UNC INSTITUTE FOR THE ENVIRONMENT

CISES Center for Integrating Statistical and Environmental Science

EAKF-CMAQ:
Ensemble based data assimilation

Alexis Zubrow, Li Chen,

and V.R. Kotamarthi

This research is funded by
U.S. EPA-Science To Achieve Results (STAR) Program
Grant # R-82940201-0

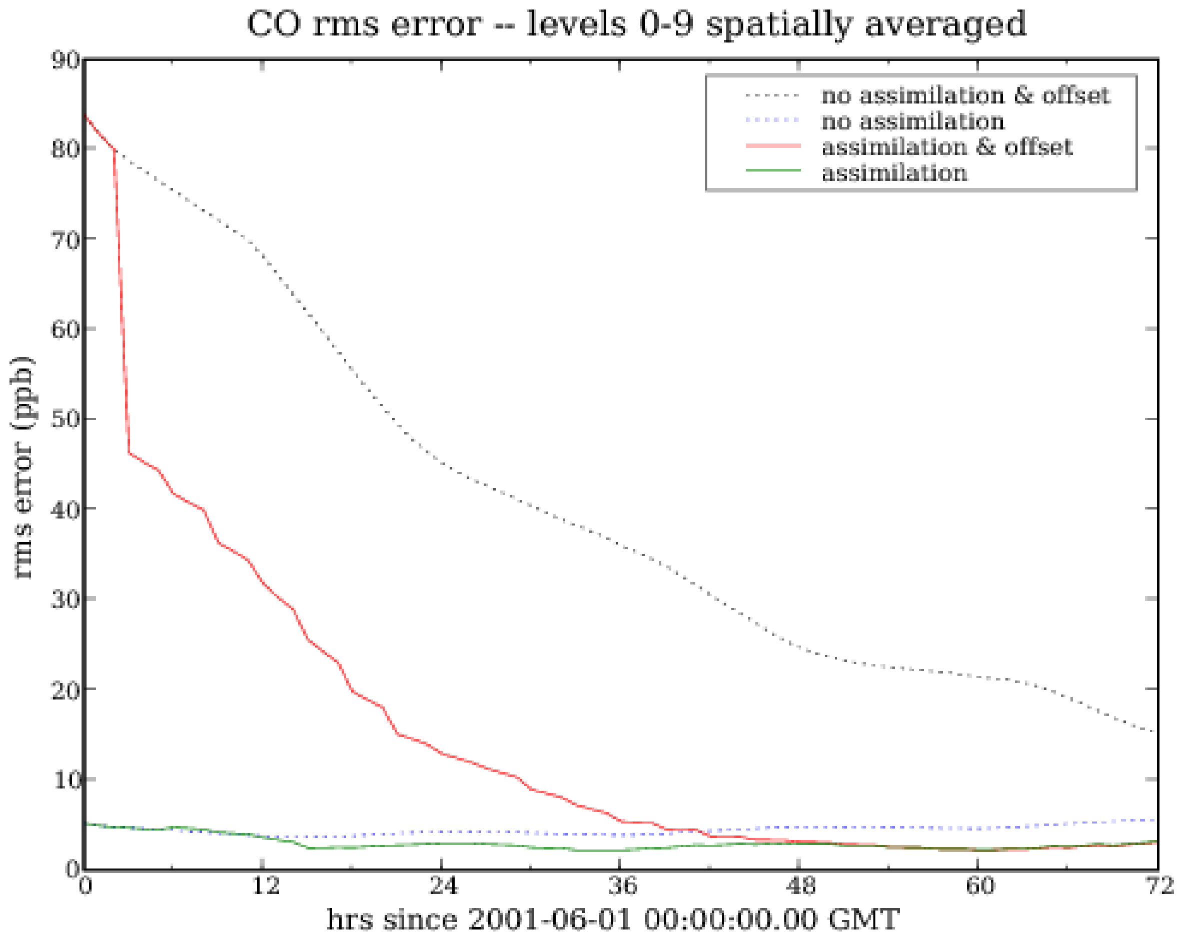
Although the research described herein has been funded wholly or in part by the United States Environmental Protection Agency through STAR Cooperative Agreement #R-82940201-0 to the University of Chicago, it has not been subjected to the Agency's required peer and policy review and therefore does not necessarily reflect the views of the Agency, and no official endorsement should be inferred.

