# Prologue – Perspective, Premises and Presentation Plan

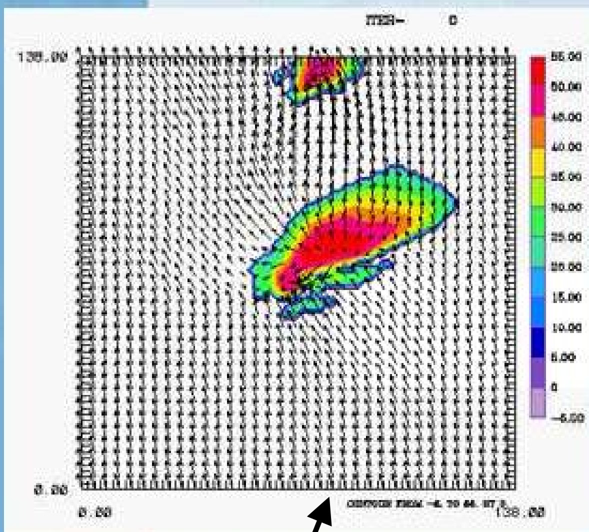
- A look at data assimilation issues from a data perspective instead of an assimilation one.
- 4D-Var context assumed:
  Mesoscale focus → Fewer valid meteorological approximations + Limited datasets → Need to look at the time evolution of fields to retrieve the many unobserved variables.
- Presentation focus is on explaining paradigm and showing results; light on methods/details.

Source: Presentation by N.A. Crook et al., 2004: Assimilation of radar observations of a supercell storm using 4DVar…. 22nd Conf. on Severe and Local Storms.

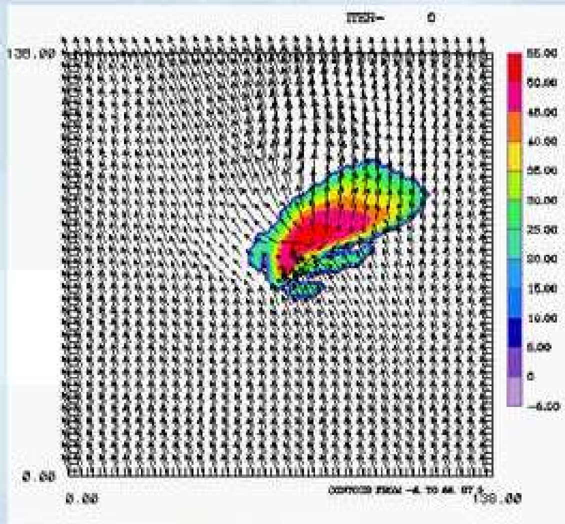
# Sensitivity to Low Level Moisture

NCAR

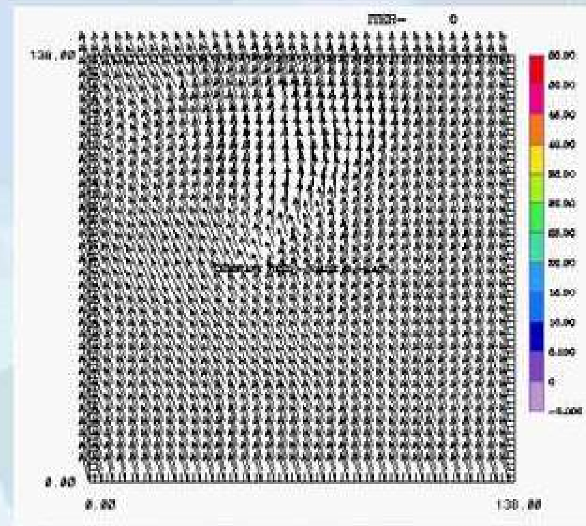
16 g/kg ($T_d \approx 21.5$°C)



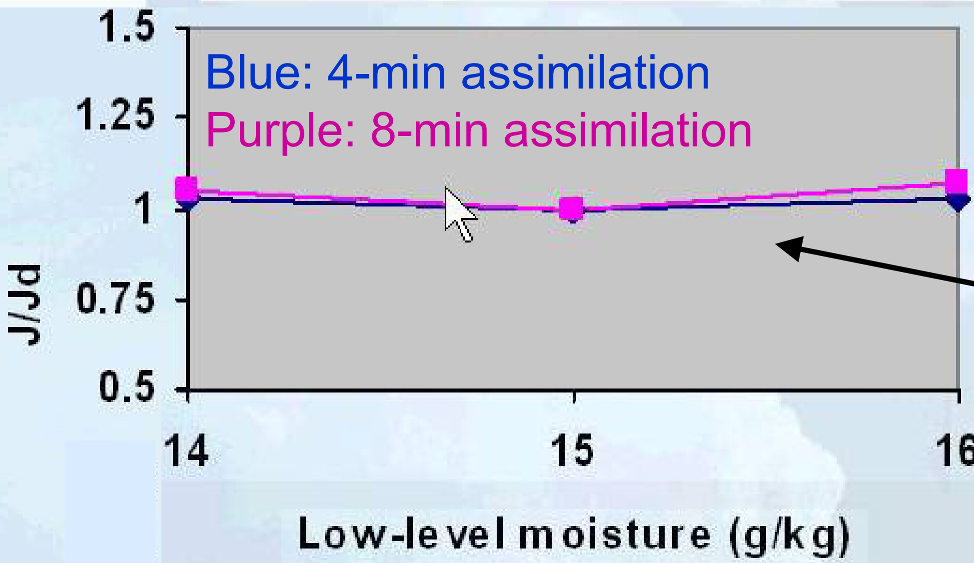
15 g/kg



14 g/kg ($T_d \approx 20$°C)



If the well observed storm intensity is so sensitive to low level moisture...



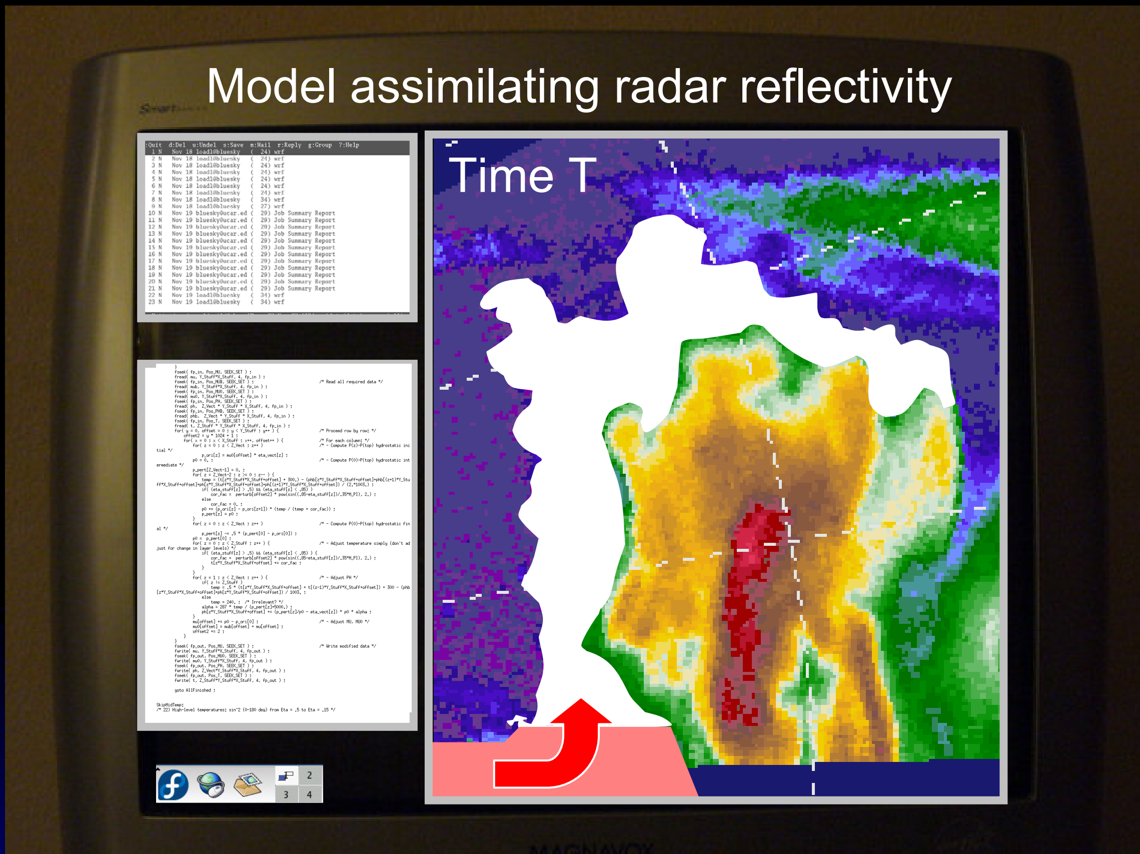

… then why can't we retrieve low-level moisture by assimilating radar data?

# Conditions for a Successful Data Assimilation

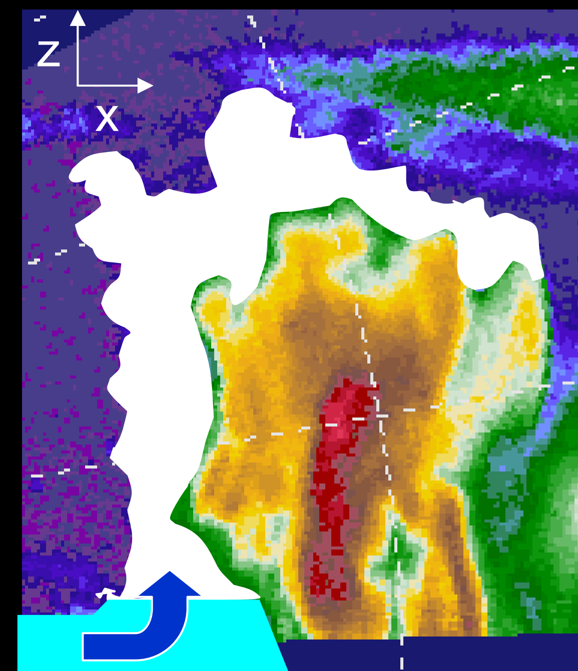
1) The difference between the expected atmospheric state **x'** and the true atmospheric state **x** results in a measurable difference between the expected observations **y'** and the true observations **y**.
2) Given **x'** and **x**, a model can reproduce the associated observations **y'** and **y**.
3) The data assimilation system can use **y'** and **y** to change the model state from **x'** to **x**.

Something "failed" in Crook et al. But what?

