Evaluation and bias-correction techniques for forecasting surface O3 and PM2.5 During the TEXAQS-II experiment of 2006

Irina Djalalova[1,2] , James Wilczak[1,]
Stuart Mckeen[3,2] ,
Georg Grell[4,2], Steven Peckham[4,2],
Mariusz Pagowski[4,2], Luca DelleMonache[5],
Jeff McQueen[6], Youhua Tang[6],
Pius Lee[7], John McHenry[8],
Wanmin Gong[9], Veronique Bouchet[10],
Rohit Mathur[11]

1 – NOAA/ESRL/Physical Sciences Division, CO, USA
2 – Also in CIRES, University of Colorado, Boulder, CO, USA
3 – NOAA/ESRL/Chemical Sciences Division, CO, USA
4 – NOAA/ESRL/Global Systems Division, CO, USA
5 – NCAR, CO, USA
6 – NOAA/NWS/Environmental Modeling Center, Maryland, USA
7 – NOAA/ARL, Maryland, USA
8 – Baron AMS, NC, USA
9 – Environment Canada, Ontario, Canada
10 – Environmental Canada, Quebec, Canada
11 – EPA/NERL, NC, USA

CIRES

TEXAQS August 12 – September 30, 2006

http://www.esrl.noaa.gov/psd/programs/2006/texaqs/verification/

OBSERVATIONS from EPA AIRNOW :

119 sites of OZONE

38 sites of PM2.5

MODELS for Verification:

