MATERHORN
Summary of Progress at NPS and NCAR*

Josh Hacker

Contributions from Sean Wile, Hank Chilcoat, Jared Lee, and Raquel Lorente-Plazas

*Does not represent work by Jason Knievel and Yubao Liu in collaboration with UU and UVA

NCAR

Topics

NCAR

- Data assimilation (ensemble)
- Ensemble sensitivities for observing strategies)
- Model error estimation
- Parameter estimation
- Observation bias estimation

Ensemble Sensitivities

Network design for fine-scale near-surface forecasts in complex terrain

Ensemble sensitivity analysis (ESA)

NCAR

How does the change in a set of initial state variables x_s change a forecast metric J?

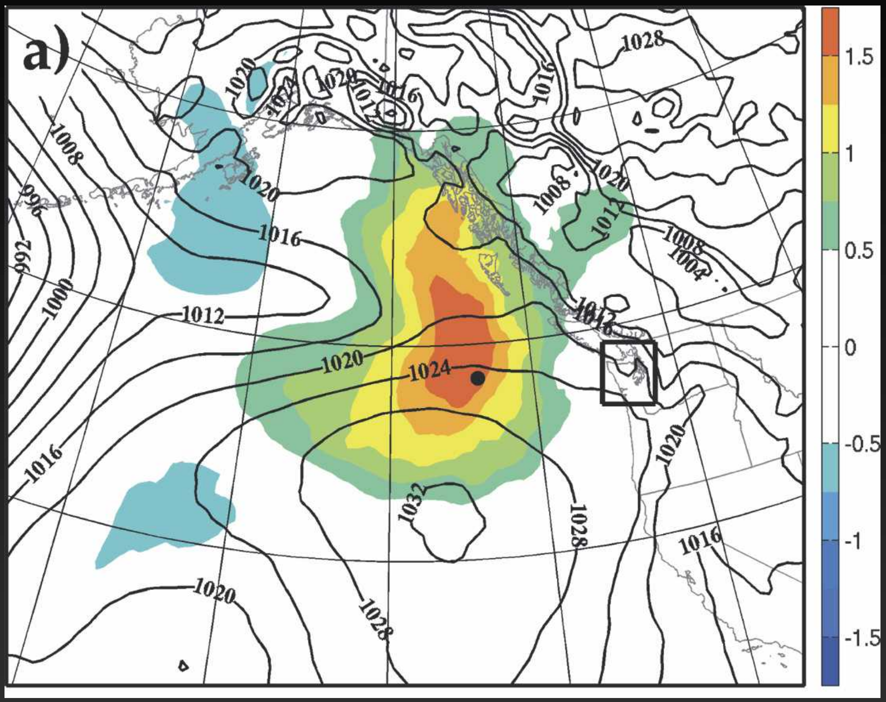
$$\frac{\partial J_e}{\partial x^a}$$

- Identify dynamically relevant covariance structures in space and time
- Propose observing strategies for mesoscale, short-range forecasts in complex terrain
- Sensitivity scales (time and space) to infer predictability scales
- Predictability of specific phenomena
- Open issues:
  - Sampling error
  - Linearity assumptions in complex terrain

Ensemble Sensitivity Background

- Ancell and Hakim (2007) showed theoretical equivalence between adjoint and ensemble sensitivity for linear perturbations and Gaussian statistics
- Relies on linearization about an ensemble-mean trajectory
- Rigorous application has so far been limited to large-scale (smooth) and integrated processes where strong linear relationships are more likely

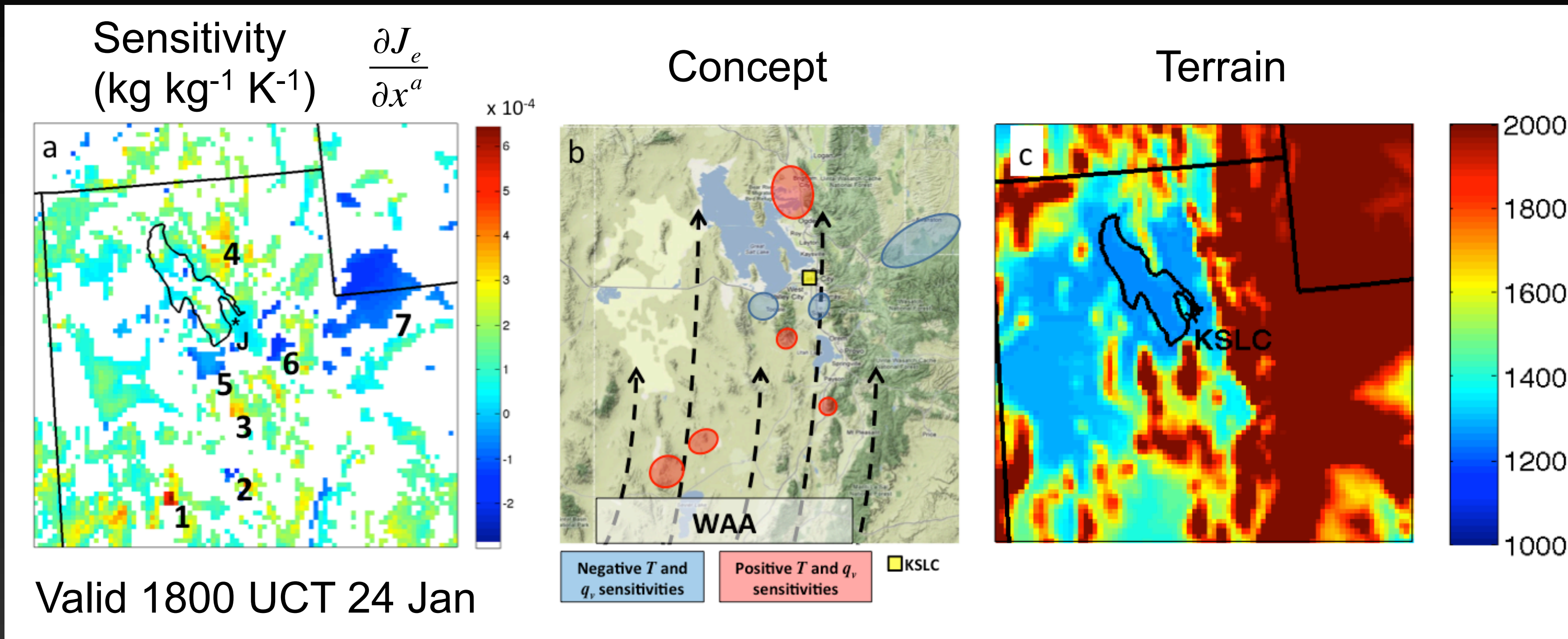
An optimal ensemble data assimilation system provides an appropriate sample



Sensitivity of 24-h sea-level pressure (SLP) over western Washington to SLP initial conditions, and ensemble-mean SLP (from Torn and Hakim 2008).

