Analog-Kalman filter based post-processing of surface PM2.5 predictions from the Community Multiscale Air Quality (CMAQ) model.

James Wilczak[1]
Irina Djalalova[1,2]
Luca Delle Monache[3]

[1] *National Oceanic and Atmospheric Administration/Earth Systems Research Laboratory (NOAA), Boulder, Colorado*
[2] *University of Colorado Cooperative Institute for Research in the Environmental Sciences (CIRES), Boulder, Colorado*
[3] *National Center for Atmospheric Research (NCAR), Boulder, Colorado*

Why use post-processing for air quality predictions?

- Virtually all weather forecast information widely disseminated to the public has some type of post-processing applied to raw model output. This is true for precipitation, temperature, humidity, etc.

- Post-processing works especially well when model forecasts have large biases.

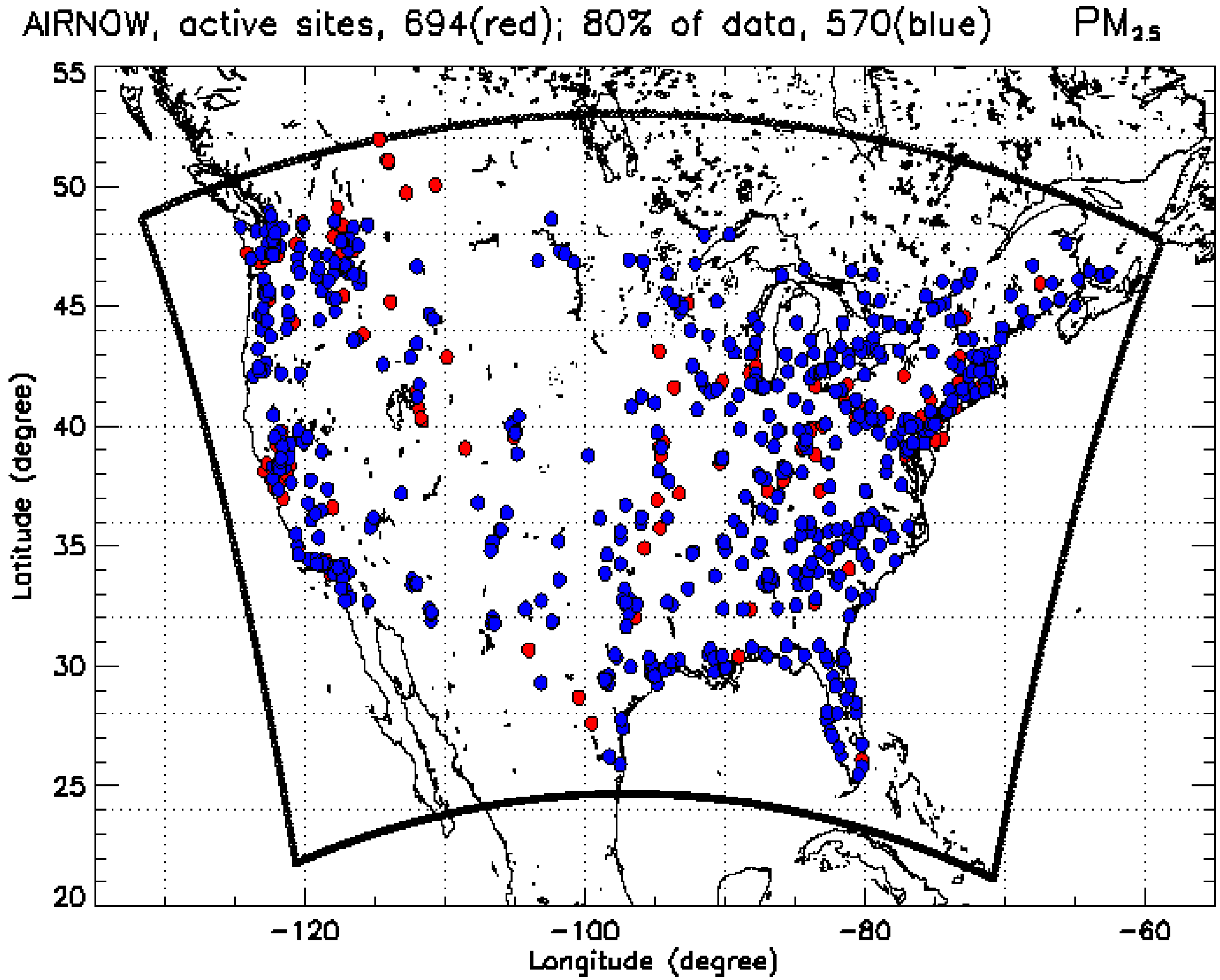
- Historically, no post-processing has been applied to NOAA's AQ forecasts.

OUTLINE

- Data set and quality control procedure for surface PM2.5 observations;
- Model post-processing at each observational site using historical forecast analogs;
- Spreading the forecast correction over the entire gridded domain.

OUTLINE

- Data set and quality control procedure for surface PM2.5 observations;
- Model post-processing at each observational site using historical forecast analogs;
- Spreading the forecast correction over the entire gridded domain.



EPA active 716 sites. Only 570 (blue) sites with at least 80% data available throughout the year of December, 2009 – November, 2010 were used.

$PM_{2.5}$ QUALITY CONTROL

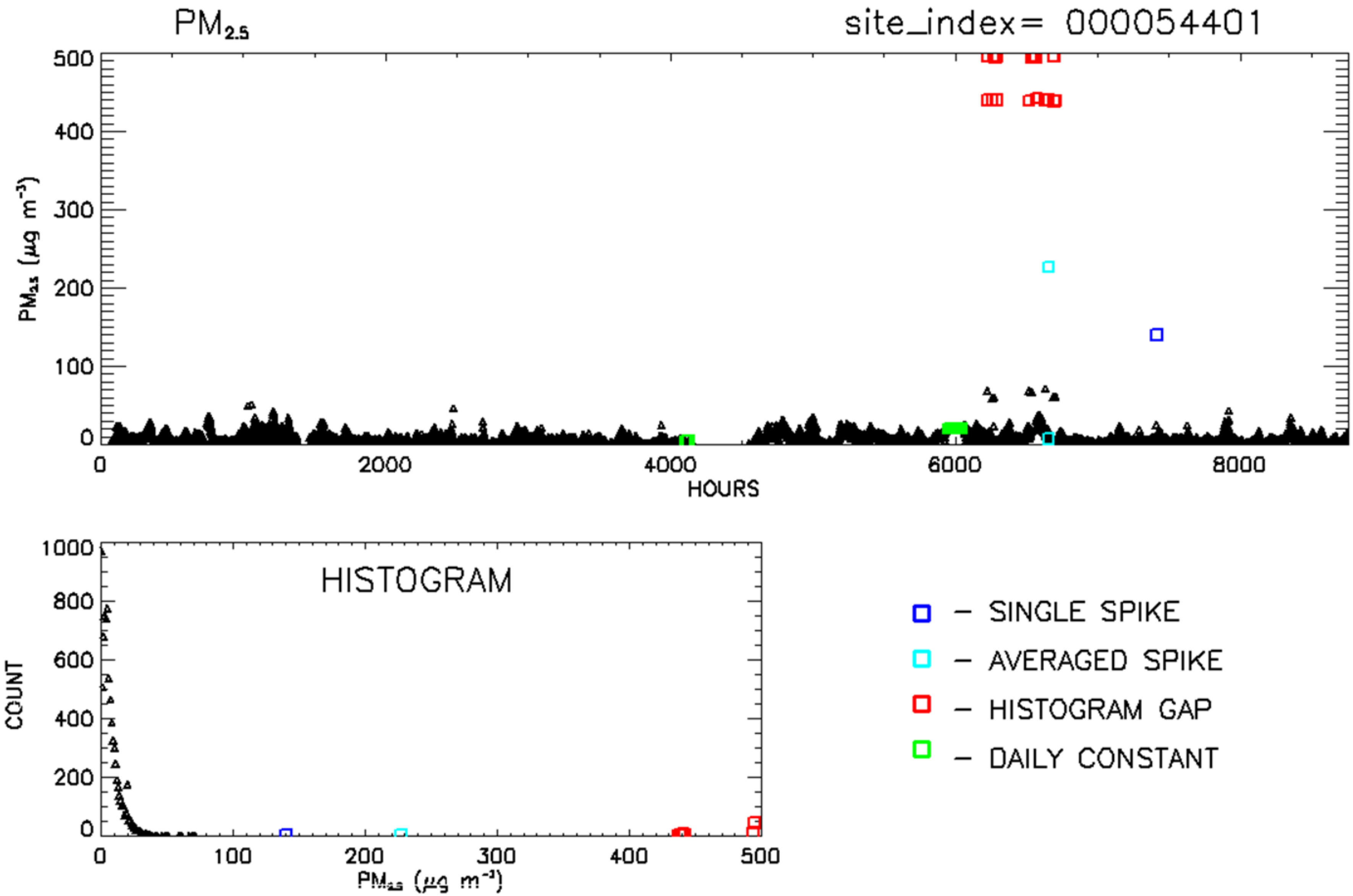
- **Constant value data**. By this method the data of one single day is checked. If MIN(PM)=MAX(PM) than the whole day is eliminated.

- The data **over 500 $\mu g/m^3$** threshed.

- **Single hour spike:** if (PM[hour]-PM[hour-1] > 50 $\mu g/m^3$ and (PM[hour]-PM[hour+1] > 50 $\mu g/m^3$

- **3-hour averaged spike:**
  if MEAN(PM[hour-1,hour,hour+1])-PM[hour-2]>100 $\mu g/m^3$ and MEAN(PM[hour-1,hour,hour+1])-PM[hour+2]>100 $\mu g/m^3$

- **Histogram check** if histogram has an empty window.
  For low PM (MAX(PM)<200 $\mu g/m^3$) an empty window >= 50 $\mu g/m^3$.
  For high PM (MAX(PM)>200 $\mu g/m^3$) an empty window >= 100 $\mu g/m^3$.