Motivation

- Expanding Obs networks
  - surface monitors
  - vertical profiles
  - satellite data
- Forecasting
- Uncertainty

Motivation

offset = 2*IC



Kalman Filter

$$K = P^f H^T (H P^f H^T + R)^{-1}$$

$$x^a = x^f + K(y^o - H x^f)$$

$$P^a = (I - KH) P^f$$

$$x^f_{t+1} = M x^a_t$$

$$P^f_{t+1} = M P^a_t M^T + Q$$

x^f = state forecast
x^a = state analysis
M = model
P^f = covariance forecast
P^a = covariance analysis
H = obsveration operator
R = covariance observation error
Q = covariance model error

Kalman Filter

$$K = P^f H^T (H P^f H^T + R)^{-1}$$

$$x^a = x^f + K(y^o - H x^f)$$

new info from obs

$$P^a = (I - K H) P^f$$

$$x^f_{t+1} = M x^a_t$$

$$P^f_{t+1} = M P^a_t M^T + Q$$

x^f = state forecast
x^a = state analysis
M = model
P^f = covariance forecast
P^a = covariance analysis
H = obsveration operator
R = covariance observation error
Q = covariance model error

Kalman Filter

$$K = P^f H^T (H P^f H^T + R)^{-1}$$

$$x^a = x^f + K(y^o - H x^f)$$

$$P^a = (I - K H) P^f$$

new info
from obs

$$x^f_{t+1} = M x^a_t$$

$$P^f_{t+1} = M P^a_t M^T + Q$$

model integration

x^f = state forecast
x^a = state analysis
M = model
P^f = covariance forecast
P^a = covariance analysis
H = obsveration operator
R = covariance observation error
Q = covariance model error

KF Key Points

Advantages:

- integration separated from filtering
- model state and errors evolve over time
- K – weights new info with uncertainty

KF Key Points

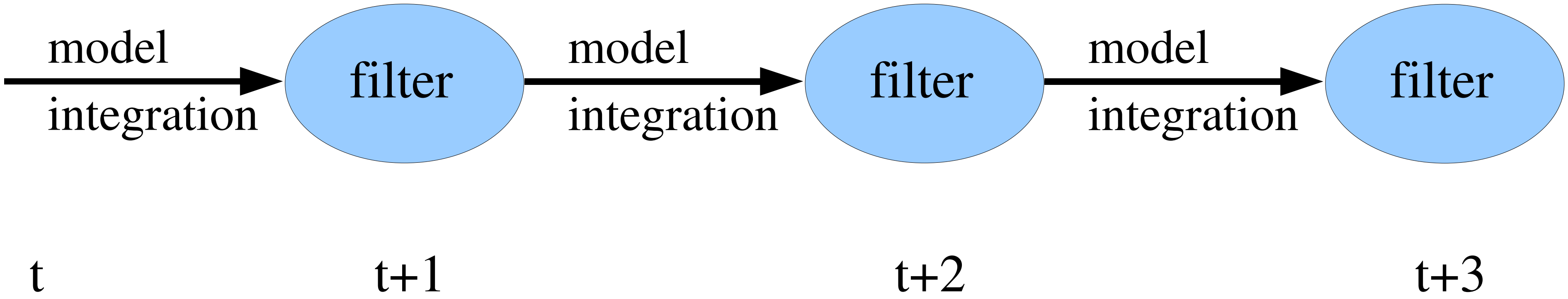
Advantages:

- integration separated from filtering
- model state and errors evolve over time
- K – weights new info with uncertainty

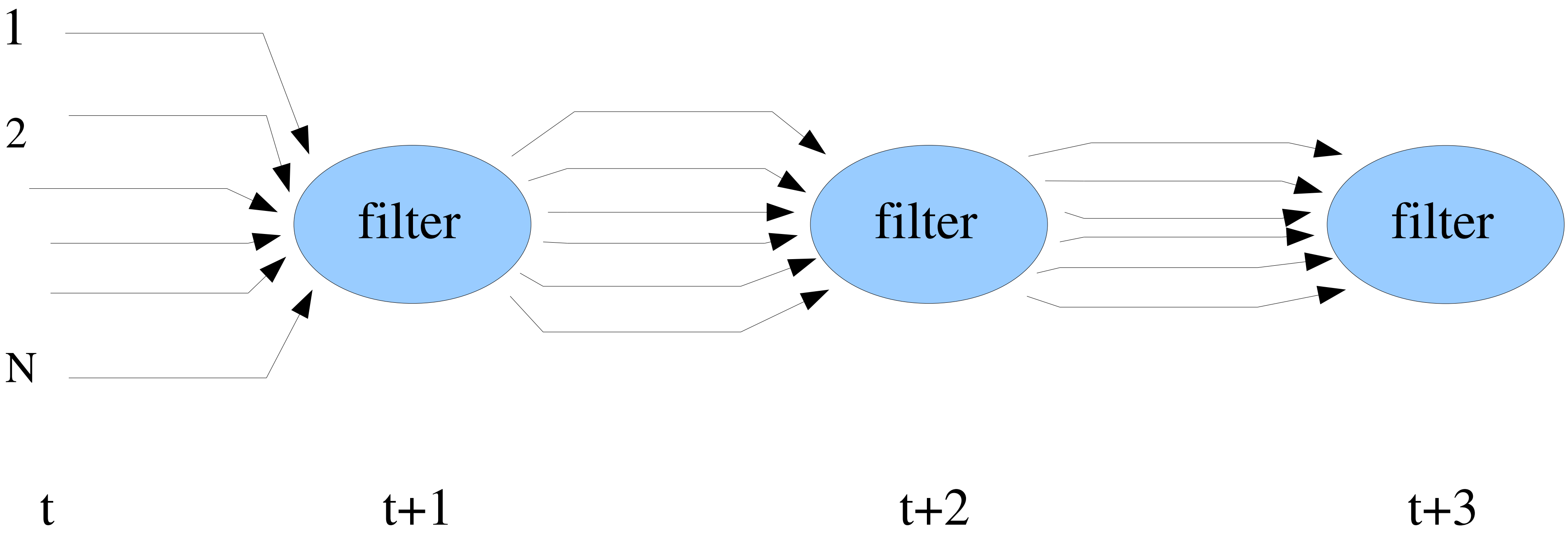
Disadvantages:

