NCAR

Kalman Filter and Analogs to Improve Numerical Weather Predictions

Luca Delle Monache*, Gregory Roux, Yubao Liu, and Thomas Warner
NCAR

Thomas Nipen, and Roland Stull
UBC

Neil Jacobs, and Peter Childs
AirDat LLC

The 4th EnKF Workshop -- April 7, 2010, Rensselaerville, NY

* lucadm@ucar.edu

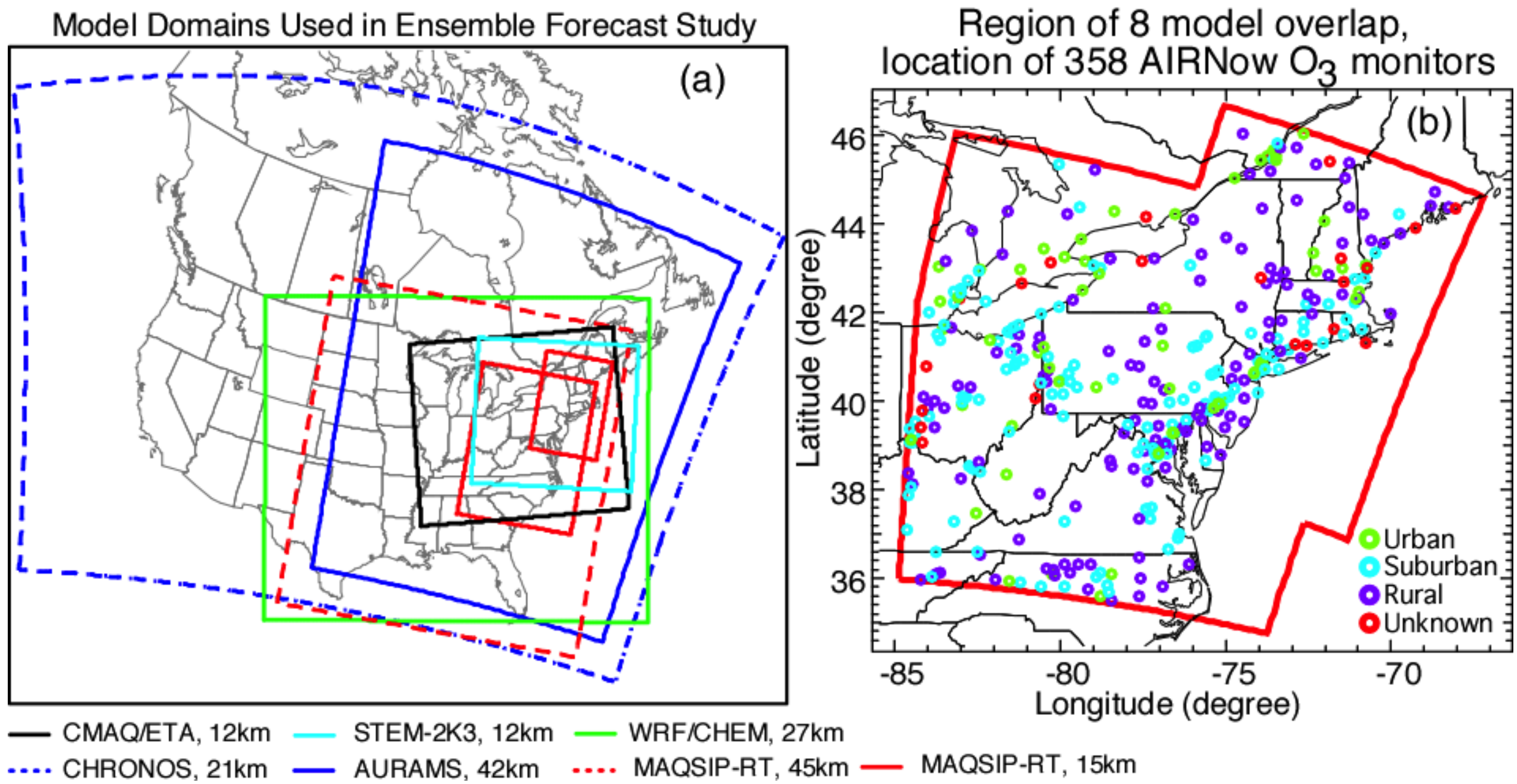
Outline

NCAR

- Ensemble and Kalman filtering (KF) for air quality predictions
- A new method based on KF and an analog approach (ANKF, AN)
- Test KF, ANKF, and AN to correct 10m wind speed
- Application of new methods within an EnDA framework
- Summary

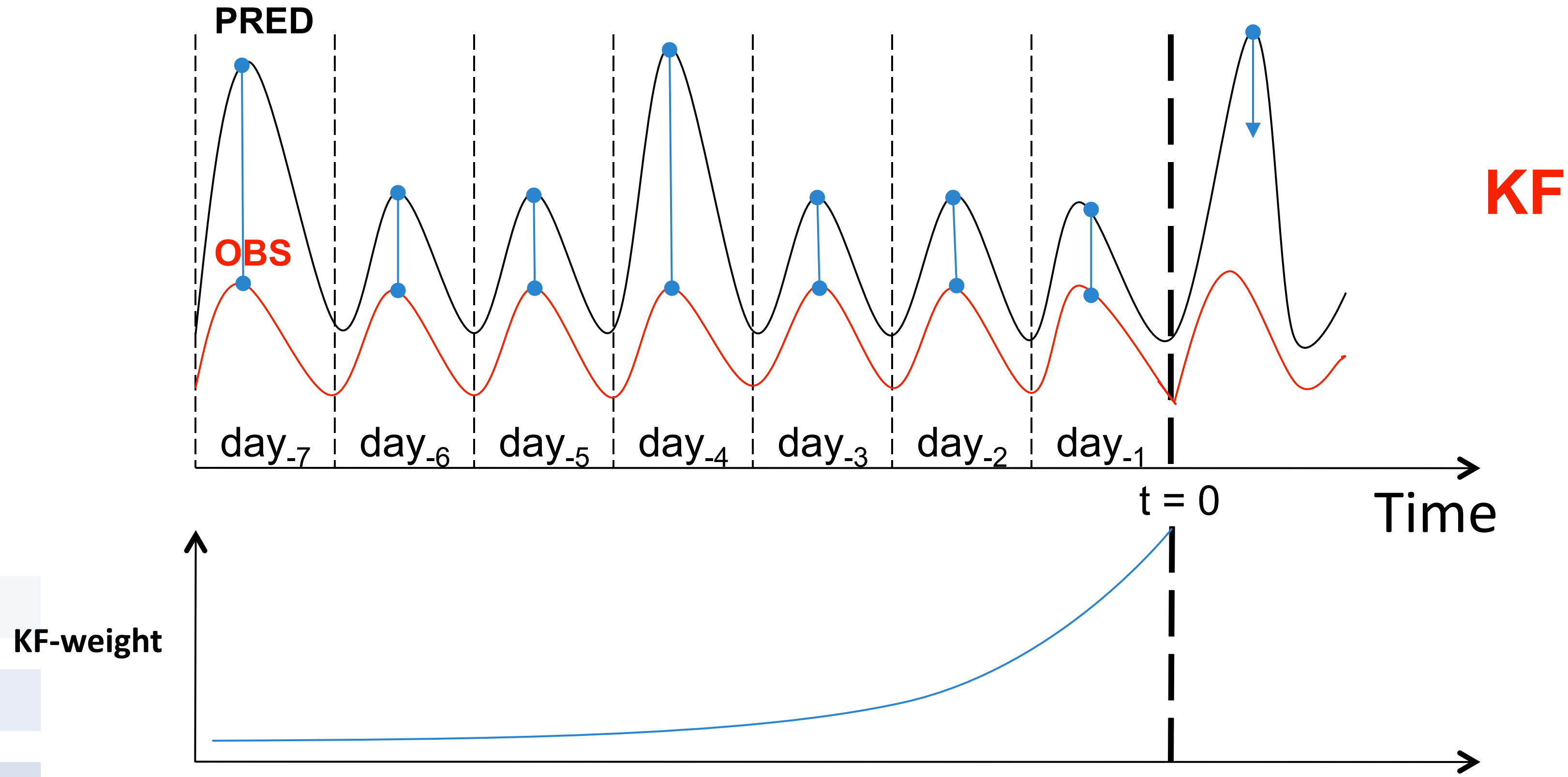
A Kalman filter bias correction for deterministic and probabilistic ozone predictions

NCAR



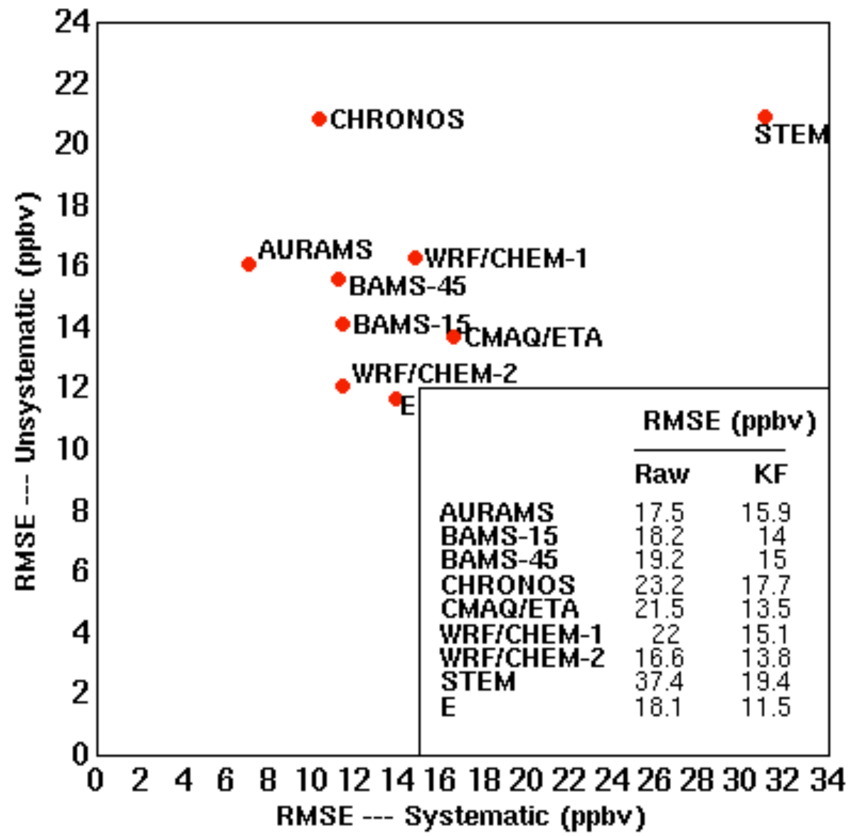
- Summer of 2004 (56 days)
- 8 photochemical models
- 360 ozone surface stations

Sources:
Delle Monache et al. (JGR, 2006b)
Delle Monache et al. (Tellus B, 2008)
Djalalova et al. (Atmospheric Environment, 2010)


PRED
OBS
KF
day-7
day-6
day-5
day-4
day-3
day-2
day-1
t = 0
Time
KF-weight

Ensemble averaging and Kalman Filtering effects on systematic and unsystematic RMSE components