# A Pathway to Failure

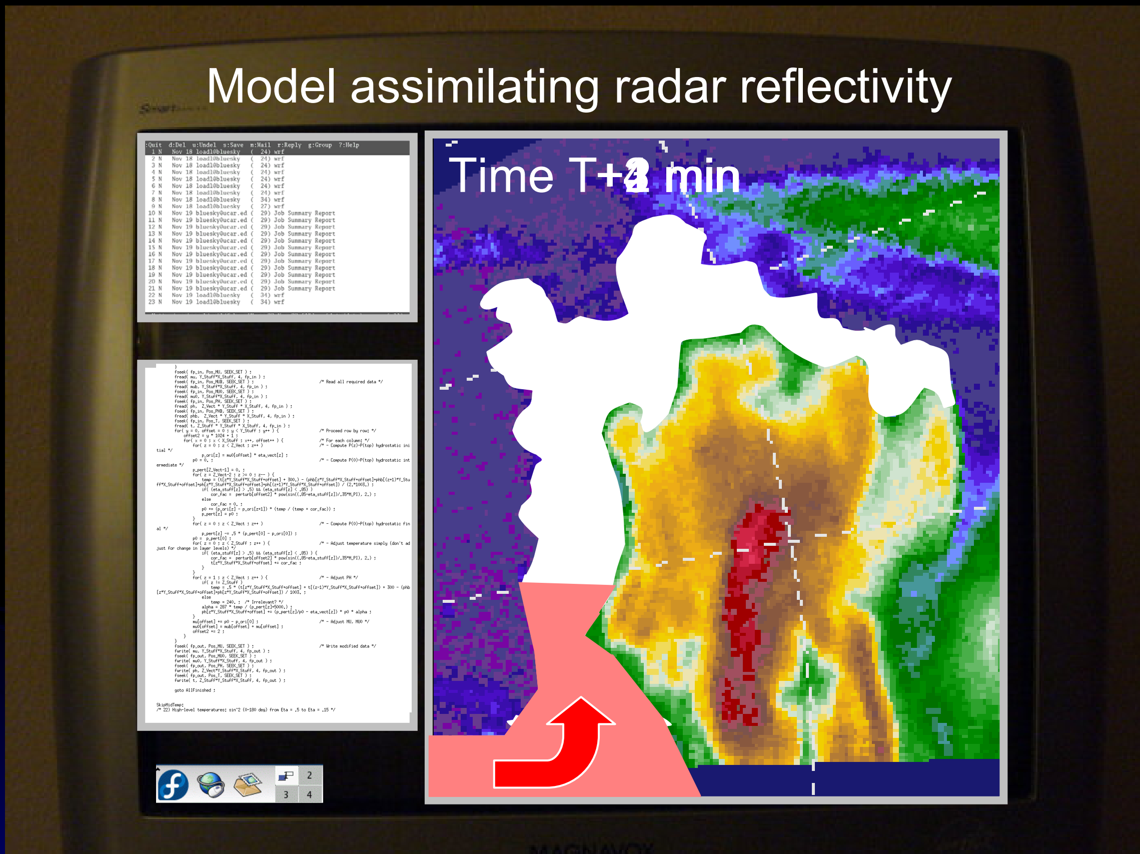



Reality

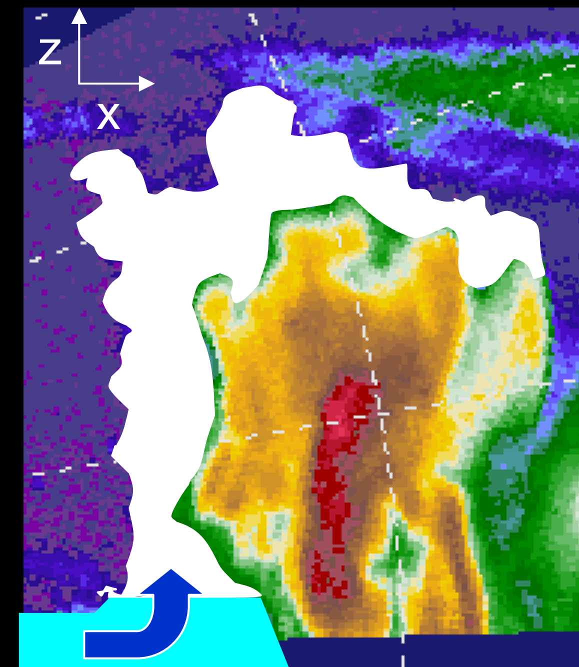



What if the model is too dry?

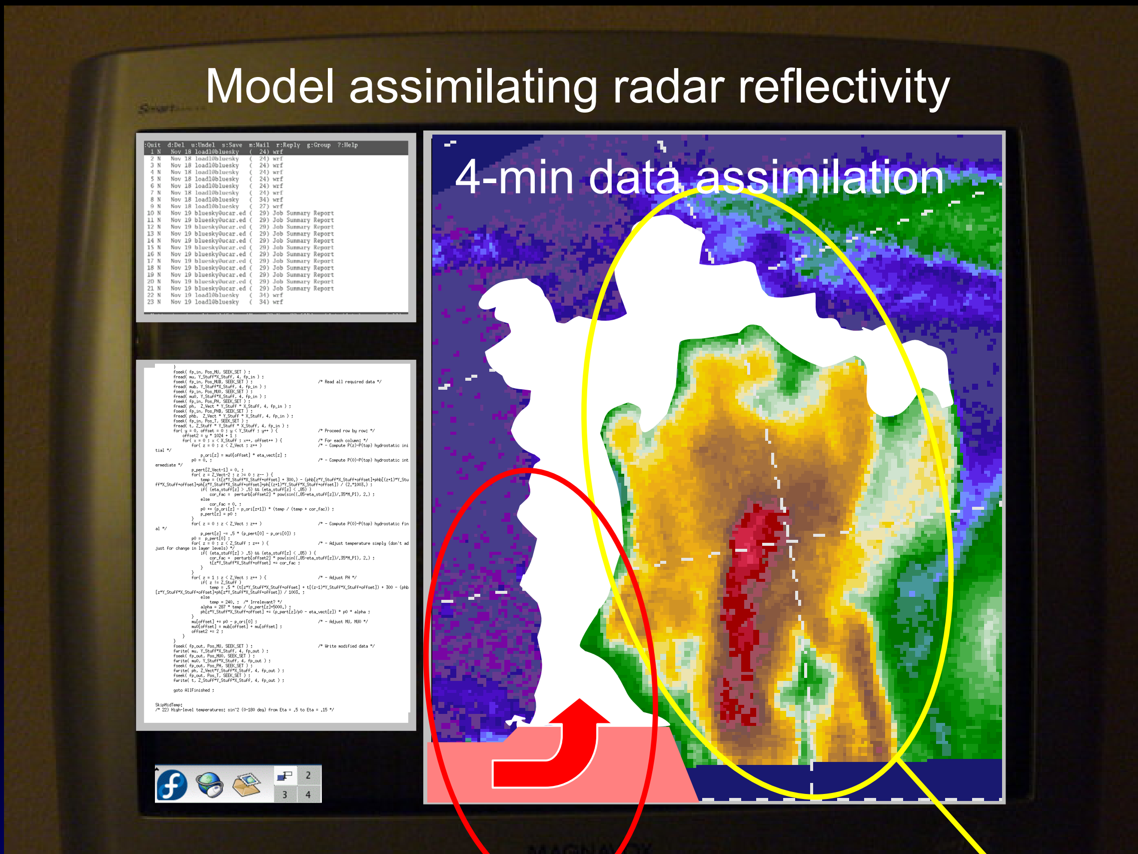
# A Pathway to Failure



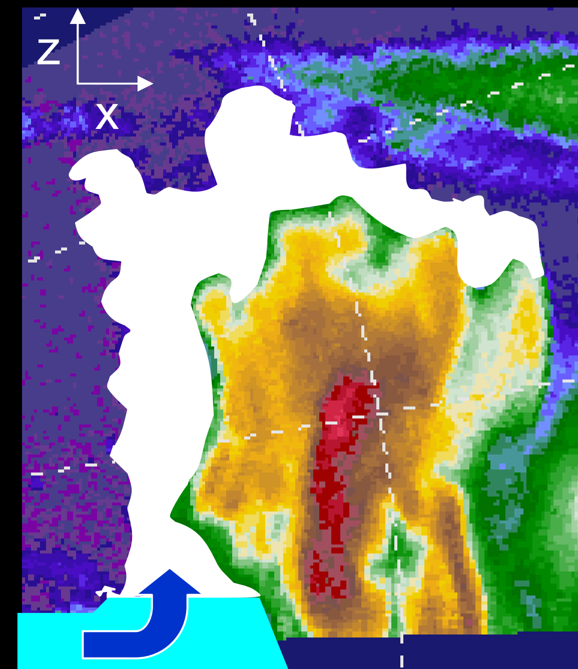

Reality



# A Pathway to Failure




Reality




Unconstrained by observations
(yet important in determining the storm's future)

Constrained by observations

In this case, only a data assimilation process exceeding 20 minutes would constrain surface moisture data.

# Arising Questions

- What type of initialization errors can be detected by observing which parameters over what duration? Focus is on the presence of a signal, and not on the specifics of its assimilation.
- What existing observing system(s) provide(s) the most information for mesoscale forecasting? Where should assimilation efforts be focused on (instrument and window duration-wise)?
- What about future observing systems?

A systematic study is needed (and doable).

# For How Long Should What Data Be Assimilated for Mesoscale Forecasting and Why?

Frédéric Fabry
McGill University, Montreal, Canada

With the help and counsel of:
Kay, Chris, Rich, Yongsheng, Wei, Michael, Morris, Joseph, Bill, David, Peggy, Don, Barry, Jenny Sun, Eunha Lim, Qingnong Xiao, and Sherrie Fredrick

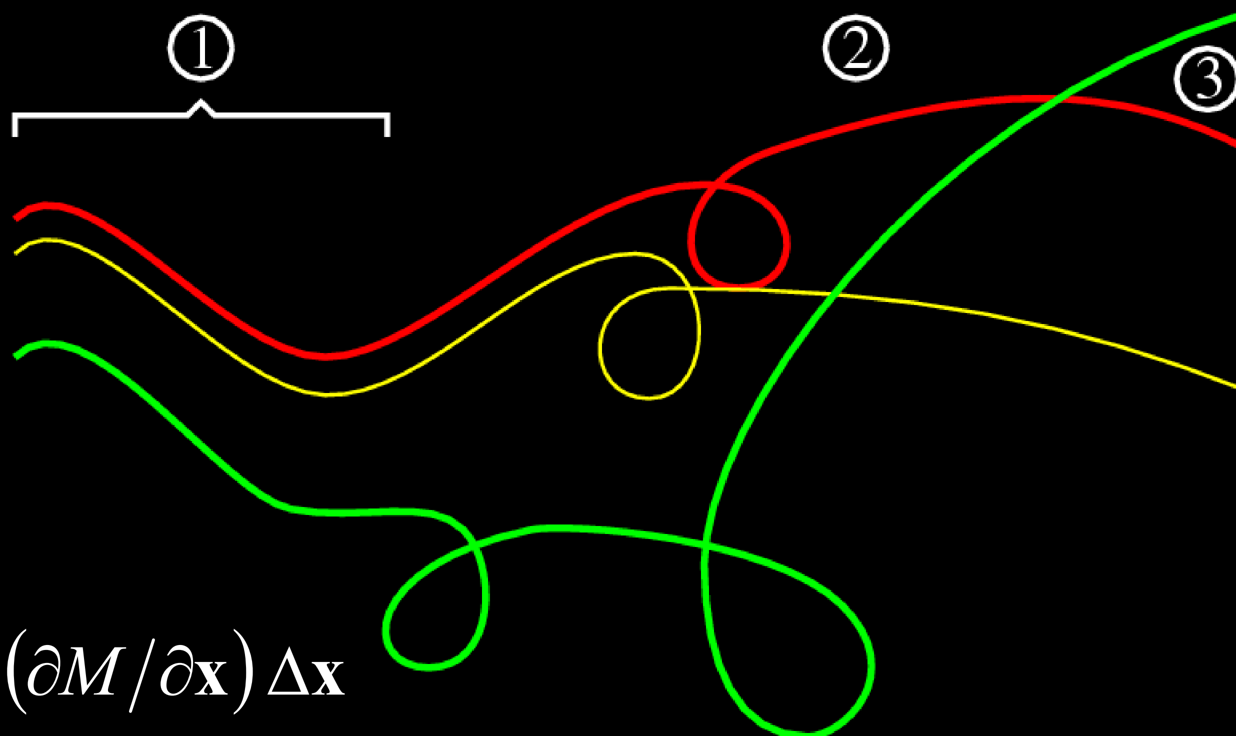
# Questions of Interest, Part I

- Issues about "*What type of initialization errors can be detected by observing which parameters over what duration?*"

  How fast do errors in one parameter move to the other parameters? How long does it take for these errors to leave a detectable signature on observations? Answer is likely scale- and parameter-dependent; Sets the minimum time required for assimilation.