PM2.5
site_index= 000054401
PM2.5 (μg m-3)
500
400
300
200
100
0
0
2000
4000
6000
8000
HOURS
HISTOGRAM
COUNT
1000
800
600
400
200
0
0
100
200
300
400
500
PM2.5 (μg m-3)
– SINGLE SPIKE
– AVERAGED SPIKE
– HISTOGRAM GAP
– DAILY CONSTANT

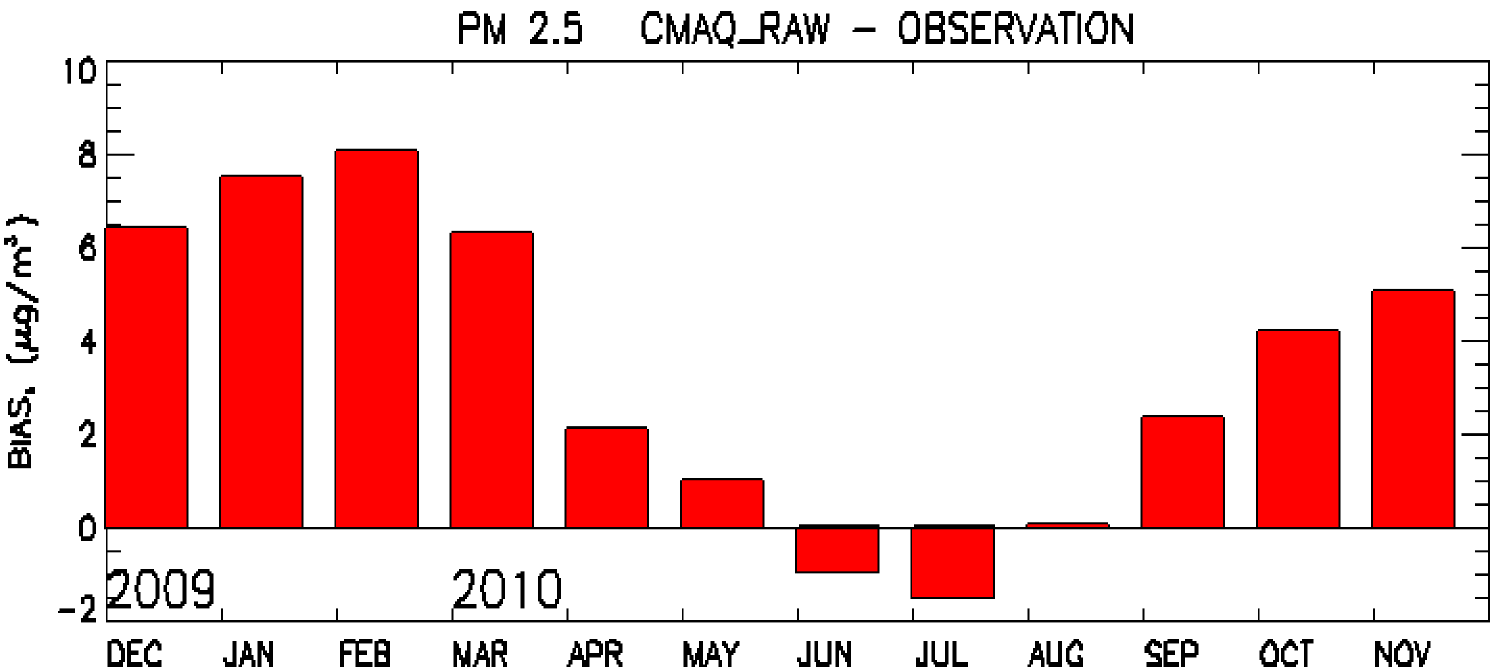
OUTLINE

- Data set and quality control procedure for surface PM2.5 observations;
- Model post-processing at each observational site using historical forecast analogs;
- Spreading the forecast correction over the entire gridded domain.

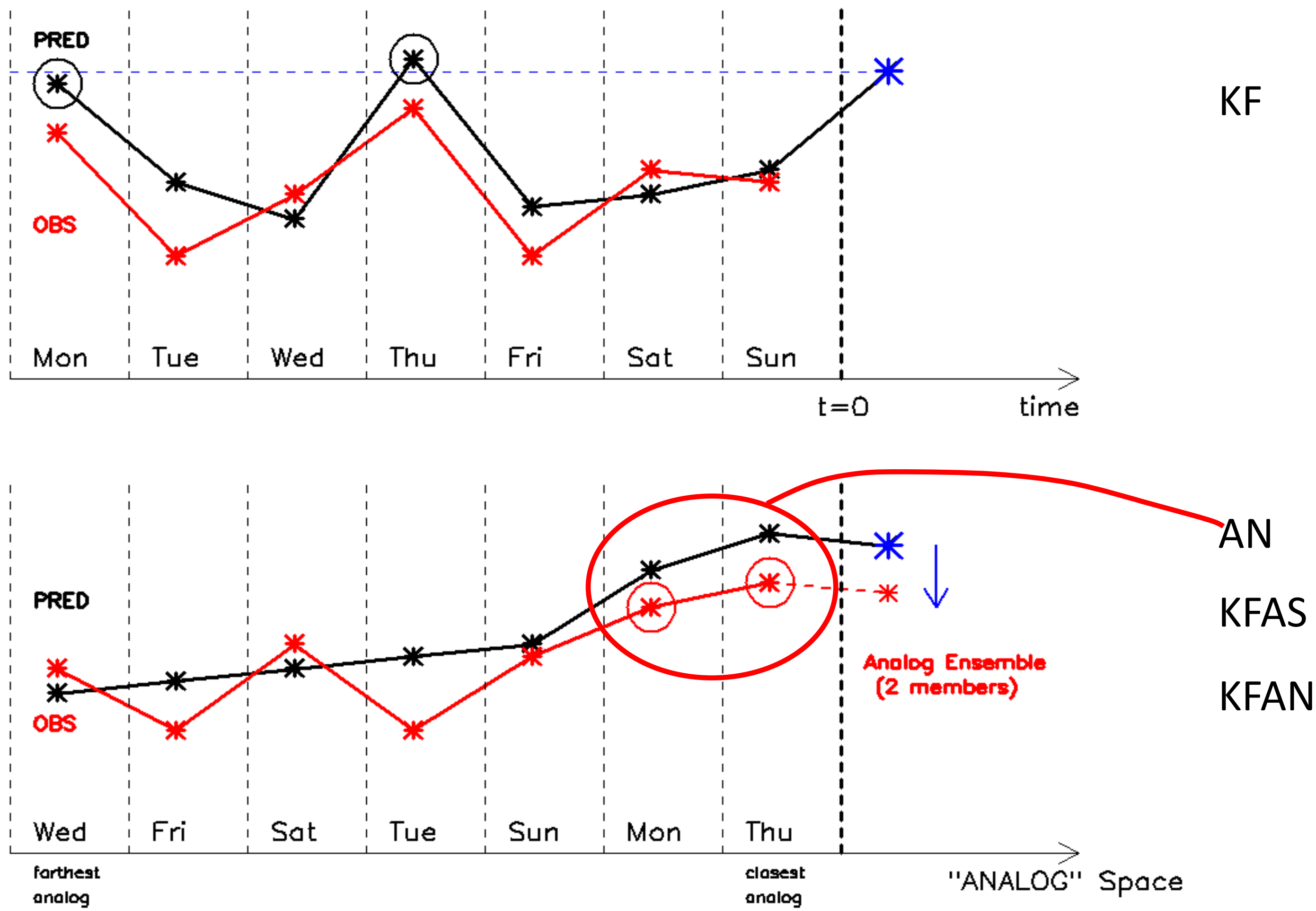
2009-2010 CMAQ Model:

- aerosol module version 4 (AERO-4)
- Carbon Bond Mechanism version IV (CBMIV) gas-phase chemical mechanism
- Sparse Matrix Operator Kernel Emissions (SMOKE) emissions

Monthly $PM_{2.5}$ Bias

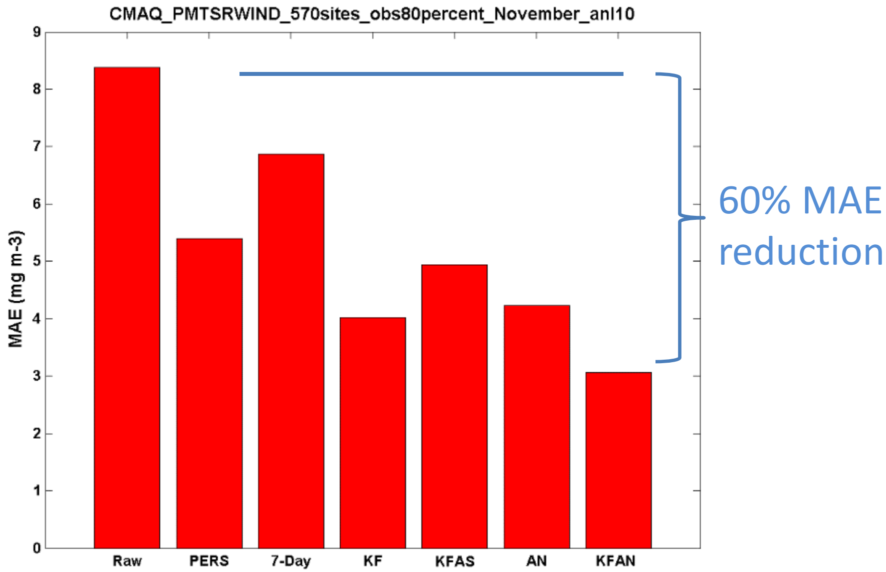


Analog Method

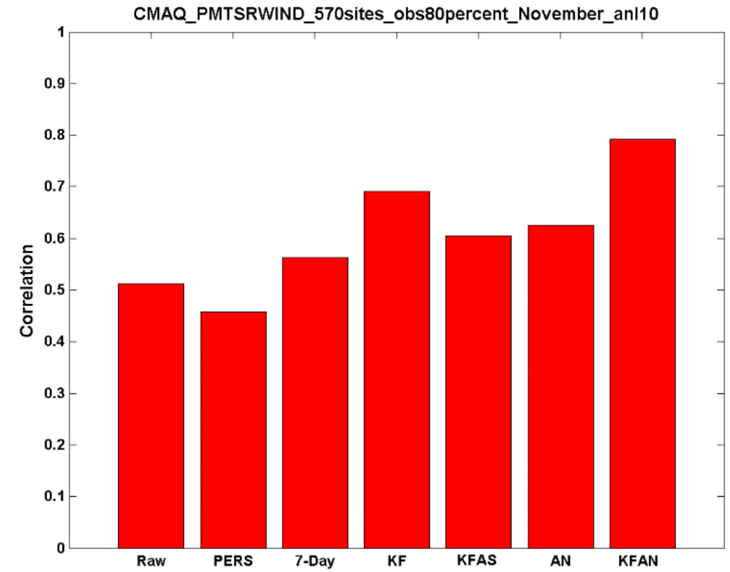


REFERENCES:
Delle Monache et al., 2011, *Monthly Weather Review*, **139**, 3554–3570.
Delle Monache, Luca, F. Anthony Eckel, Daran L. Rife, Badrinath Nagarajan, Keith Searight, 2013: *Monthly Weather Review,* **141**, 3498–3516.