- model and observation operator are linear
- size of P (covariance matrices) leads to computational burden
- P at IC

Kalman Filter



Ensemble Filters



Ensemble Kalman Filter

$$x_i{}^f(t+1) = \mathcal{M}(x_i^a(t)), i = 1, \ldots, N$$

$$x_i^a = x_i^f + K(y_i{}^o - \mathcal{H}x_i{}^f), i = 1, \ldots, N$$

$$P^f = \frac{1}{N-1}\sum (x_i^f - \overline{x^f})(x_i^f - \overline{x^f})^T$$

$$P^f\mathcal{H}^T = \frac{1}{N-1}\sum (x_i^f - \overline{x^f})(\mathcal{H}x_i^f - \overline{\mathcal{H}x^f})^T$$

$$\mathcal{H}P^f\mathcal{H}^T = \frac{1}{N-1}\sum (\mathcal{H}x_i^f - \overline{\mathcal{H}x^f})^T(\mathcal{H}x_i^f - \overline{\mathcal{H}x^f})^T$$

$$K = P^f\mathcal{H}(\mathcal{H}P^f\mathcal{H}^T + R)^{-1}$$

EnKF Key Points

<u>KF</u>

- linear M,H
- large P
- P at IC

EnKF Key Points

<u>KF</u>

- linear M,H
- large P
- P at IC

<u>EnKF</u>

- nonlinear M,H
- smaller P
- ensemble approximate P at IC
- perturbed obs

Ensemble Adjusted Kalman Filter

$$y = H(x) + \epsilon$$

$$\Delta y_i = {y_i}^a - \overline{y^a} = \alpha({y_i}^f - \overline{y^f})$$

$$\alpha = \sqrt{r(r+\sigma^f)^{-1}}$$

$\sigma^f = forecast\ \ variance$

$r = observation\ \ variance$

$y_i^f = forecast\ \ i^{th}\ \ ensemble\ \ member\ \ obs\ \ space$

$y_i^a = analysis\ \ i^{th}\ \ ensemble\ \ member\ \ obs\ \ space$

EAKF Key Points

KF

- linear M,H
- large P
- P at IC

EAKF Key Points

KF

- linear M,H
- large P
- P at IC

EAKF

- nonlinear M,H
- scalar r,σ
- ensemble approximate P at IC
- convert to obs space

Comparison of Filters

EnKF

- stochastic
- perturbed obs
- under predicts ensemble spread (N~20-40)
- captures non-Gaussian errors (N~80+)