NOAA NWS/NCEP,12 km NAM/CMAQ

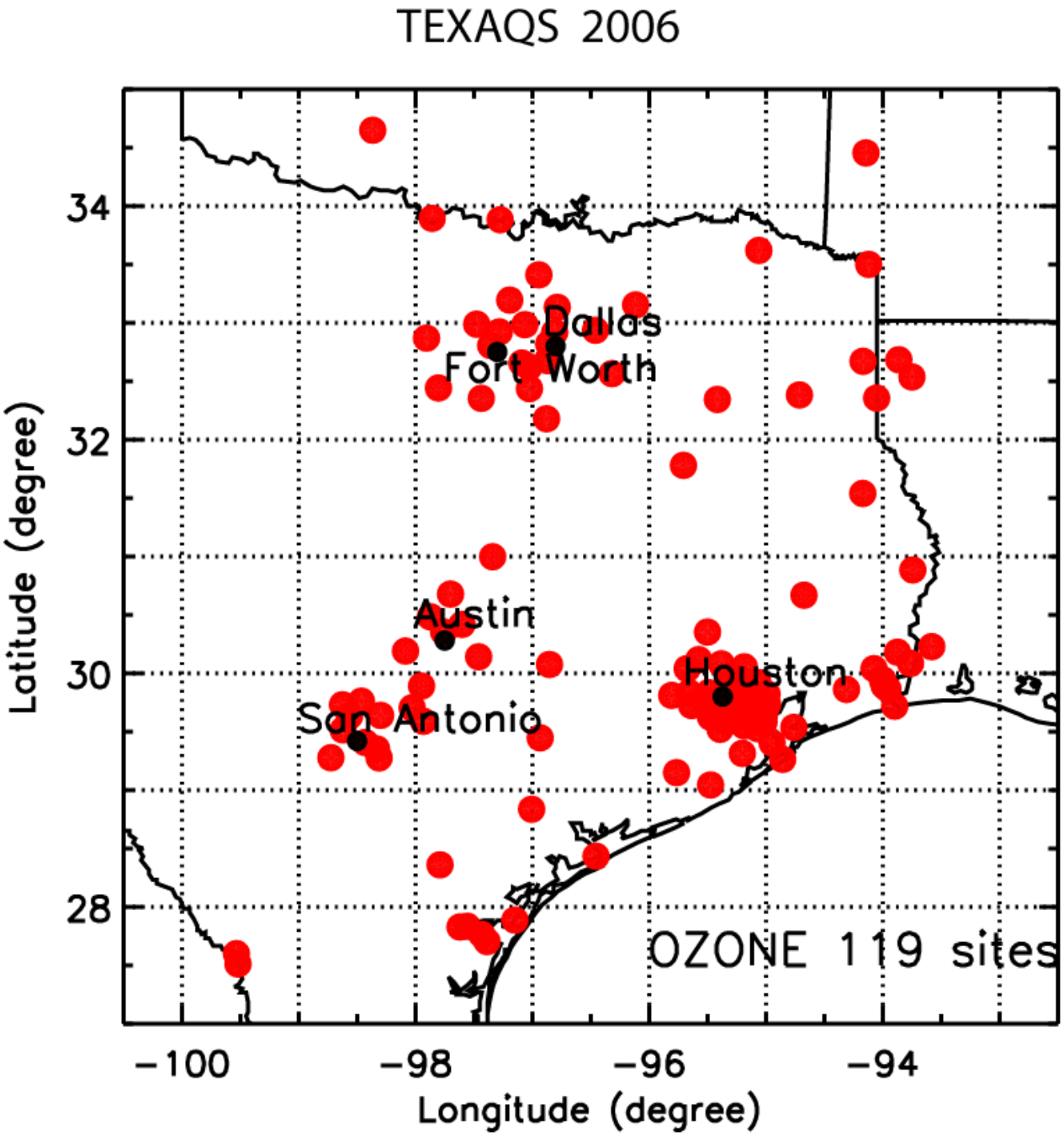
NOAA ESRL/GSD 12 & 36 km, WRF/CHEM

Canadian CMC, 21 km CHRONOS

Canadian CMC, 28 km, AURAMS

Baron AMS, 15 km, MM5/MAQSIP-RT

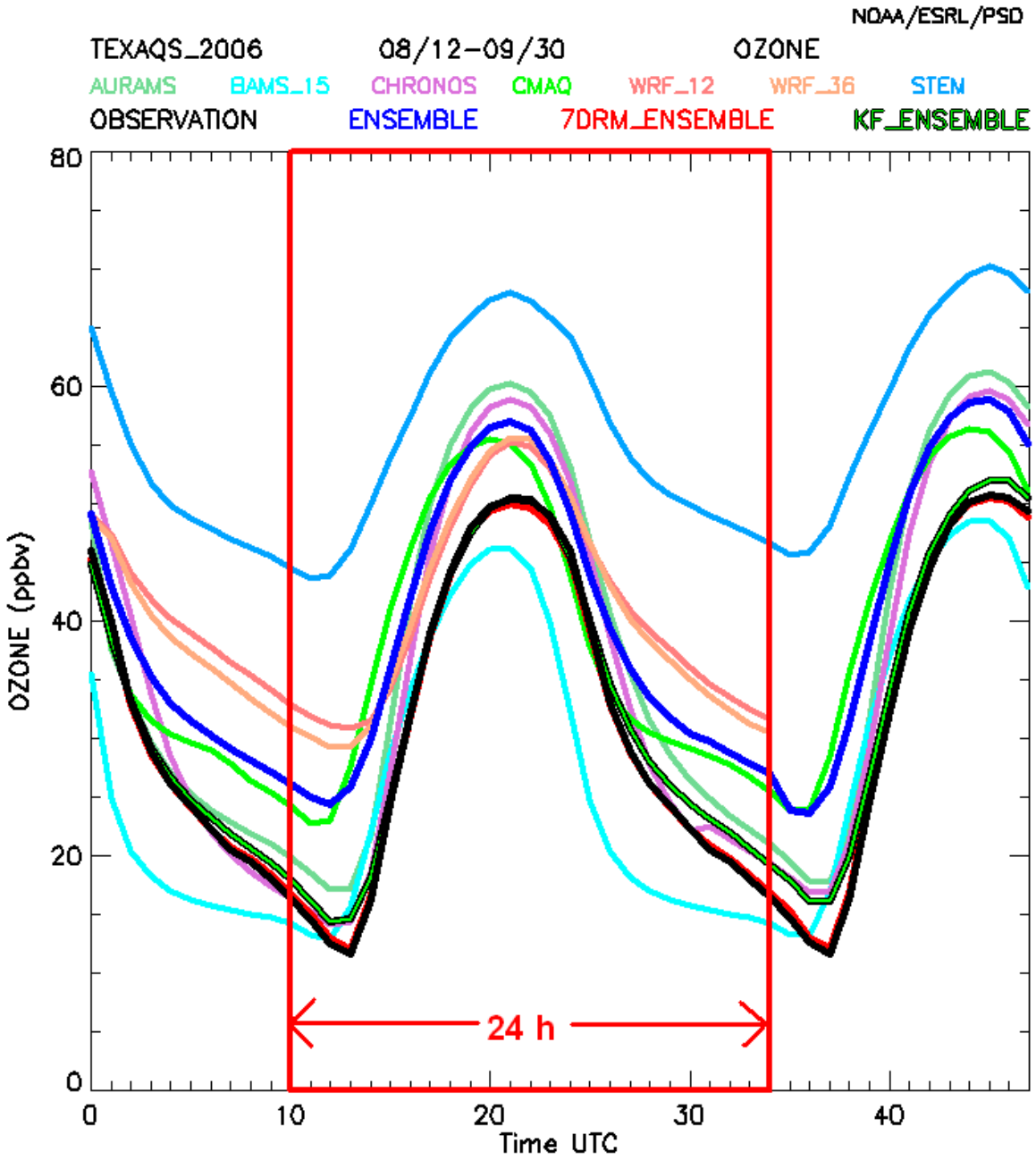
University of Iowa, 12 km WRF/STEM


TEXAQS 2006
Dallas
Fort Worth
Austin
Houston
San Antonio
OZONE 119 sites
34
32
30
28
−100
−98
−96
−94
Latitude (degree)
Longitude (degree)

ENSEMBLES

For each site, day and hour:

- Ensemble = Σ(models)
- 7DRM_Ensemble = Σ(7days_bias_corrected models)
- KF_Ensemble = Σ (7days_KF_models)



As the EPA standard, daily 8-hour maximum ozone is calculated for each day by using a sliding window to produce a time-series of 8-hour averaged ozone and then selecting the maximum of these values in the 24 hour window corresponding to 10-34 UTC.

Singular Value Decomposition (SVD) Method

$$\begin{bmatrix} m11 & m12 & \dots & m17 & 1 \\ m21 & m22 & \dots & m27 & 1 \\ \dots & \dots & \dots & \dots & \dots \\ mN1 & mN2 & \dots & mN7 & 1 \end{bmatrix} \begin{bmatrix} weight1 \\ weight2 \\ \dots \\ weight7 \\ weight8 \end{bmatrix} = \begin{bmatrix} Obs\ 1 \\ Obs\ 2 \\ \dots \\ \dots \\ \dots \\ Obs\ N \end{bmatrix}$$

M – number of models = 7

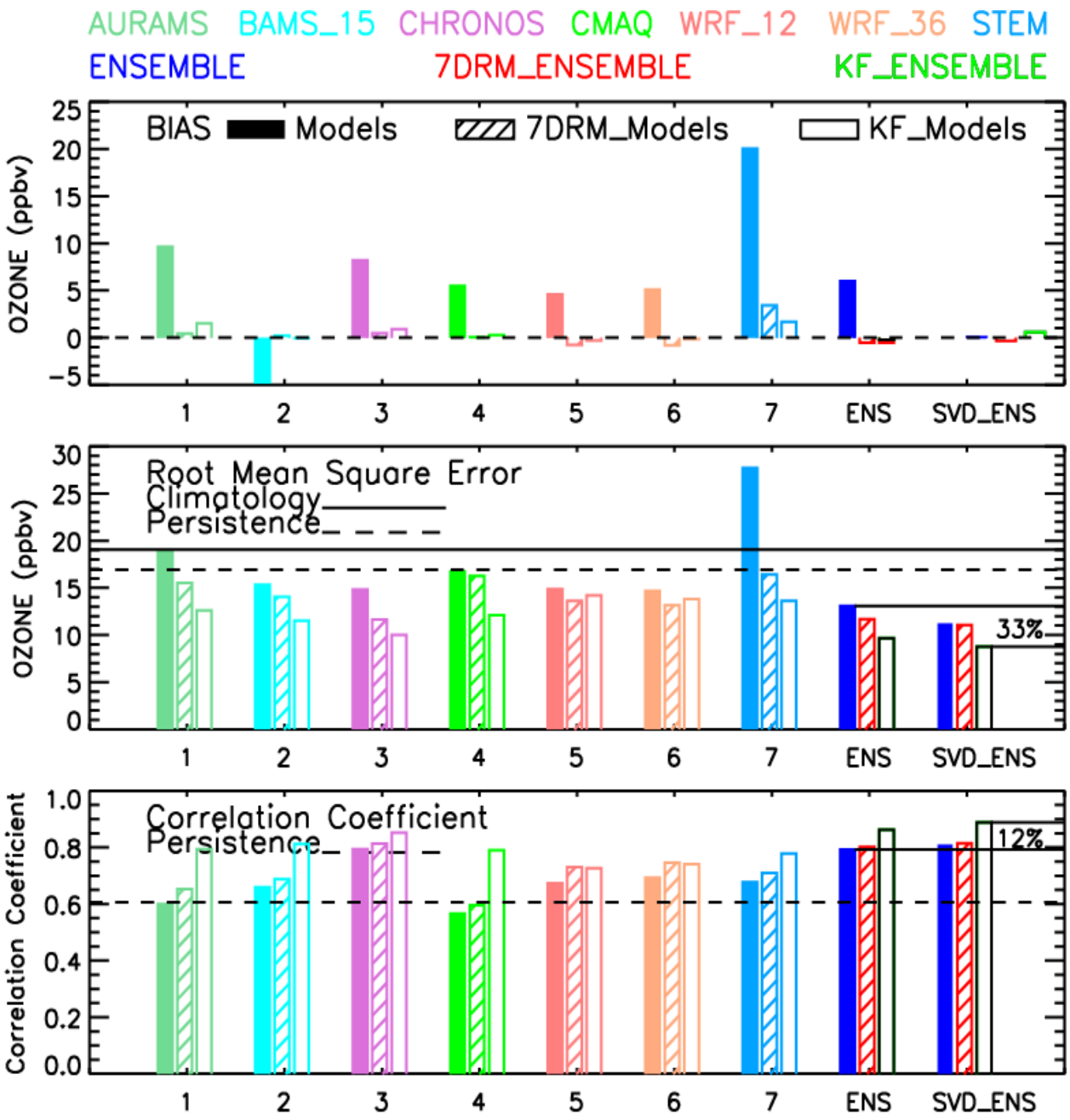
N – number of the observation points, N =7days*119sites ~800 >>7