MAE



CORRELATION



Statistics for Raw model, persistence, plus 5 p-p methods, averaged over the month of November 2010 with the rest of the year used as the training period. Methods from left to right:

1 – raw model;

2 – persistence (at the same hour one day previously);

3 – 7-day running mean;

4 – KF Kalman Filter with recent days weighted more heavily;

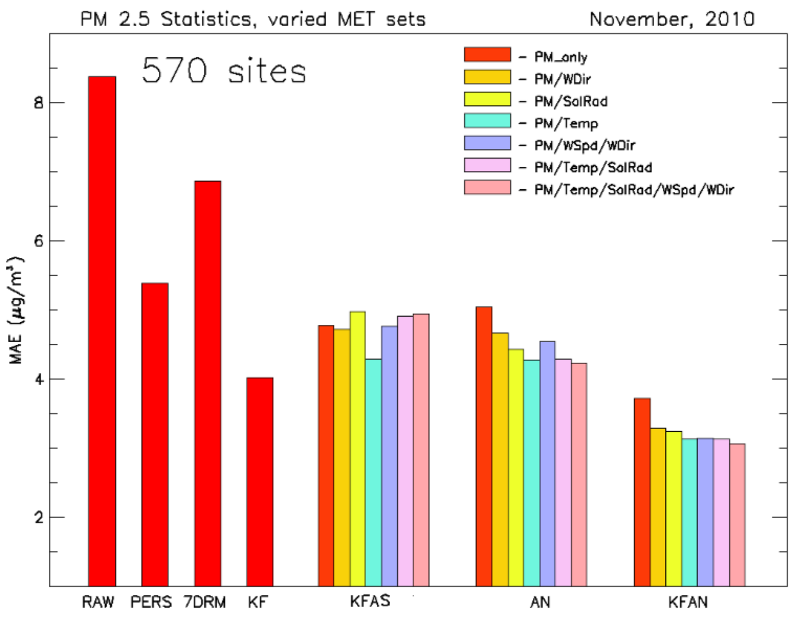
5 – KFAS a Kalman Filter in analog space using ordered model PM analogs;

6 – AN Analog Ensemble using weighted ensemble mean of 10 best analogs;

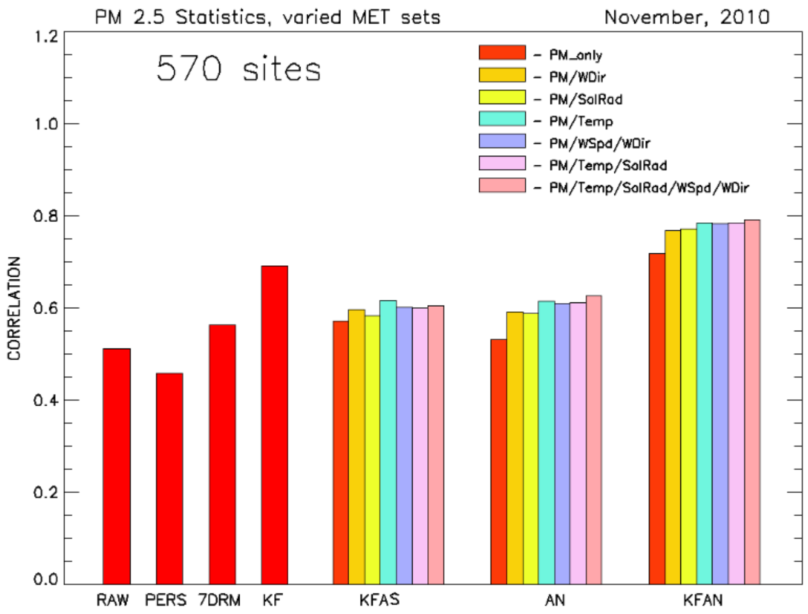
7 – KFAN - a Kalman Filter applied to the AN time series;

Analog Sensitivity to Meteorological Variables

MAE



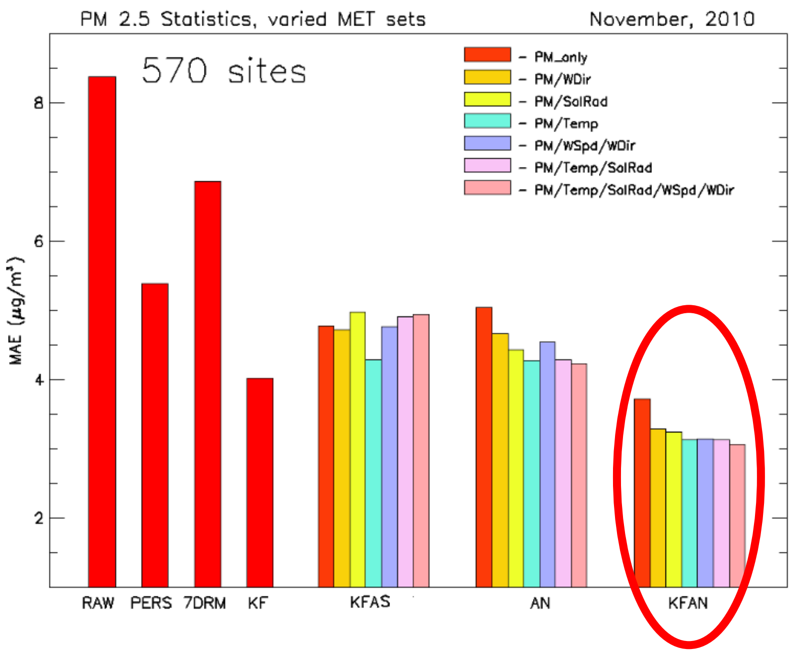
CORRELATION



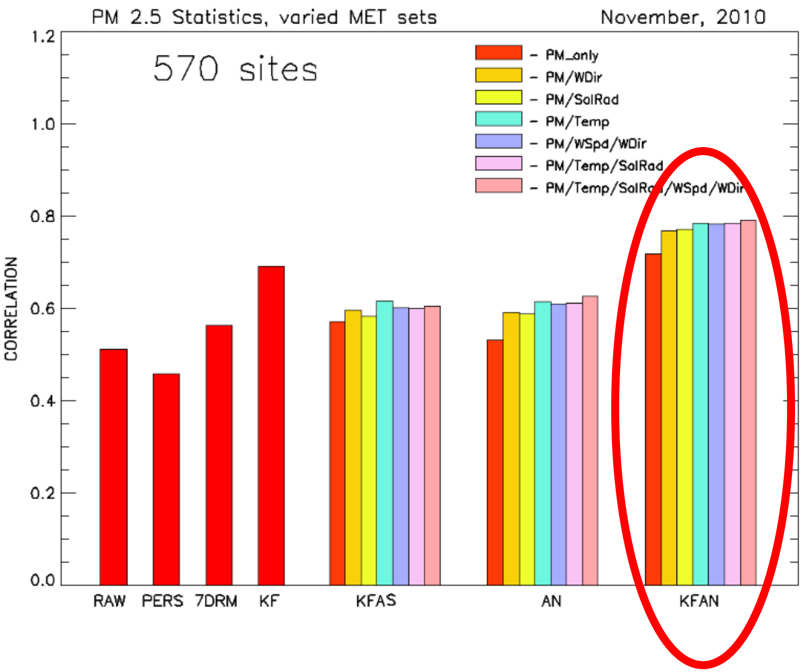
PM_only
PM/WindDirection
PM/SolarRadiation
PM/Temperature
PM/Wind(Speed&Direction)
PM/Temperature&SolarRadiation
PM/ Temperature&SolarRadiation& Wind(Speed&Direction)

Analog Sensitivity to Meteorological Variables

MAE



CORRELATION



PM_only
PM/WindDirection
PM/SolarRadiation
PM/Temperature
PM/Wind(Speed&Direction)
PM/Temperature&SolarRadiation
PM/ Temperature&SolarRadiation& Wind(Speed&Direction)

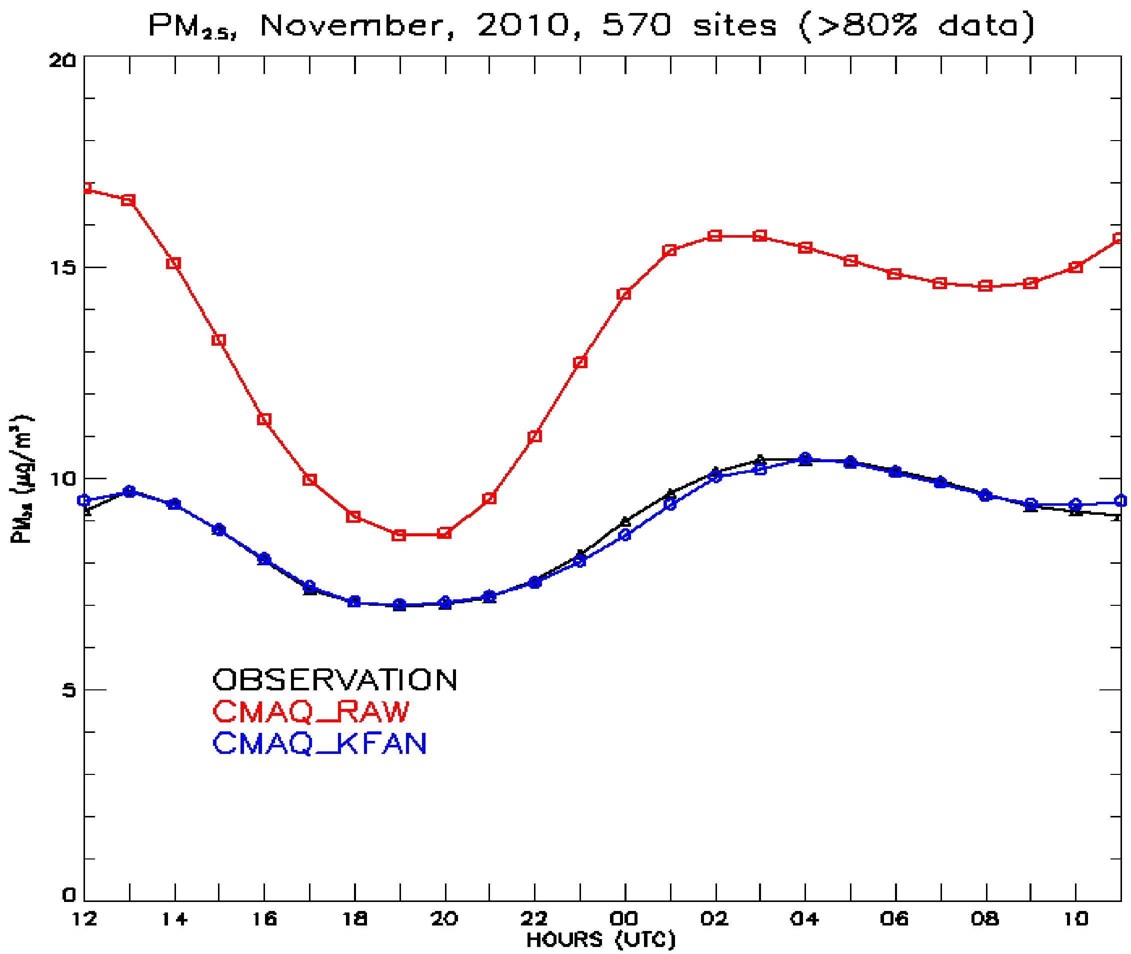
MAE Sensitivity to Number of Analogs

Analog numbers	3	5	10	20	30
KFAS	4.946	4.946	4.946	4.946	4.946
AN	4.415	4.311	4.234	4.266	4.289
KFAN	3.446	3.247	3.064	2.961	2.919