Moisture sensitivity to temperature

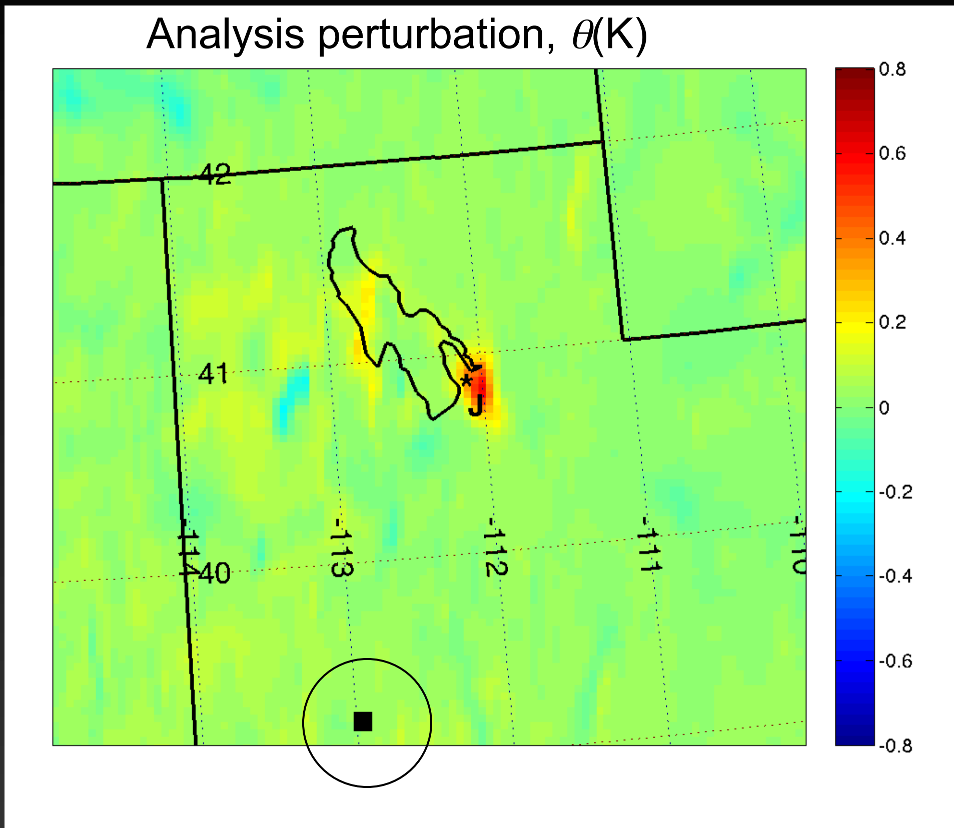
NCAR

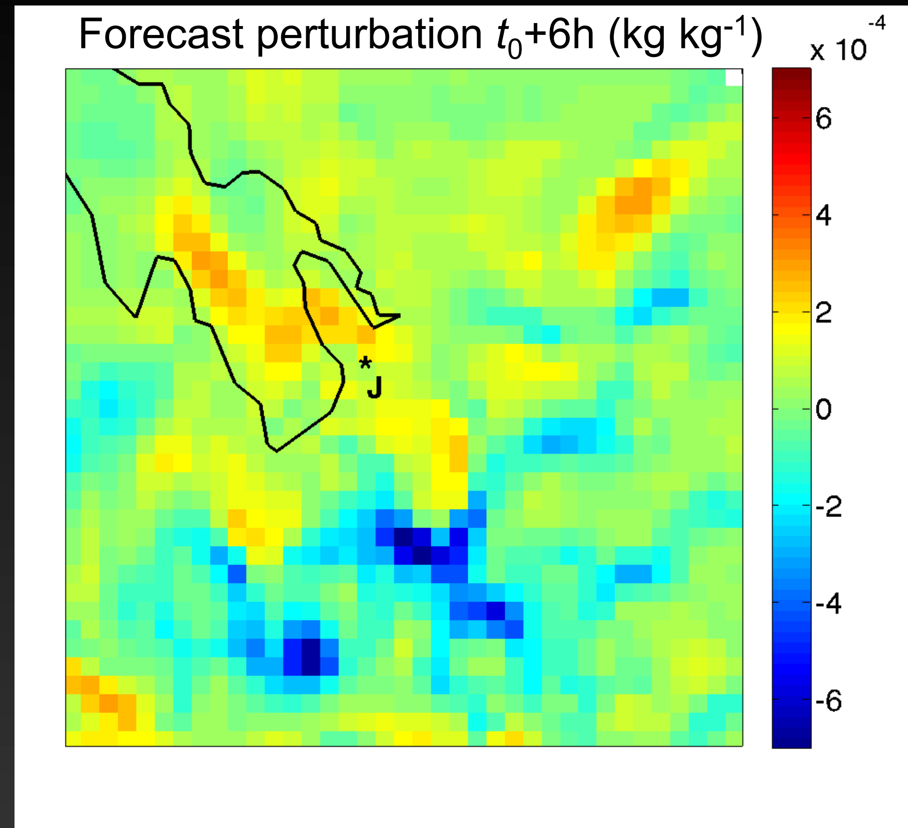


J = 2x2x2 box-mean water vapor mixing ratio over Salt Lake City airport

x=Potential temperature (here on model first layer)

