speed & direction to generate the mean, spread, mode, 10%, 50% and 90% forecasts
  - No change for wind speed & direction calculation, no change for probabilistic wind direction calculation

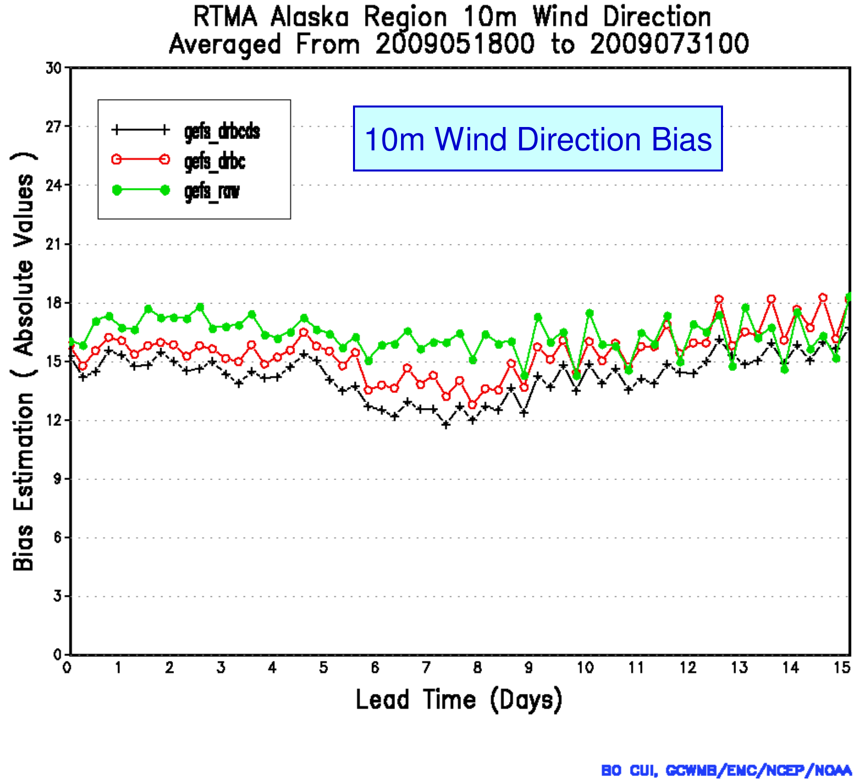
Probabilistic Wind Direction Calculation