NCAR



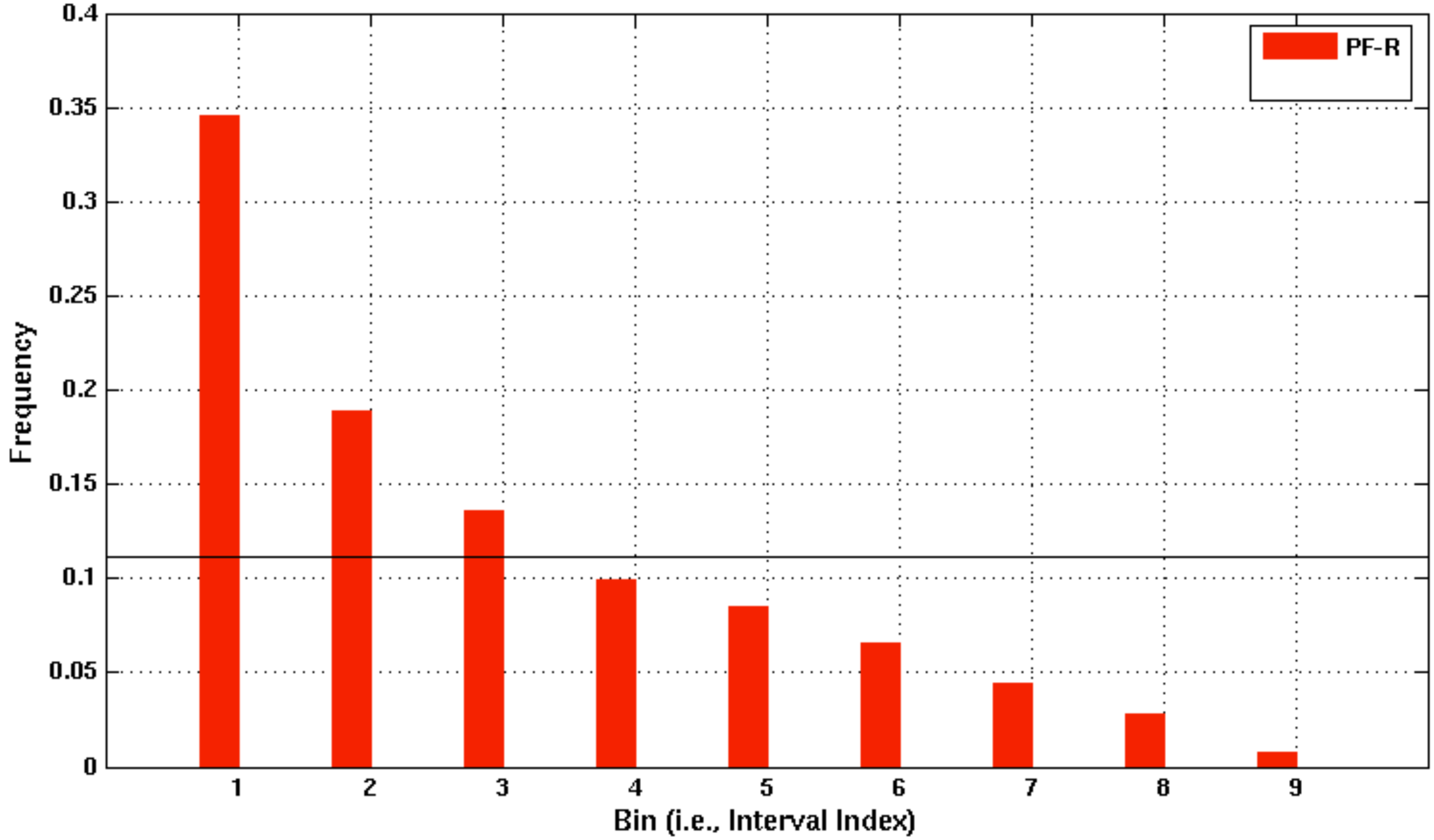
RMSE (ppbv)	Raw	KF
AURAMS	17.5	15.9
BAMS-15	18.2	14
BAMS-45	19.2	15
CHRONOS	23.2	17.7
CMAQ/ETA	21.5	13.5
WRF/CHEM-1	22	15.1
WRF/CHEM-2	16.6	13.8
STEM	37.4	19.4
E	18.1	11.5

RMSE decomposition (Willmott, Physical Geography 1981):

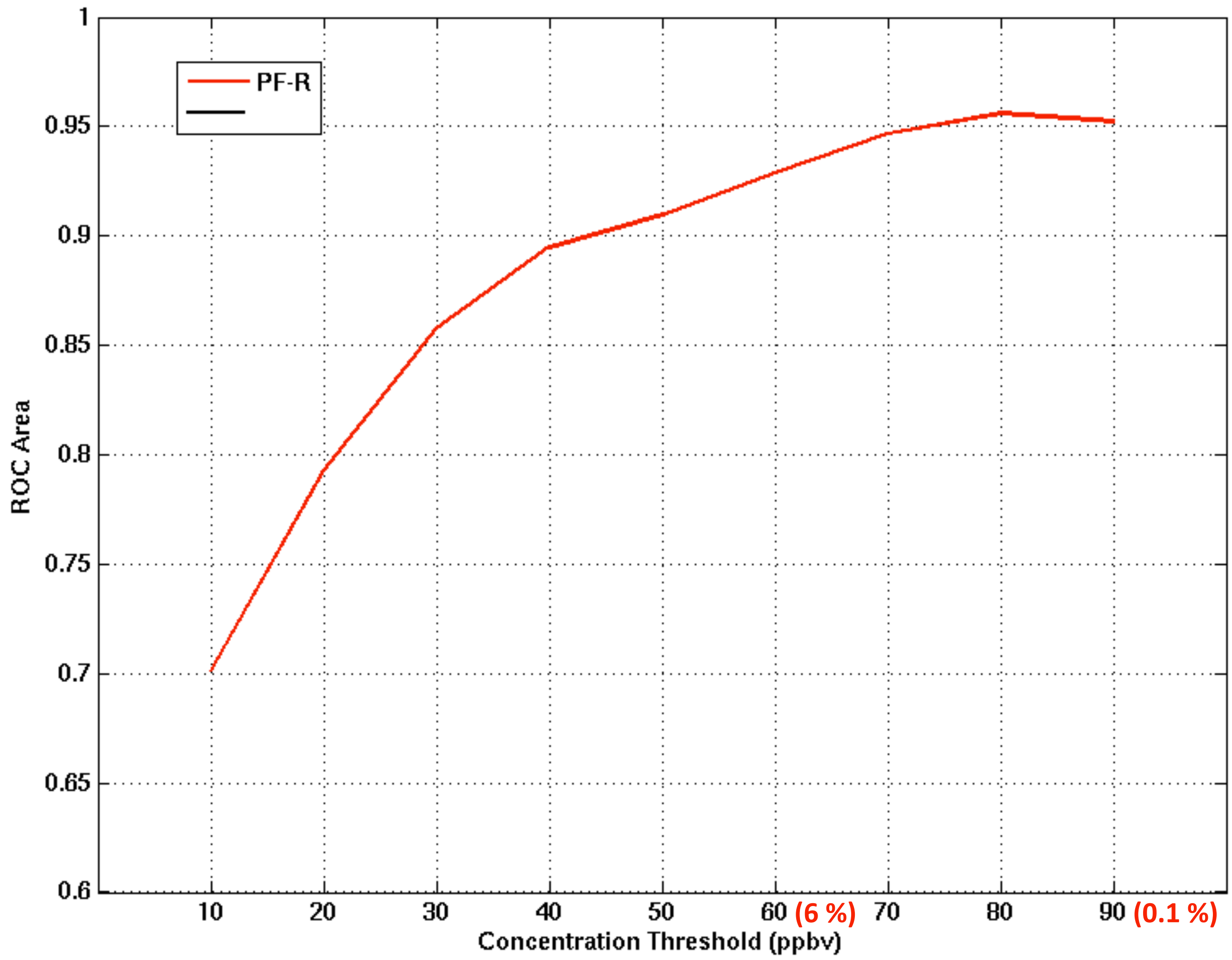
$$RMSE = \sqrt{RMSE_s^{\ 2} + RMSE_u^{\ 2}}, \quad RMSE_s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{P}_i - O_i)^2}, \quad RMSE_u = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{P}_i - P_i)^2}$$

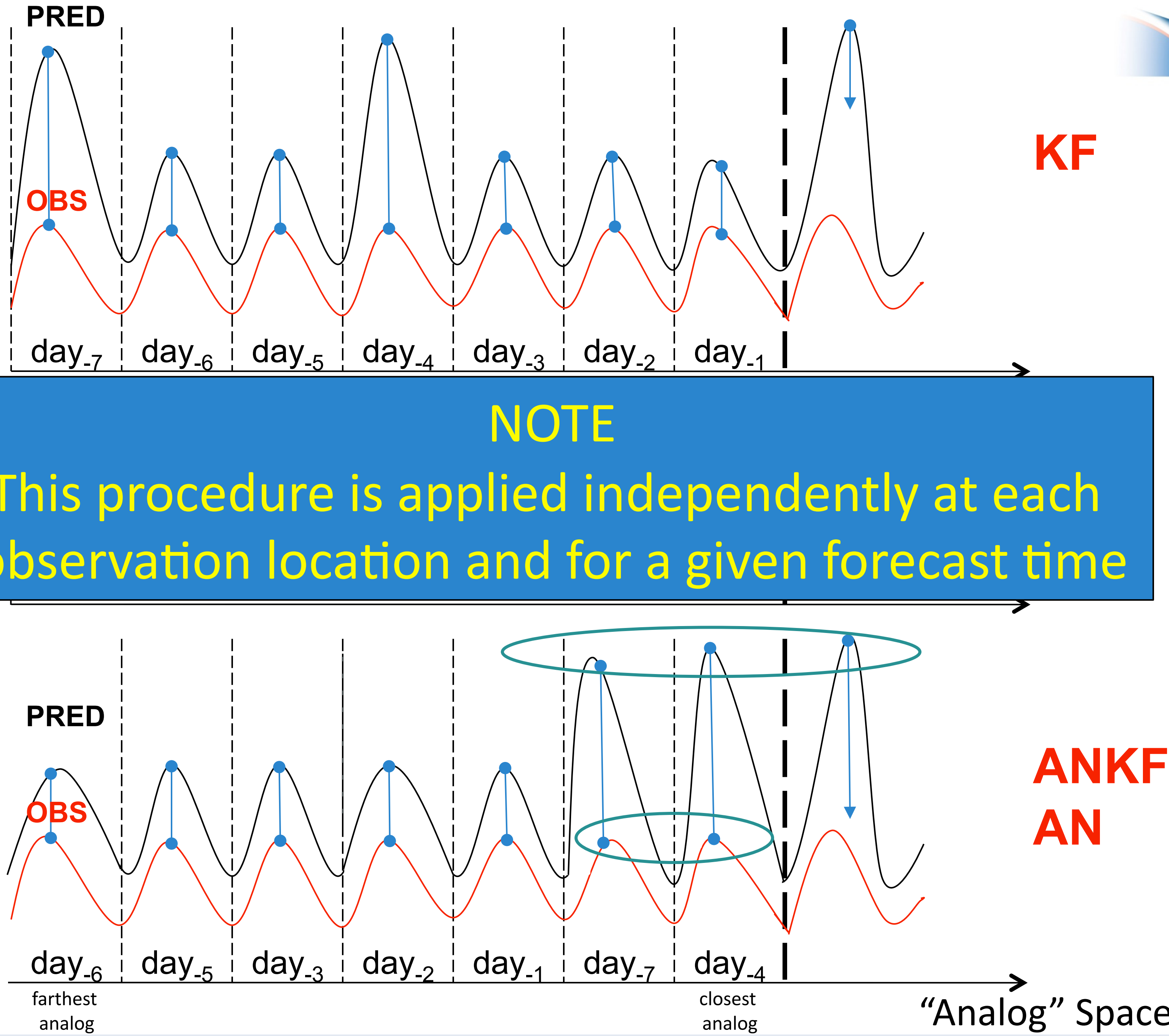
$\hat{P}_i = a + bO_i$, a and b least-squares regression coefficients of P_i and O_i