  How long before assimilation becomes challenging because of limits in predictability? Answer is likely parameter-dependent; sets maximum assimilation time.

# Assimilation and Predictability




(1): Linear region:

$$M(\mathbf{x}+\Delta\mathbf{x}) \approx M(\mathbf{x}) + (\partial M/\partial\mathbf{x})\,\Delta\mathbf{x}$$

$$\rightarrow\ M(\mathbf{x}+k\Delta\mathbf{x}) - M(\mathbf{x}) \approx k[M(\mathbf{x}+\Delta\mathbf{x}) - M(\mathbf{x})]$$

(2): Nonlinear region: $M(\mathbf{x}+k\Delta\mathbf{x}) - M(\mathbf{x}) \neq k[M(\mathbf{x}+\Delta\mathbf{x}) - M(\mathbf{x})]$

(3): "Contradictory" region: Adjusting **x**(initial) towards **x**+*k*Δ**x**(initial) worsens the fit to M(**x**+Δ**x**) → Assimilation of these (perfect) data will *worsen* the fit with the truth.

# Questions of Interest, Part II

- Issues about "*What existing, planned, or unplanned observing system(s) provide(s) the most information for mesoscale forecasting?*"

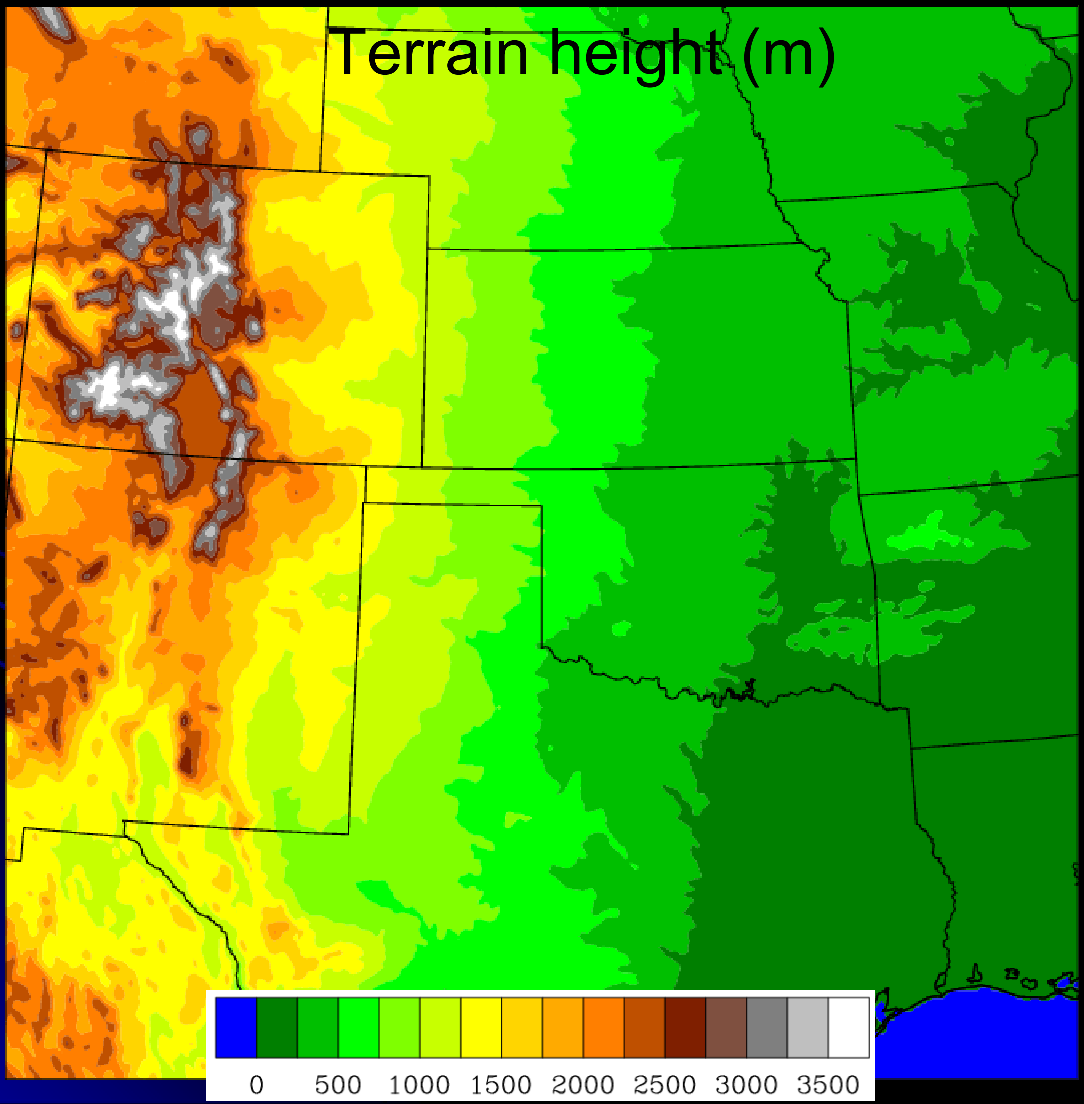
  How much do typical errors in each parameter affect forecast quality? Could significant gains in forecast accuracy be made by focusing on a few parameters? Of interest for assimilation and for instrument design.

  Which instruments or technologies provide the most information? Answer depends on variable(s) targeted by instrument because of all the aforementioned issues, data coverage, measurement accuracy, and strength of link between the quantity observed and the variables.

# Overall Approach: Identical Twins Experiment

- "*Truth*": Simulate a series of plausible convective events (12-hr long control runs);
- *Error growth experiment*: Perturb initial conditions by a plausible "error"), and run a "forecast";
- Check the magnitude of the resulting forecast error as a function of the type of errors in the initial conditions;
- Study the properties of the transfer of errors from one parameter to the next; focus on magnitude, predictability;
- Evaluate the ability of different sensors of detecting early signs of the forecast going astray.

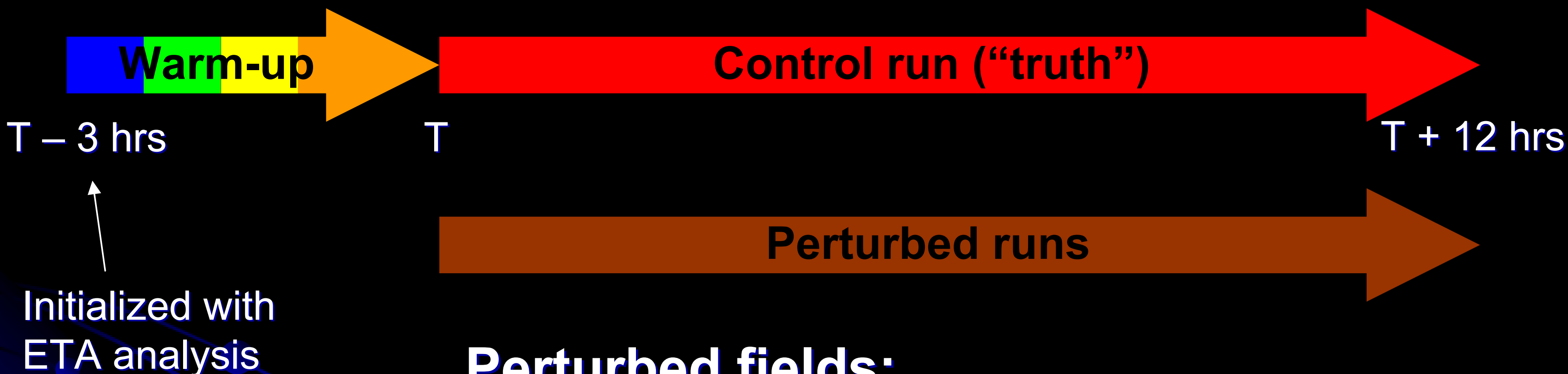
# In Detail:
# 1) Model and Domain Used




- WRF v. ß2.2
- 1600 x 1600 km domain
- 4 km resolution, 28 levels
- Thompson et al. microphysics, RRTM & Dudhia radiation, Noah land surface, YSU PBL…

# In Detail:
# 2) The Model Runs

Sixteen 12-hr runs; one every 9 hrs for a 6-day period (10-16 June 2002)