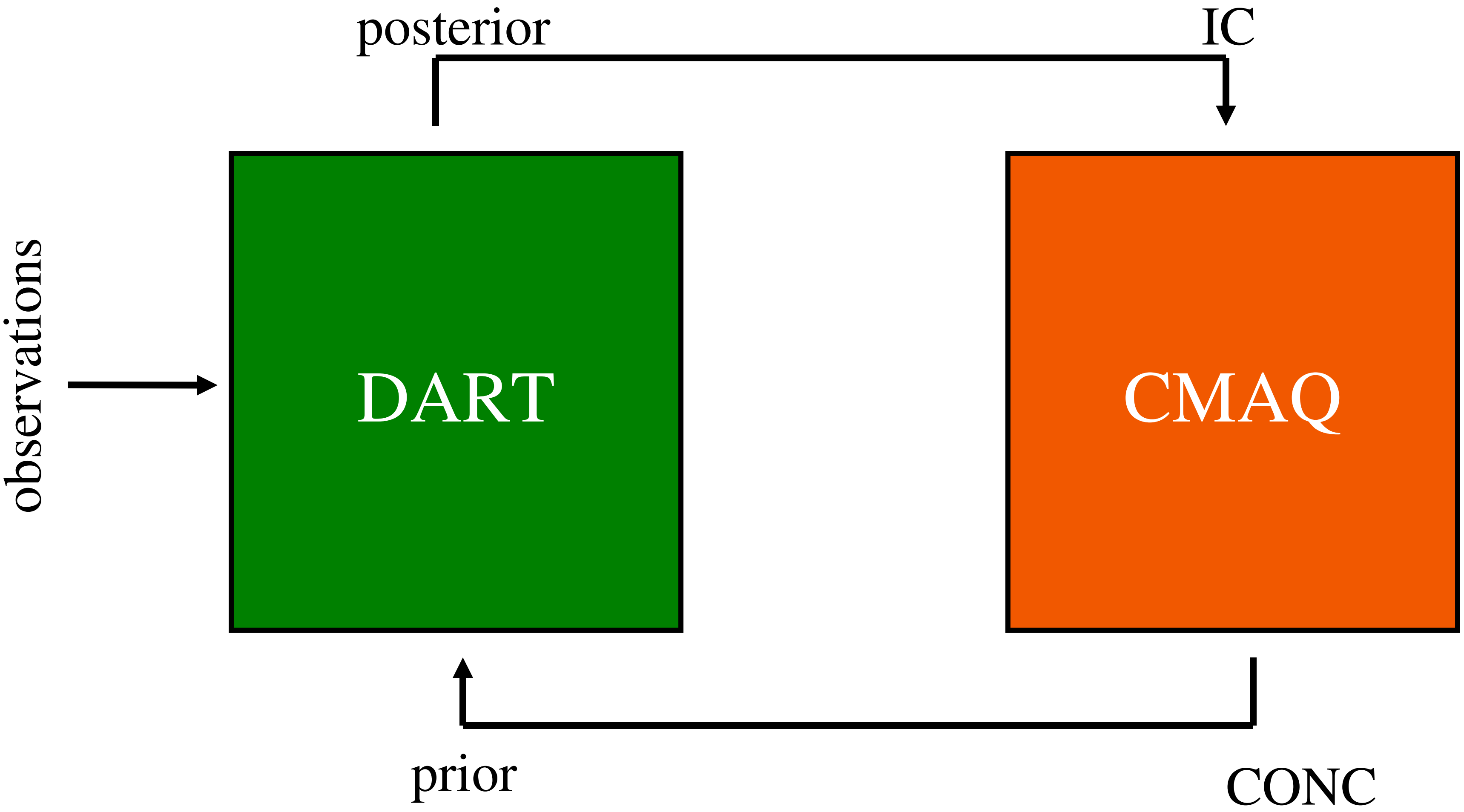
EAKF

- deterministic
- unperturbed obs
- preserves higher order moments
- scalar in calculations

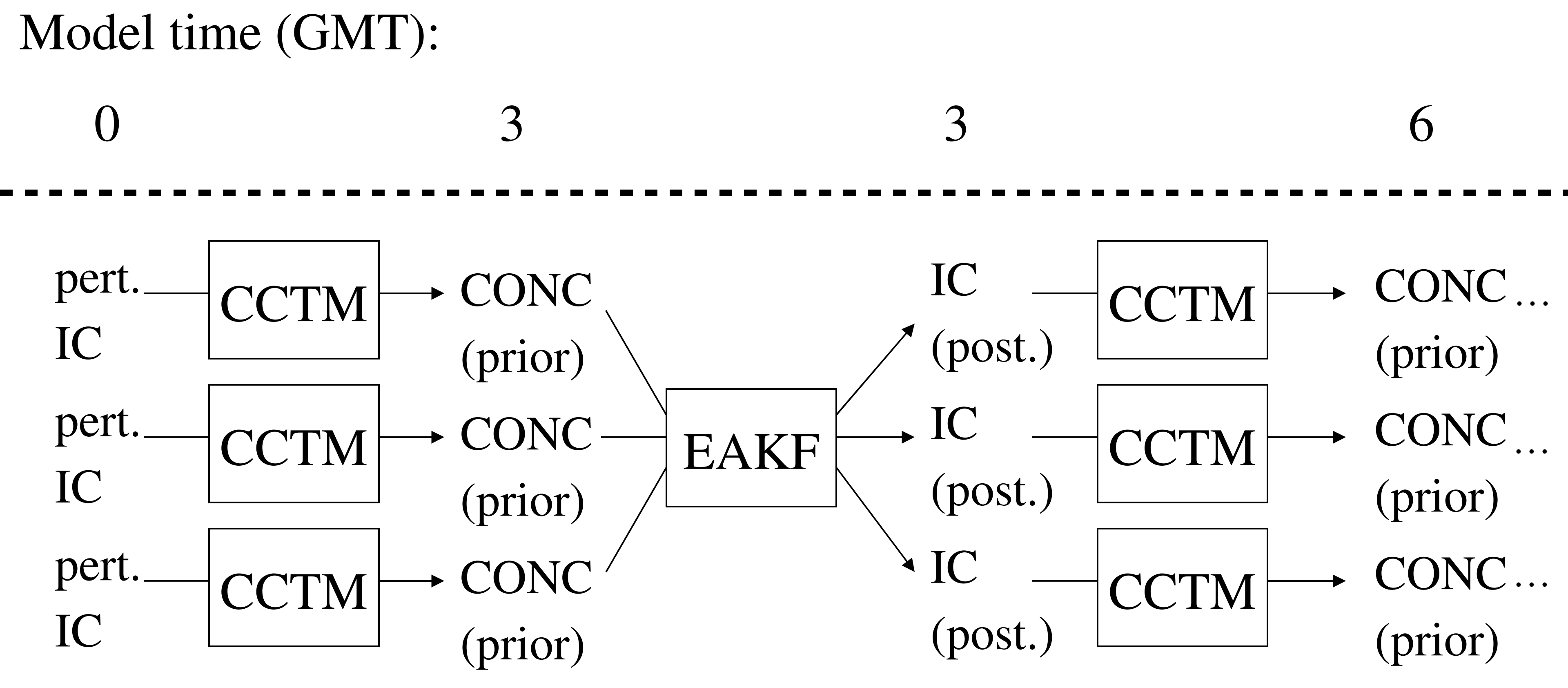
DART

- Data Assimilation Research Testbed
- NCAR
- Bayesian approach
- Ensemble - multiple realizations
- multiple filters