weight 1:7 – weights of the models in the Ensemble

weight8 – bias of the Ensemble

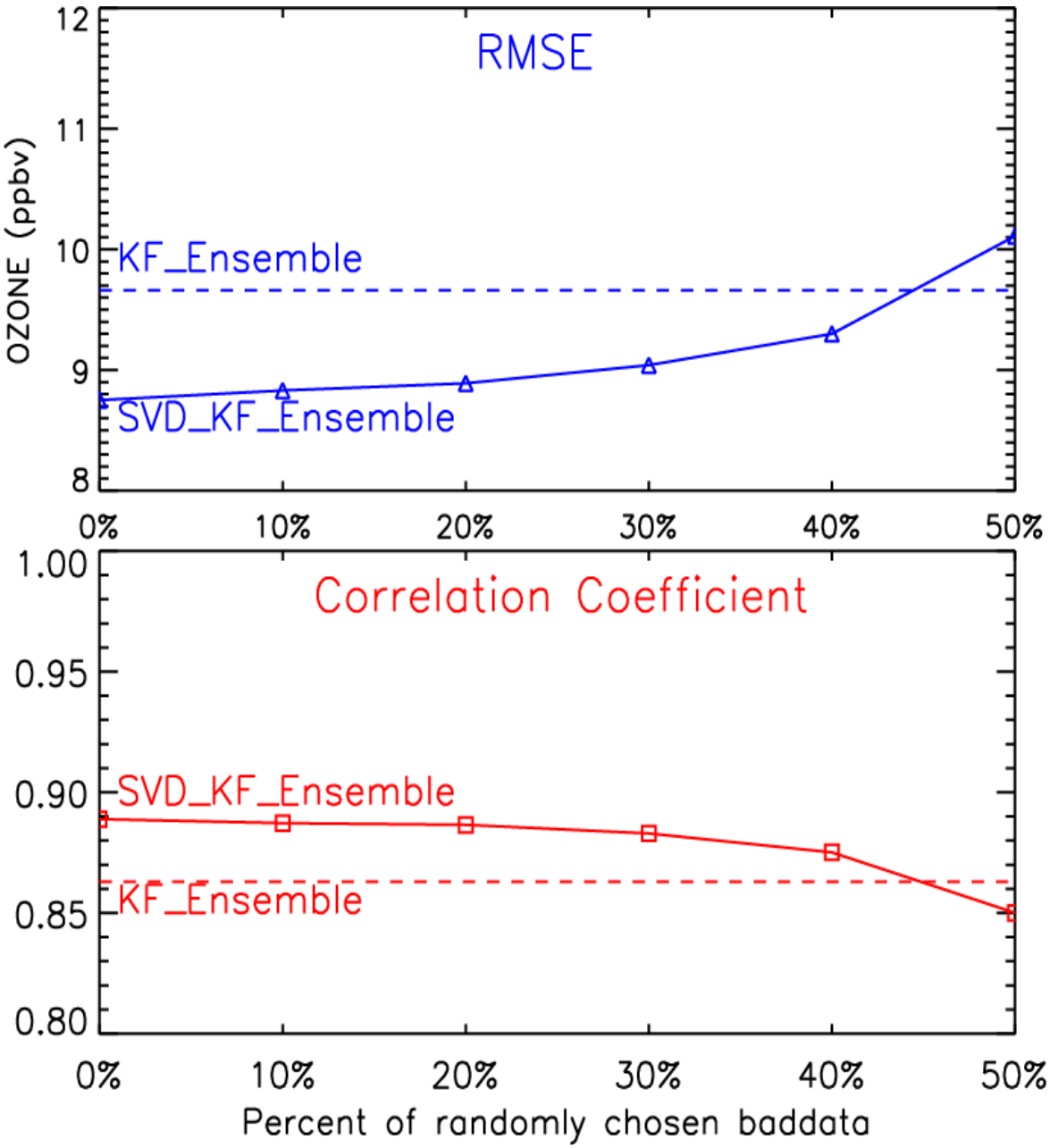
Note: 1) weights are different for each hour of the forecast cycle
2) weights are determined using the previous 7 days of data
3) single set of weights is determined for all sites

SVD_Ensemble = Σ(weight * Model) + bias
SVD_7DRM_Ensemble = Σ(weight * 7DRM_Model) + bias
SVD_KF_Ensemble = Σ(weight * KF_Model) + bias



The two bias correction schemes improve the statistics for all models, but larger improvements for most models occur with the KF scheme.