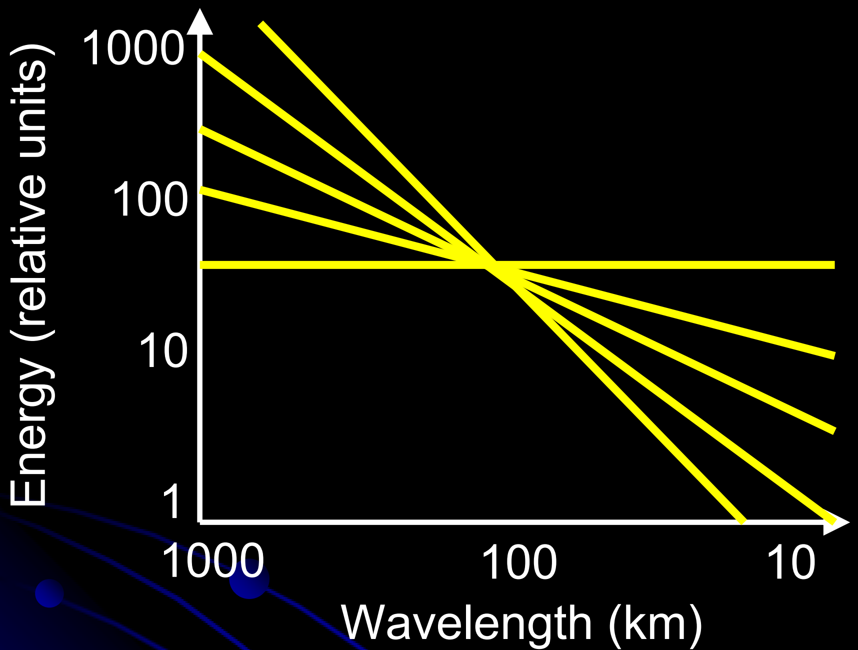


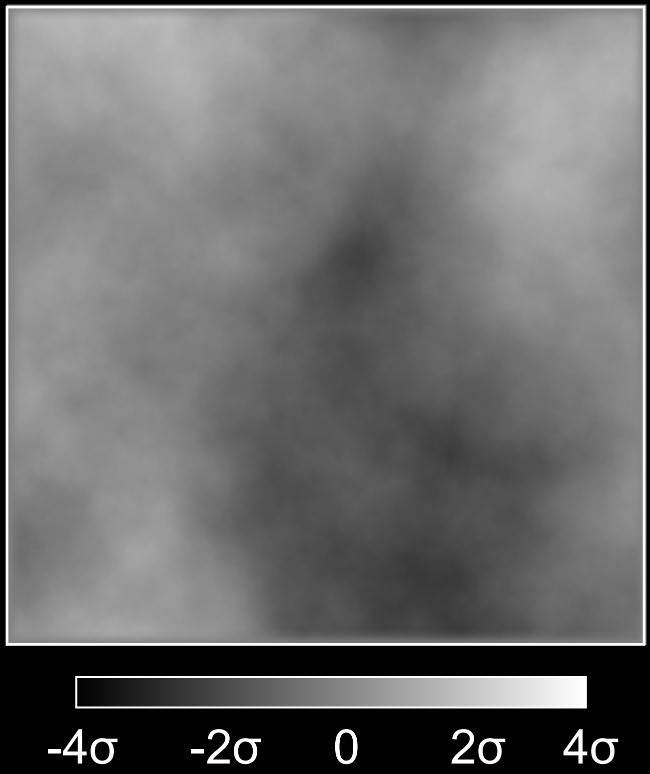

**Perturbed fields:**

- Low-level
- Mid-level
- Hi-level
- Condensates
- Soil moisture

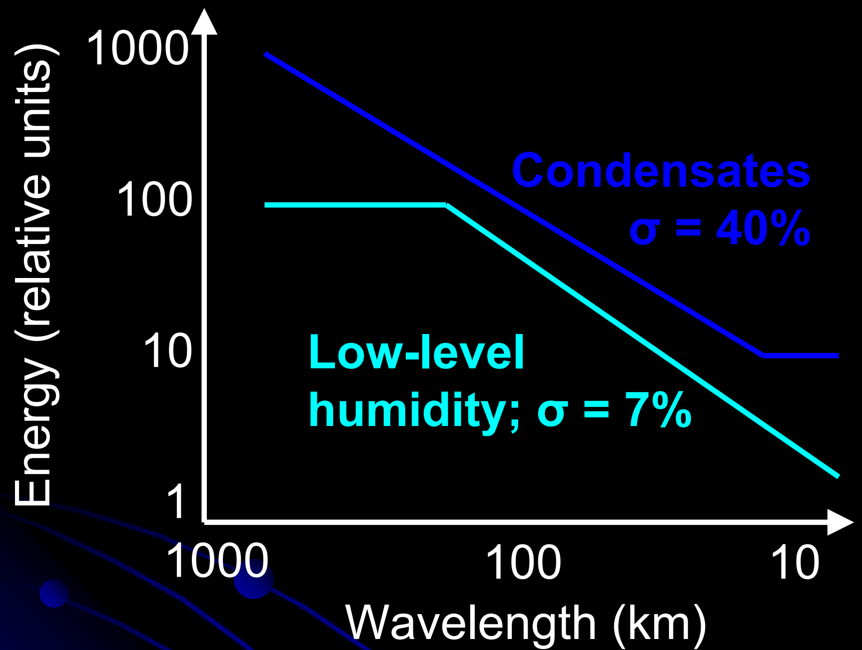
Humidity
Temperature
Winds

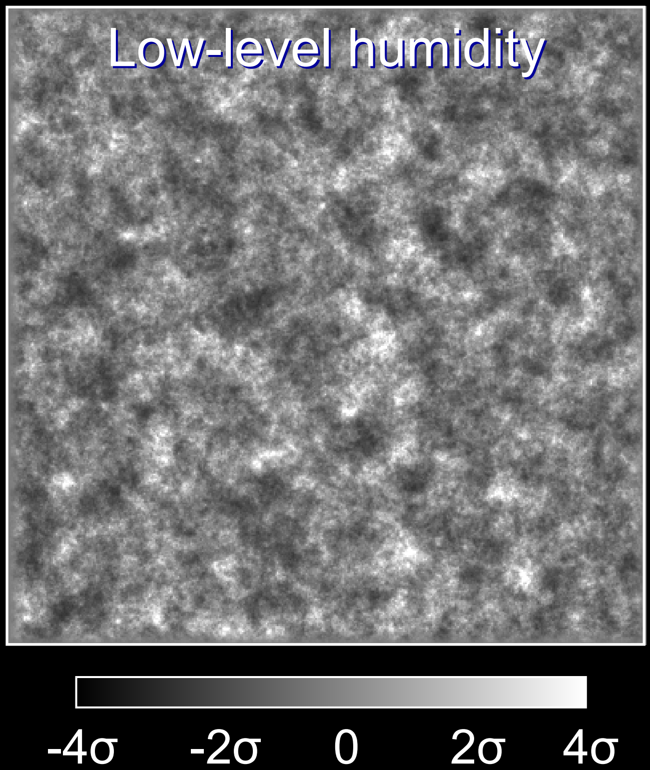
# In Detail:
# 3) The Perturbations

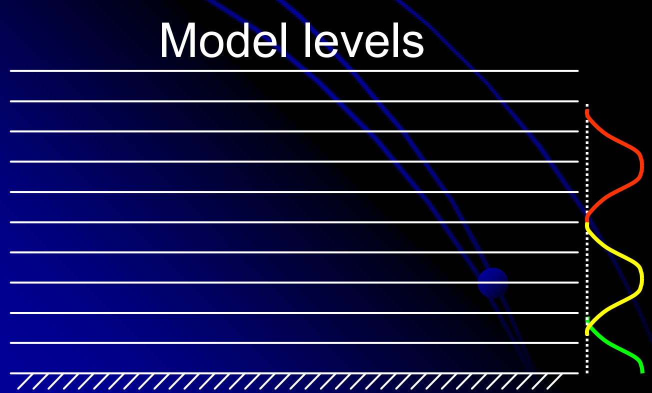


# In Detail:
# 3) The Perturbations



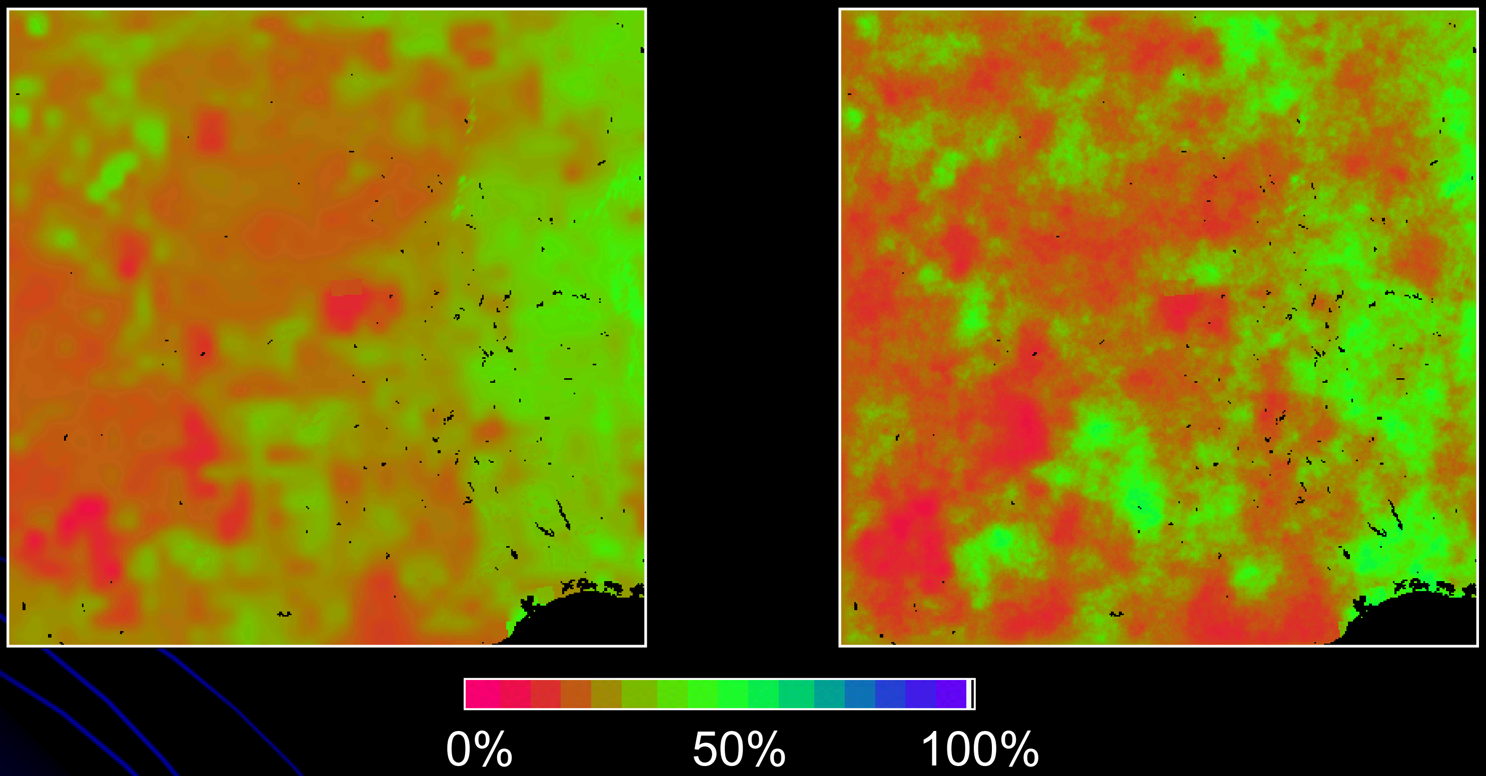




High-level perturbations: from η = .15 to η = .5

Mid-level perturbations: from η = .5 to η = .85

Low-level perturbations: from η = .825 to η = 1

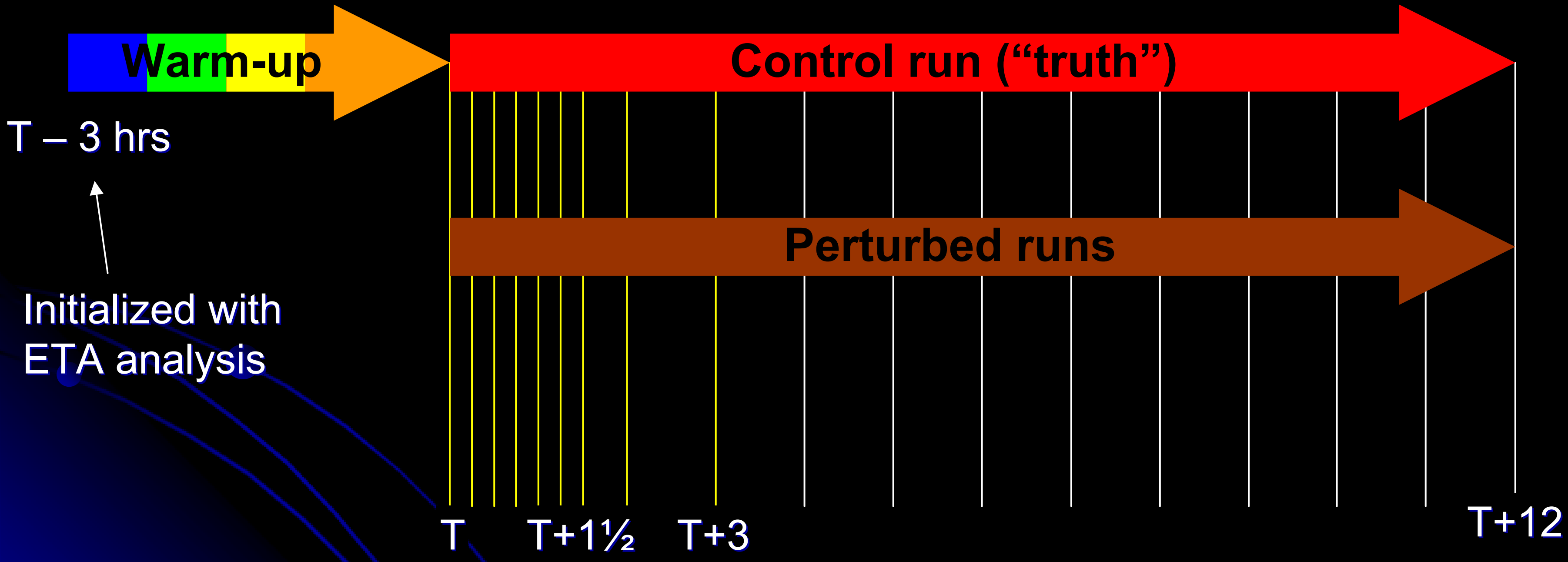
# In Detail:
# 3) The Perturbations




Soil moisture before (left) and after (right) ±25% perturbation

# In Detail:
# 4) The Data Obtained

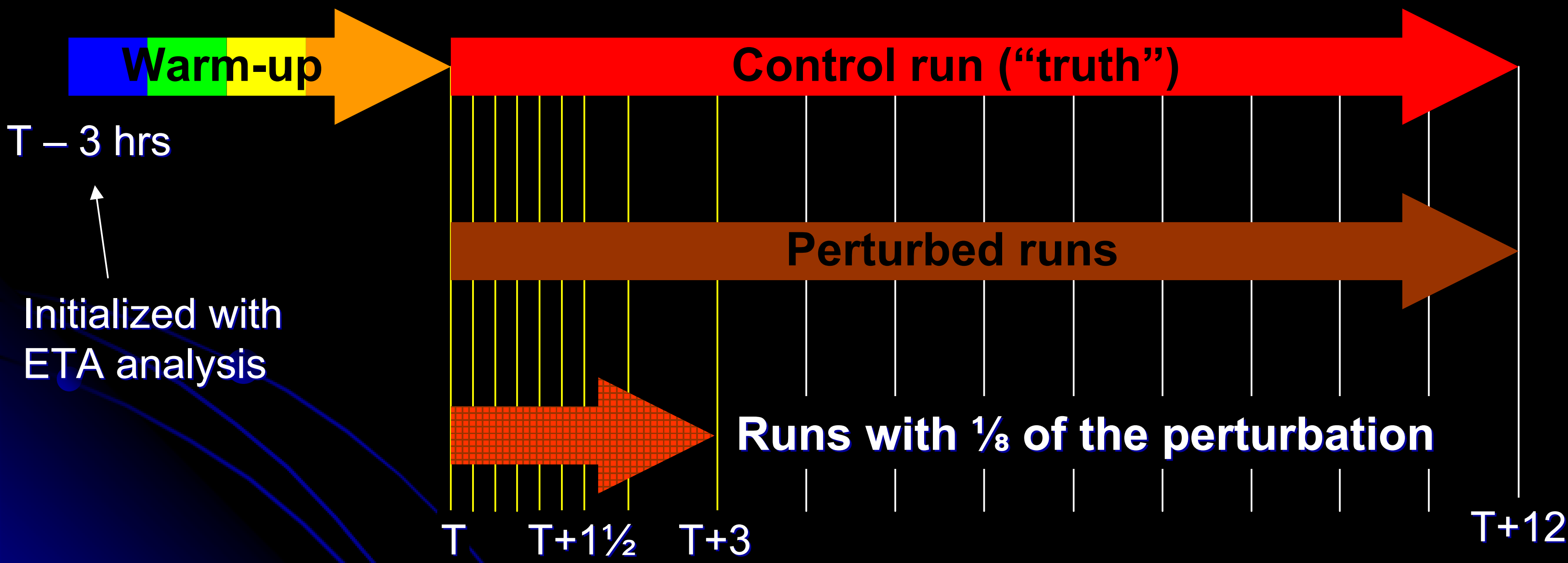