Perturbation experiments



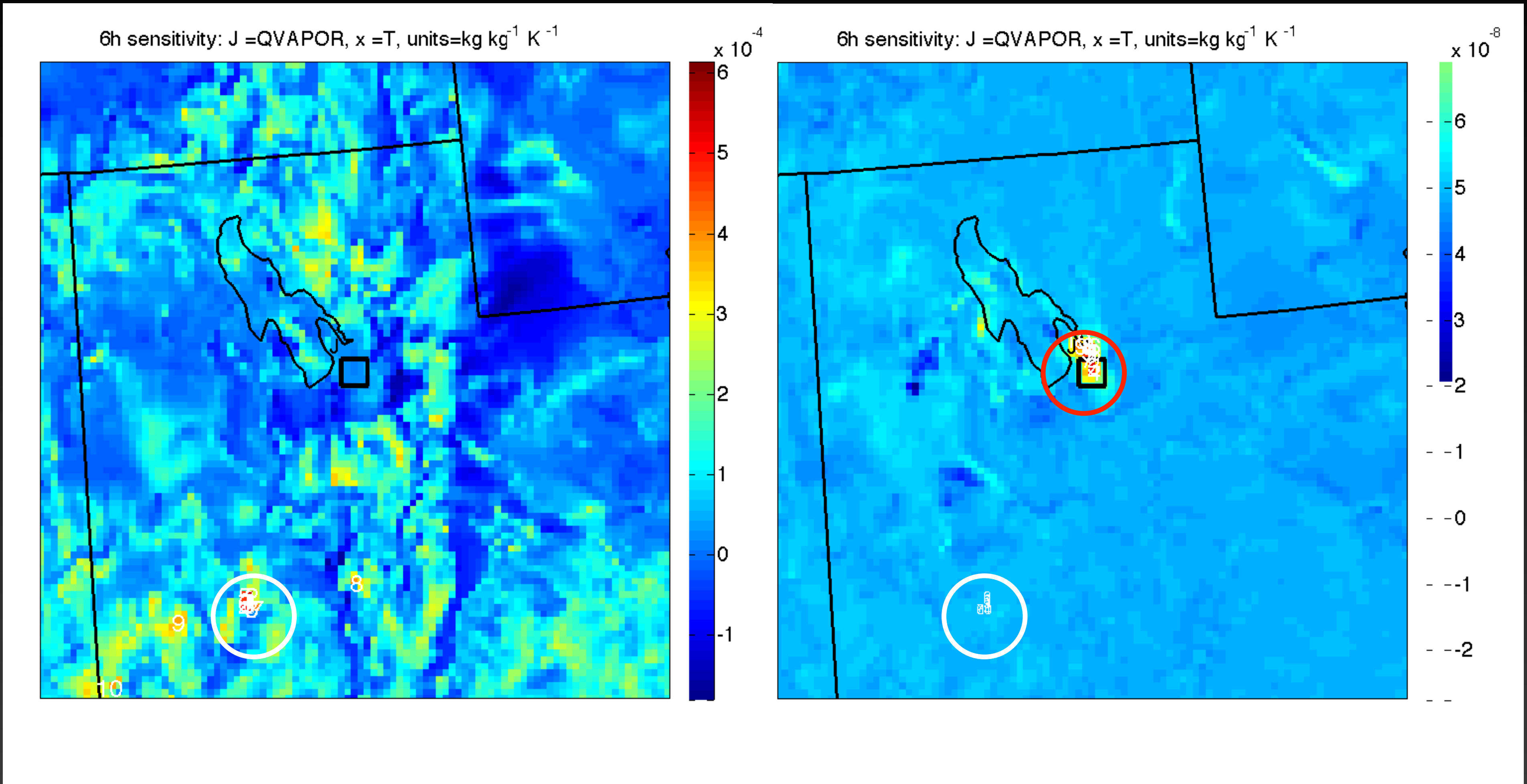


Perturbation of one analysis standard deviation in θ at the most sensitive location, *assimilated* with ensemble filter.

Effect of approximation

Diagonal approximation | Full covariance

NCAR



Approximation under-emphasizes sensitivities local to the response. Agreement on some sensitive points (numbered) to southwest of response.

Ensemble Sensitivity with Localization

$$\delta J = \alpha \circ \left\{ \mathbf{J}_{\mathbf{e}}^{\mathrm{T}} \left[\mathbf{X}_i^{\mathbf{a}} \left(\mathbf{X}_i^{\mathbf{a}\,\mathrm{T}} \mathbf{X}_i^{\mathbf{a}} \right)^{-1} \right]^{\mathrm{T}} \rho \circ \mathbf{P}_i^{\mathbf{a}} \mathbf{h}_{i+1}^{\mathrm{T}} \left(\mathbf{h}_{i+1} \rho \circ \mathbf{P}^{\mathbf{a}} \mathbf{h}_{i+1}^{\mathrm{T}} + \mathbf{R} \right)^{-1} \left(y_{i+1}^{o} - \mathbf{h}_{i+1} \mathbf{x}_i^{\mathbf{a}} \right) \right\}$$

$$= \alpha \circ \left\{ \mathbf{J}_{\mathbf{e}}^{\mathrm{T}} \left[\mathbf{X}_i^{\mathbf{a}} \left(\mathbf{X}_i^{\mathbf{a}\,\mathrm{T}} \mathbf{X}_i^{\mathbf{a}} \right)^{-1} \right]^{\mathrm{T}} \delta \mathbf{x}^{\mathbf{a}} \right\}$$

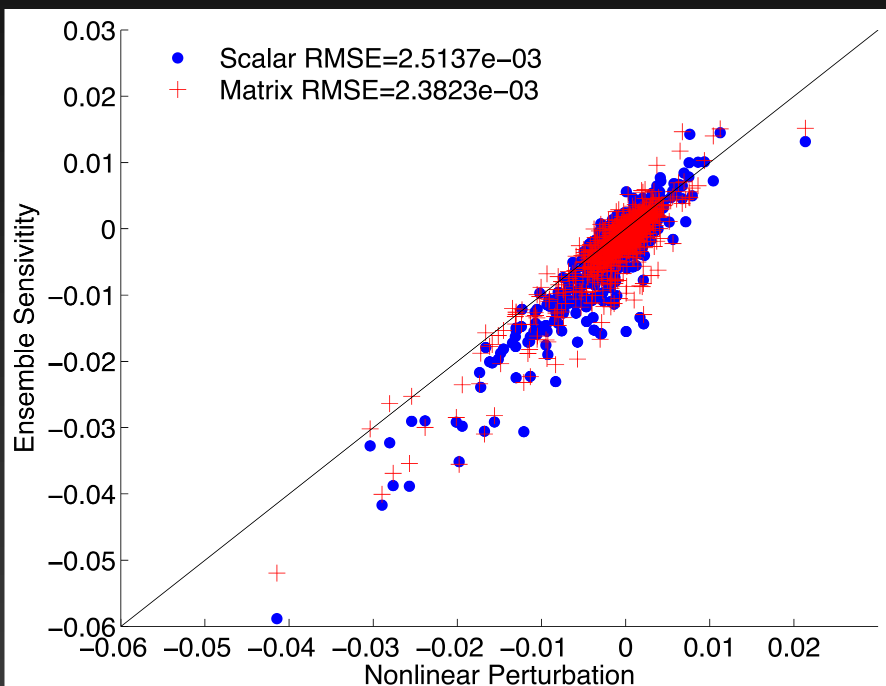
- Covariance localization, or tapering, can be applied
  - at the assimilation step with ρ
  - to the regressions with α
- ρ is typically a function of space alone
- α is function of space and time, here from a Bayesian hierarchical estimate (Anderson 2007)

Perfect Model (two scales)