MAE Sensitivity to Number of Analogs

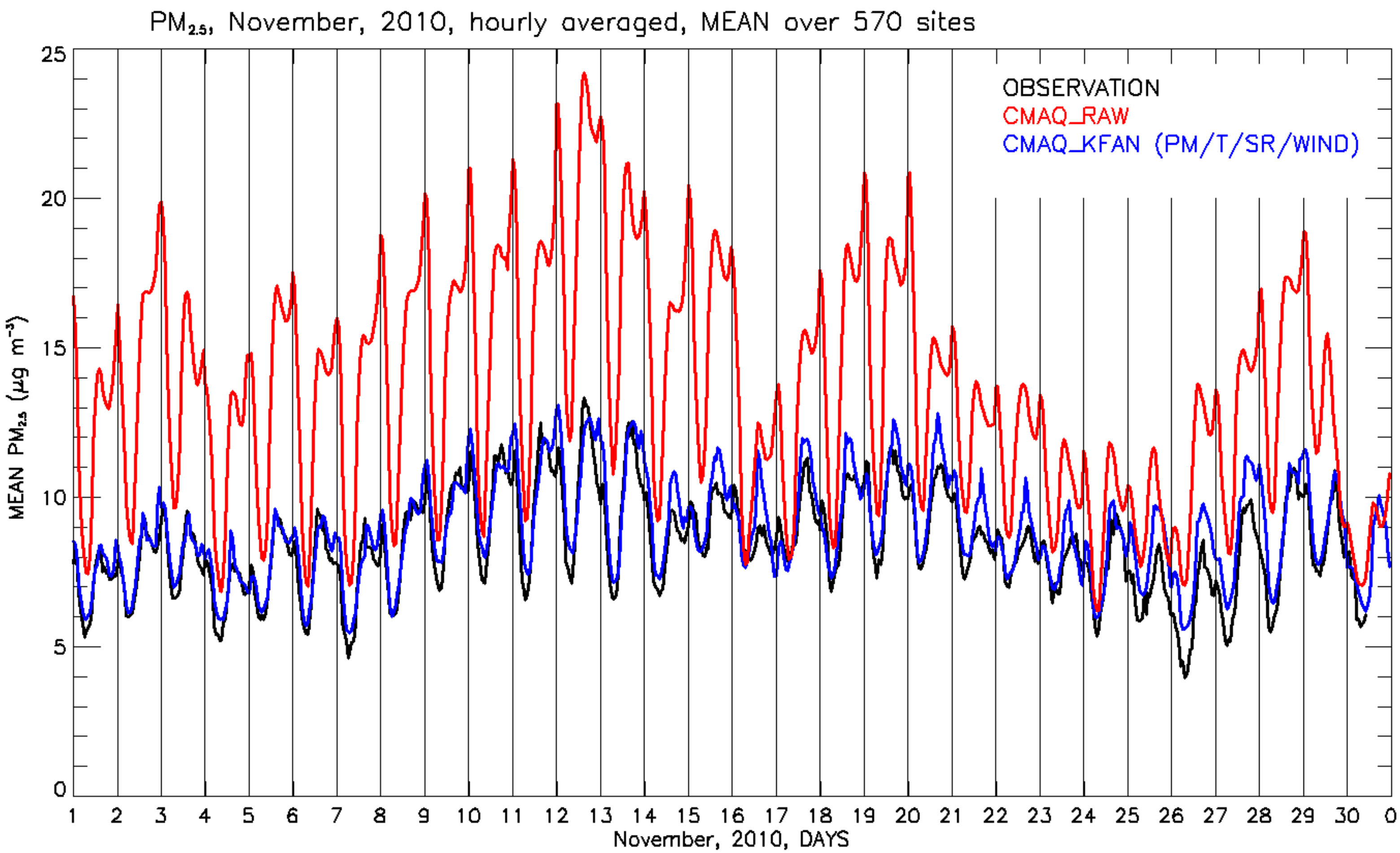
Analog numbers	3	5	10	20	30
KFAS	4.946	4.946	4.946	4.946	4.946
AN	4.415	4.311	4.234	4.266	4.289
KFAN	3.446	3.247	3.064	2.961	2.919

PM2.5 Diurnal Cycle At Obs Sites

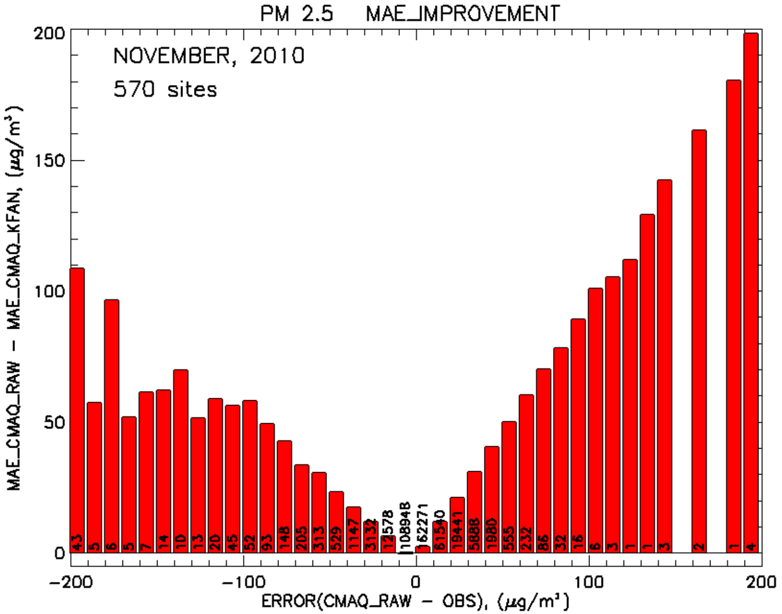


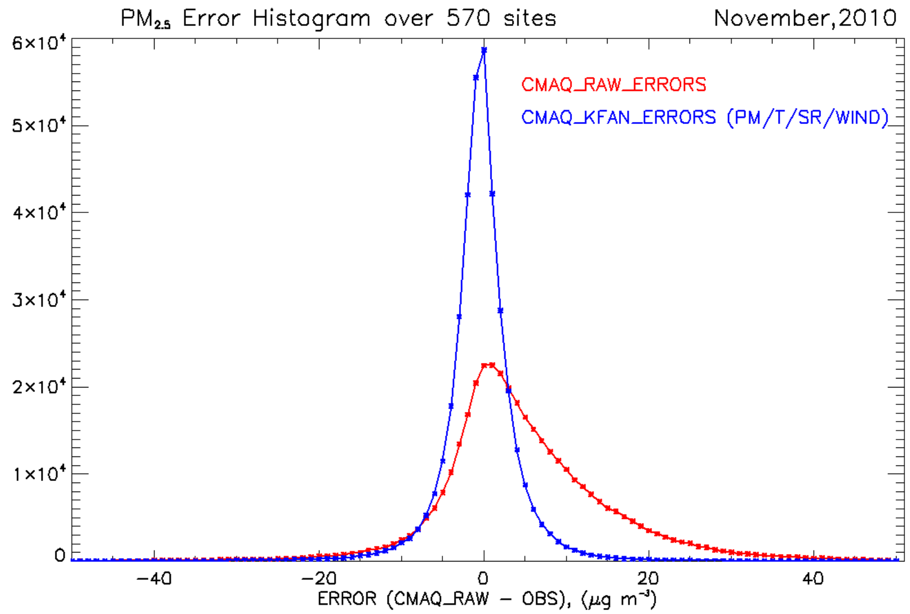
PM diurnal cycle, averaged for the month of November and all 570 sites.