Sixteen 12-hr runs; one every 9 hrs for a 6-day period (10-16 June 2002)




- Model outputs every 15 min up to T+1½; more model outputs on the hour up to T+12. "Observations" are evaluated up to T+3.

# In Detail:
# 4) The Data Obtained

Sixteen 12-hr runs; one every 9 hrs for a 6-day period (10-16 June 2002)

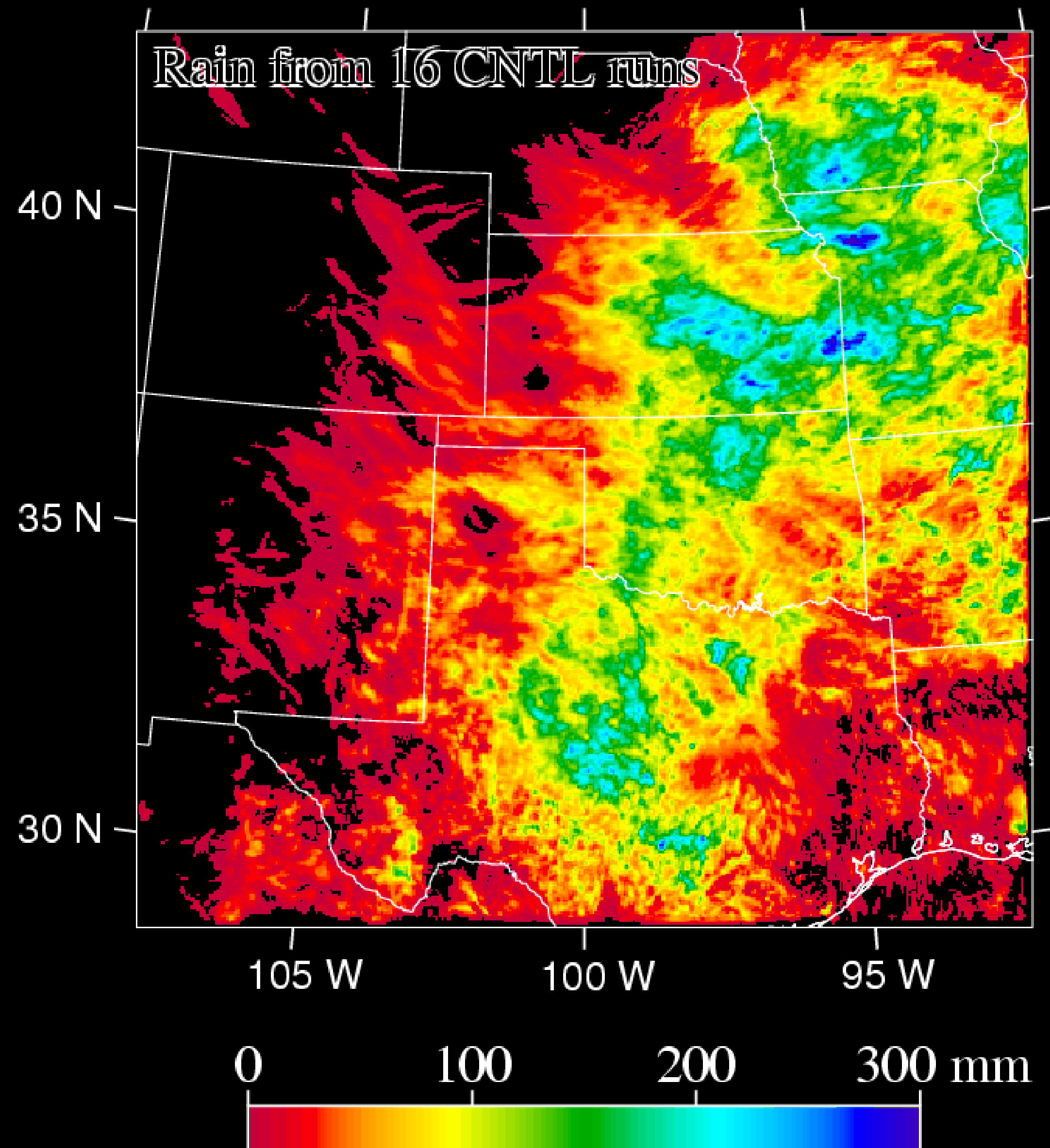



• Partially perturbed run up to T+3 allows us to test for the linearity of the forecast errors and of the change in the measurements.

# The Simulated Weather

Right: Rainfall accumulation over the 16 control runs.

Despite the limited period, much weather occurred with different types of forcing mechanisms. Hopefully the results will be representative.



# Forecast Errors: Analysis Approach

How do we compare errors in winds, temperature, humidity, and precipitation?