Kalman Filtering effects on probabilistic prediction reliability



Kalman Filtering effects on probabilistic prediction resolution





NOTE

This procedure is applied independently at each observation location and for a given forecast time

KF-wei

NCAR

How to find analogs? (1)

We can define a metric:

$$d_{t_i} = \left\| f_{t_n} - A_{t_i} \right\| = \sum_{var=1}^{N_{var}} w_{var} \frac{1}{\sigma_{f^{var}}} \sqrt{\sum_{k=-\tilde{t}}^{+\tilde{t}} (f_{t_n+k}^{var} - A_{t_i+k}^{var})^2}$$

Where,

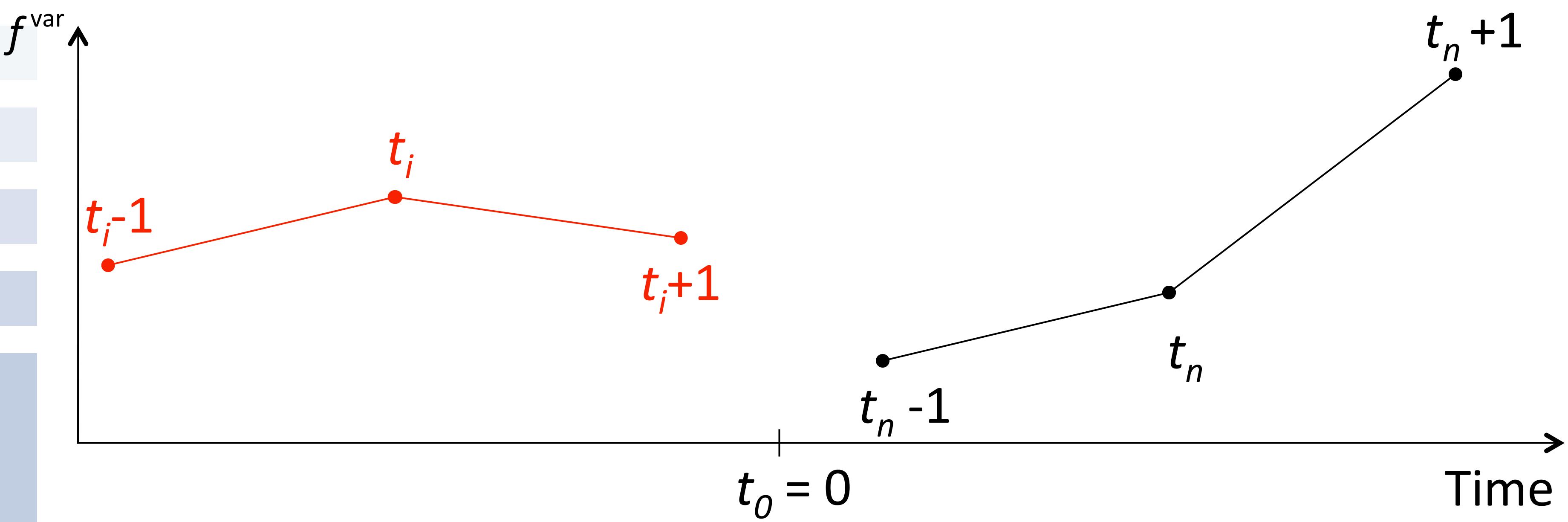
- N_{var} is the number of variable to compute the "distance" between f_{t_n} and A_{t_i}
- w_{var} is the weight given to each variable while computing the metric
- $\sigma_{f^{var}}$ is the standard deviation of the set $\{f_t^{var}\}$ with $0 \leq t \leq t_0$
- $\tilde{t}$ is the half-width of the time window over which differences are computed

How to find analogs? (2)

NCAR

We can define a metric:

$$d_{t_i} = \left\| f_{t_n} - A_{t_i} \right\| = \sum_{\mathrm{var}=1}^{N_{\mathrm{var}}} w_{\mathrm{var}} \frac{1}{\sigma_{f^{\mathrm{var}}}} \sqrt{\sum_{k=-\tilde{t}}^{+\tilde{t}} \left(f^{\mathrm{var}}_{t_n+k} - A^{\mathrm{var}}_{t_i+k} \right)^2}$$

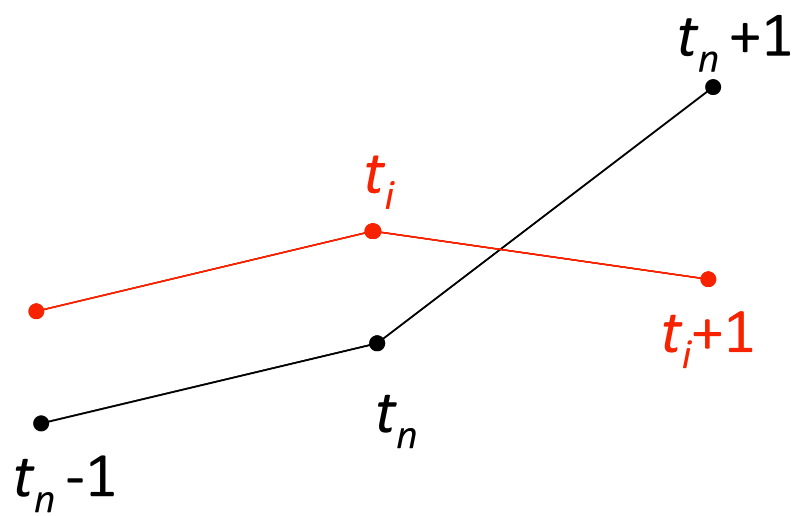


How to find analogs? (3)

NCAR

We can define a metric:

$$d_{t_i} = \left\| f_{t_n} - A_{t_i} \right\| = \sum_{var=1}^{N_{var}} w_{var} \frac{1}{\sigma_{f^{var}}} \sqrt{\sum_{k=-\tilde{t}}^{+\tilde{t}} \left(f_{t_n+k}^{var} - A_{t_i+k}^{var} \right)^2}$$

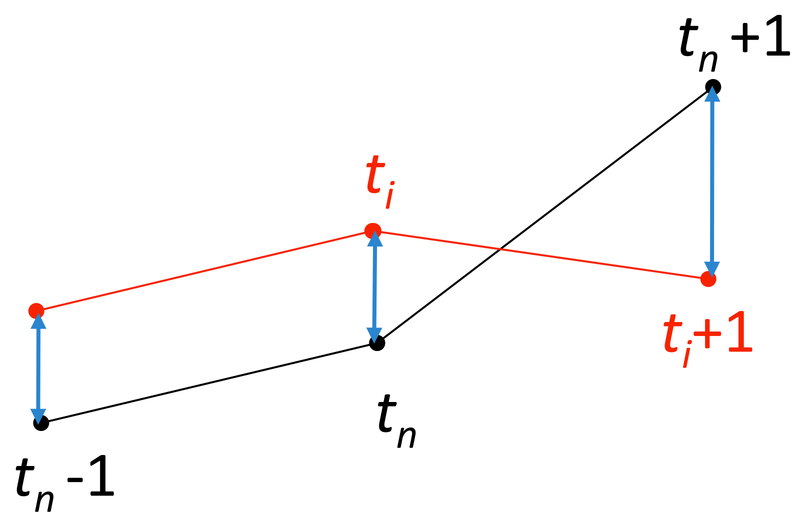


How to find analogs? (4)

NCAR

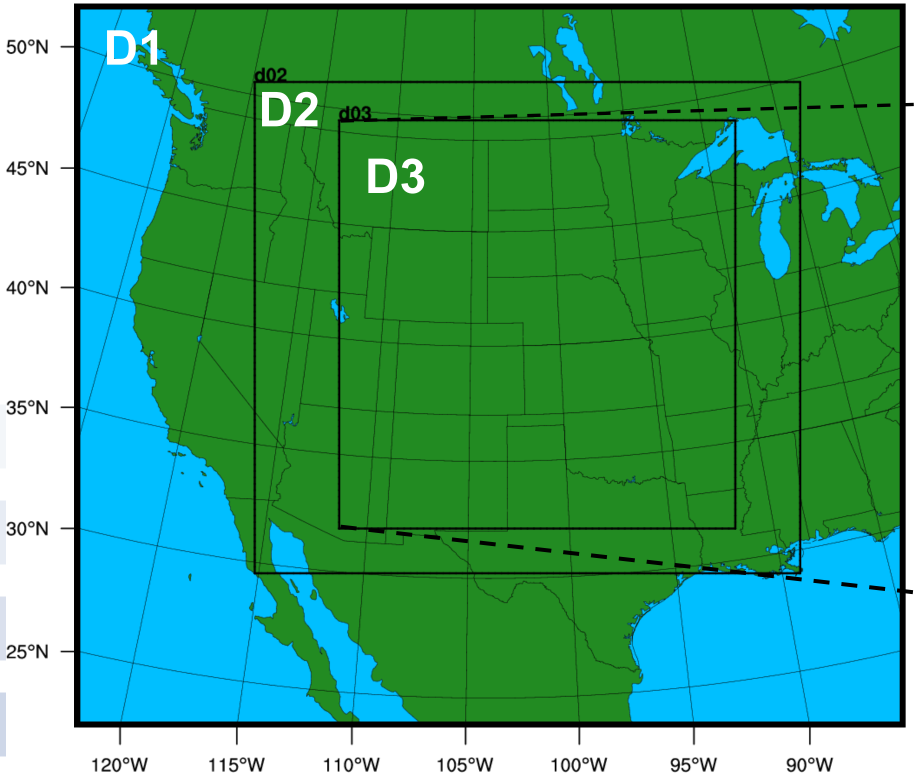
We can define a metric:

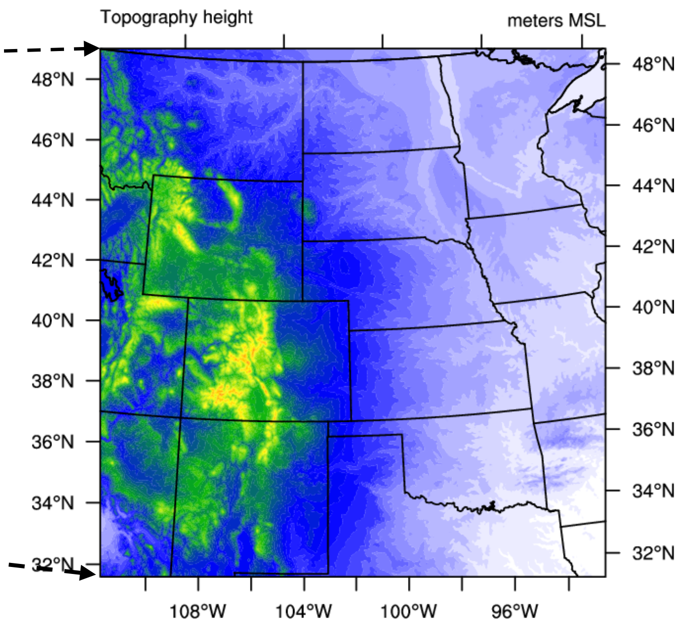
$$d_{t_i} = \left\| f_{t_n} - A_{t_i} \right\| = \sum_{\text{var}=1}^{N_{\text{var}}} w_{\text{var}} \frac{1}{\sigma_{f^{\text{var}}}} \sqrt{\sum_{k=-\tilde{t}}^{+\tilde{t}} \left(f_{t_n+k}^{\text{var}} - A_{t_i+k}^{\text{var}}\right)^2}$$