November 2010 Hourly Data

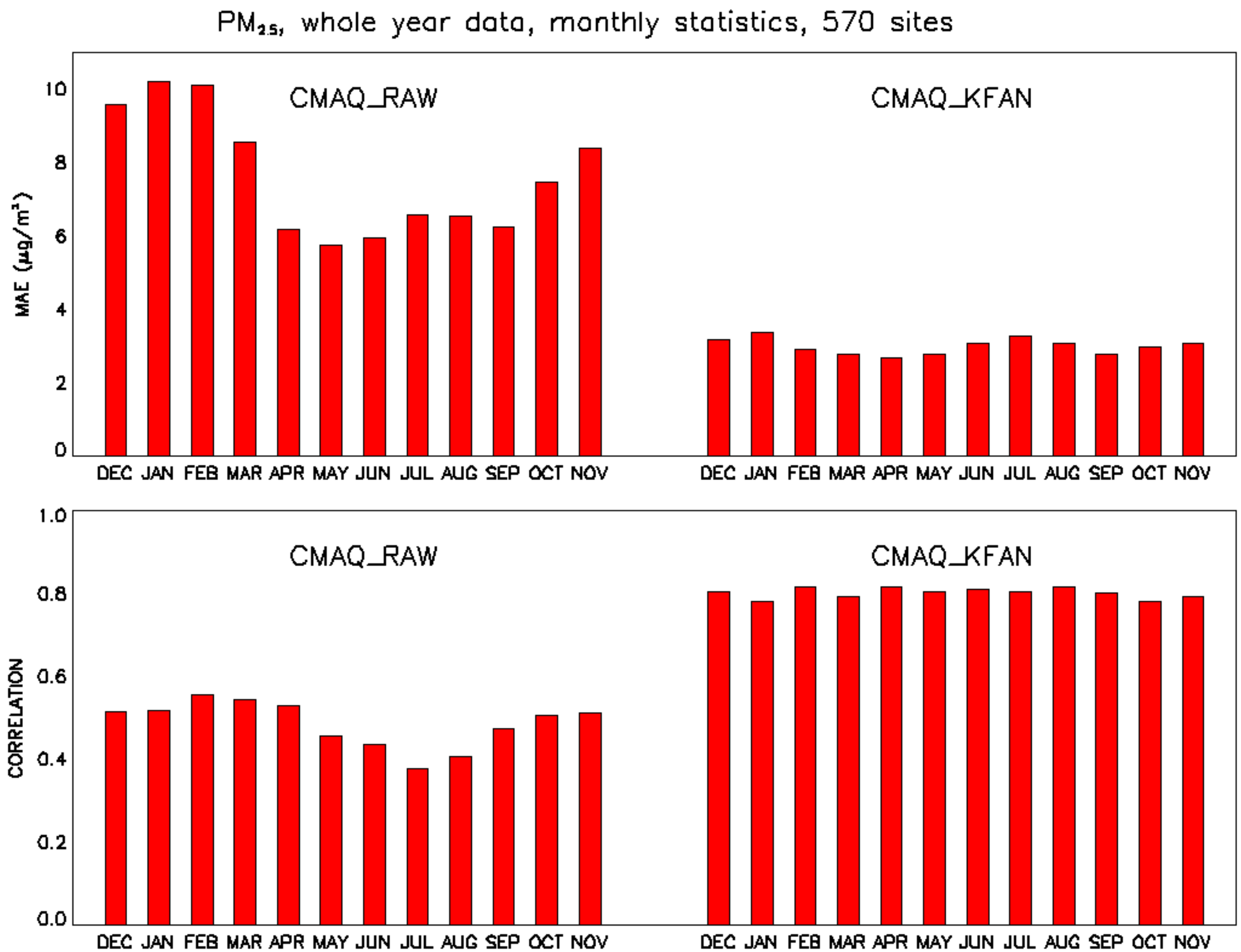


MAE Improvement of Large Errors





Monthly Statistics



MAE

CORRELATION

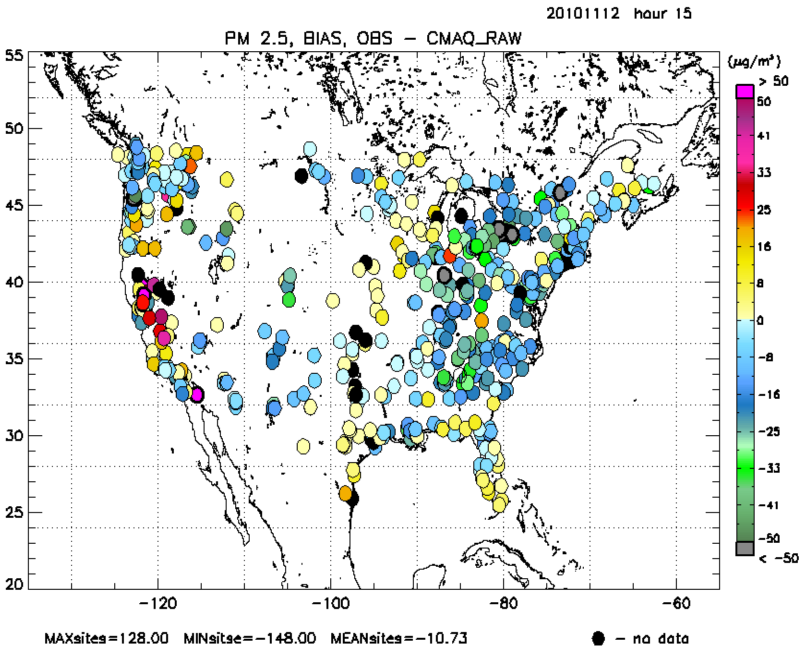
RAW KFAN

OUTLINE

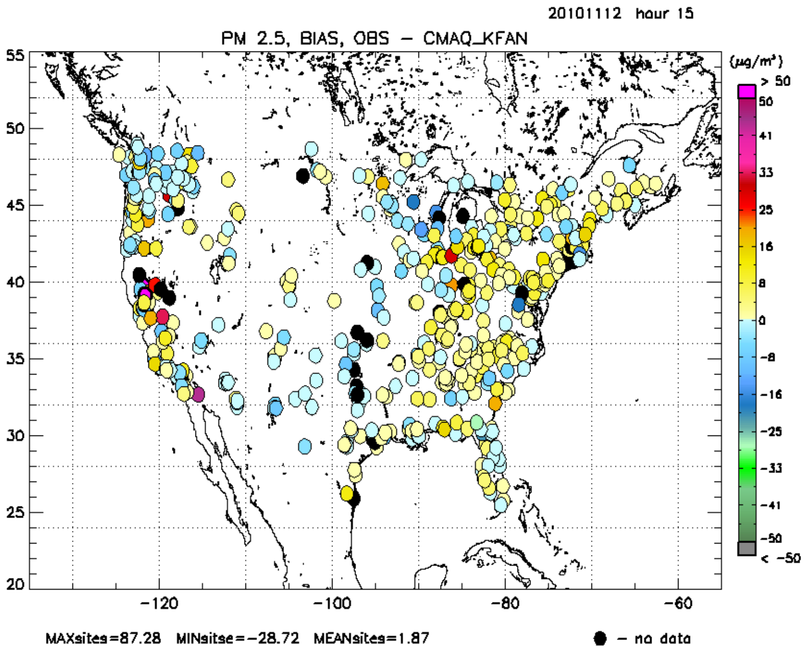
- Data set and quality control procedure for surface PM2.5 observations;
- Model post-processing at each observational site using historical forecast analogs;
- Spreading the forecast correction over the entire gridded domain.

AIRNow Site Corrections

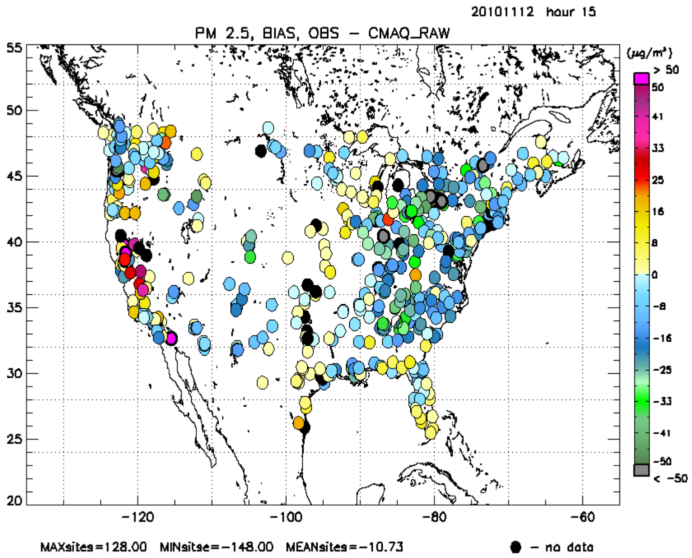
Example data from 12 Nov 2012, hour 15



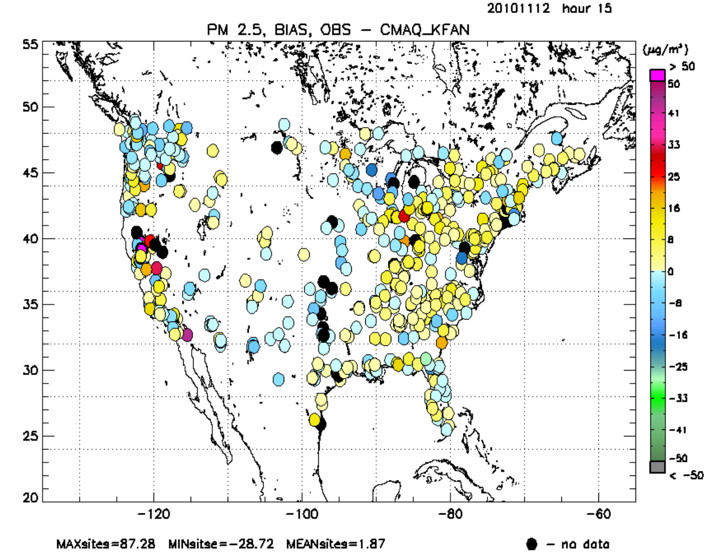
OBS – CMAQ_RAW



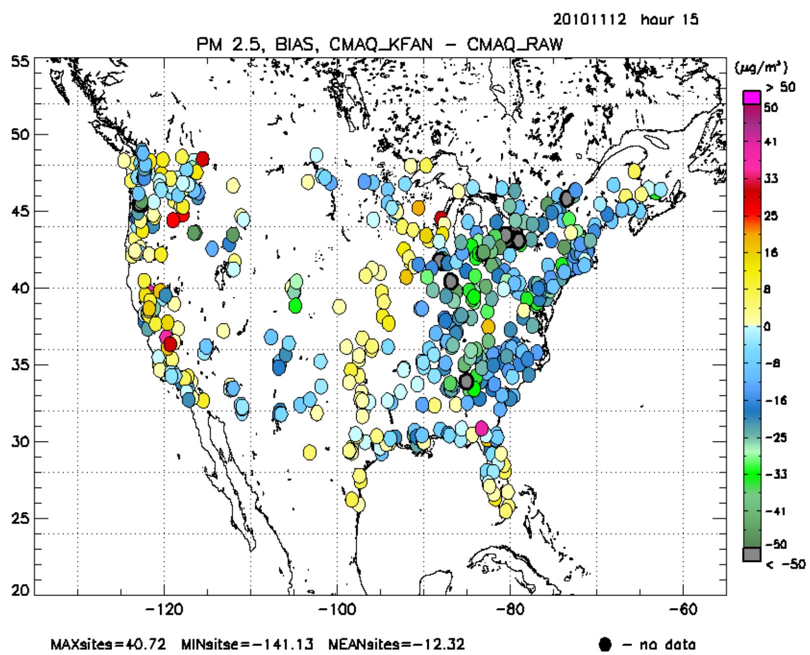
OBS – CMAQ_KFAN



OBS – CMAQ_RAW



OBS – CMAQ_KFAN



CMAQ_KFAN – CMAQ_RAW => Forecasted bias

Spreading Method

- To create graphical images of corrected CMAQ PM2.5 forecasts, the forecasted bias calculated at each AIRNow obs location must be spread to every model gridpoint.

- An iterative objective analyses method is used which starts with a very large radius of influence (R=2000km).

- $Sbias_k$ = $CMAQ\_KFAN_k$ - $CMAQ\_RAW_k$, $Mbias_{i,j} = 0$

- At each grid point the correction value is calculated as

  $C_{i,j} = \frac{1}{n} \Sigma \frac{R*R-d*d}{R*R+d*d} (Sbias_k - Mbias_{i,j})$, $d<R$,where

  R is radius of influence;

  d is the distance from a grid point to the site k inside the circle R;

  $C_{i,j}$ is the correction at a grid point;

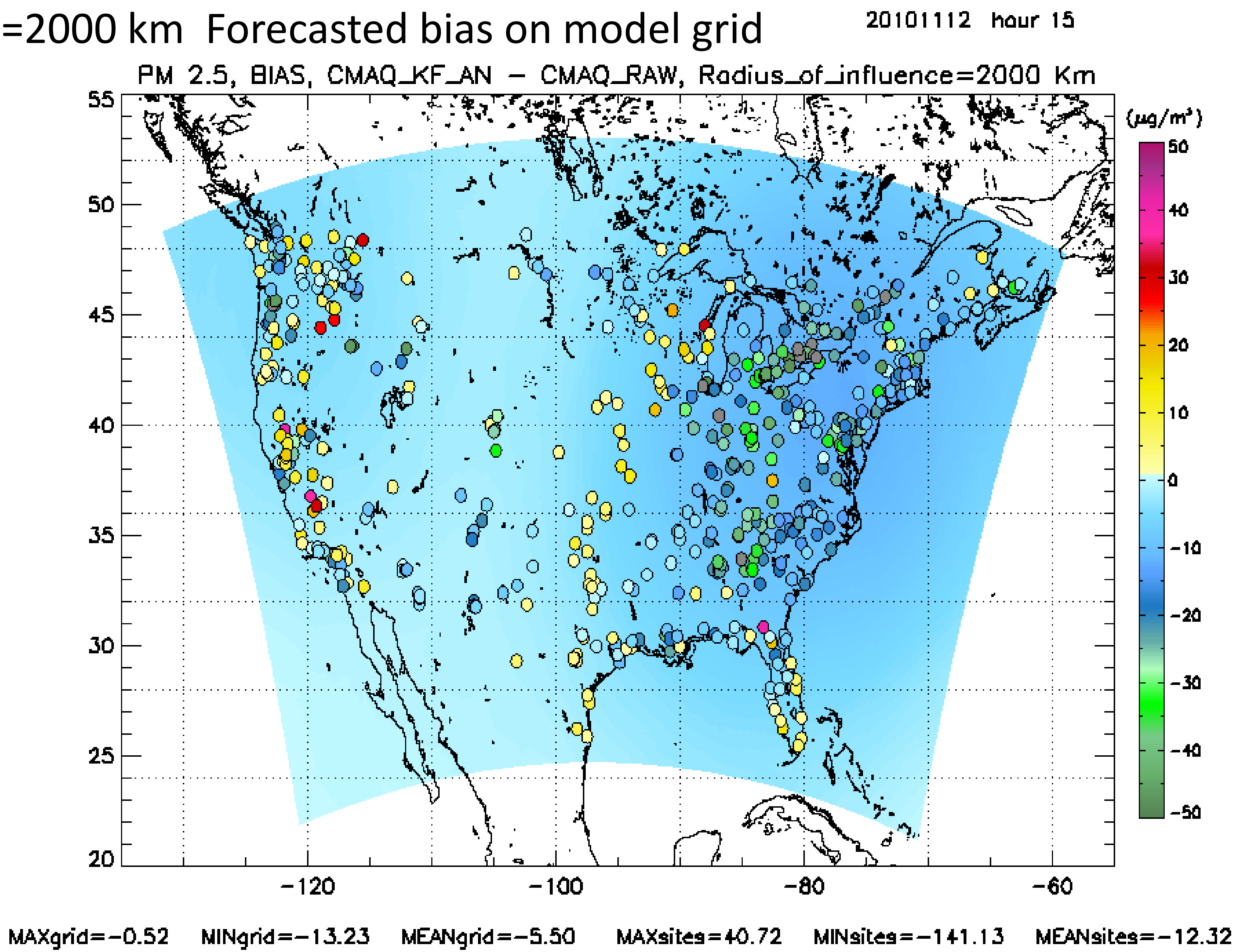
  $Mbias_{i,j} = Mbias_{i,j} + C_{i,j}$

  Summation is done over ALL obs sites k inside the circle R.