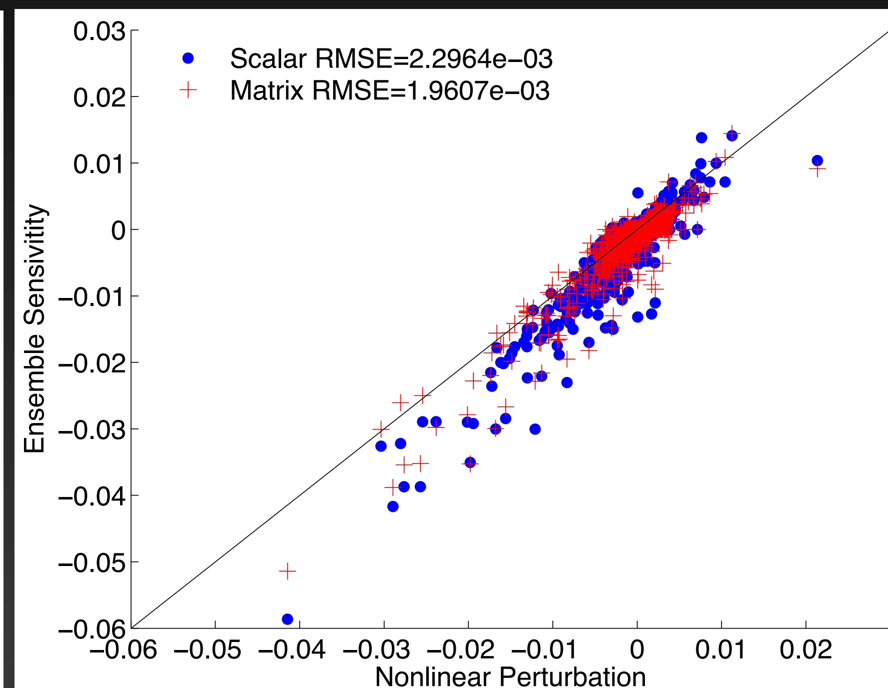
NCAR

When only smooth/slow scales present, little difference between univariate (scalar) and multivariate (matrix) predictions of response to perturbation.

Sensitivity *without* localization



Sensitivity *with* localization



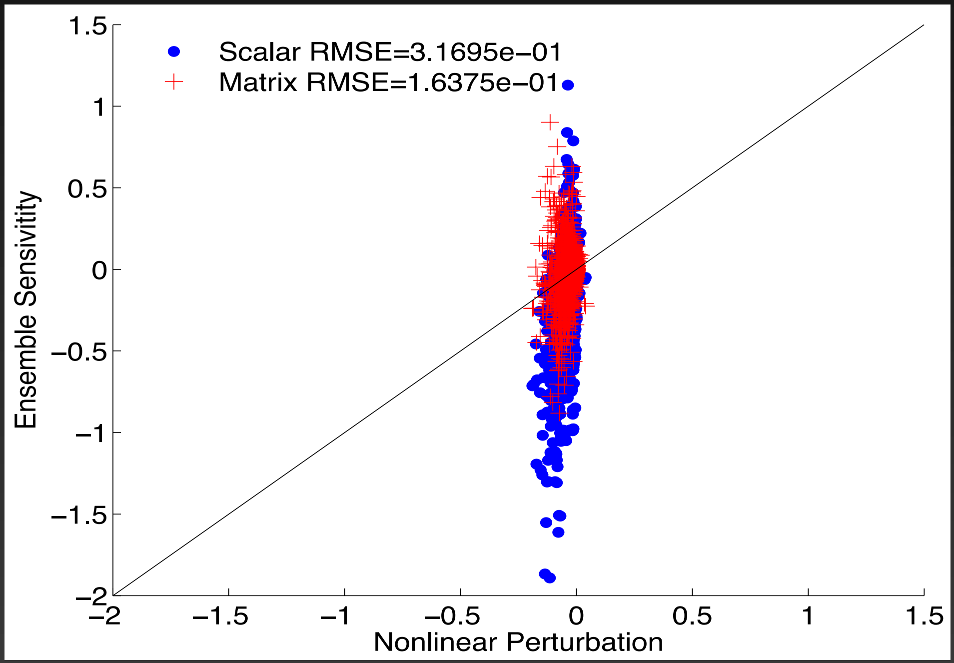
Here observations are randomly chosen from every other gridpoint (which are un-observed for sensitivity calculations).

Imperfect Model

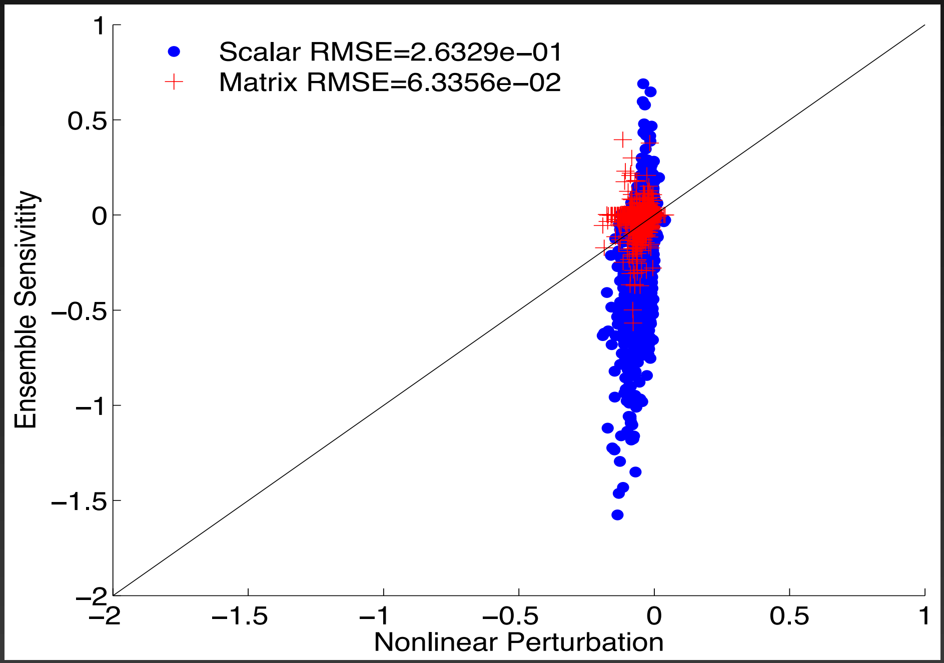
For imperfect model, diagonal approximation results in greater over-prediction of response; multivariate sensitivities account for presence of fast scales in real system, which appears as noise.

NCAR

Sensitivity *without* localization



Sensitivity *with* localization



Here observations are assimilated on half of domain that is data void; more impact from observations because greater uncertainty in analysis.