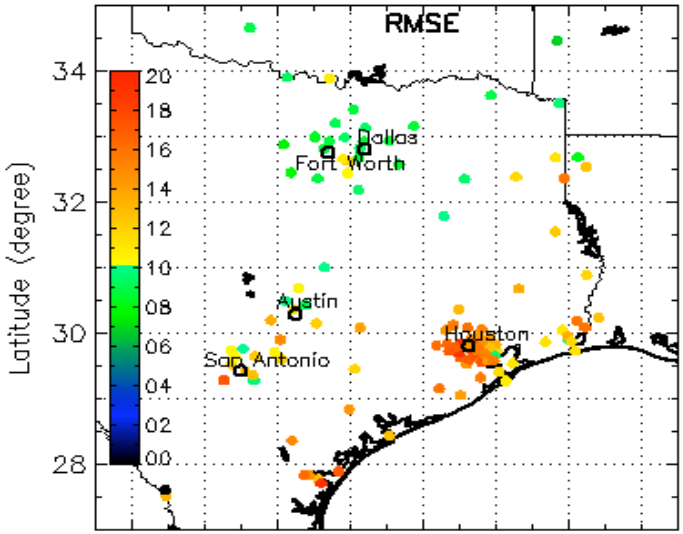
For ensembles the highest skill comes from the combination of both the SVD and KF techniques.

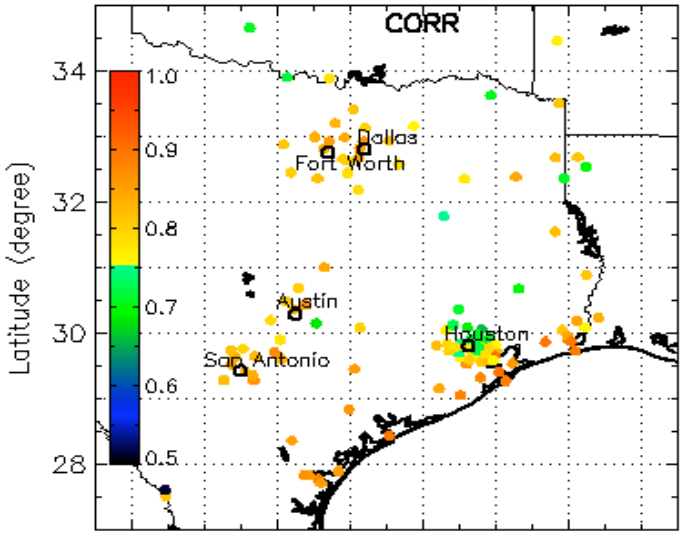


Using random number generator we artificially shrink the data up to 50%.

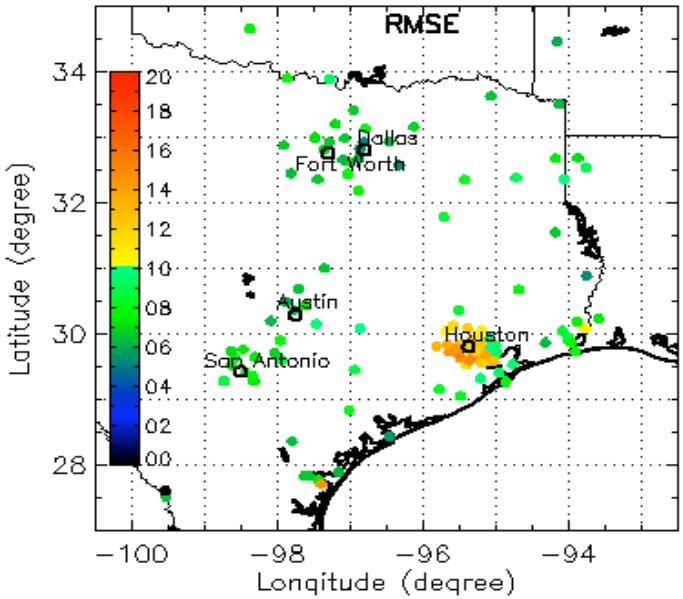
Even with 60% of reliable data SVD_KF ensemble over-performed original KF_ensemble in terms of RMSE and Correlation Coefficient.