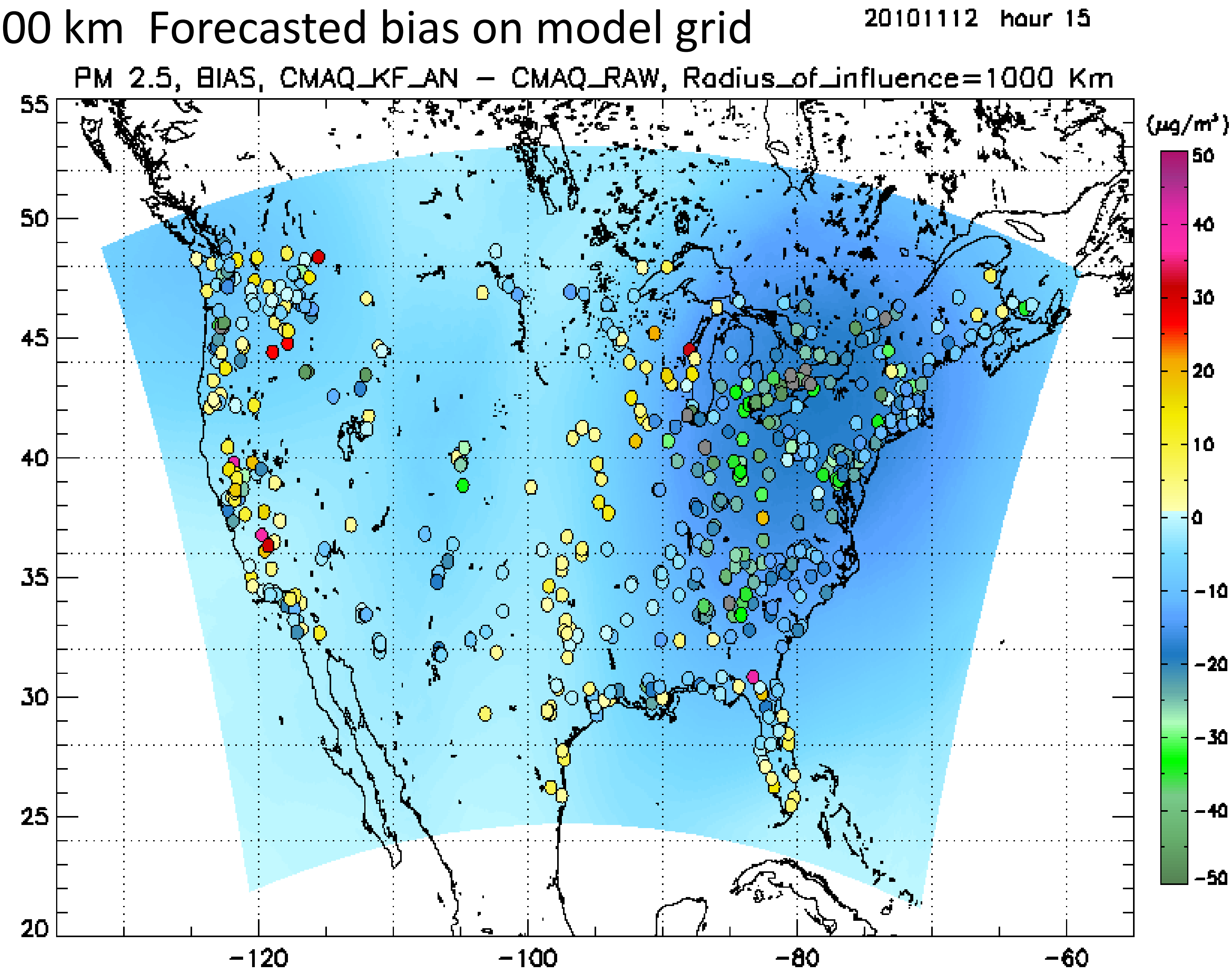
- 8 passes with R=2000, 1000, 500, 250, 125, 62, 31 & 15 km are used.