An approach: Use as inspiration Talagrand's energy difference of perturbations (ΔE) per unit mass:

$$\Delta E = \tfrac{1}{2} \int [\Delta u^2 + \Delta v^2 + c_p/T_{ref}\, \Delta T^2 + R T_{ref}\, (\Delta p/p_{ref})^2] dV$$

Kinetic E    Thermal E    Pressure E

Vapor: $LH = L\, \Delta r_v = c_p\, (L \Delta r_v / c_p) = c_p \Delta T_{latent}$

$$\rightarrow \Delta E_v = c_p/T_{ref}\, \Delta T_{latent}^2 = L^2/(c_p T_{ref})\, \Delta r_v^2$$

Condensates: $\Delta E_c = \Delta PE + \Delta KE \approx \Delta PE = \sum gh\, \Delta r_{r,s,g}$

# Forecast Errors: Analysis Approach

How do we compare errors in winds, temperature, humidity, and precipitation? Energy differences:

[wind] KED = $\frac{1}{2}\left(\overline{\Delta u^2} + \overline{\Delta v^2}\right)$

+ [temperature] TED = $\frac{1}{2} \sum c_p/T_{ref}\, \overline{\Delta T^2}$

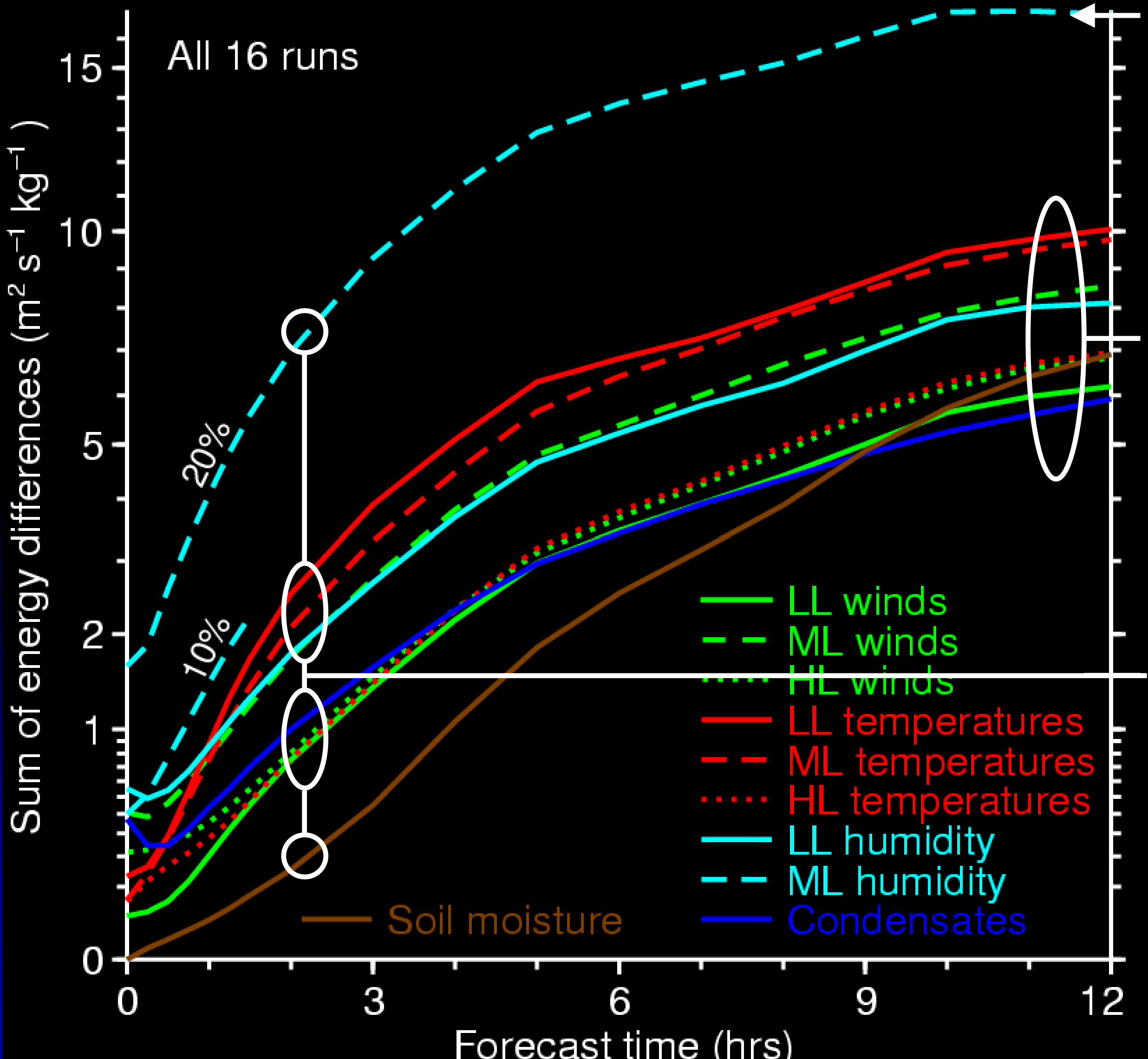
+ ~~[pressure] PED = $\frac{1}{2} \sum R T_{ref}\, (\overline{\Delta p}/p_{ref})^2$~~