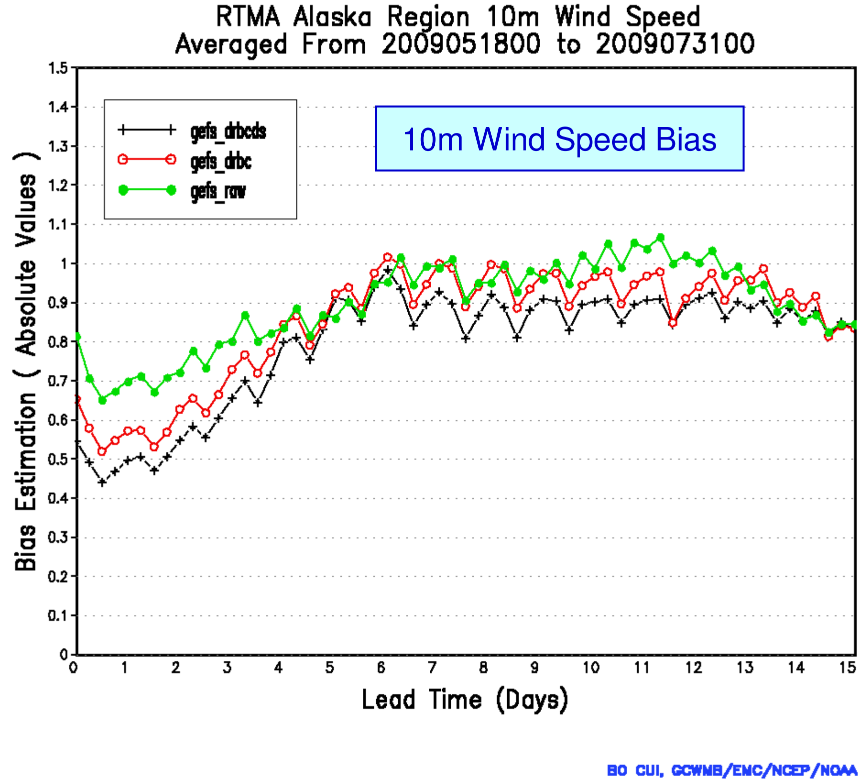


- Divide (0,360) into 6 units, choose the closed 2 units where Wdir data (equal weight currently, different weight by wind speed as a option ?) fall most
- Rearrange the data to allow 2 units in the middle of the distribution
- Set a 60 degree window, move the window through the 2 units, mode is the center of the window with the most members
- Calculate the average wind direction using 6 units data
- Calculate probability 10%, 50% and 90%, mode and spread by using full data
- Adjust wind direction phase in [0,360]


RTMA Alaska Region 10m U Component
Averaged From 2009051800 to 2009073100
10m U Bias
gefs_drbcds
gefs_drbc
gefs_raw
Bias Estimation (Absolute Values)


RTMA Alaska Region 10m V Component
Averaged From 2009051800 to 2009073100
10m V Bias
gefs_drbcds
gefs_drbc
gefs_raw
Bias Estimation (Absolute Values)


RTMA Alaska Region 10m Wind Direction
Averaged From 2009051800 to 2009073100
10m Wind Direction Bias
gefs_drbcds
gefs_drbc
gefs_raw
Bias Estimation (Absolute Values)
Lead Time (Days)
BO CUI, GCWMB/EMC/NCEP/NOAA


RTMA Alaska Region 10m Wind Speed
Averaged From 2009051800 to 2009073100
10m Wind Speed Bias
gefs_drbcds
gefs_drbc
gefs_raw
Bias Estimation (Absolute Values)
Lead Time (Days)
BO CUI, GCWMB/EMC/NCEP/NOAA

Summary and Future Plan

- Bias corrected NAEFS products
  - Benefit for 35 variable for short range forecasts
  - NAEFS is better than lone GEFS, more members is better
  - No skill improvement of week-2 forecast for some season and variables
    - need hindcast for the calibration of week-2 forecast?
  - NCEP plans to include FNMOC ensemble in NAEFS
    - Preliminary evaluations after FNMOC bias correction and combination to NAEFS
    - Future FNMOC ensemble upgrade and improvements

- Downscaling NAEFS products
  - Statistical down-scaling data adds value
  - Bias correction with downscaling adds significant value to the forecasts
    - Reduced mean absolute errors (by 4+ days)
    - Improved probabilistic skills (by 8+ days)

- Apply statistical downscaling method to Alaska region
  - Ready for submit RFC to NCO for implementation

- Apply statistical downscaling method to other regions, Hawaii, Puerto Rico and Guam, when RTMA is available
  - Enhance products by improvements to RTMA

Background

Downscaling Method with Decaying Averaging Algorithm

- True = high resolution analysis
  - Operational North American Real-Time Mesoscale Analysis (RTMA)
    - 5x5 km National Digital Forecast Database (NDFD) grid (e.g. G. DiMego et al.)
    - 4 variables available: surface pressure, T2m, 10m U and V
  - Other data can also be used

- Downscaling method: apply decaying averaging algorithm

$$\text{Downscaling Vector}^{5km}(t_0) = (1-w) * \text{prior DV}^{5km}(t_{-1}) + w * (\text{GDAS}^{5km}(t_0) - \text{RTMA}^{5km}(t_0))$$

> *$GDAS^{5km}$: GDAS 1x1 analysis interpolated to $RTMA^{5km}$ grids by bilinear interpolation*
>
> *4 cycles, individual grid point, DV^{5km} = Downscaling Vector on 5km grids*
>
> *choose different weight: 0.5%, 1%, 2%, 5%, 10%*

- Downscaling Process

$$\text{Downscaled Forecast}^{5km}(t) = \text{Bias-corrected Forecast}^{5km}(t) - DV^{5km}(t_0)$$

> *Bias-corrected $Forecast^{5km:}$ interpolated to $RTMA^{5km}$ grids by bilinear interpolation*
>
> *subtract DV^{5km} from bias-corrected $forecast^{5km}$ valid at analysis time*

Bias Before/After Bias Correction (NCEP NH)