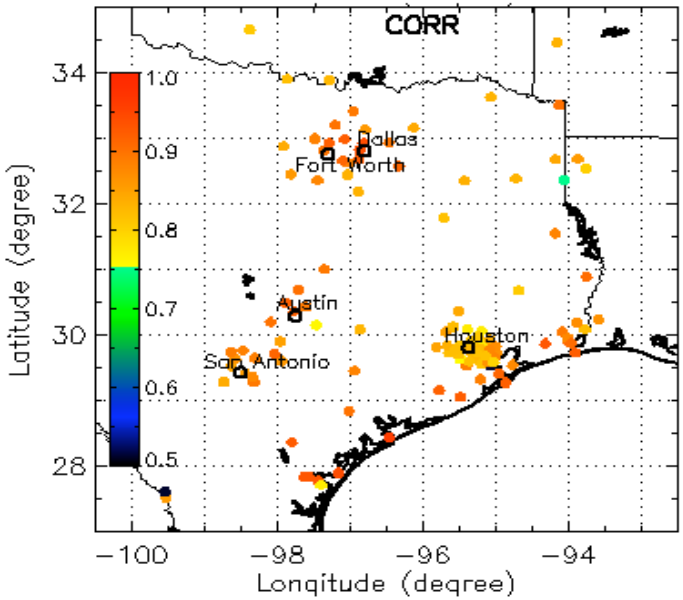
Ensemble 8 hours MAX OZONE

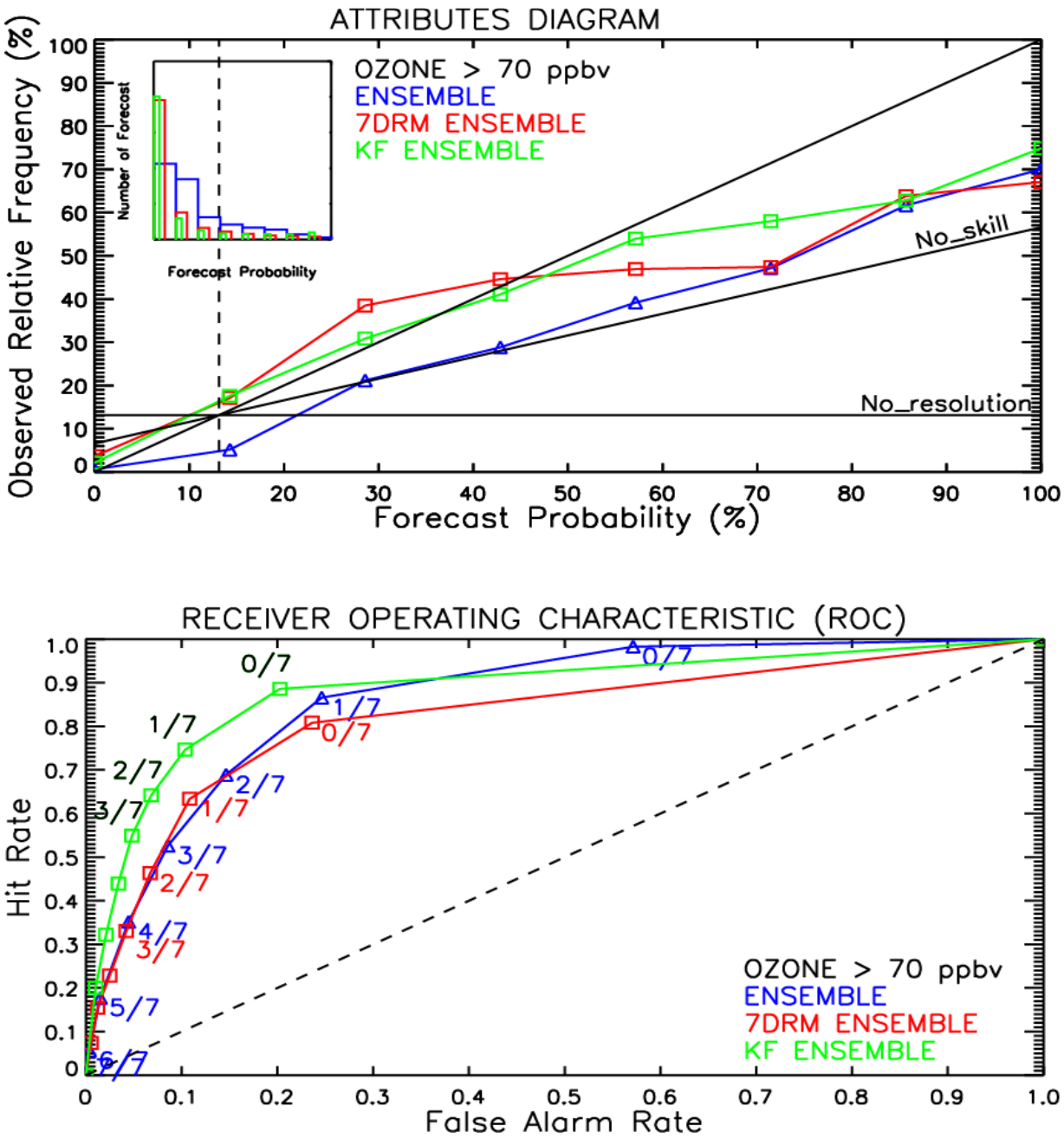


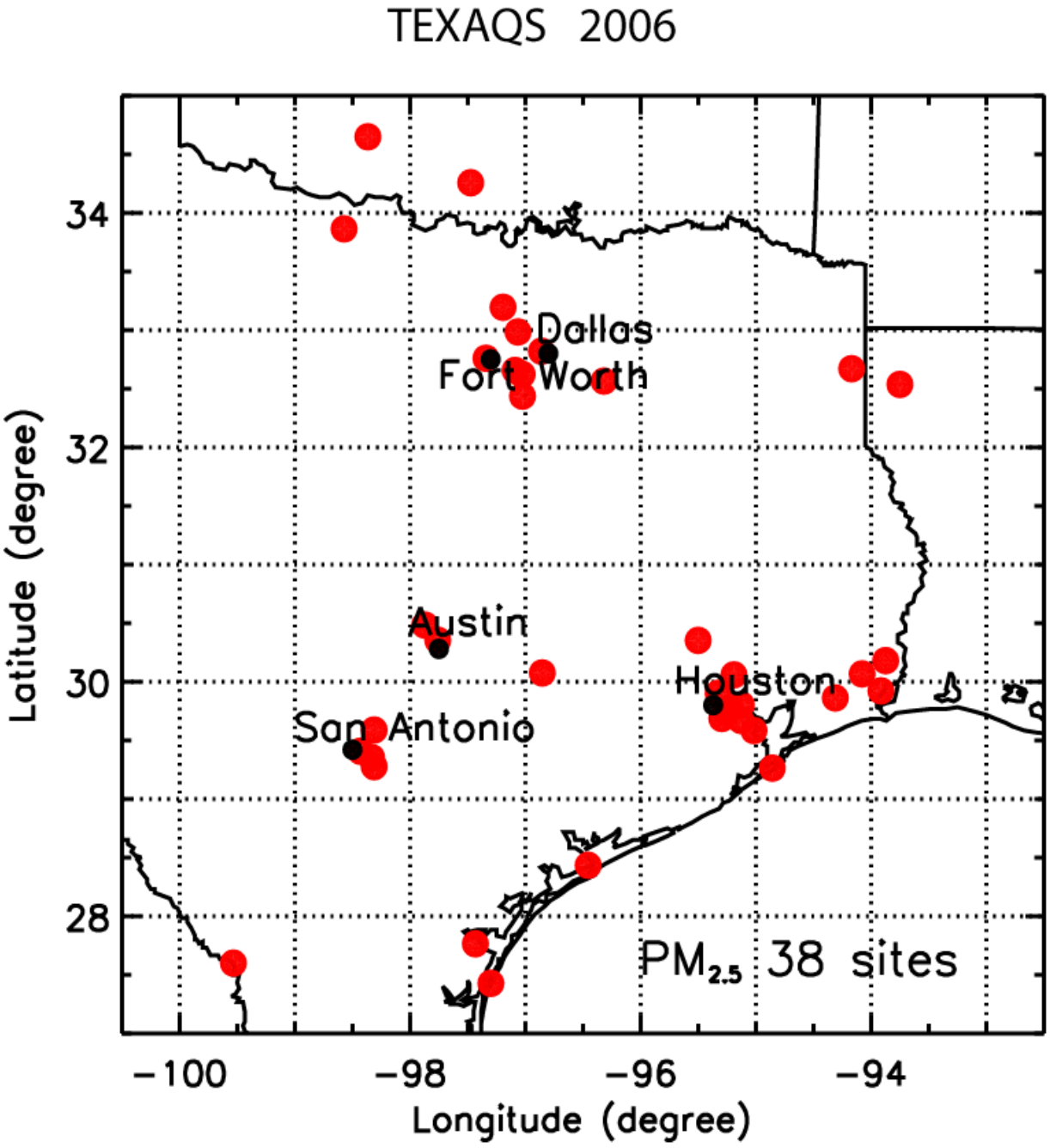


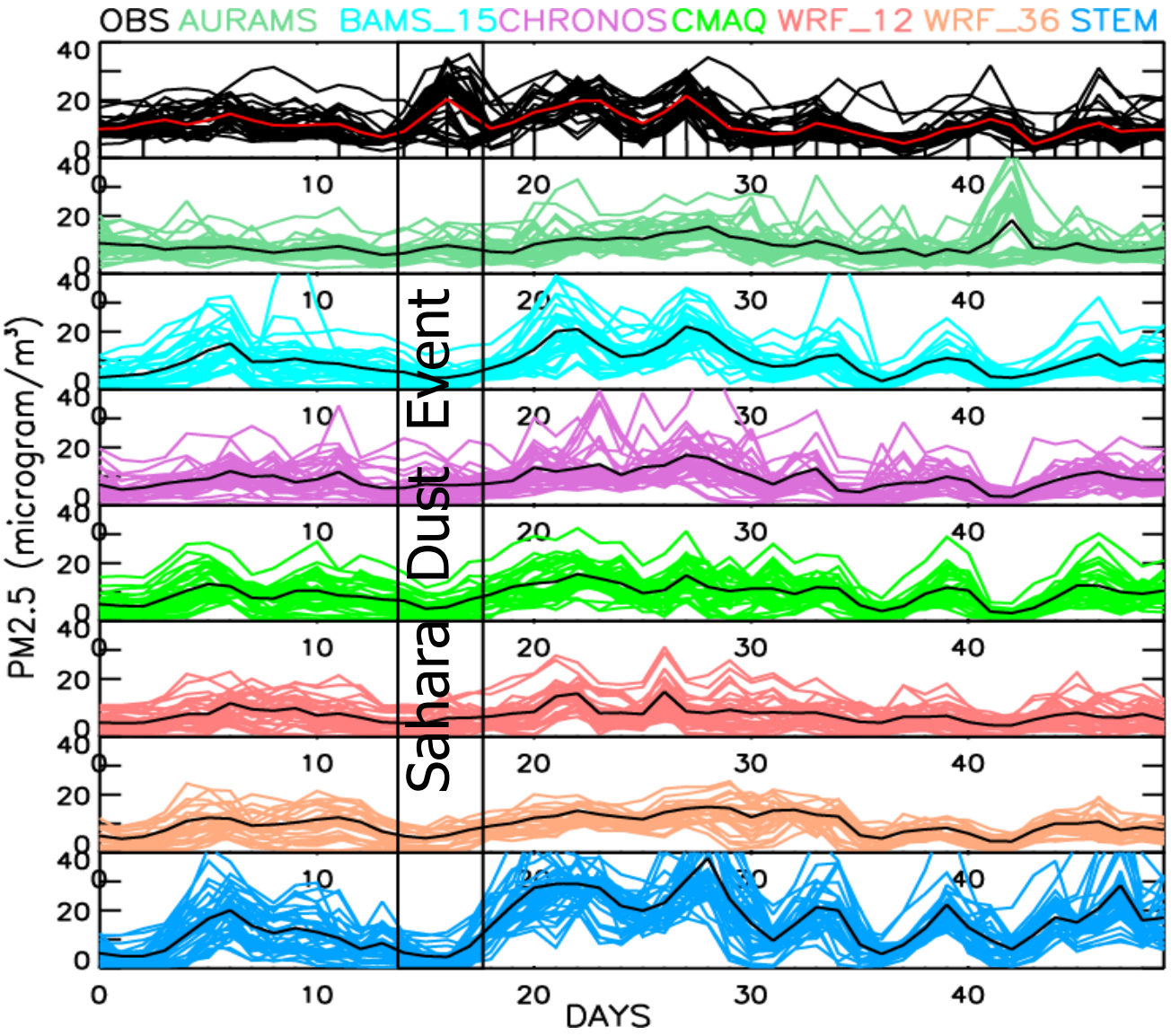
SVD_KF_Ensemble 8 hours MAX OZONE



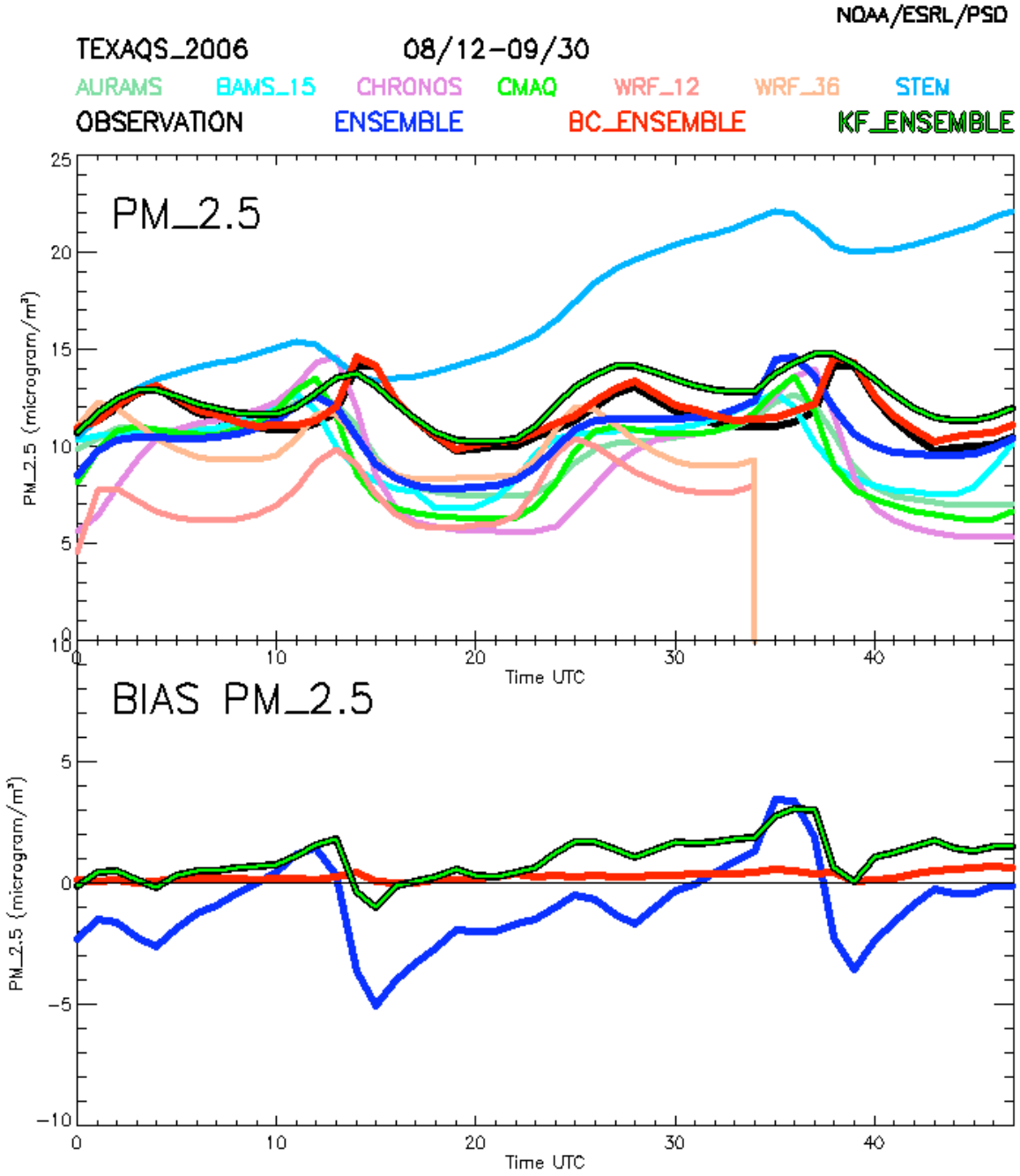



ATTRIBUTES DIAGRAM
OZONE > 70 ppbv
ENSEMBLE