Model Inadequacy

Quantifying model error with ensemble data assimilation

Bias in state estimation

NCAR

A scalar example (adapted from Dee and DaSilva 1998):

Consider

$$\beta_f = E\left(\varepsilon_f\right) \neq 0$$
$$\beta_o = E\left(\varepsilon_f\right) = 0$$

The background/forecast is biased but the observation is not.

Then

$$\sigma_f^2 = E\left[\left(\varepsilon_f - \beta_f\right)^2\right]$$
$$\sigma_o^2 = E\left(\varepsilon_o^2\right)$$

The background error variance is biased, but the observation error variance is not.

If β_b is considered, the analysis is unbiased:

$$\text{for } \sigma_f^2 = \sigma_o^2 = \sigma^2 :$$
$$x_a = \frac{1}{2}\left(x_f - \beta_f + y_o\right)$$
$$\beta_a = 0, \quad E\left(\varepsilon_a^2\right) = \frac{1}{2}\sigma^2$$

If β_b is ignored, the analysis is biased:

$$\text{for } \sigma_f^2 = \sigma_o^2 = \sigma^2 :$$
$$x_a = \frac{1}{2}\left(x_f + y_o\right)$$
$$\beta_a = \frac{1}{2}\beta_f, \quad E\left(\varepsilon_a^2\right) = \frac{1}{4}\beta_f^2 \frac{1}{2}\sigma^2$$

Systematic model errors

NCAR

- Persistent structures or time-means in the innovations or increments in data assimilation, or tendencies in the model, result from systematic error.
- Objective data assimilation cannot eliminate bias.

For $\sigma_f^2 = \sigma_o^2 = \sigma^2$ and an unbiased observation:

$$x_a - x_f = \frac{1}{2}\left(x_f + y_o\right) - x_f = \frac{1}{2}\left(y_o - x_f\right)$$

$$2E\left(x_a - x_f\right) = E\left(y_o - x_f\right) = E\left(y_o - x_t - \varepsilon_f\right) = \left[E\left(y_o\right) - E\left(x_t\right) - \beta_f\right]$$

$$= \left[E\left(x_t\right) + E\left(\varepsilon_o\right) - E\left(x_t\right) - \beta_f\right] = \beta_f$$

- Bias can β be estimated from analysis increments or background.
- Not immediately clear whether the observations or the model is biased, but we can extract the statistics given an observation network containing unbiased observations.

Systematic observation errors

- Persistent structures or time-means in the innovations or increments in data assimilation, or tendencies in the model, result from systematic error.
- Objective data assimilation cannot eliminate bias.

For $\sigma_f^2 = \sigma_o^2 = \sigma^2$ and an unbiased forecast:

$$x_a - x_f = \frac{1}{2}\left(x_f + y_o\right) - x_f = \frac{1}{2}\left(y_o - x_f\right)$$

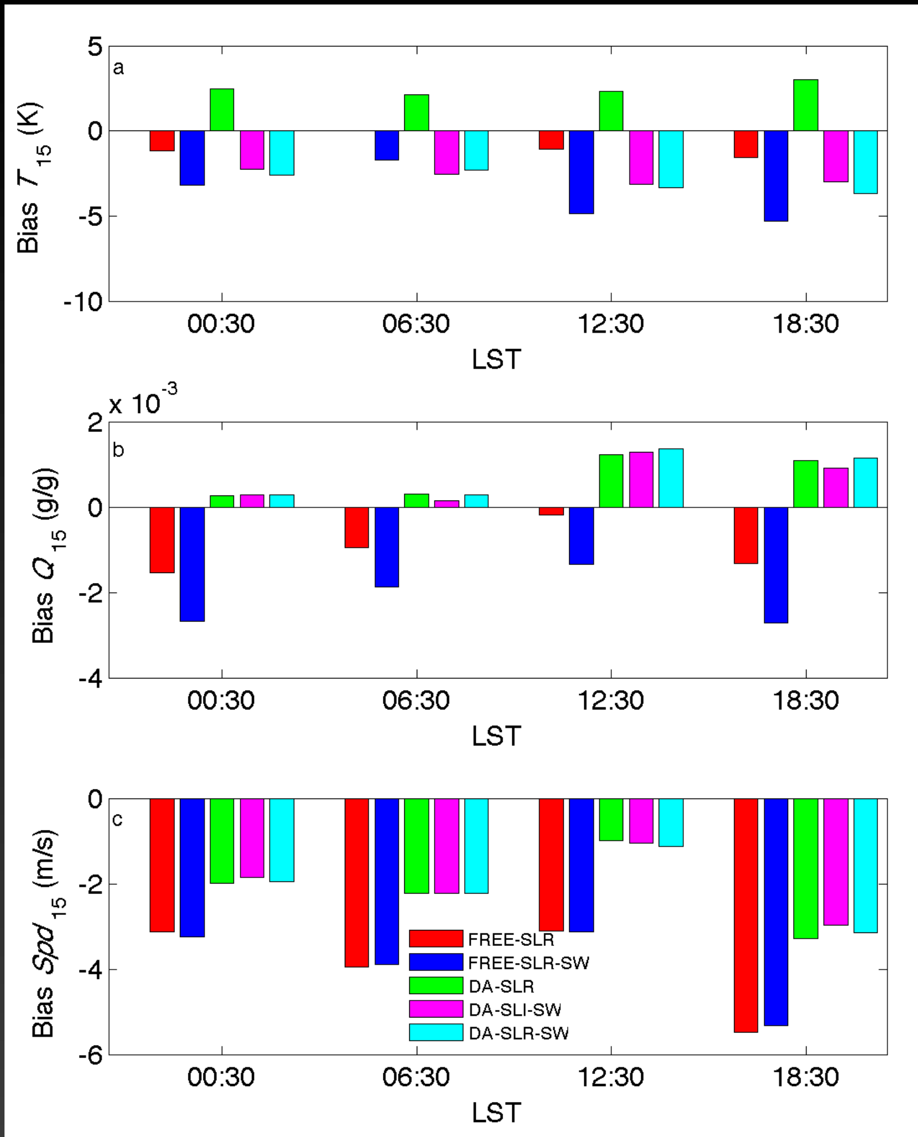
$$2E\left(x_a - x_f\right) = E\left(y_o - x_f\right) = E\left(y_t - \varepsilon_f - x_f\right) = \left[E\left(y_t\right) + \beta_o - E\left(x_f\right)\right]$$

$$= \left[E\left(x_t\right) + \beta_o - E\left(x_t\right) - E\left(\varepsilon_f\right)\right] = \beta_o$$

- Bias can β be estimated from analysis increments or background.
- Not immediately clear whether the observations or the model is biased, but we can extract the statistics given an observation network containing unbiased observations.