Modeling settings for wind predictions (wind energy project)

NCAR





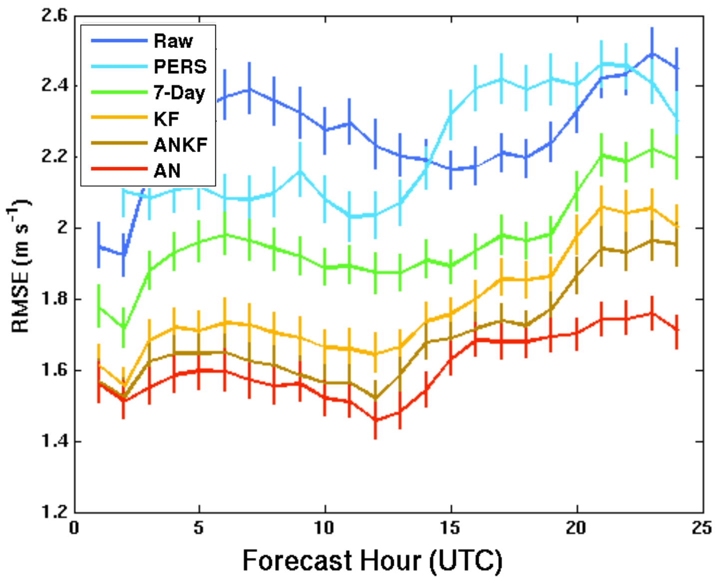
D1: 30 km, 128x114
D2: 10 km, 253x232
D3: 3.3 km, 541x571
NOTE: 37 vertical levels, with 12 levels in the lowest 1-km

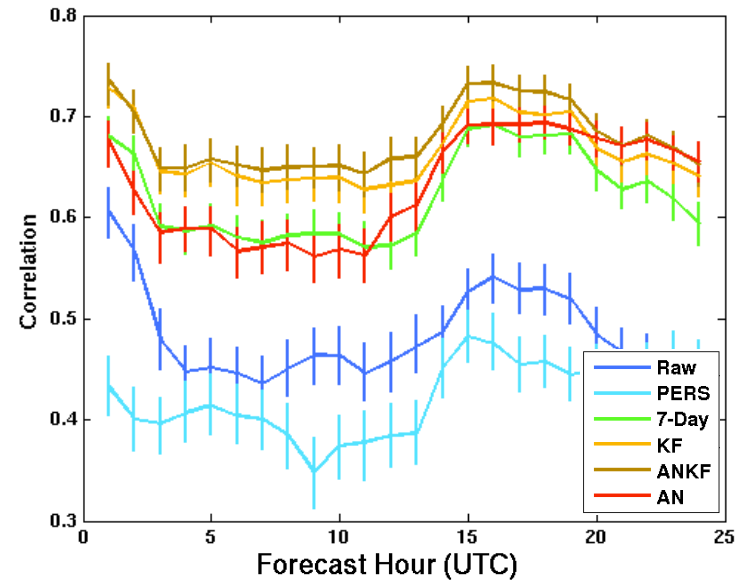
WRF Model physics:

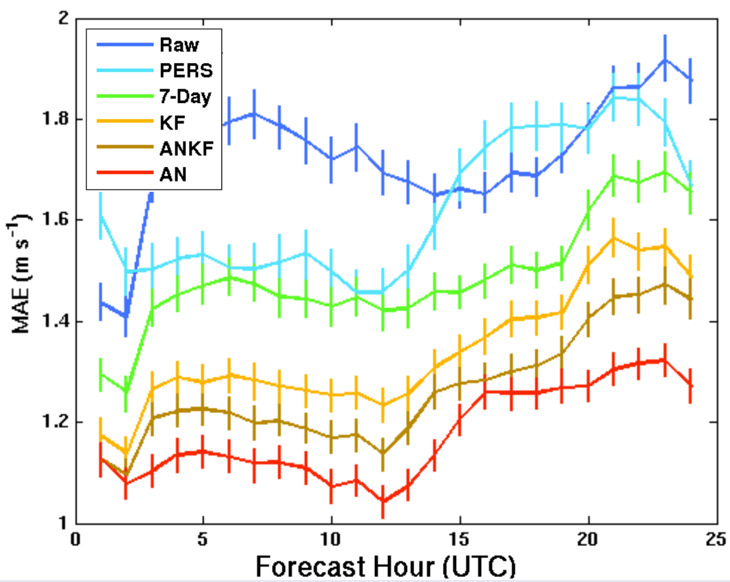
- Lin et al. microphysics
- YSU for PBL
- Monin-Oboukov for SL
- Kain-Fritsch CUP (Domain 1/2)
- Noah Land Surface Model

Wind speed: statistics as function of time

NCAR

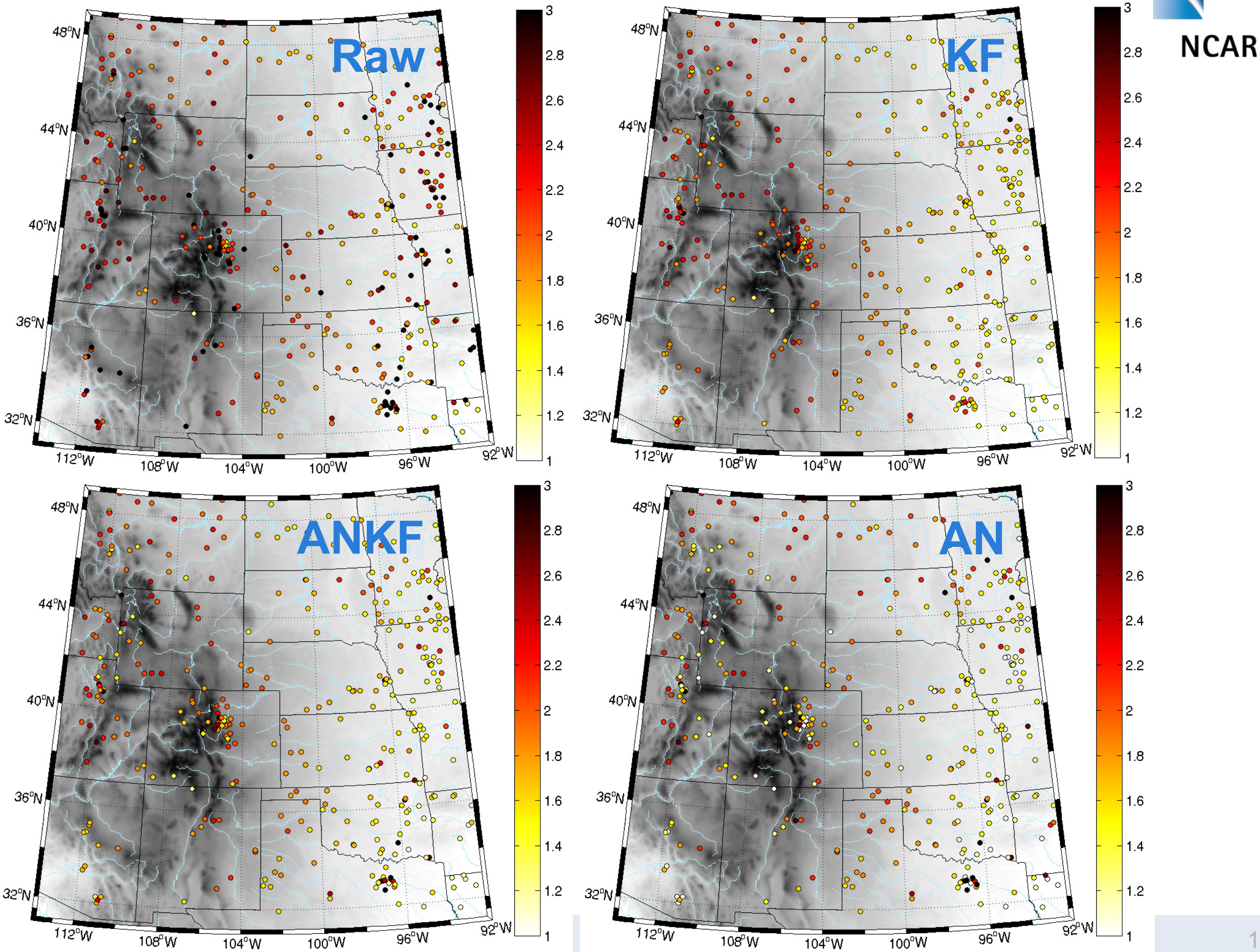






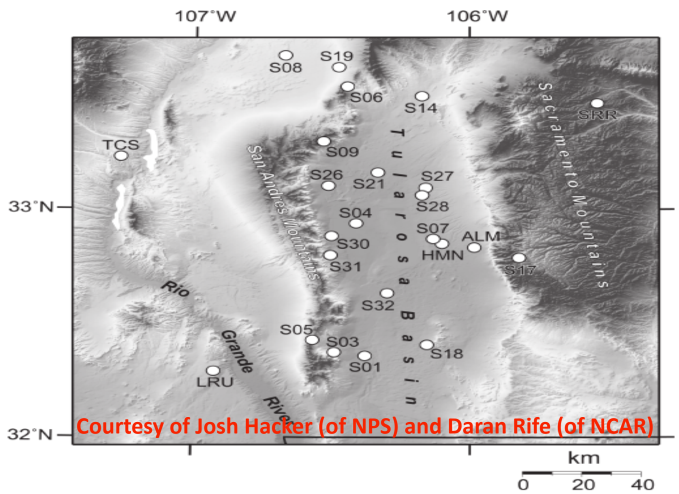
- 6 months period
- 500 surface stations
- $\tilde{t} = 1$
- $w_{\text{var}} = 1, \forall \text{var}$
- Variables: u, v, T, P, Q @ surface