Spreading technique based on Glahn et al., 2012

R=2000 km Forecasted bias on model grid

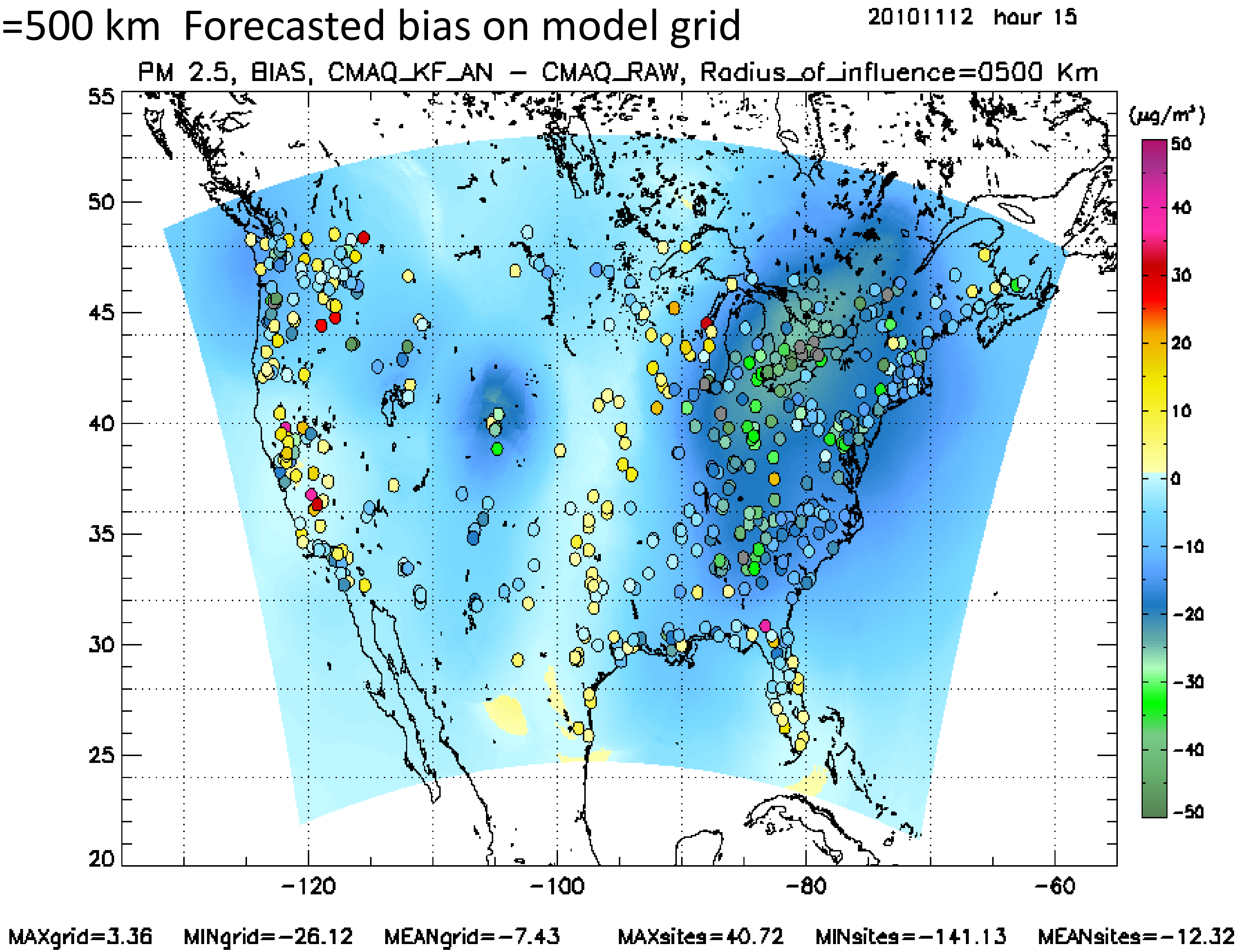


R=1000 km Forecasted bias on model grid

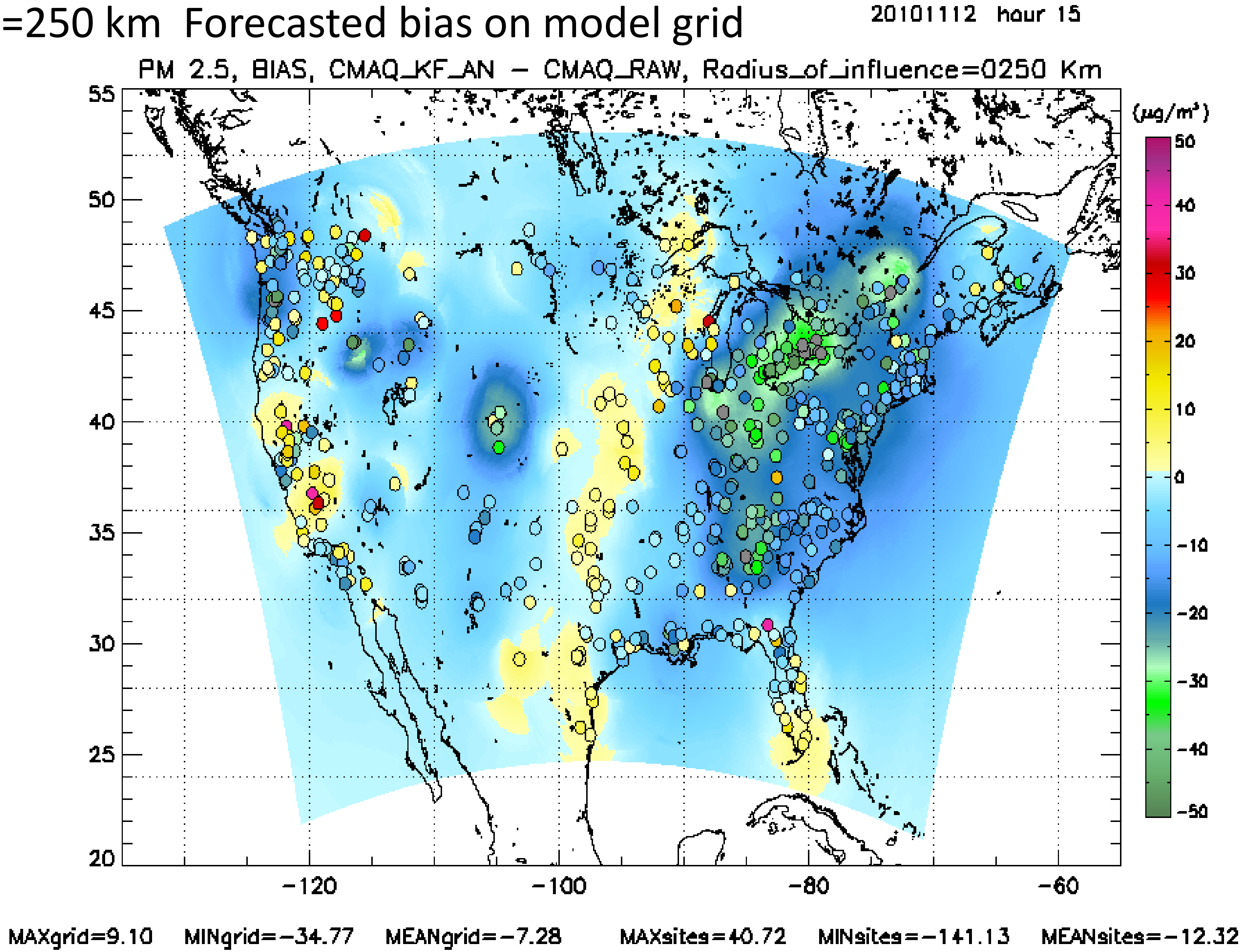


MAXgrid=−0.52   MINgrid=−19.21   MEANgrid=−7.05   MAXsites=40.72   MINsites=−141.13   MEANsites=−12.32