+ [vapor] LED = $L^2/(c_p T_{ref})\, \overline{\Delta r_v^{\,2}}$

+ [condensates] CED = $|gh\, \overline{\Delta r_{r,s,g}}|$

= [sum] SED = KED + TED + LED + CED

# Forecast Errors: Main Results



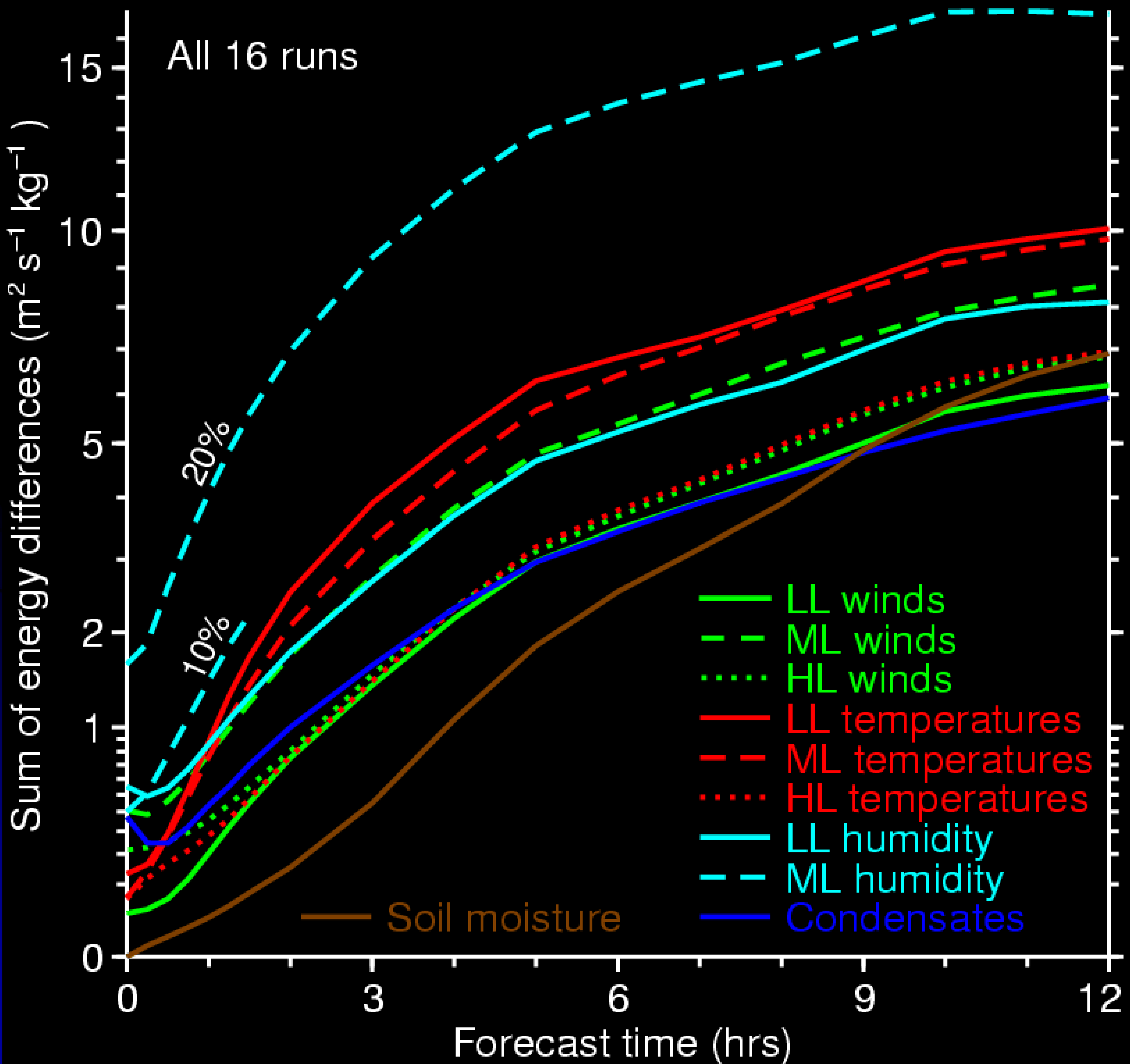

Mid-level humidity uncertainty dominates

For long mesoscale forecasts: all other errors are comparable

For short mesoscale forecasts: 4 groups of different magnitudes

→ Assimilation priorities should differ for different forecasts

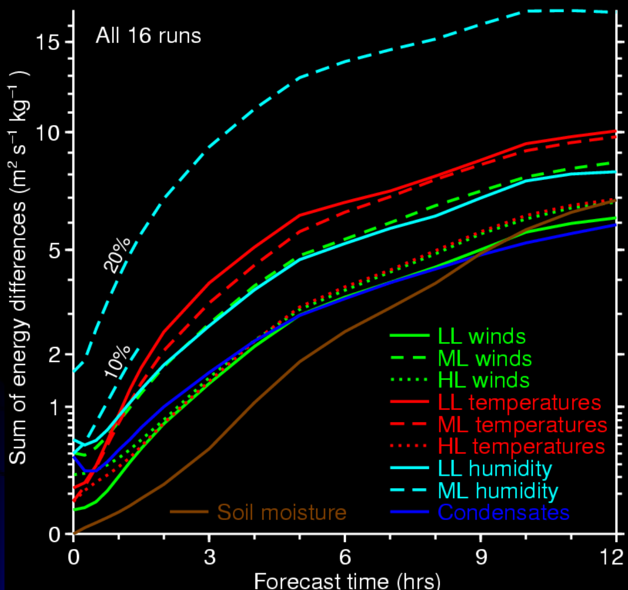
# Forecast Errors: Other Results

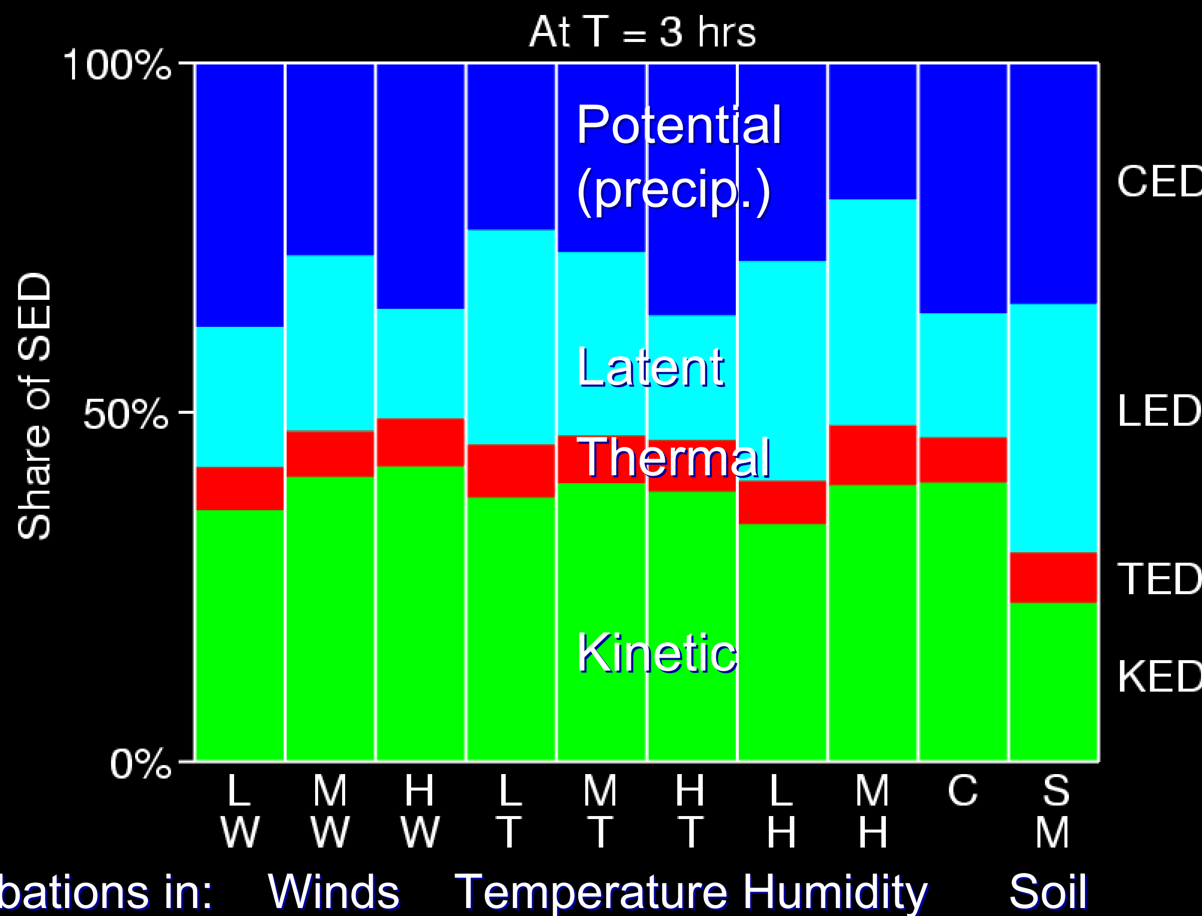



- Day vs. night runs: some changes in the overall magnitude (daytime differences are larger), but minor changes in the ranking between perturbations;

- Significant run-to-run differences, with forecast error growth being correlated with mean rainfall rate.

# Forecast Errors: Results

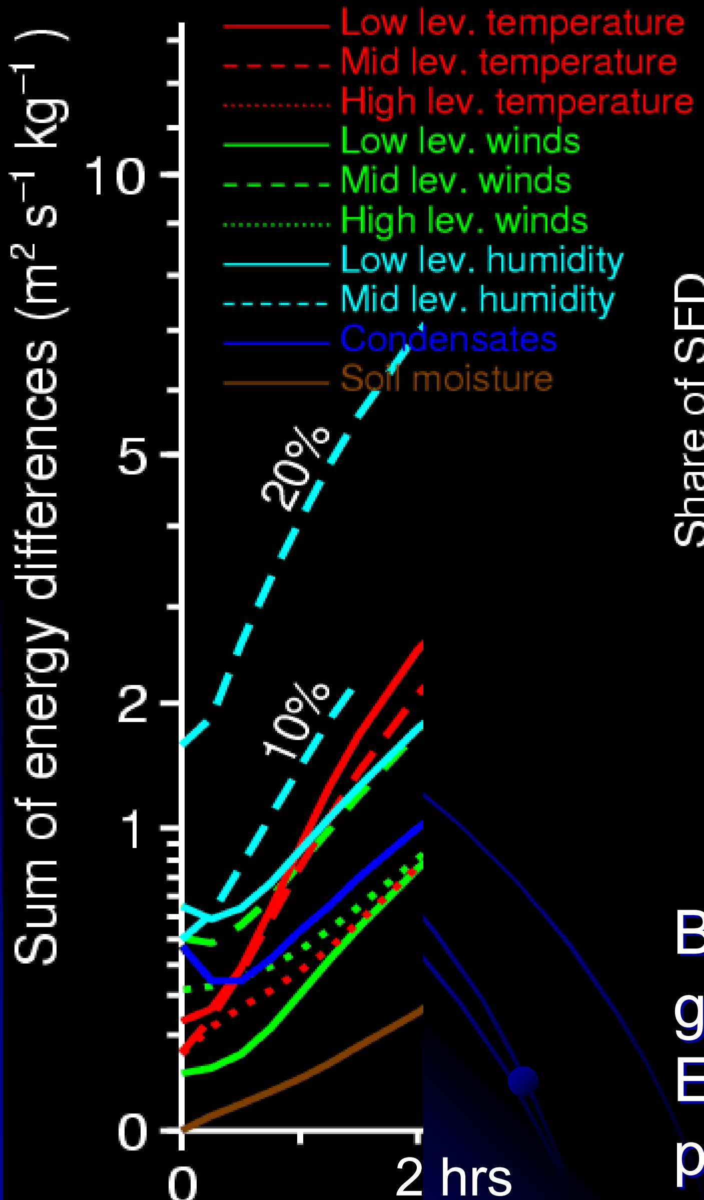


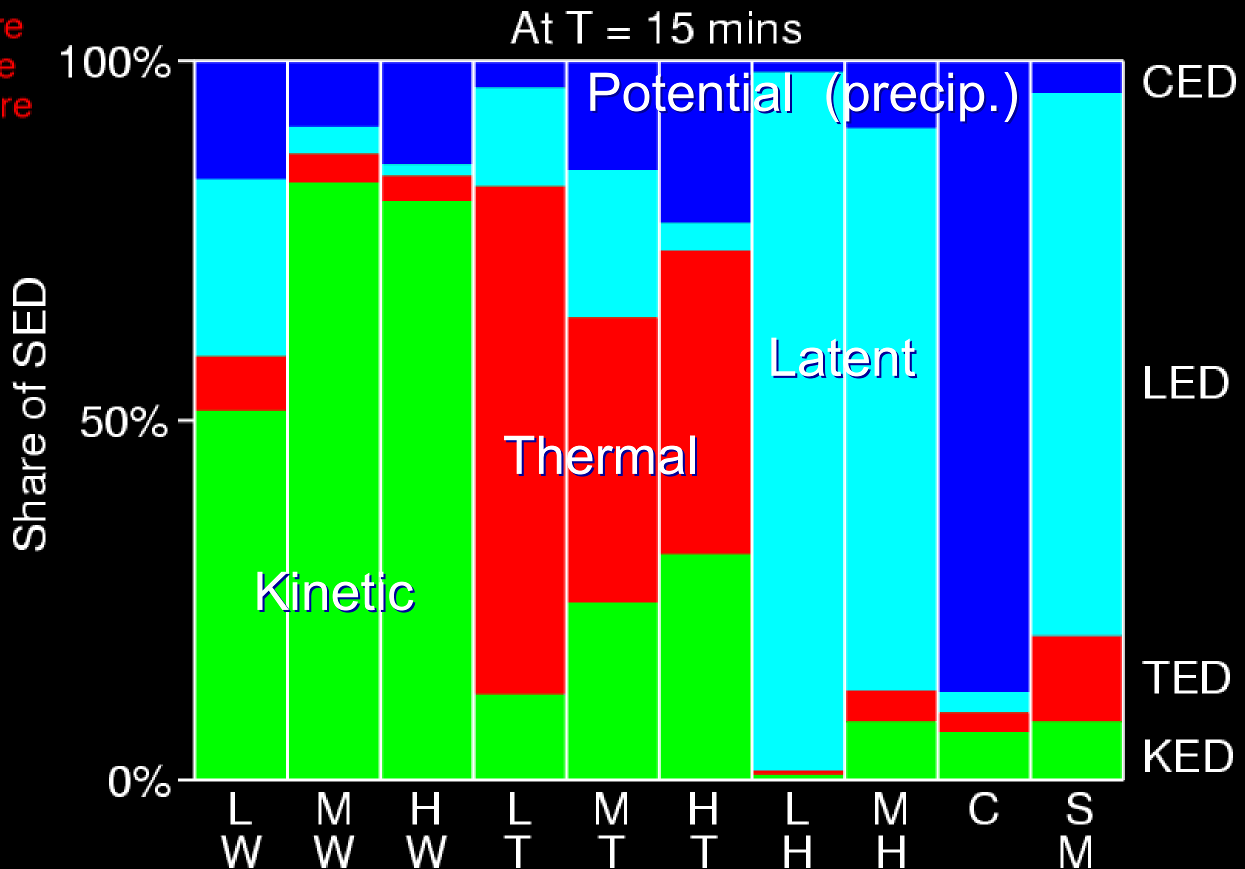




The origin of the perturbations is mostly forgotten by T = 3 hours: Each term is a near-constant fraction of SED.

→ Other measures of forecast mismatch would give similar results.

# Forecast Errors: Results



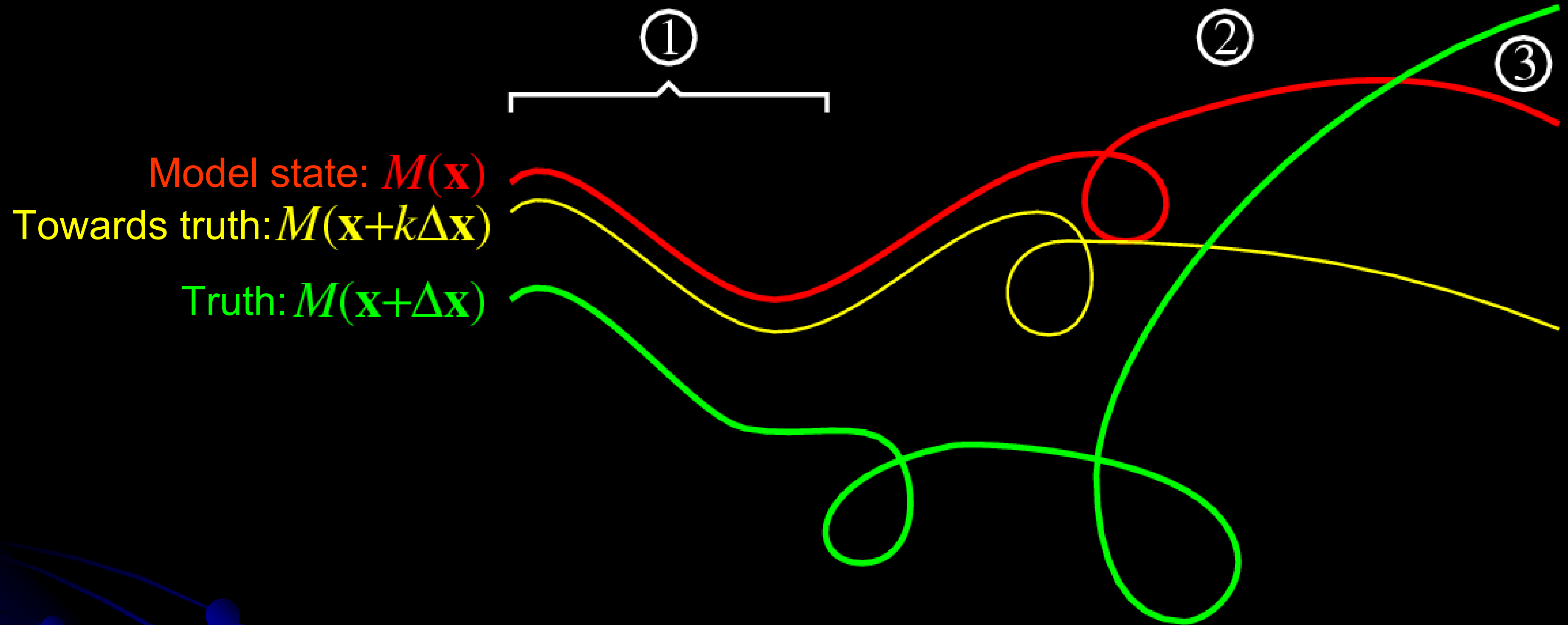

By T = 15 min, perturbation redistribution is generally well underway → Easier detection. Exceptions: low-level humidity, condensates perturbations that undergo slow starts.