Wind speed: RMSE (m/s) as function of space



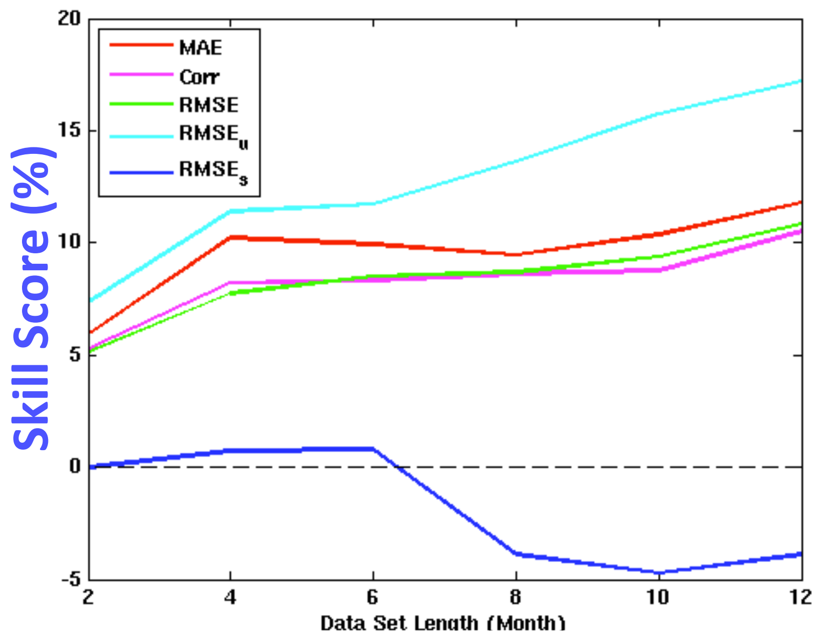
Sensitivity to data set length

NCAR

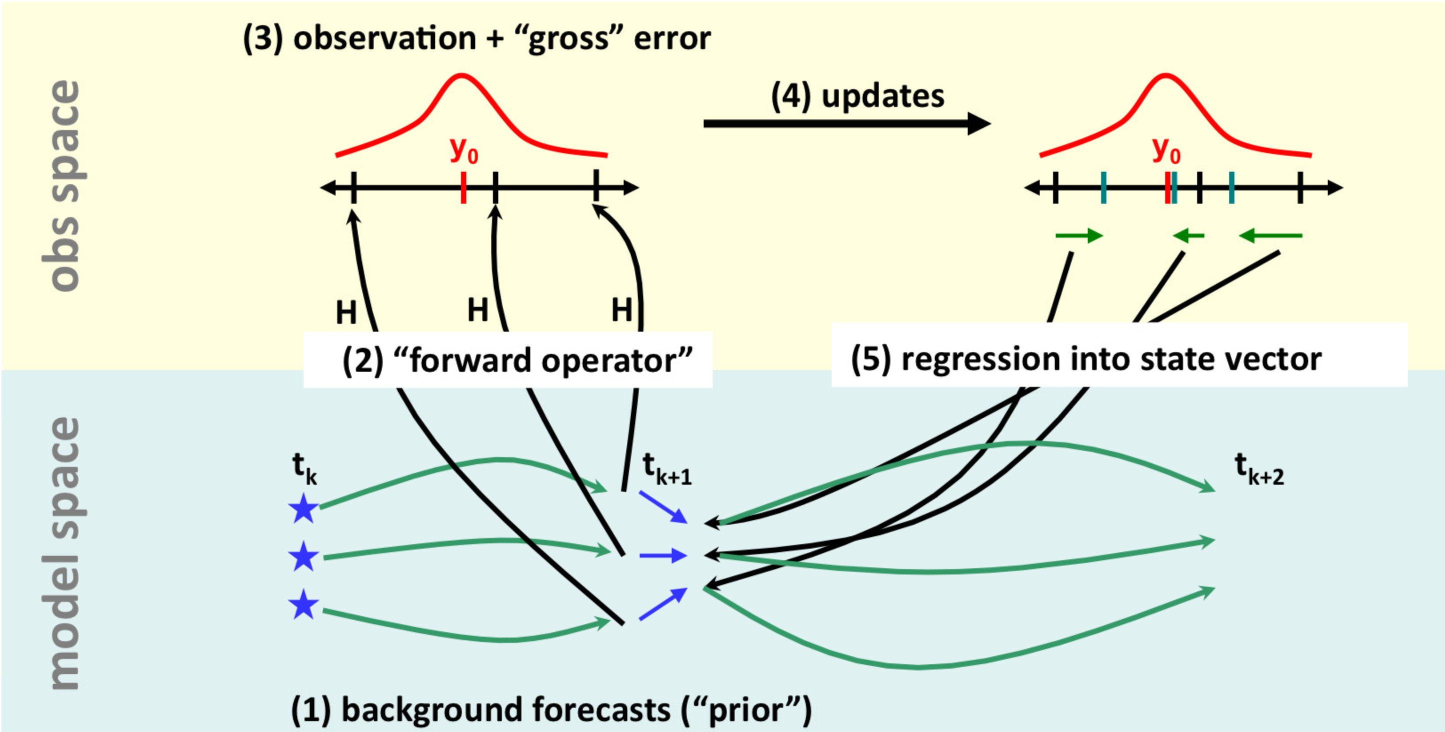


- NWP model: MM5
- 1 year period
- 22 surface stations in NM
- $\tilde{t} = 1$
- $w_{var} = 1, \forall \, var$
- Variables: u, v, T @ surface

KF vs ANKF



Ensemble filters main steps: Possible sources of error



Sources of error:

- Model error: (1)
- **Forward operator (H) errors (e.g., interpolation, flow-dependent behavior of representativeness errors, etc.): (2)**
- Instrumental errors, retrieval and transmission of obs: (3)
- Gaussian and other (depending on the method) assumptions: (4)
- **Sampling errors (analogs as a way to populate an ensemble): (4), (5)**
- **Errors from linear regression between obs and state variables increments: (5) (next slide)**

Source: Adapted from Anderson (Fig. 1 – Physica D, 2007)

Summary

NCAR

- Ensemble and Kalman filtering (KF) for air quality predictions
- A new method based on KF and an analog approach (ANKF, AN)
- Test KF, ANKF, and AN to correct 10m wind speed
  - ANKF and AN beats KF over a range of metrics
  - ANKF gain vs KF grows with length of data set
- The analog concept can be explored also within an ensemble data assimilation framework


GRAZIE!
(lucadm@ucar.edu)

KF in analog space (1)

NCAR

- f_n is a forecast at time t_n and at a given location, with $t_n > t_0$
- $d_i = \|f_n - A_i\|$ is a metric to measure the "distance" between f_n and A_i
- $\{A_i\}$ is a set of "analog" forecasts at a time t_i, with $t_i < t_0$
  - $\{A_i\}$ are ordered with respect to d_i : $d_{i-1} > d_i$, and $\{i, N \in \mathbb{N} : 2 \leq i \leq N = |\{A_i\}|\}$

We can now introduce the Kalman filter bias correction procedure as follows:

- The true unknown forecast bias at time t_n can me modeled by

$$x_{t_n|t_{n-1}} = x_{t_{n-1}|t_{n-2}} + \eta_{t_{n-1}}, \eta \sim N(0, \sigma^2_{\eta_{t_{n-1}}})$$

- And the actual forecast error can be expressed as

$$y_{t_i} = A_{t_i} - O_{t_i} = x_{t_i} + \varepsilon_{t_i} = x_{t_{i-1}} + \eta_{t_{i-1}} + \varepsilon_{t_i}, \varepsilon_{t_i} \sim N(0, \sigma^2_{\varepsilon_{t_i}})$$

KF in analog space (2)

The optimal recursive predictor of x_t can be written as

$$\hat{x}_{t_n|t_{n-1}} = \hat{x}_{t_{n-1}|t_{n-2}} + K_{t_{n-1}|t_{n-2}}(y_{t_{n-1}} - \hat{x}_{t_{n-1}|t_{n-2}})$$

Where *K*, the "Kalman gain" is

$$K_{t_n|t_{n-1}} = \frac{p_{t_{n-1}|t_{n-2}} + \sigma^2_{\eta_{t_{n-1}}}}{(p_{t_{n-1}|t_{n-2}} + \sigma^2_{\eta_{t_{n-1}}} + \sigma^2_{\varepsilon_{t_{n-1}}})}$$

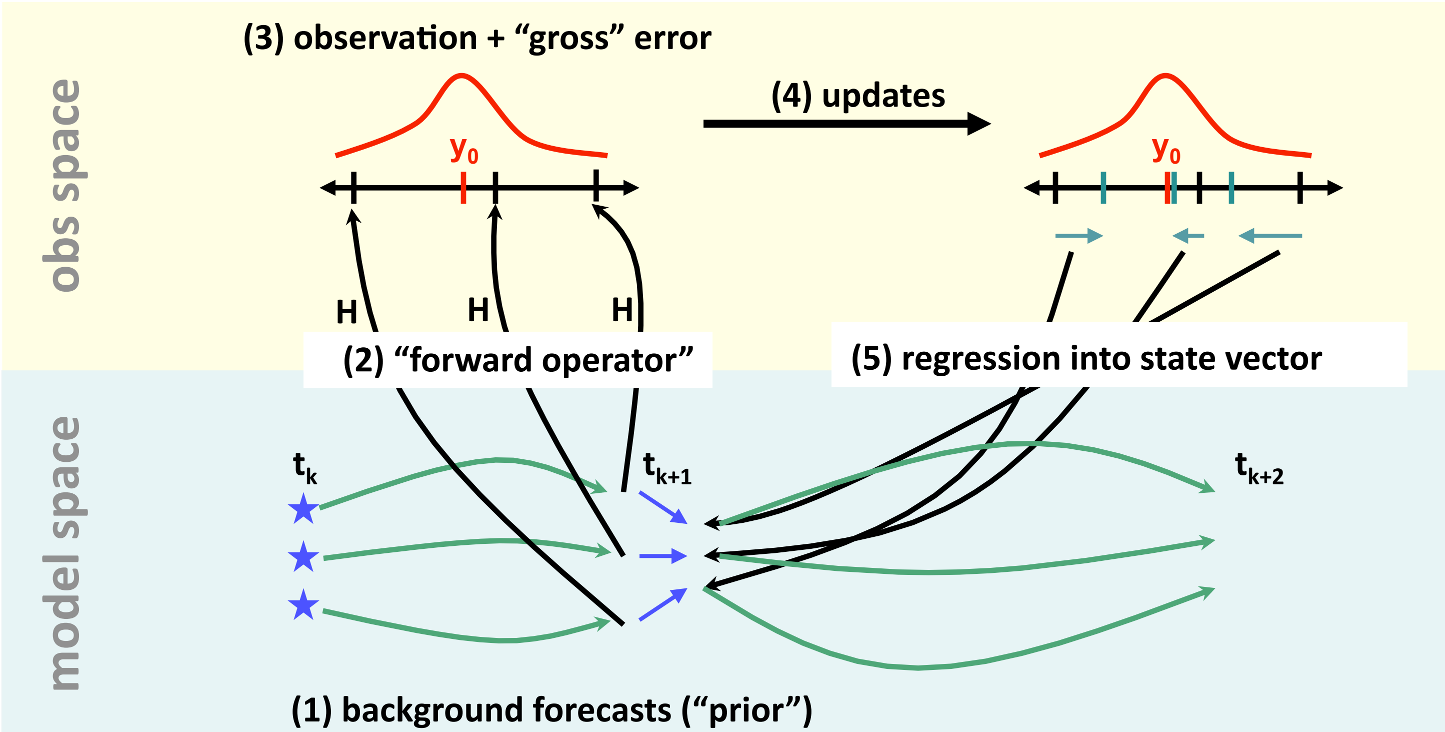
And *p*, the expected mean square error is

$$p_{t_n|t_{n-1}} = (p_{t_{n-1}|t_{n-2}} + \sigma^2_{\eta_{t_{n-1}}})(1 - K_{t_{n-1}|t_{n-2}})$$

NOTE: The system of equations is closed by:

- first running the filter for σ^2_{ε} (with $\sigma^2_{\sigma^2_{\varepsilon}}$ and $\sigma^2_{\sigma^2_{\eta}}$ constant)
- $r = \sigma^2_{\eta} / \sigma^2_{\varepsilon}$