DART-CMAQ



DART-CMAQ


Model time (GMT):
0
3
3
6
pert.
IC
CCTM
CONC
(prior)
pert.
IC
CCTM
CONC
(prior)
pert.
IC
CCTM
CONC
(prior)
EAKF
IC
(post.)
CCTM
CONC ...
(prior)
IC
(post.)
CCTM
CONC ...
(prior)
IC
(post.)
CCTM
CONC ...
(prior)

Simplified CCTM

- Focus on CO (CCTM_CO):

$$\frac{\delta[CO]}{\delta t} = K_{form}[OH][HCOH] - K_{OH}[CO][OH]$$

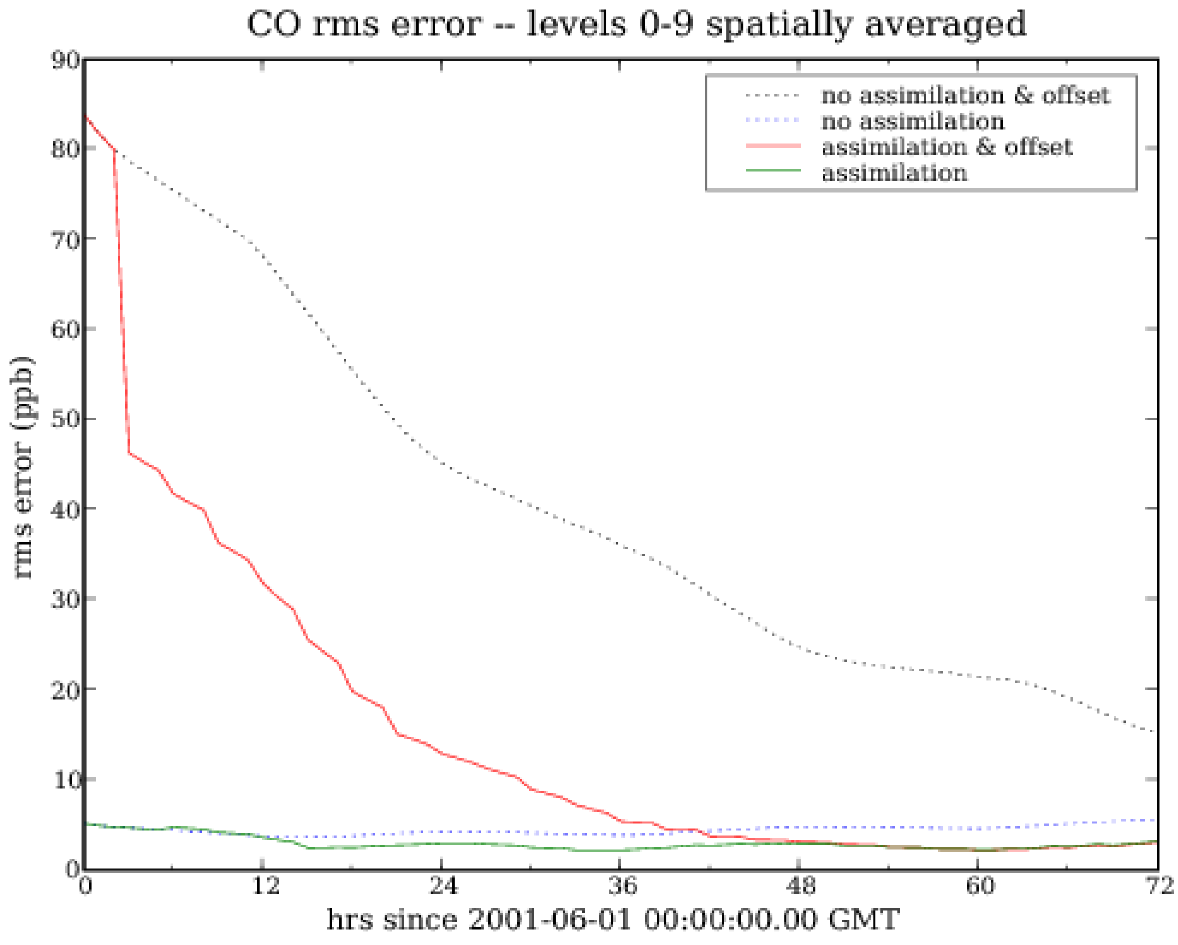
- Full CMAQ - OH, FORM (1 realization)
- Multiple realizations CCTM_CO (20)
- Computational efficiency:
  - Full CMAQ: 55 min/model day (4 processors)
  - CCTM_CO: 3-4 min/model day (1 processor)