7DRM ENSEMBLE
KF ENSEMBLE
No_skill
No_resolution
Observed Relative Frequency (%)
Forecast Probability (%)
Number of Forecast
Forecast Probability
RECEIVER OPERATING CHARACTERISTIC (ROC)
Hit Rate
False Alarm Rate
0/7
1/7
2/7
3/7
4/7
5/7
6/7
OZONE > 70 ppbv
ENSEMBLE
7DRM ENSEMBLE
KF ENSEMBLE

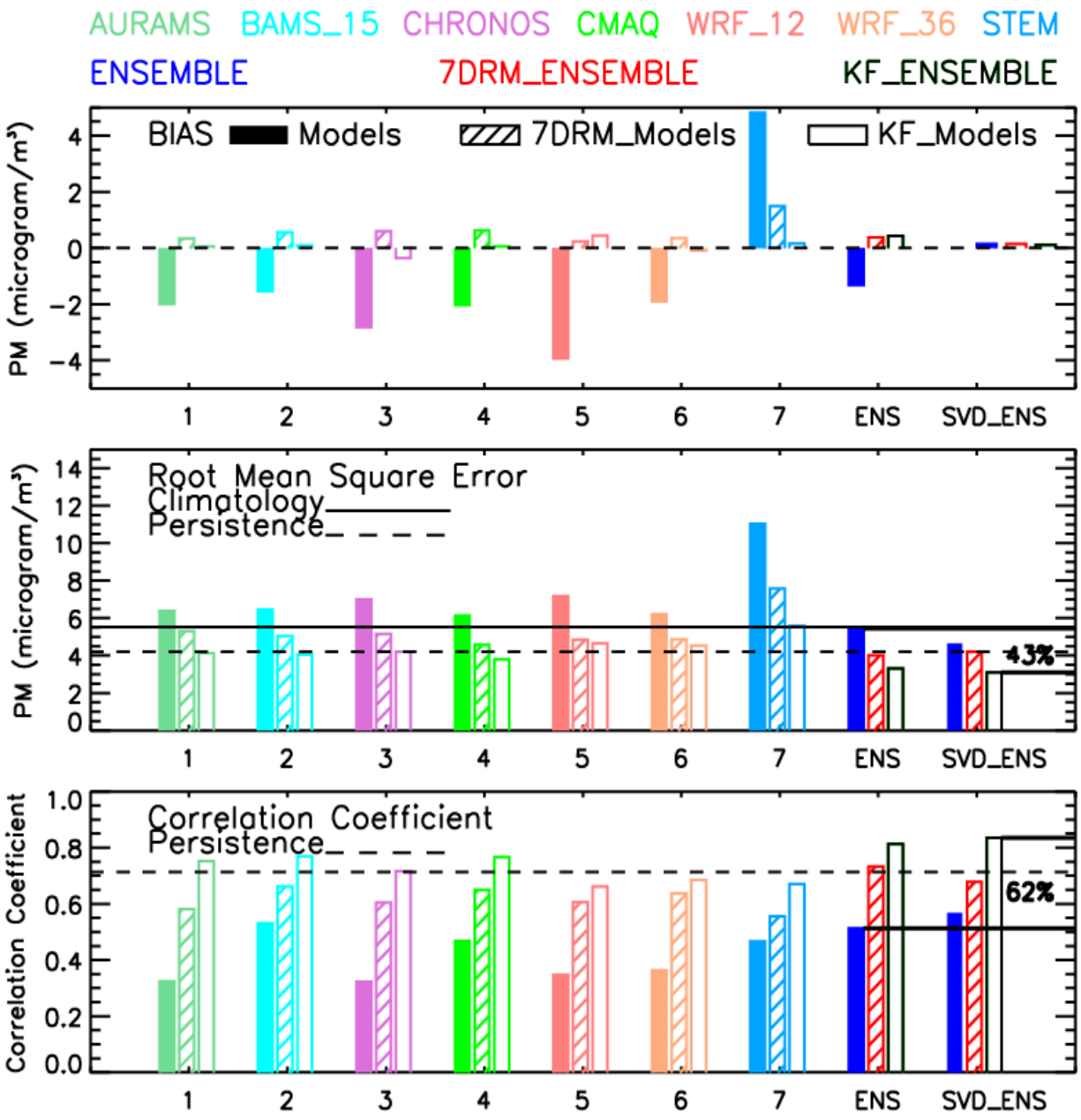

TEXAQS 2006
Dallas
Fort Worth
Austin
San Antonio
Houston
PM2.5 38 sites
Latitude (degree)
Longitude (degree)
34
32
30
28
−100
−98
−96
−94



To eliminate the Sahara dust influence in the data, we omitted $PM_{2.5}$ values between August 27-30, 2006, as shown in the black box, for all sites south of 31 degrees latitude.