Ensemble filters main steps: Possible sources of error

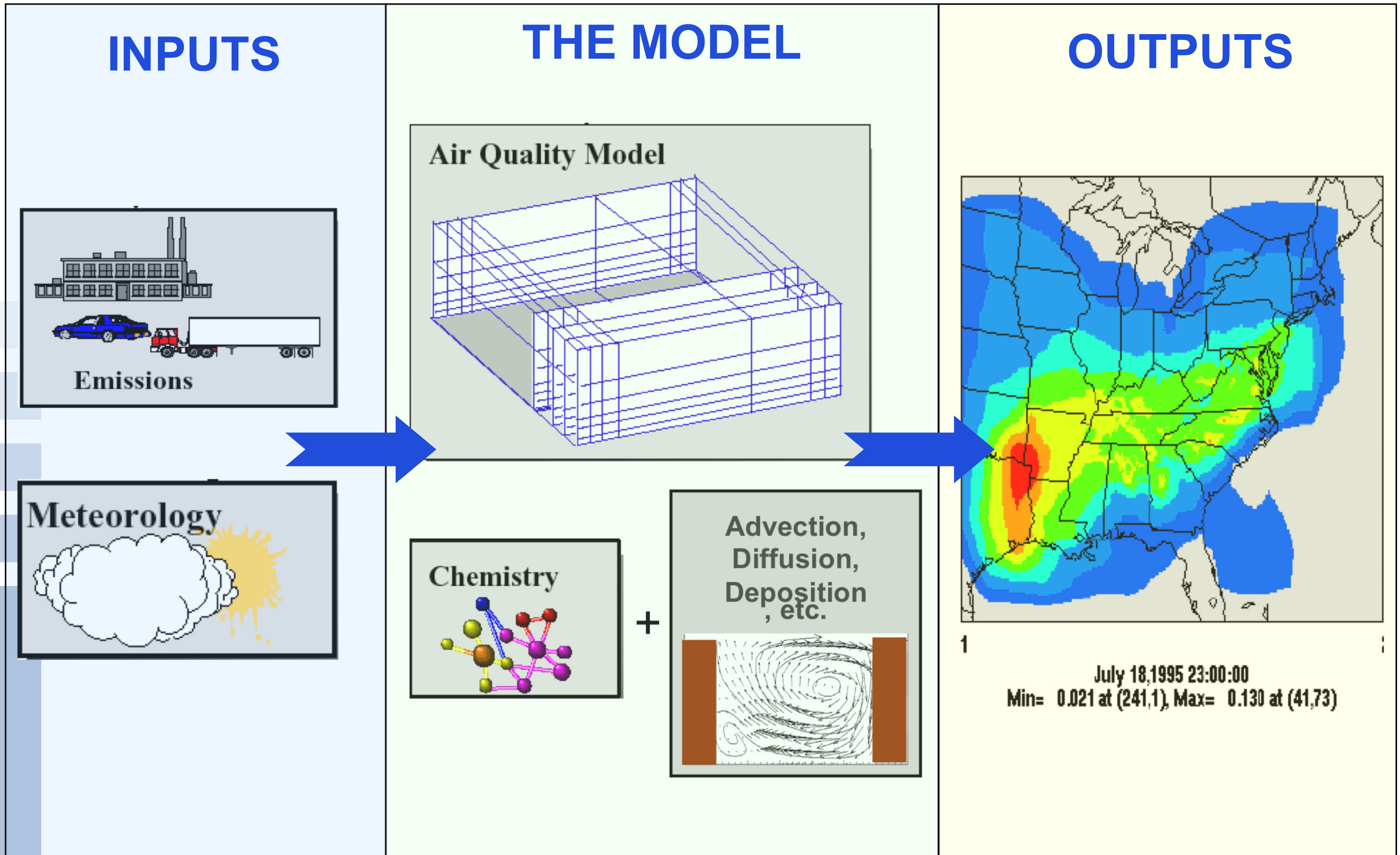


Sources of error:

- Model : (1)
- Forward operator (**H**): (2)
- Observation: (3)
- Gaussian and other assumptions: (4)
- Sampling: (4), (5)
- Linear regression: (5)

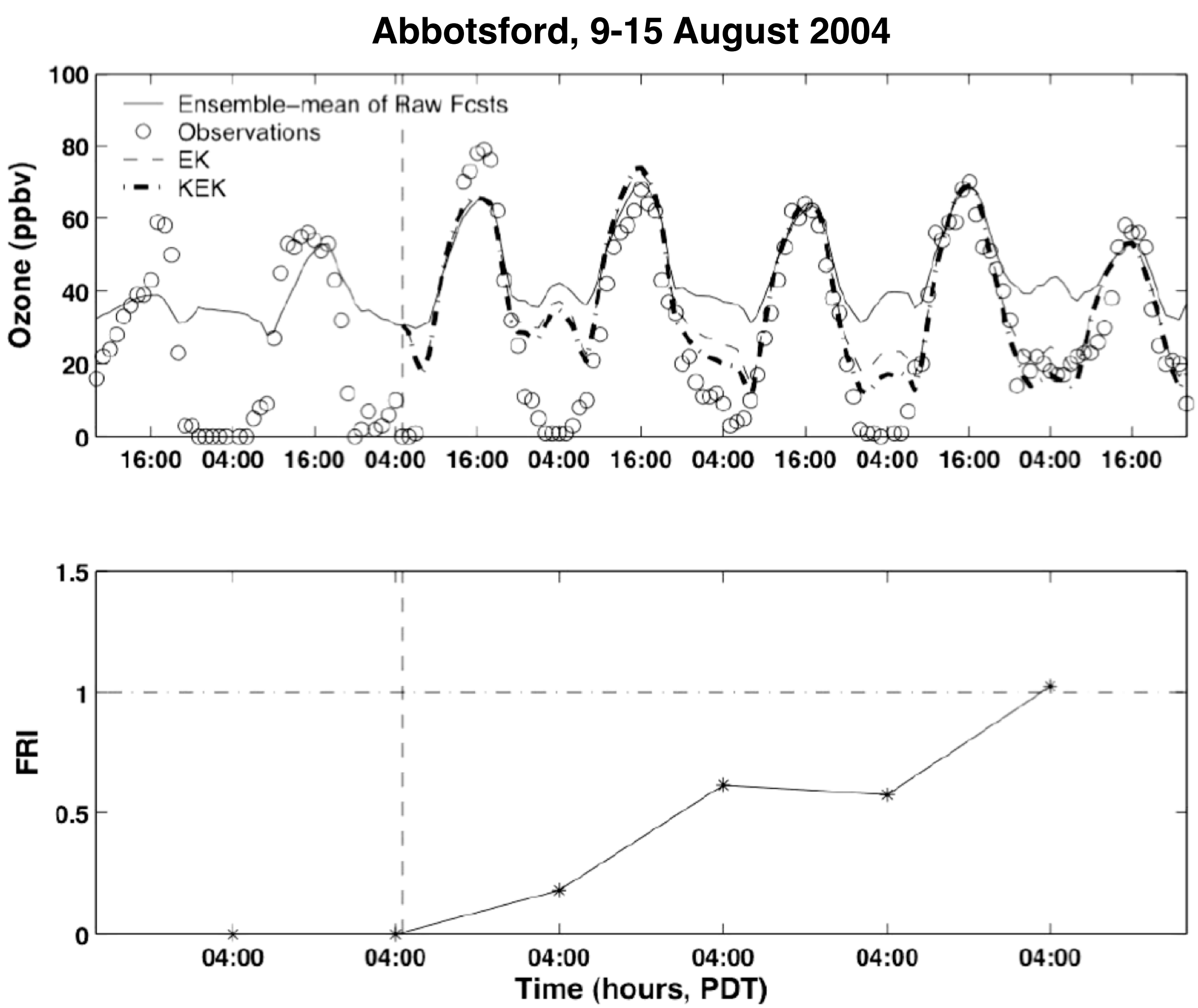
Source: Adapted from Anderson (Fig. 1 – Physica D, 2007)

Photochemical modeling



Kalman Filter predictor bias correction

NCAR

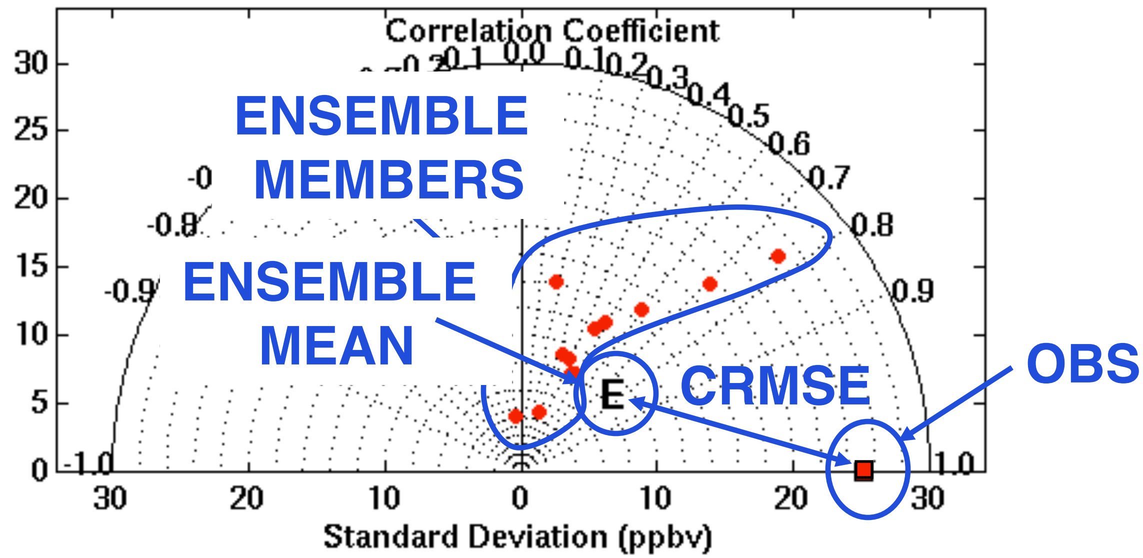


$$FRI = \frac{|RawFcsts - KEK|}{|RawFcsts - Obs|}$$

"Fractional Relative Improvement"

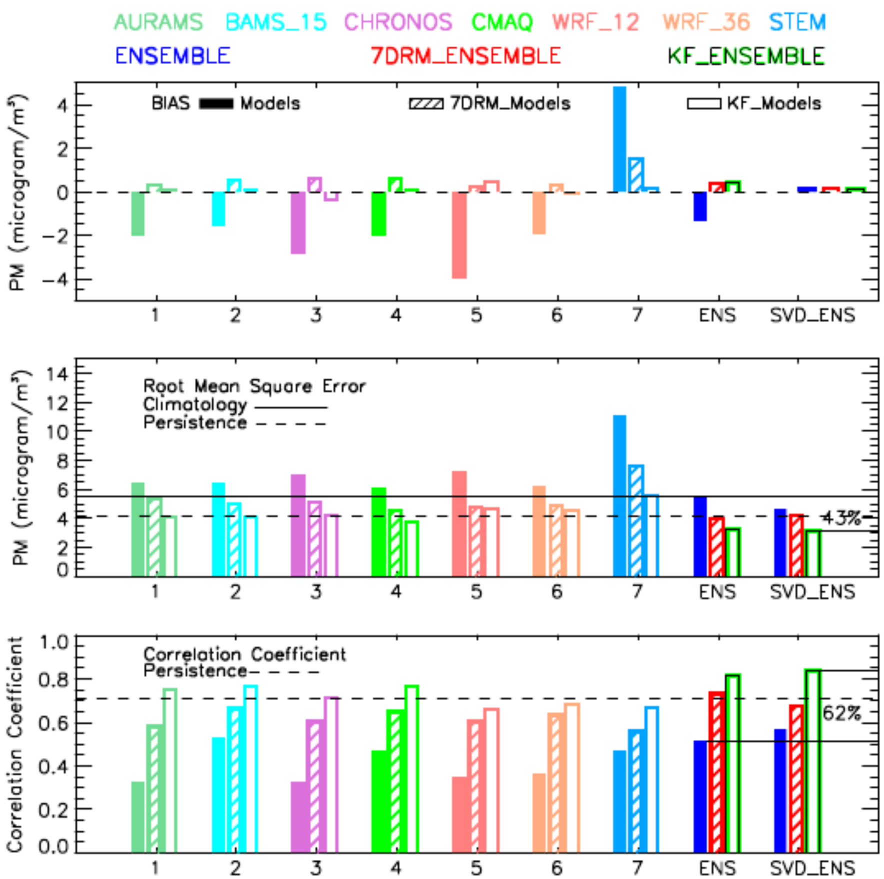
Delle Monache et al., Journal of Geophysical Research (2006b)