Synthetic Experiments

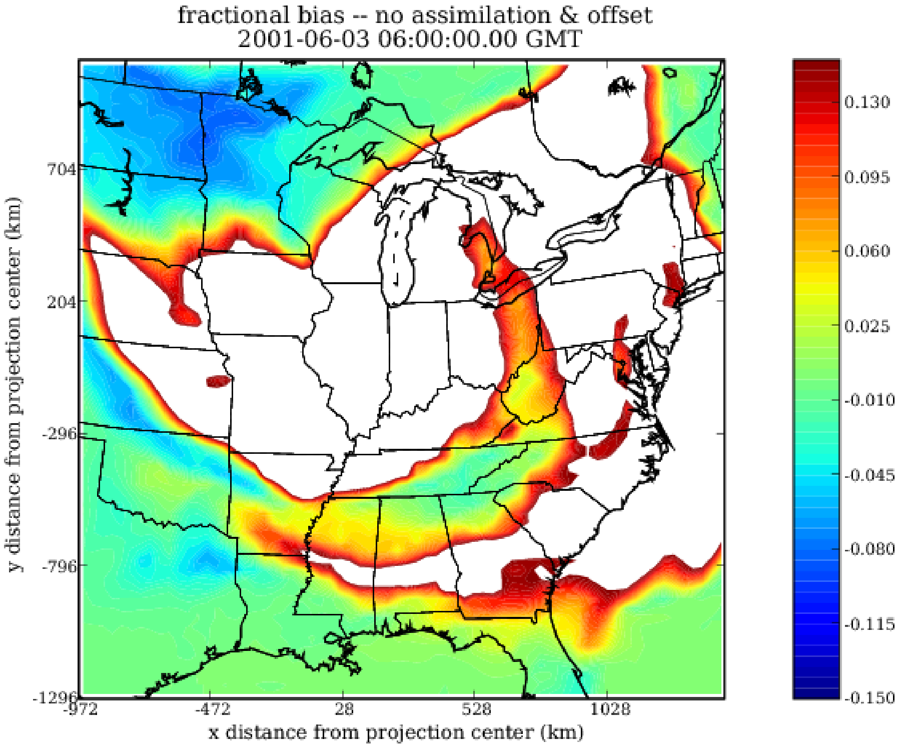
- Observations: Full CMAQ at AQS monitors
- Perturbed IC
- Perturbed OH and FORM
- EAKF-CMAQ w/ CCTM_CO