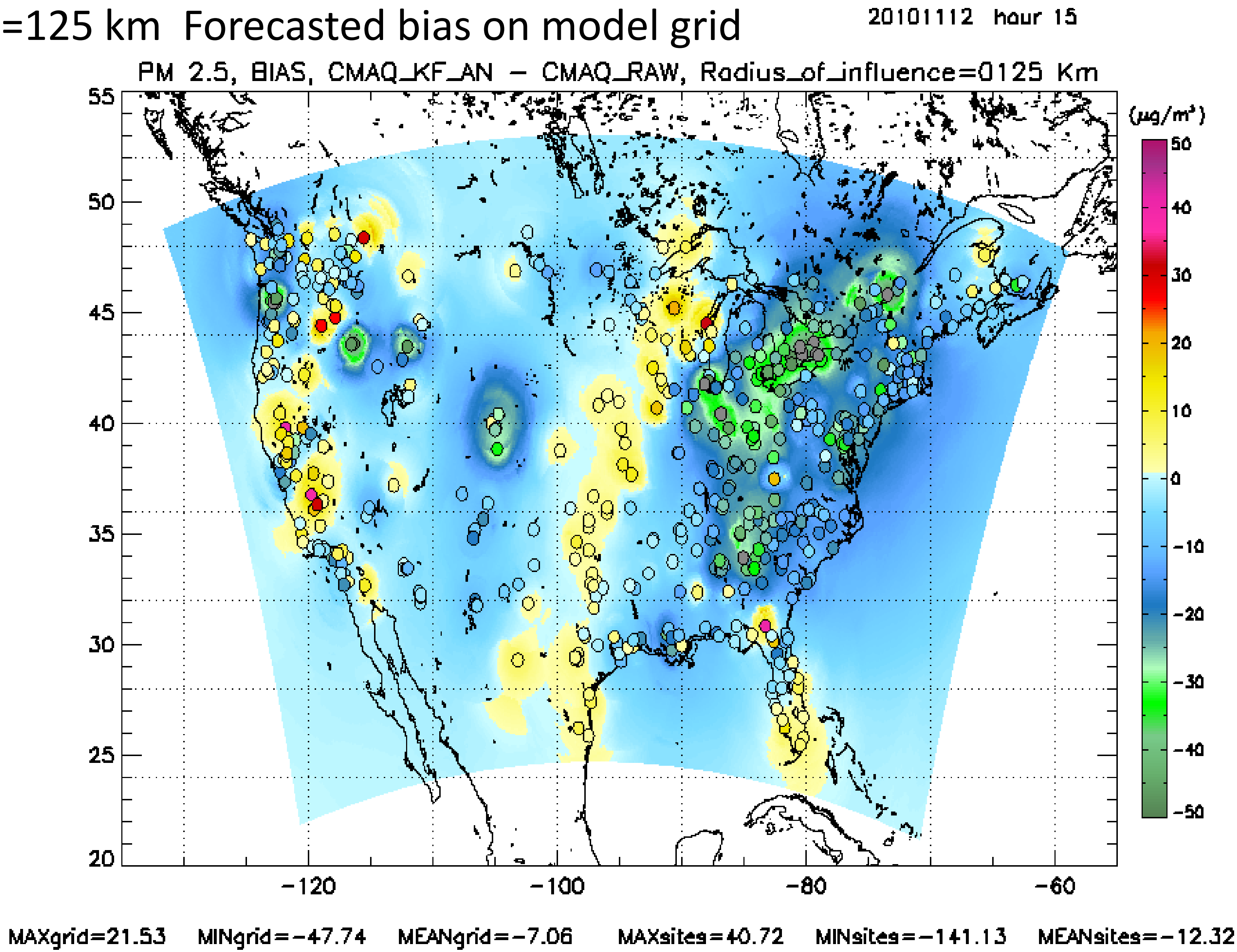
R=500 km Forecasted bias on model grid



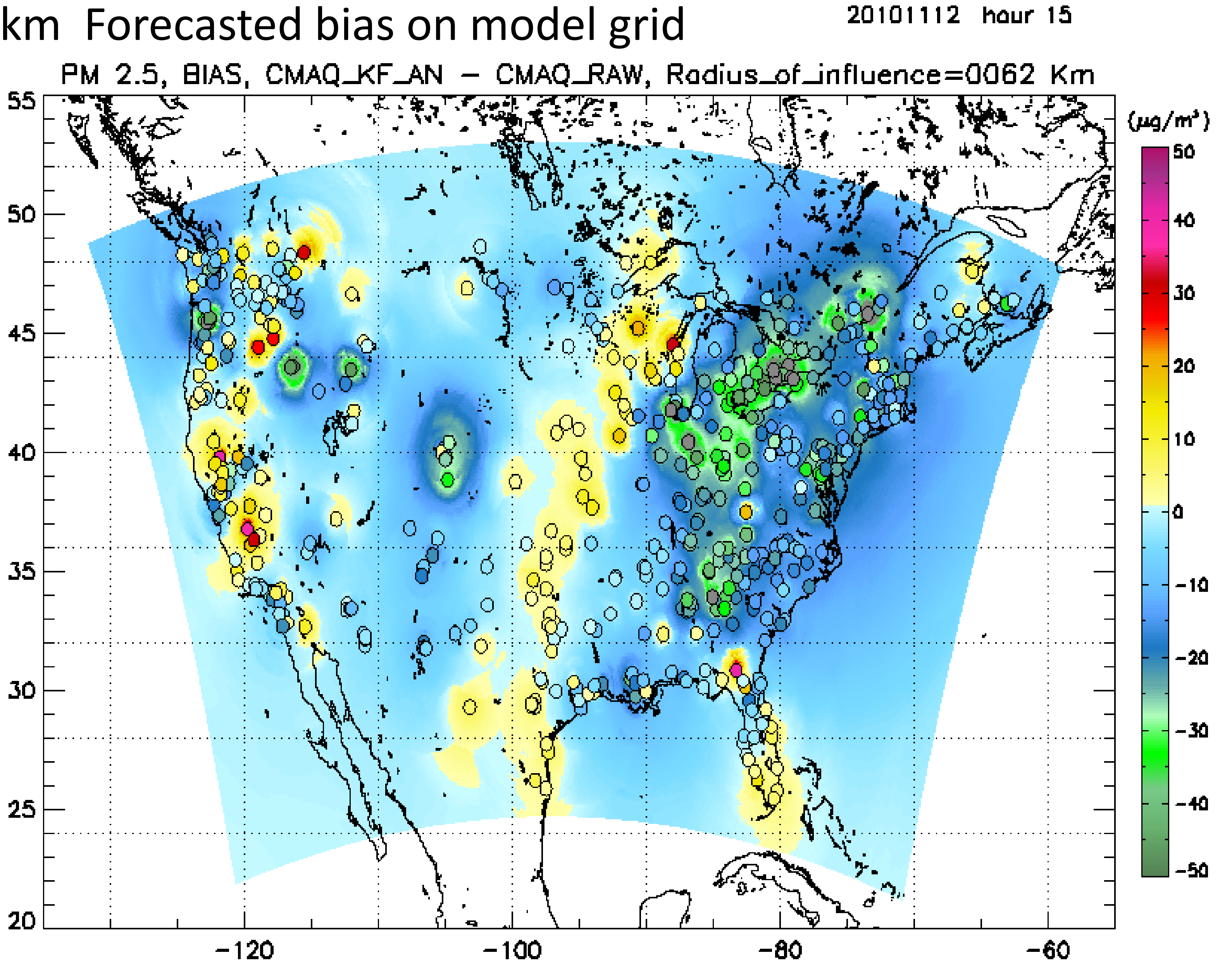
R=250 km Forecasted bias on model grid



R=125 km Forecasted bias on model grid

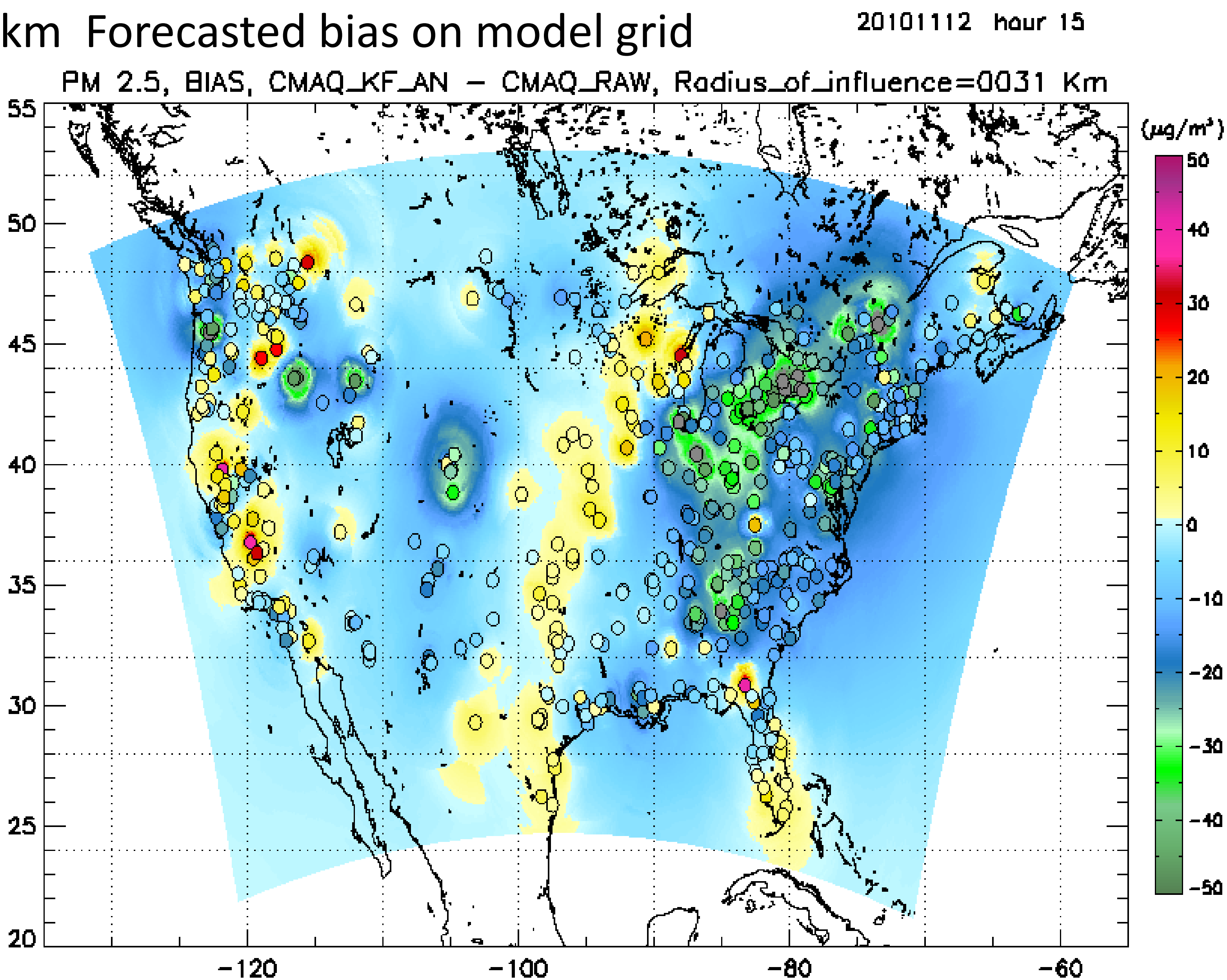


R=62 km Forecasted bias on model grid



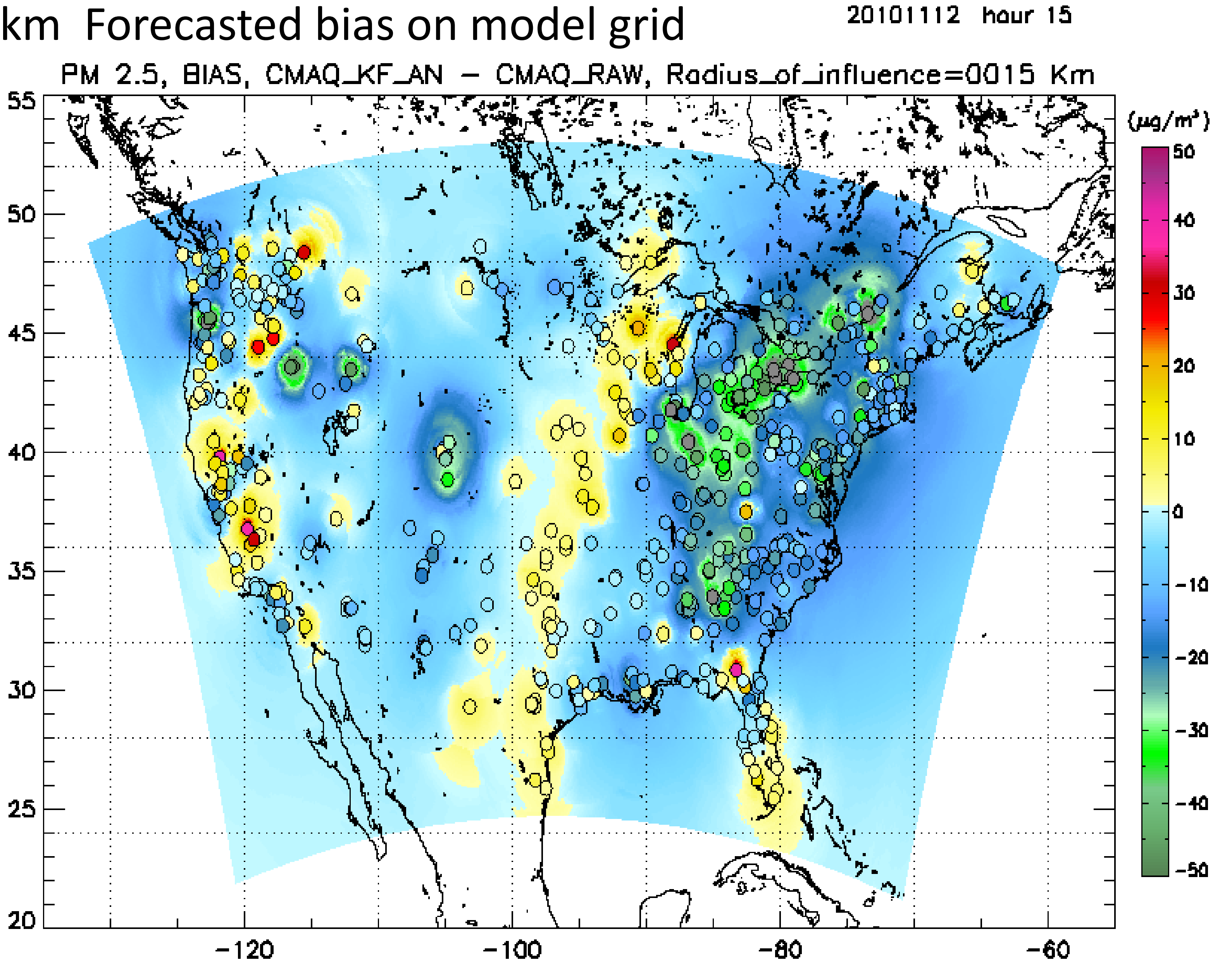
MAXgrid=39.57 MINgrid=-140.19 MEANgrid=-6.99 MAXsites=40.72 MINsites=-141.13 MEANsites=-12.32

R=31 km Forecasted bias on model grid



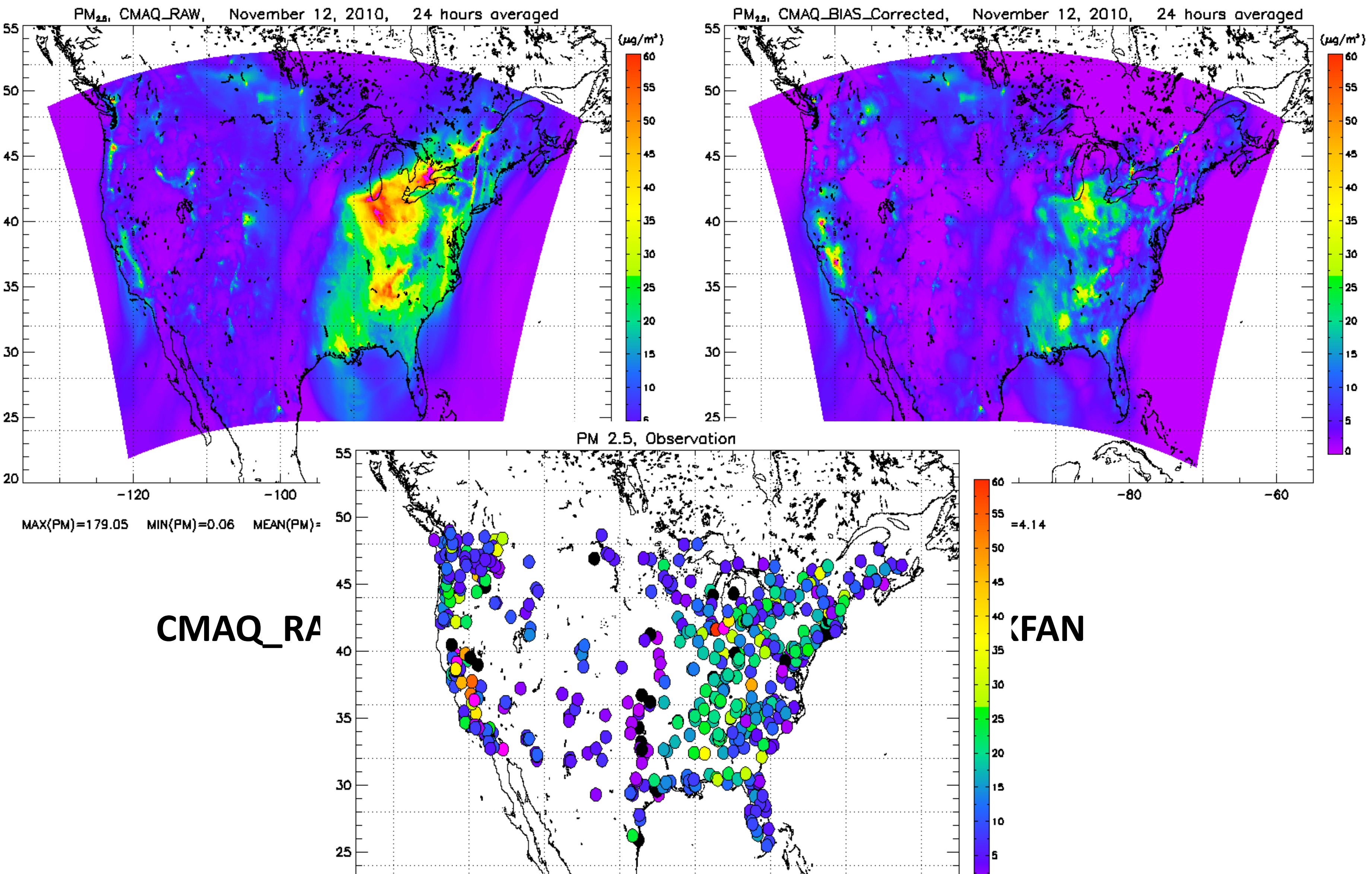
MAXgrid=40.31   MINgrid=−141.09   MEANgrid=−6.98   MAXsites=40.72   MINsites=−141.13   MEANsites=−12.32

R=15 km Forecasted bias on model grid



MAXgrid=40.69 MINgrid=-141.12 MEANgrid=-6.98 MAXsites=40.72 MINsites=-141.13 MEANsites=-12.32

24 hour Averaged PM2.5 Concentrations



CONCLUSIONS

- Automatic quality control procedures have been developed to eliminate spurious measurement values. This technique could be used in real-time operational CMAQ model runs.

- Implementation and evaluation of several post-processing bias-correction techniques has been completed. It is found that the KFAN approach that first searches for analogs and then applies the Kalman filter has the best overall skill.

- Although a significant improvement is found when using analogs based only on $PM_{2.5}$, combining $PM_{2.5}$, temperature, solar radiation, wind speed and direction together gives the most accurate analog forecasts.

- A technique has been developed for interpolating the station corrected model forecasts to the entire CMAQ grid.