Predicting (independent) observations at layer-1 (~15 m AGL)

NCAR



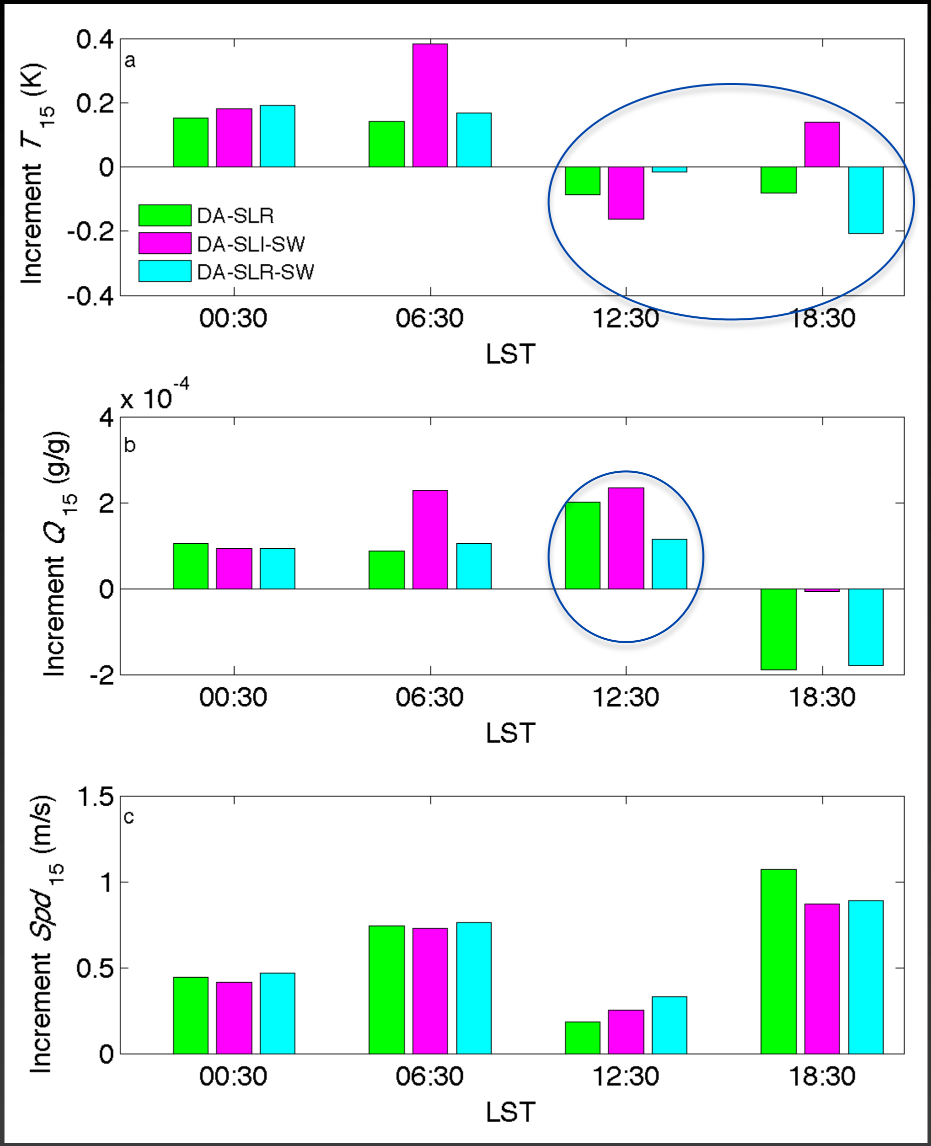
← Night Day →

- Bias magnitudes of winds smaller when assimilating
- Biases here are the generally the same sign as biases in observation space

Nothing looks too strange

- Bias generally of the same sign as at shelter or anemometer
- Magnitude of bias not reduced as clearly as at shelter or anemometer
- Except: FREE-SLR less biased in temperature than DA-SLR (maybe a little strange)

Analysis increments



← Night Day →

Signs that covariances can be biased:

- Some increments are acting to increase the bias

Possible causes:

- Nonlinear statistics in regressions underlying the assimilation (not the case here)
- Poor parameterization of surface-layer profiles

Model Inadequacy

Parameter estimation to address parametric model error (and systematic observation error)

Parametric errors

NCAR

Theory leads to state augmentation approach:

$$\mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \mathbf{p} \end{bmatrix}$$