Offset Experiment

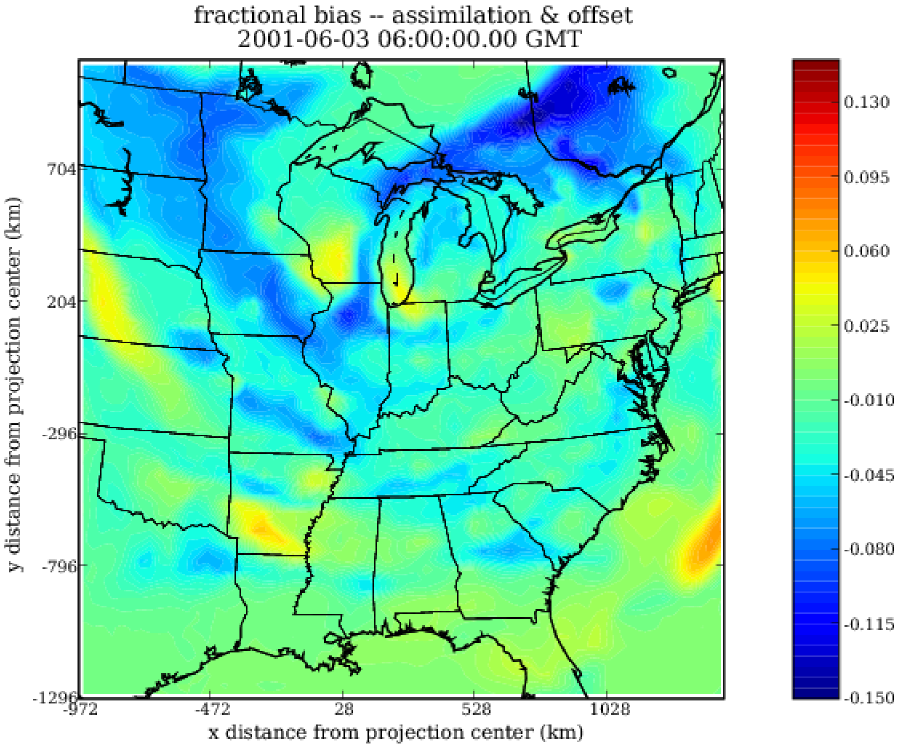
offset = 2*IC



Offset Experiment

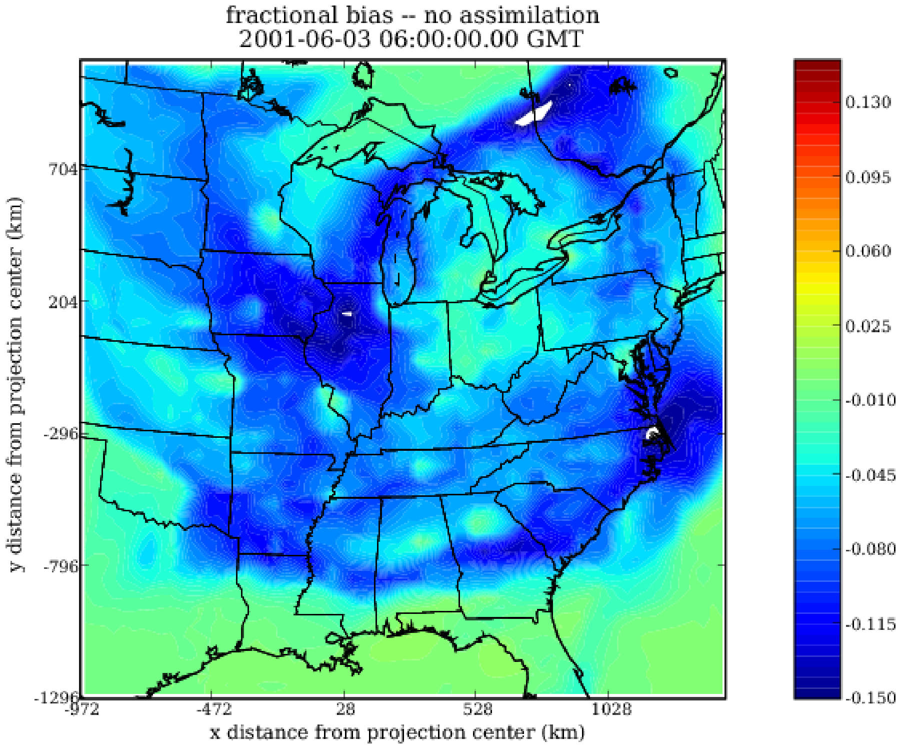


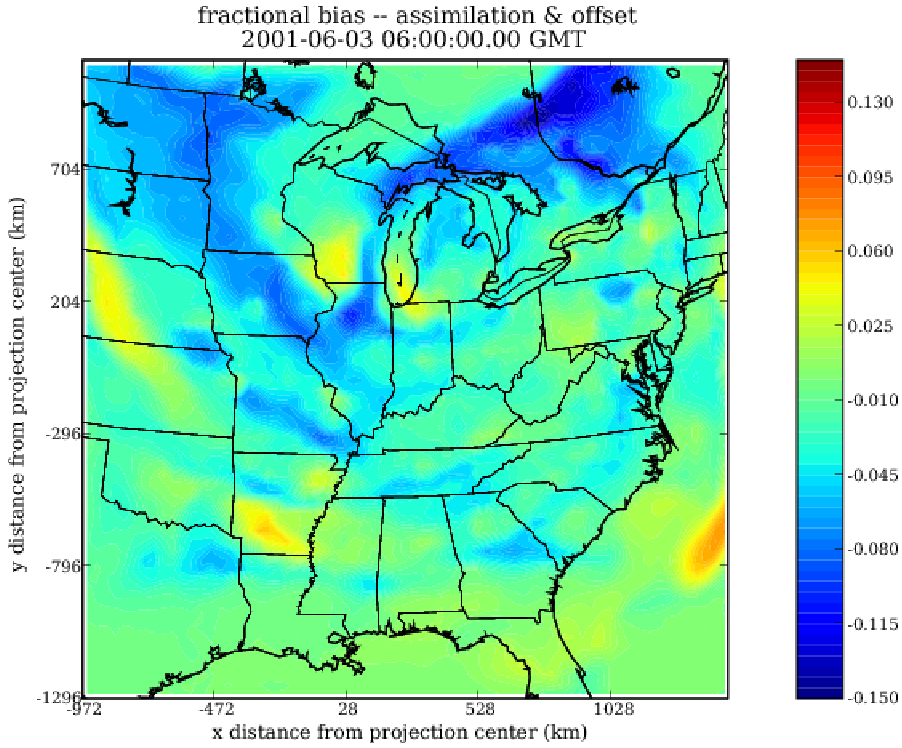
bias, no DA



bias, DA

Offset Experiment





no bias, no DA

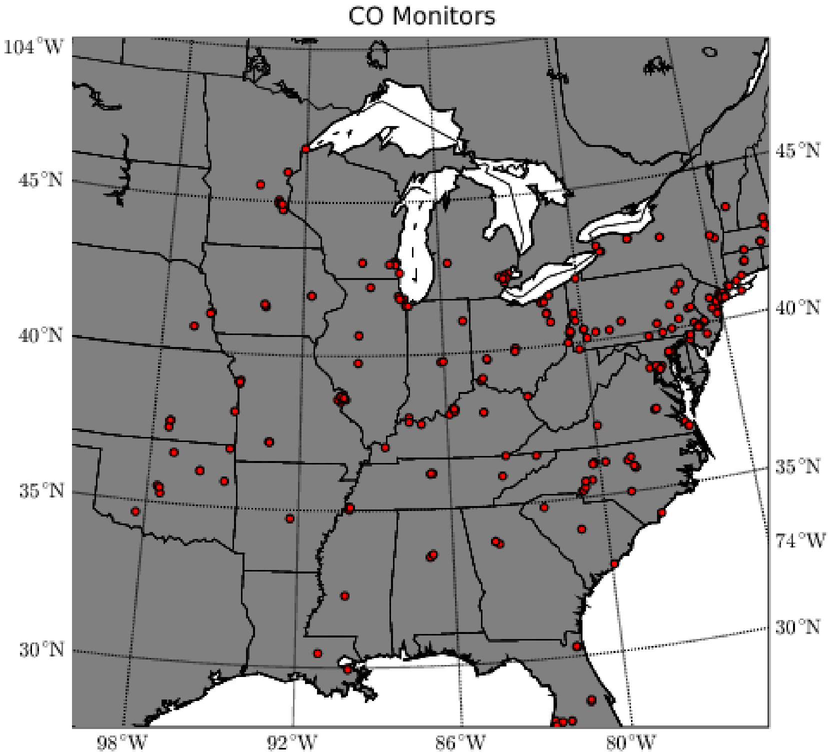
bias, DA

Future Research/Lessons

- Perturbing emissions, meteorology, & BC
- Assimilating for emissions
- MOPITT data
- DA reduces bias
- KF family of filters feasible
- Fully coupled models
- Ensembles approximate uncertainty

EXTRA SLIDES

Monitor Locations



Ensemble Kalman Filter

$$x_i{}^f(t+1) = \mathcal{M}(x_i^a(t)), i = 1, \ldots, N$$

$$x_i^a = x_i^f + K(y_i{}^o - \mathcal{H}x_i{}^f), i = 1, \ldots, N$$

$$P^f = \frac{1}{N-1}\sum(x_i^f - \overline{x^f})(x_i^f - \overline{x^f})^T$$

$$P^f\mathcal{H}^T = \frac{1}{N-1}\sum(x_i^f - \overline{x^f})(\mathcal{H}x_i^f - \overline{\mathcal{H}x^f})^T$$

$$\mathcal{H}P^f\mathcal{H}^T = \frac{1}{N-1}\sum(\mathcal{H}x_i^f - \overline{\mathcal{H}x^f})^T(\mathcal{H}x_i^f - \overline{\mathcal{H}x^f})^T$$

$$K = P^f\mathcal{H}(\mathcal{H}P^f\mathcal{H}^T + R)^{-1}$$

Ensemble Kalman Filter

$$x_i{}^f(t+1) = \mathcal{M}(x_i^a(t)), i = 1, \ldots, N$$