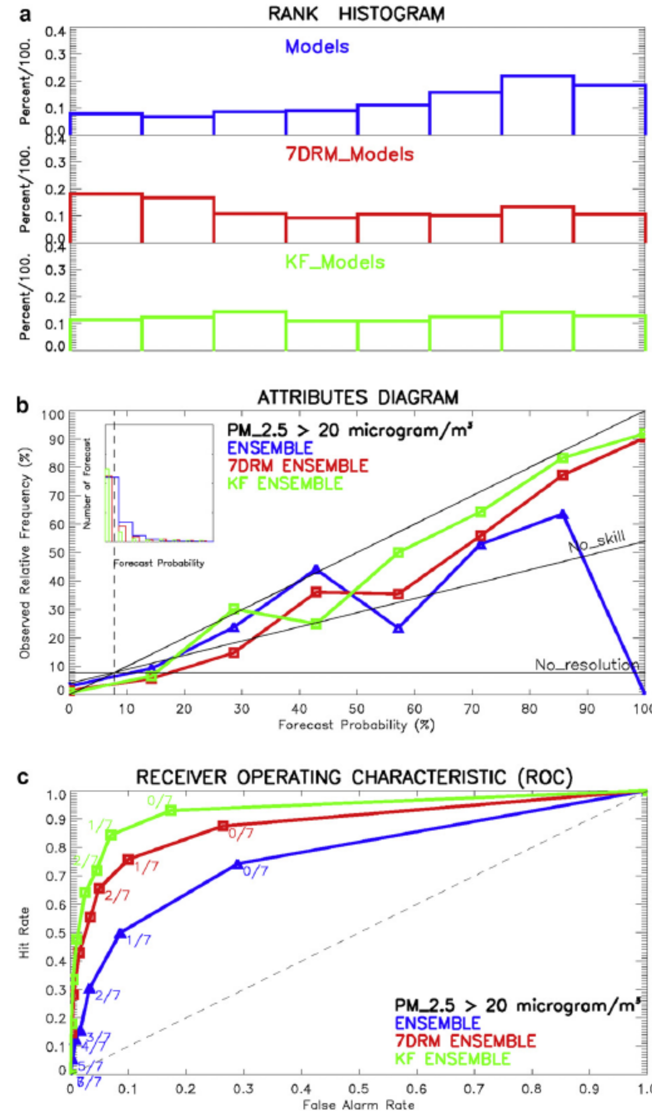
KF results: Chilliwack



A Kalman filter bias correction for deterministic and probabilistic $PM_{2.5}$ predictions

NCAR

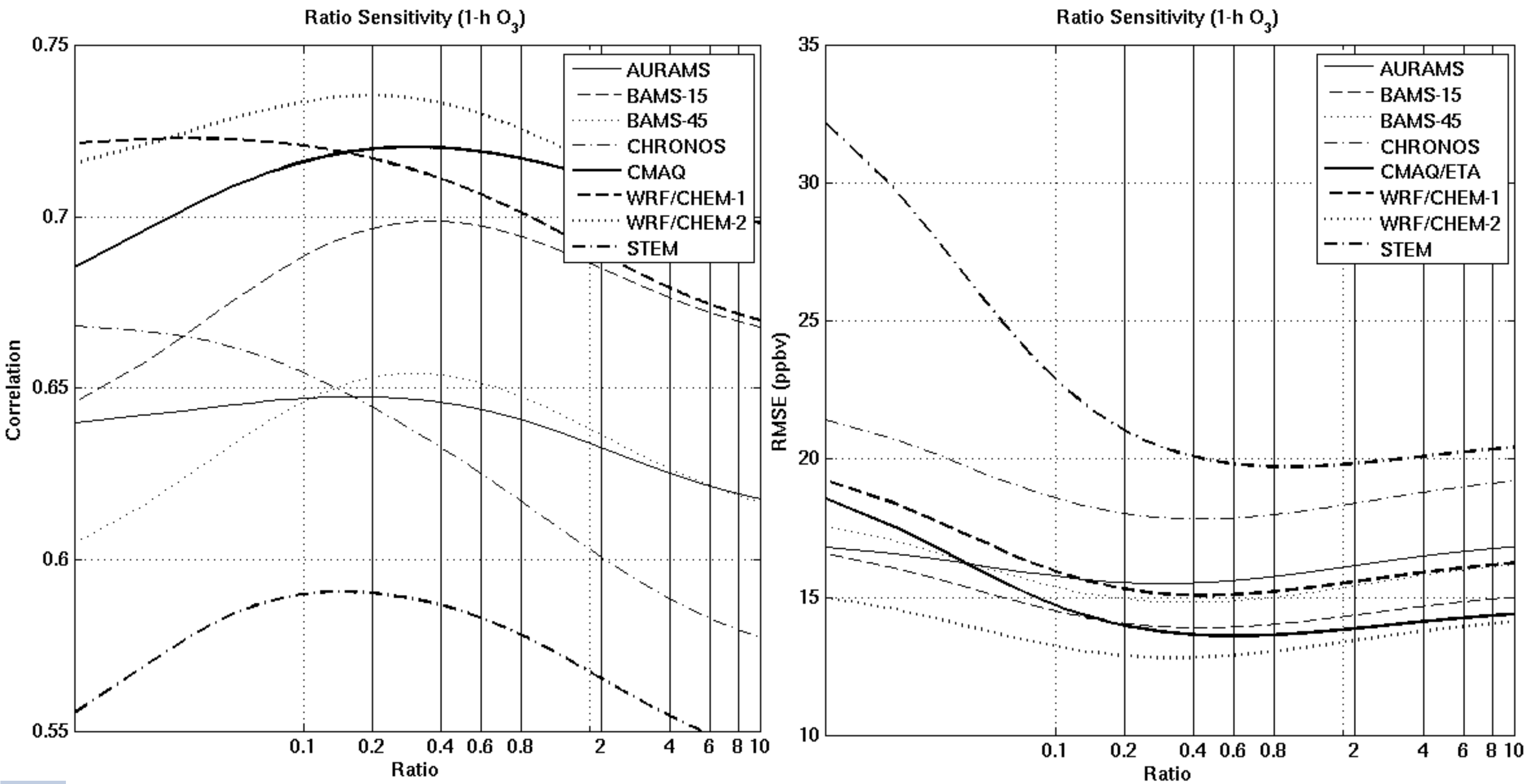




Source: Djalalova et al. (Atmospheric Environment, 2010)