"Augmented" state vector

$\mathbf{x}$ State vector

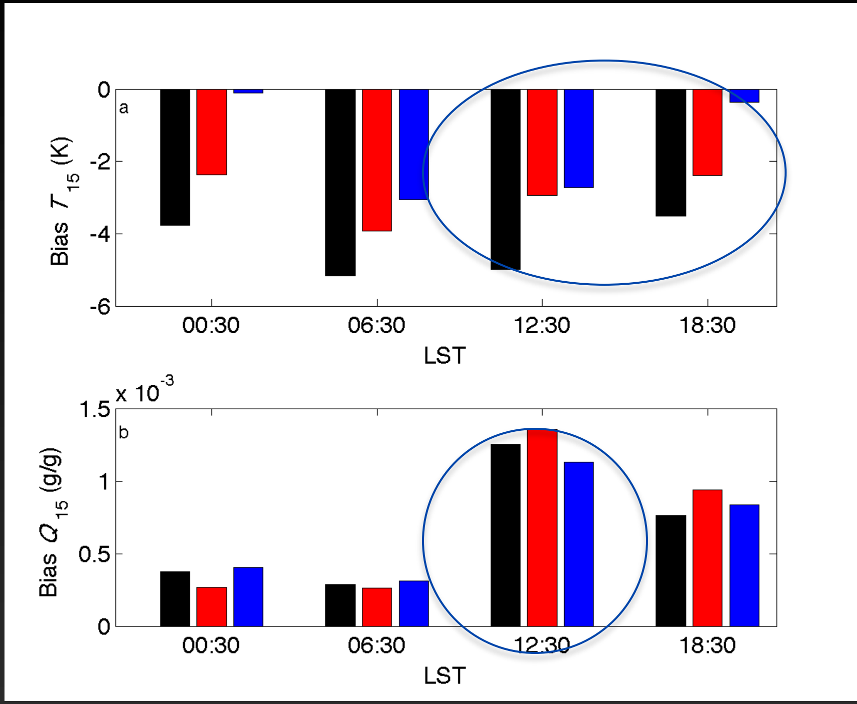
$\mathbf{p}$ Vector of parameters

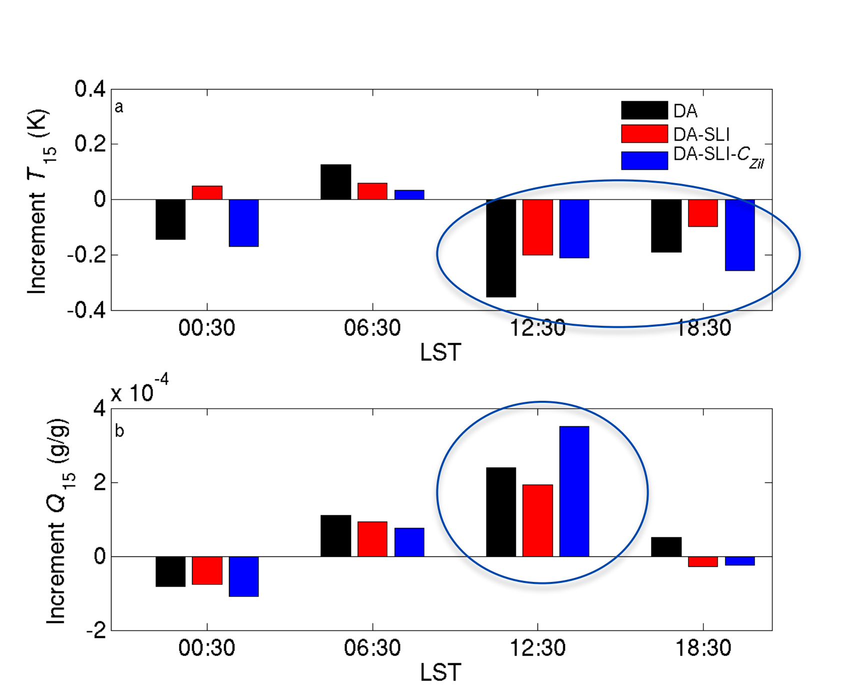
The statistical analysis equation is still valid; forward operators and covariance matrices are modified accordingly

- If a model prediction is sensitive to a parameter value, then covariances between parameters and predicted observations ($\mathbf{Gz}_b$) will result ($\mathbf{G}$ is the augmented forward operator).

- Parameter values can be modified by observations.
- Parameters can be in the model or otherwise arbitrary.

Bias and analysis increments

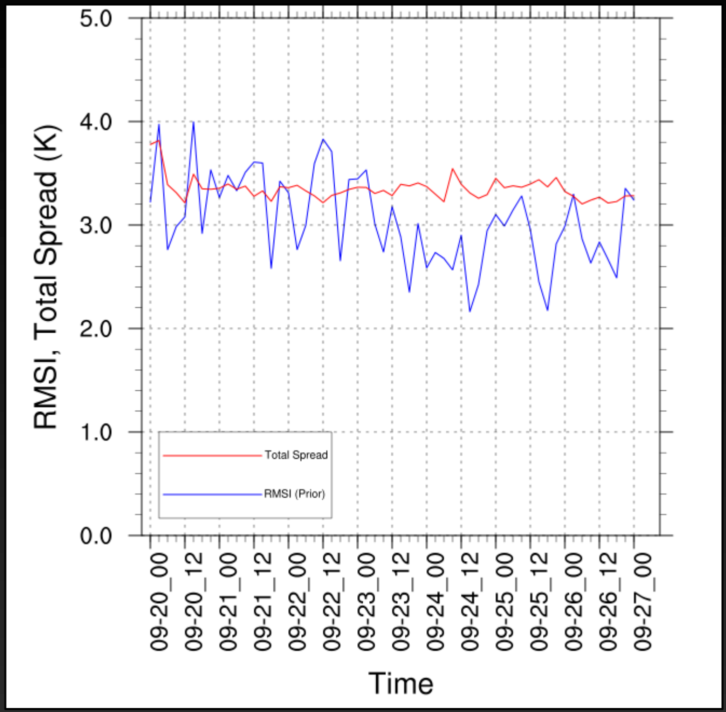


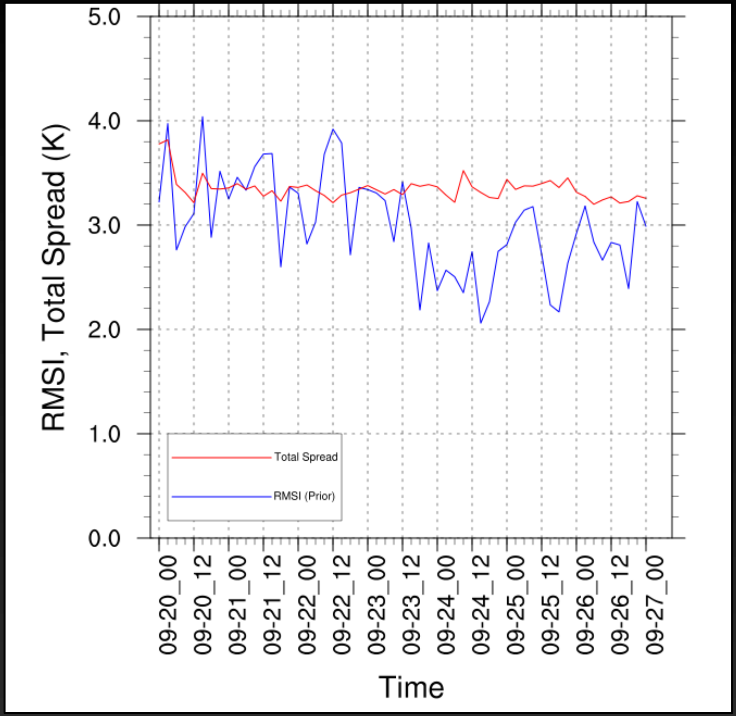


- Even with parameter estimation, analysis increments can be systematically in the wrong direction.
- DA increasing the biases via the analysis increments.
- Suggests structural errors.