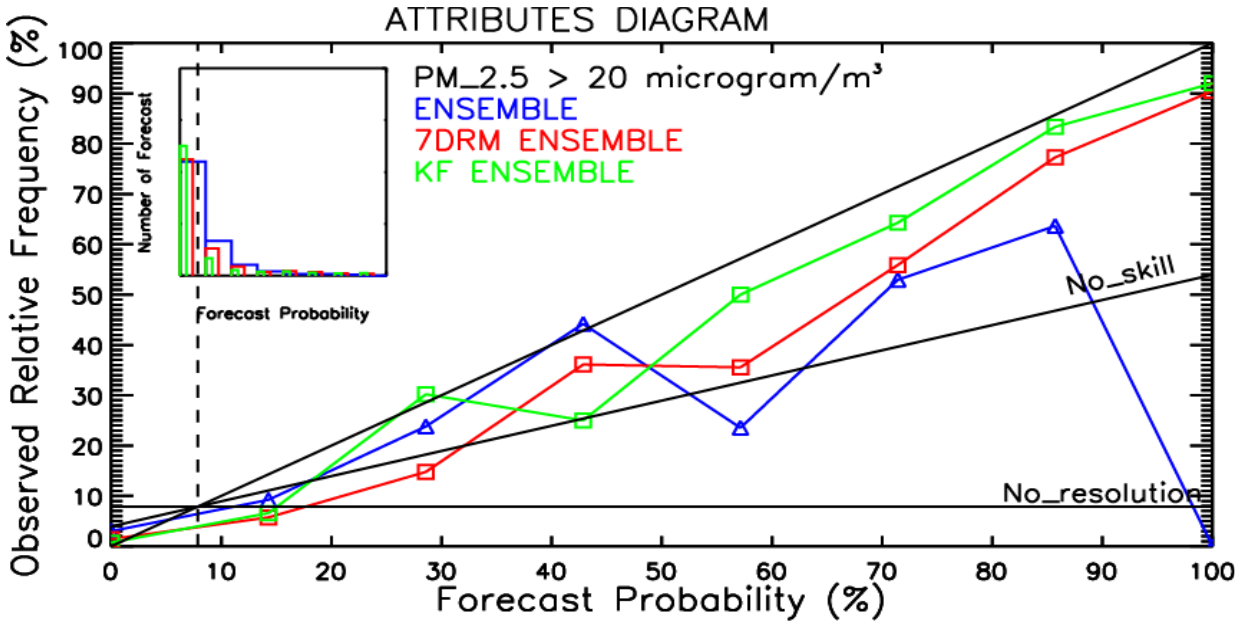


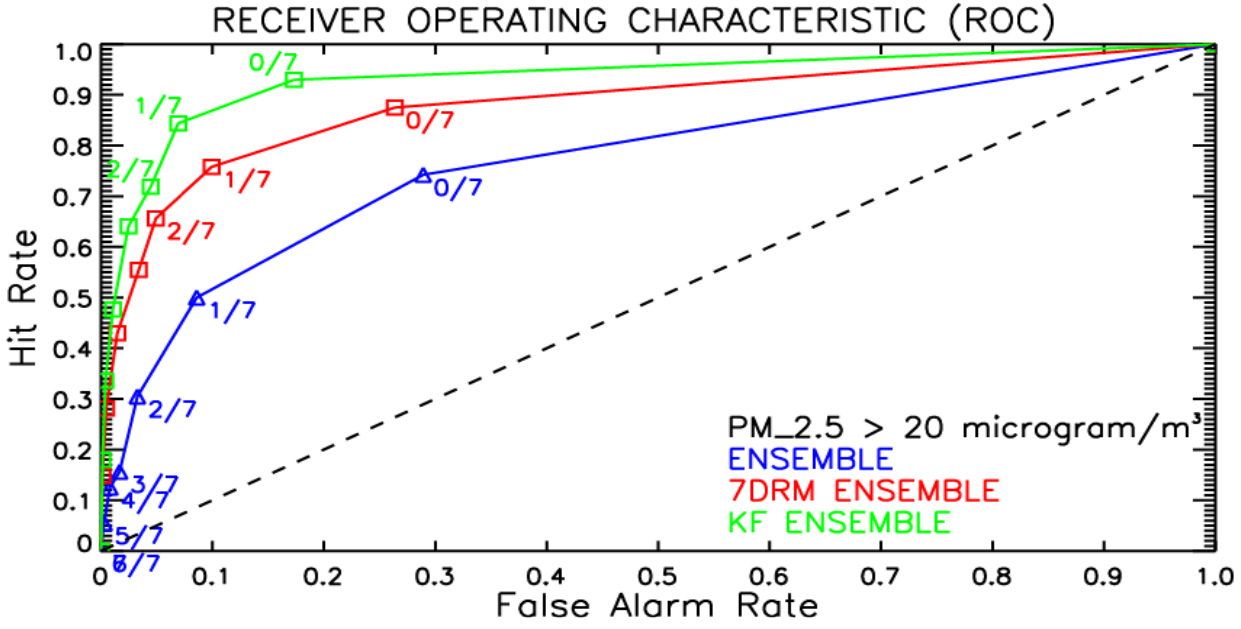
Pm 2.5 has a double spike in the diurnal cycle, which is following by all ensembles and individual models.



Not a single raw or 7DRM bias corrected model is able to perform better than persistence.

Only the KF ensemble and SVD_KF ensemble are capable of significantly beating the persistence forecast.


ATTRIBUTES DIAGRAM
PM_2.5 > 20 microgram/m³
ENSEMBLE
7DRM ENSEMBLE
KF ENSEMBLE
No_skill
No_resolution
Observed Relative Frequency (%)
Forecast Probability (%)
Number of Forecast
Forecast Probability
0
10
20
30
40
50
60
70
80
90
100


RECEIVER OPERATING CHARACTERISTIC (ROC)
PM_2.5 > 20 microgram/m³
ENSEMBLE
7DRM ENSEMBLE
KF ENSEMBLE
Hit Rate
False Alarm Rate
0/7
1/7
2/7
3/7
4/7
5/7
6/7
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

CONCLUSIONS

- Ensemble beats all individual models.
- Bias Corrected models have better skill for RMSE and for correlation than uncorrected ones.
- 7DRM_Ensemble, KF_Ensemble and especially SVD_KF_Ensemble significally improve all skills.
- For PM2.5, 7DRM_Ensemble and especially KF_Ensemble and SVD_KF_Ensemble are the only models that perform better than persistence in terms of RMSE and correlation coefficient.
- All Ensembles use data only from 7 previous days so can be calculated on a daily basis during future experiments.