$$x_i^a = x_i^f + K(y_i{}^o - \mathcal{H}x_i{}^f), i = 1, \ldots, N$$

ensemble of model runs

$$P^f = \frac{1}{N-1}\sum (x_i^f - \overline{x^f})(x_i^f - \overline{x^f})^T$$

$$P^f\mathcal{H}^T = \frac{1}{N-1}\sum (x_i^f - \overline{x^f})(\mathcal{H}x_i^f - \overline{\mathcal{H}x^f})^T$$

$$\mathcal{H}P^f\mathcal{H}^T = \frac{1}{N-1}\sum (\mathcal{H}x_i^f - \overline{\mathcal{H}x^f})^T(\mathcal{H}x_i^f - \overline{\mathcal{H}x^f})^T$$

$$K = P^f\mathcal{H}(\mathcal{H}P^f\mathcal{H}^T + R)^{-1}$$

Ensemble Kalman Filter

$$x_i{}^f(t+1) = \mathcal{M}(x_i^a(t)), i = 1, \ldots, N$$

$$x_i^a = x_i^f + K(y_i{}^o - \mathcal{H}x_i{}^f), i = 1, \ldots, N$$

perturbed obs

$$P^f = \frac{1}{N-1} \sum (x_i^f - \overline{x^f})(x_i^f - \overline{x^f})^T$$

$$P^f \mathcal{H}^T = \frac{1}{N-1} \sum (x_i^f - \overline{x^f})(\mathcal{H}x_i^f - \overline{\mathcal{H}x^f})^T$$

$$\mathcal{H}P^f \mathcal{H}^T = \frac{1}{N-1} \sum (\mathcal{H}x_i^f - \overline{\mathcal{H}x^f})^T (\mathcal{H}x_i^f - \overline{\mathcal{H}x^f})^T$$

$$K = P^f \mathcal{H}(\mathcal{H}P^f \mathcal{H}^T + R)^{-1}$$

$$b = \frac{CO_{full} - CO_{mean}}{\left[\frac{CO_{full} + CO_{mean}}{2}\right]}$$