Parameter estimation (3D model)

NCAR





METAR 2-m T

Constant C_{ZIL}

Mean RMS Innov. = 3.01 K

Mean Tot. Spread = 3.33 K

(Means discard 20 Sep)

METAR 2-m T

Variable C_{ZIL}

Mean RMS Innov. = 2.96 K

Mean Tot. Spread = 3.33 K

(Means discard 20 Sep)

Forward operators

NCAR

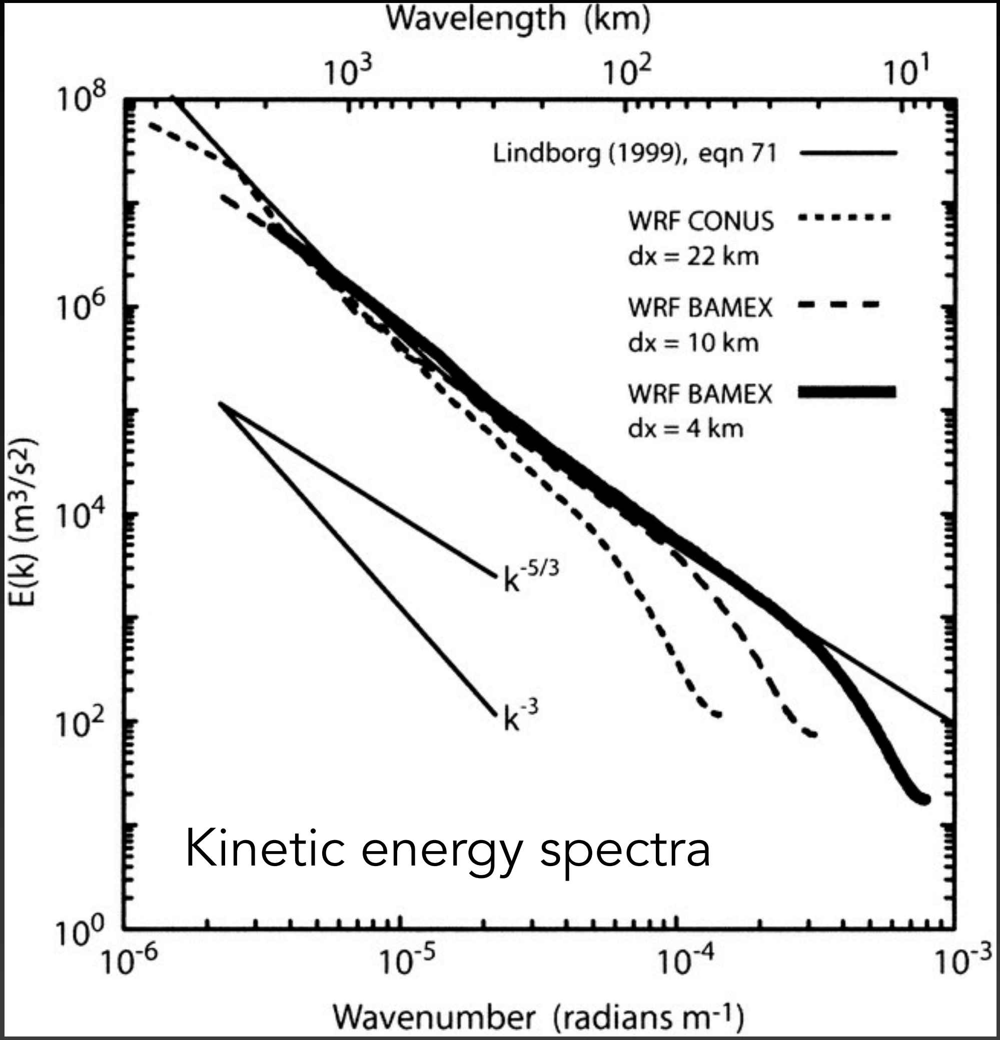
- Give predicted observations from forecasts model prognostic variables: $\mathbf{Hx}^{\mathrm{f}} \approx h(\mathbf{x}^{\mathrm{f}})$
- Simple:
  - Interpolation to a radiosonde observation
  - Diagnostic surface-layer variables (2-m T, 10-m winds) and interpolation to horizontal location)
- Complex:
  - Satellite radiances (also nonlinear)
  - GPS radio occultation (also nonlinear)
  - Doppler winds
  - Radar reflectivity (also nonlinear)

Observing scales

- An observation "sees" all scales of motion slower than its sampling rate.
- Most observations sample well below the numerically diffusive range.
- Most useful observations contain energy projecting onto scales above the diffusive range.

Examples:
- 10-minute average surface obs
- Packages on small UAVs.

Skamarock (2004)



Progress summary

NCAR

- Ensemble sensitivities for observation network design in complex terrain
- Model error quantification and parameter estimation for land-atmosphere coupling errors
- Parameter estimation for near-surface observation error estimation (Raquel Lorente-Plazas)

References

- **Hacker, J. P.** and L. Lei, 2015: Multivariate ensemble sensitivity with localization. *Mon. Wea. Rev.*, **143**, 2013–2027. doi: http://dx.doi.org/10.1175/MWR-D-14-00309.1
- Wile, S.M., **J. P. Hacker**, and K.H. Chilcoat, 2015: The potential utility of high-resolution ensemble sensitivities during weak flow in complex terrain. Accepted for publication in *Wea. and Forecast.*, May 2015.
- Ryerson, W., and **J. Hacker**, 2014: The potential for mesoscale visibility predictions with a multi-model ensemble. *Wea. and Forecast.*, **29**, 543–562. doi: http://dx.doi.org/10.1175/WAF-D-13-00067.1
- **Hacker, J.** and W. Angevine, 2013: Ensemble data assimilation to characterize surface-layer errors in numerical weather prediction models. *Mon. Wea. Rev.*, **141**, 1804–1821. doi: http://dx.doi.org/10.1175/MWR-D-12-00280.1

Statistical-analysis equation

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}\left(\mathbf{y}_o - \mathbf{H}\mathbf{x}_b\right)$$

$$\mathbf{K} = \frac{\mathbf{B}\mathbf{H}^T}{\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}}$$

$$\mathbf{B} = \left\langle \mathbf{e}_b \mathbf{e}_b^T \right\rangle, \quad \mathbf{R} = \left\langle \mathbf{e}_o \mathbf{e}_o^T \right\rangle$$

$$\mathbf{P}_a = \left\langle \mathbf{e}_a \mathbf{e}_a^T \right\rangle = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B}$$

Under assumptions of Gaussian, unbiased errors, and linear operator $\mathbf{H}$, this equation holds for:

- Optimal interpolation
- Variational minimization
- Kalman filters
- Other linear filters

Differences arise when trying to estimate the error statistics, or when additional constraints are imposed.

Effect of hypothetical θ observation

NCAR

$$\delta J_e = \frac{\partial J_e}{\partial x^a}\mathbf{K}(\mathbf{y}^o - \mathbf{h}\mathbf{x}^a)$$

$$\mathbf{K} = \mathbf{P}^a\mathbf{h}^{\mathrm{T}}\left(\mathbf{h}\mathbf{P}^a\mathbf{h}^{\mathrm{T}} + \mathbf{R}\right)^{-1}$$

Can test use of sensitivities to predict the change in forecast metric resulting from a hypothetical observation. Analysis increment can come from:

- assimilating synthetic obs
- approximation with univariate linear regression