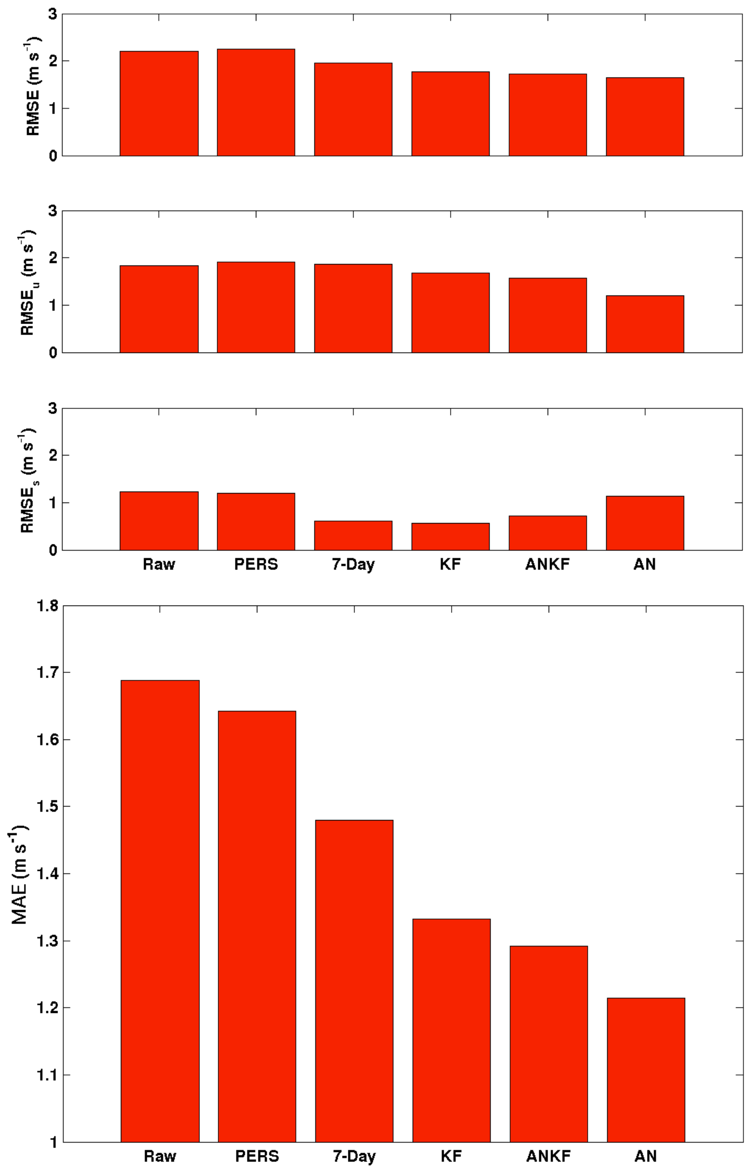
Error-ratio sensitivity tests

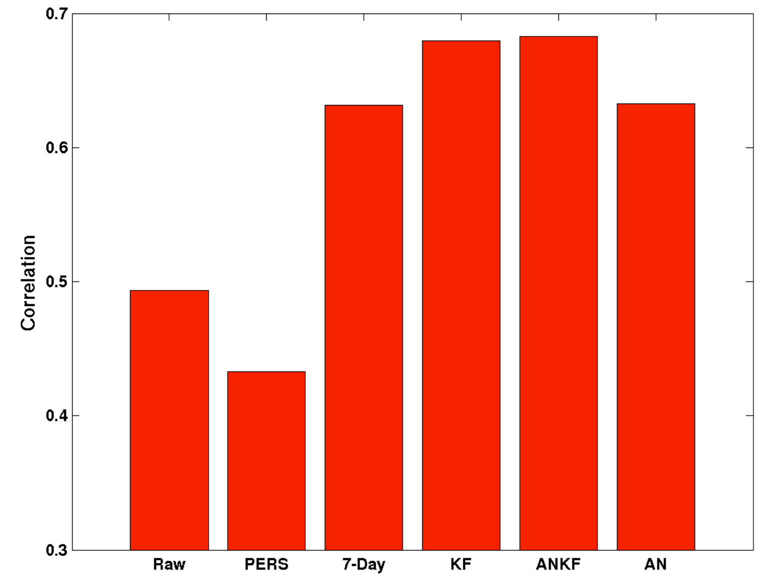
NCAR



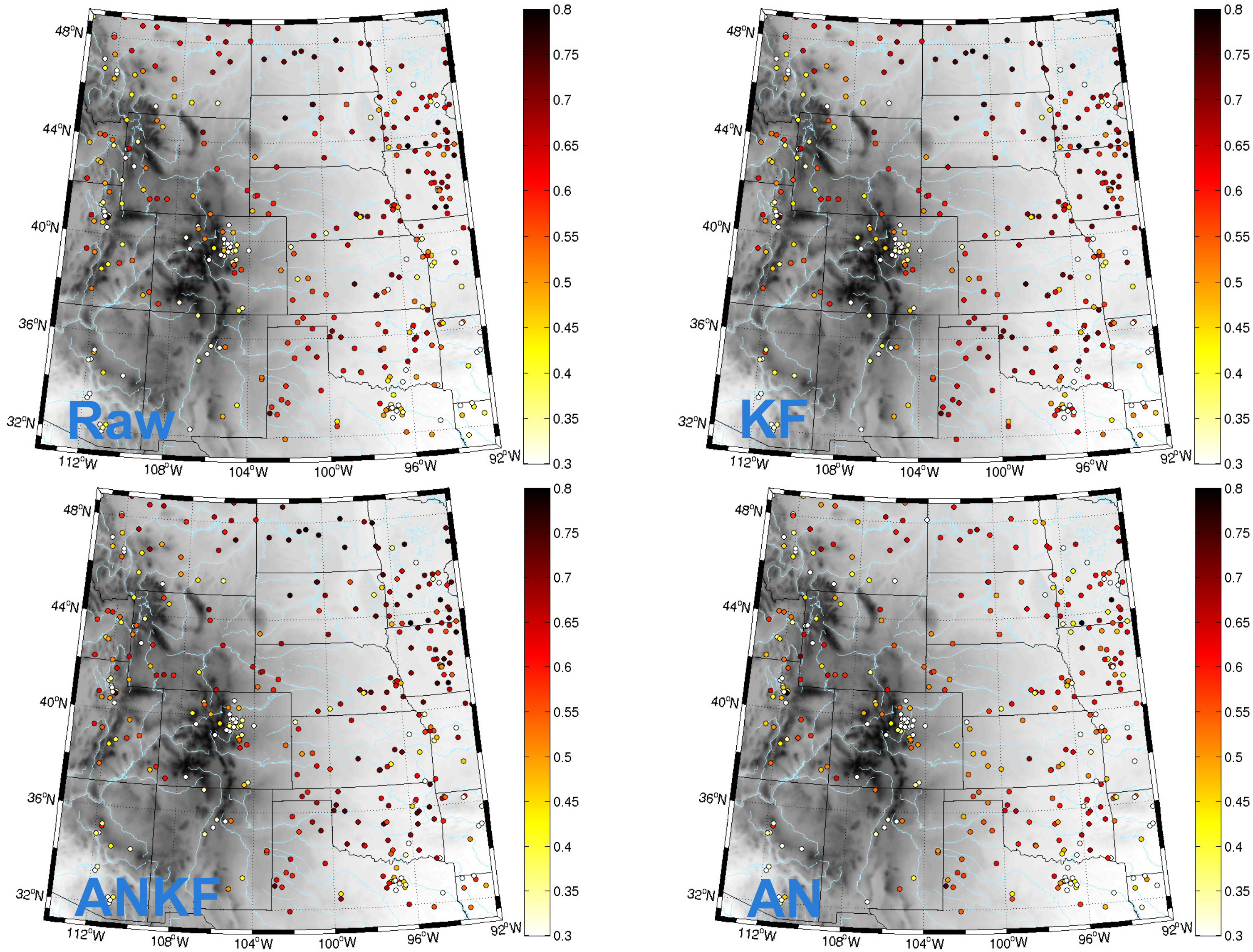
“Global” Statistics

NCAR



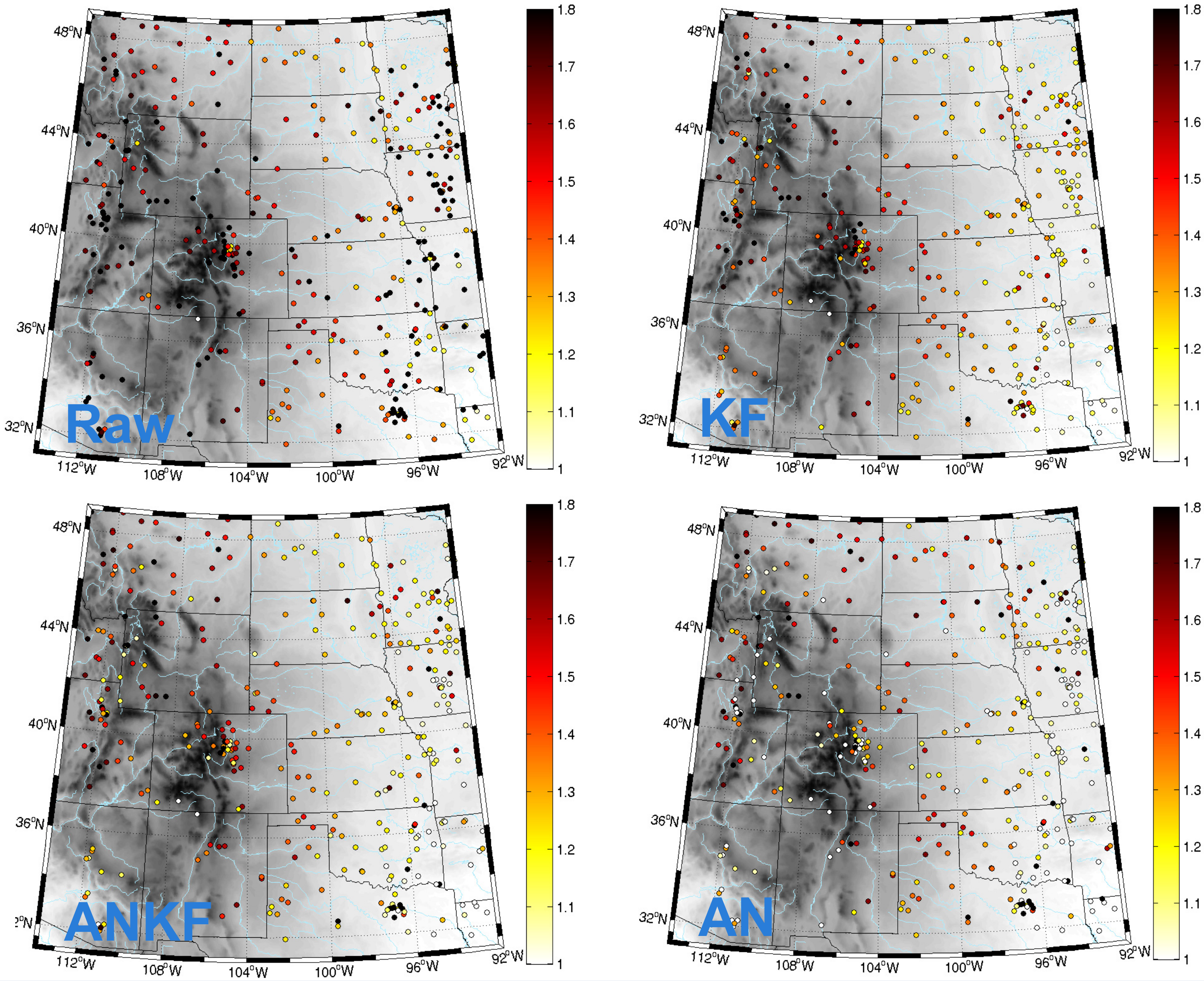


Statistics in space: Correlation

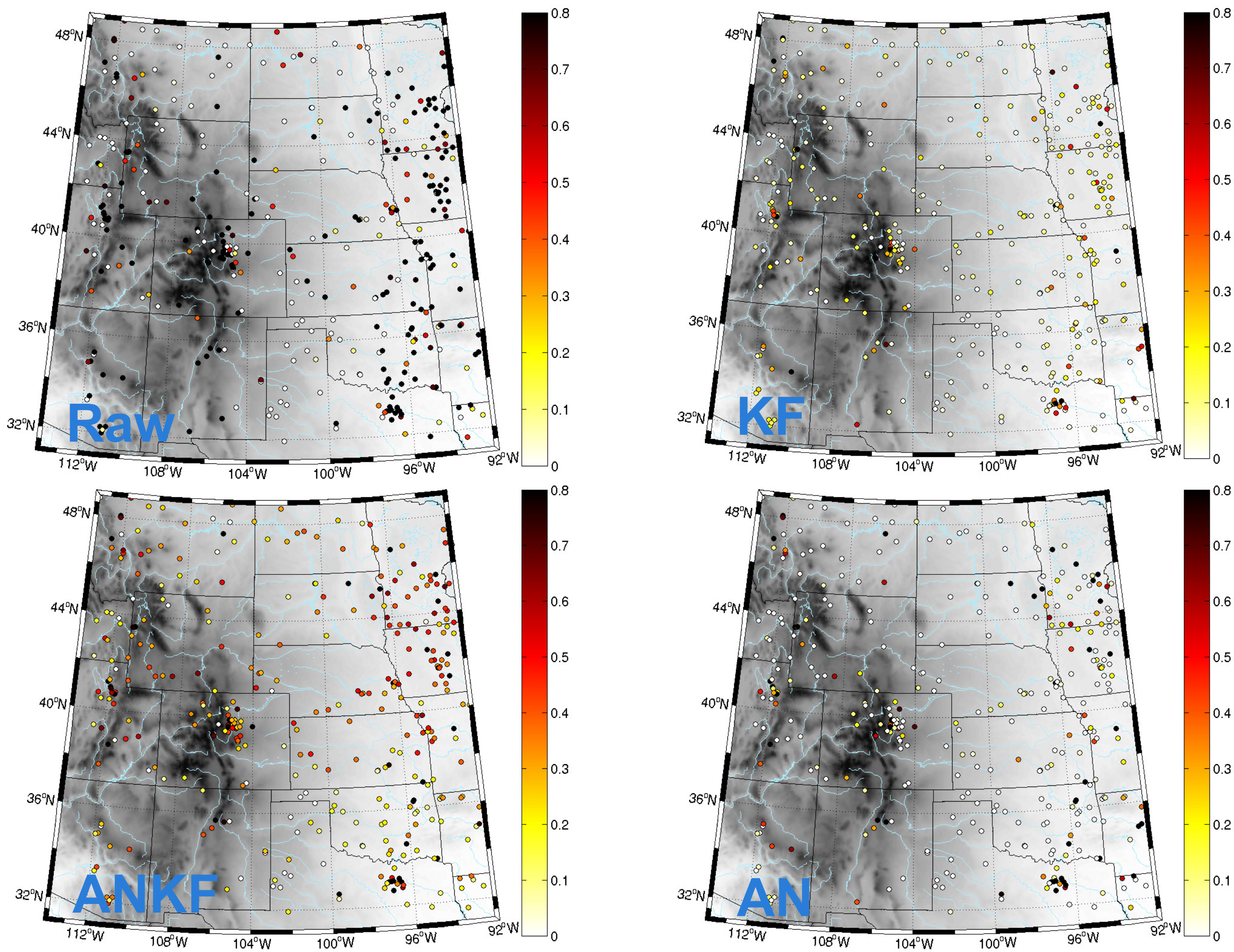


Statistics in space: MAE (m/s)

NCAR



Statistics in space: BIAS (m/s)



Wind speed: PDFs (observations vs. predictions)

NCAR