# Forecast Errors: Partial Conclusions

- Mid-level (centered on 700 mb) moisture uncertainty seems to have the largest impact on forecasts; other uncertainties have lesser impacts, but these vary with forecast timescale.
- Most perturbations, especially the ones growing fast, get transferred rapidly to other variables.

→ If the assimilation is long enough, most types of instruments have the opportunity to get a signal resulting from the perturbations.
But how useful a signal?

# Linearity of Perturbations




Many data assimilation systems work best if the perturbations to the model state are linear:

$$M(\mathbf{x}_o + \Delta\mathbf{x}_o) \approx M(\mathbf{x}_o) + (\partial M / \partial \mathbf{x})_{\mathbf{x}_o} \Delta\mathbf{x}_o$$

Or: $M(\mathbf{x}_o + k\Delta\mathbf{x}_o) - M(\mathbf{x}_o) \approx k[M(\mathbf{x}_o + \Delta\mathbf{x}_o) - M(\mathbf{x}_o)]$

# Linearity of Perturbations

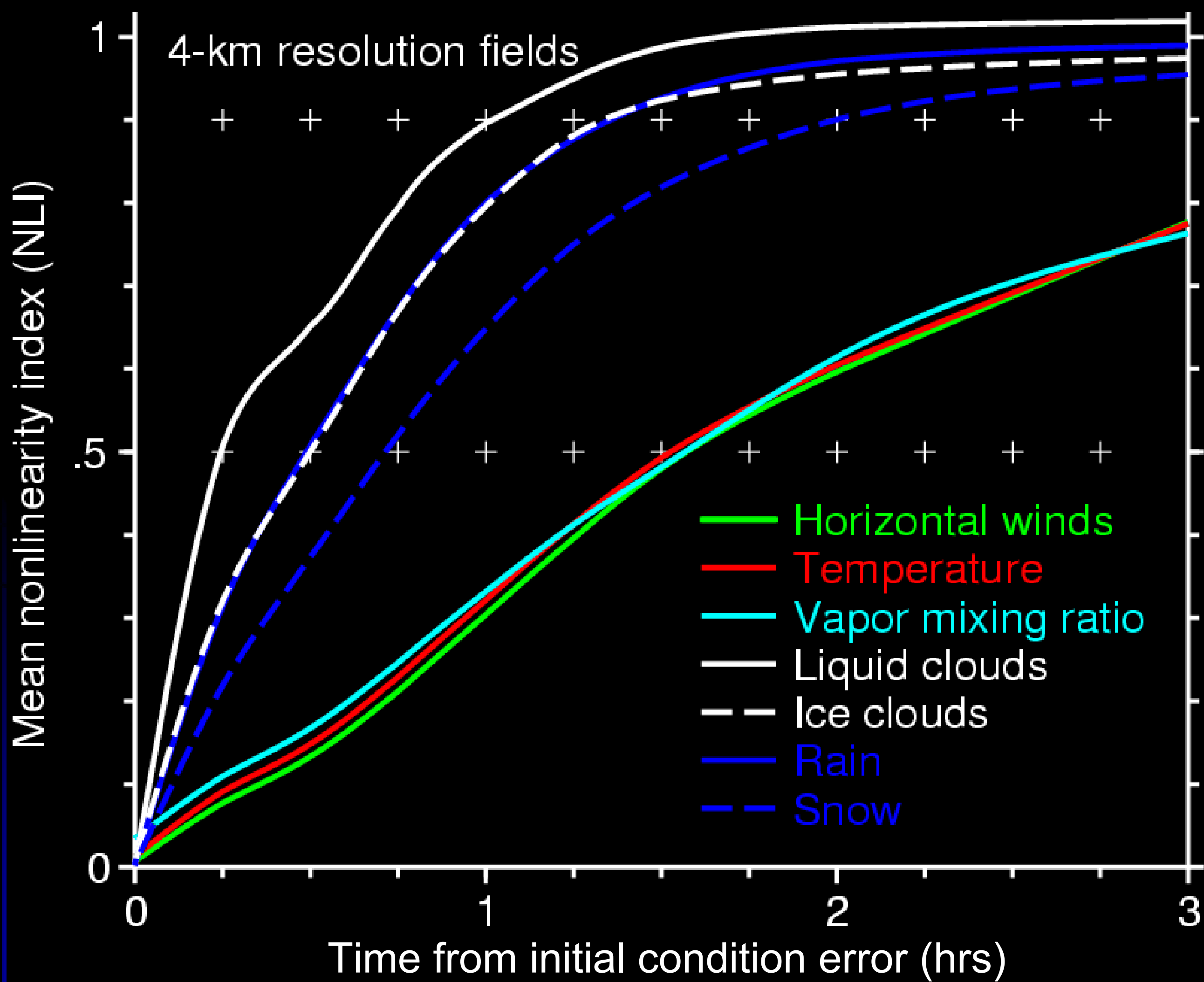
Linear perturb.: $M\left(\mathbf{x}_0 + \Delta\mathbf{x}\right) \approx M\left(\mathbf{x}_0\right) + \left(\frac{\partial M}{\partial \mathbf{x}}\right)_{\mathbf{x}_0} \Delta\mathbf{x}$

Or: $M\left(\mathbf{x}_0 + k\Delta\mathbf{x}\right) - M\left(\mathbf{x}_0\right) \approx k\left[M\left(\mathbf{x}_0 + \Delta\mathbf{x}\right) - M\left(\mathbf{x}_0\right)\right]$

We can use this property to define a nonlinearity index NLI for each variable **V**:

$$NLI\left(\mathbf{V}, \Delta\mathbf{x}, t\right) = \frac{\sum_{x,y,z}\left|\left[\mathbf{V}\left(\mathbf{x}_0 + k\Delta\mathbf{x}\right) - \mathbf{V}\left(\mathbf{x}_0\right)\right] - k\left[\mathbf{V}\left(\mathbf{x}_0 + \Delta\mathbf{x}\right) - \mathbf{V}\left(\mathbf{x}_0\right)\right]\right|}{\sum_{x,y,z}\left|\mathbf{V}\left(\mathbf{x}_0 + k\Delta\mathbf{x}\right) - \mathbf{V}\left(\mathbf{x}_0\right)\right|}$$

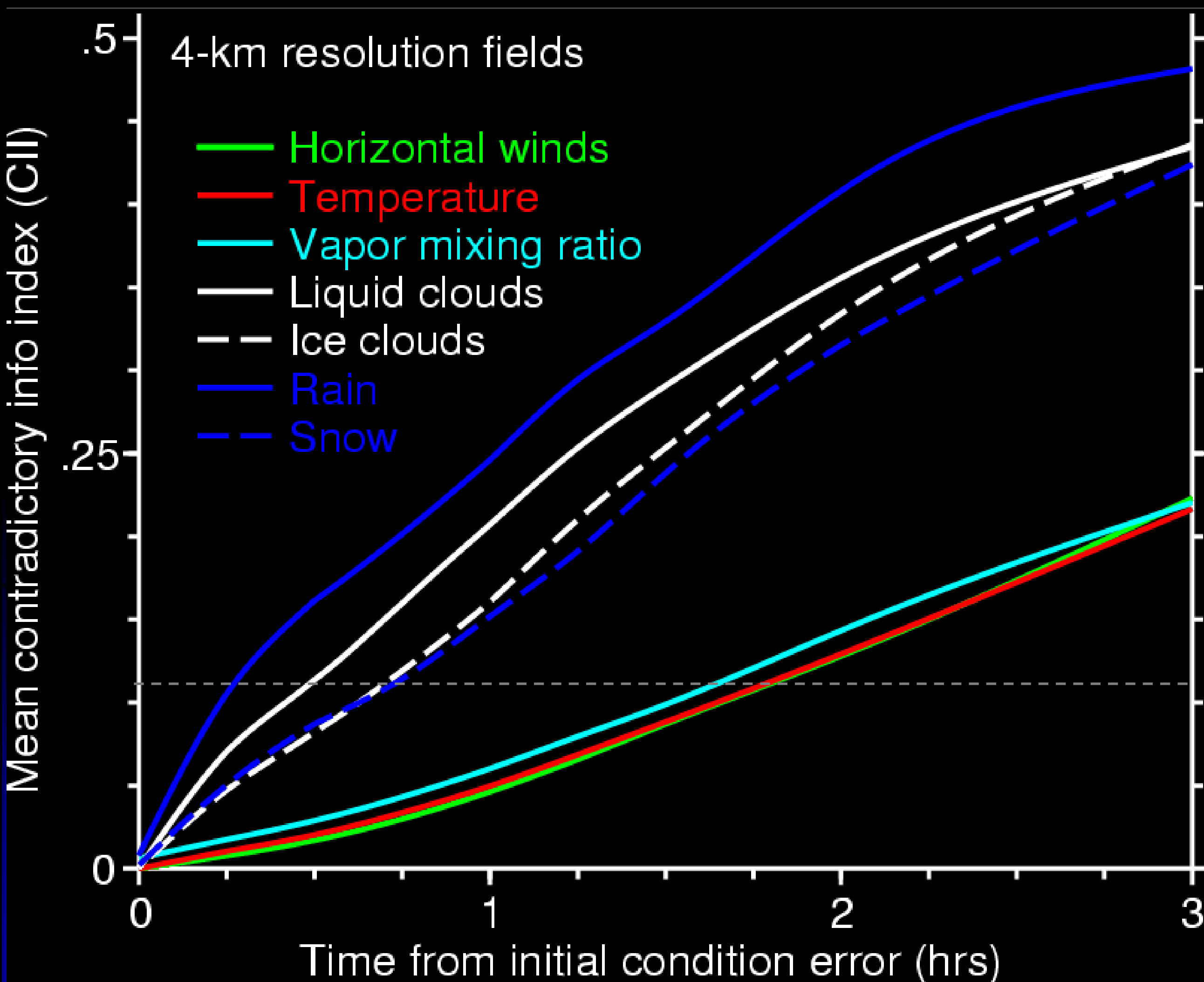
# Linearity of Perturbations: Results




Nonlinearities express themselves more in some variables than others:

- Winds, humidity, temperature evolves “more” linearly;
- Precipitation is more nonlinear: liquid is worse than ice, and (surprise) clouds seem worse than precipitation.

# Contradictory Information: Results



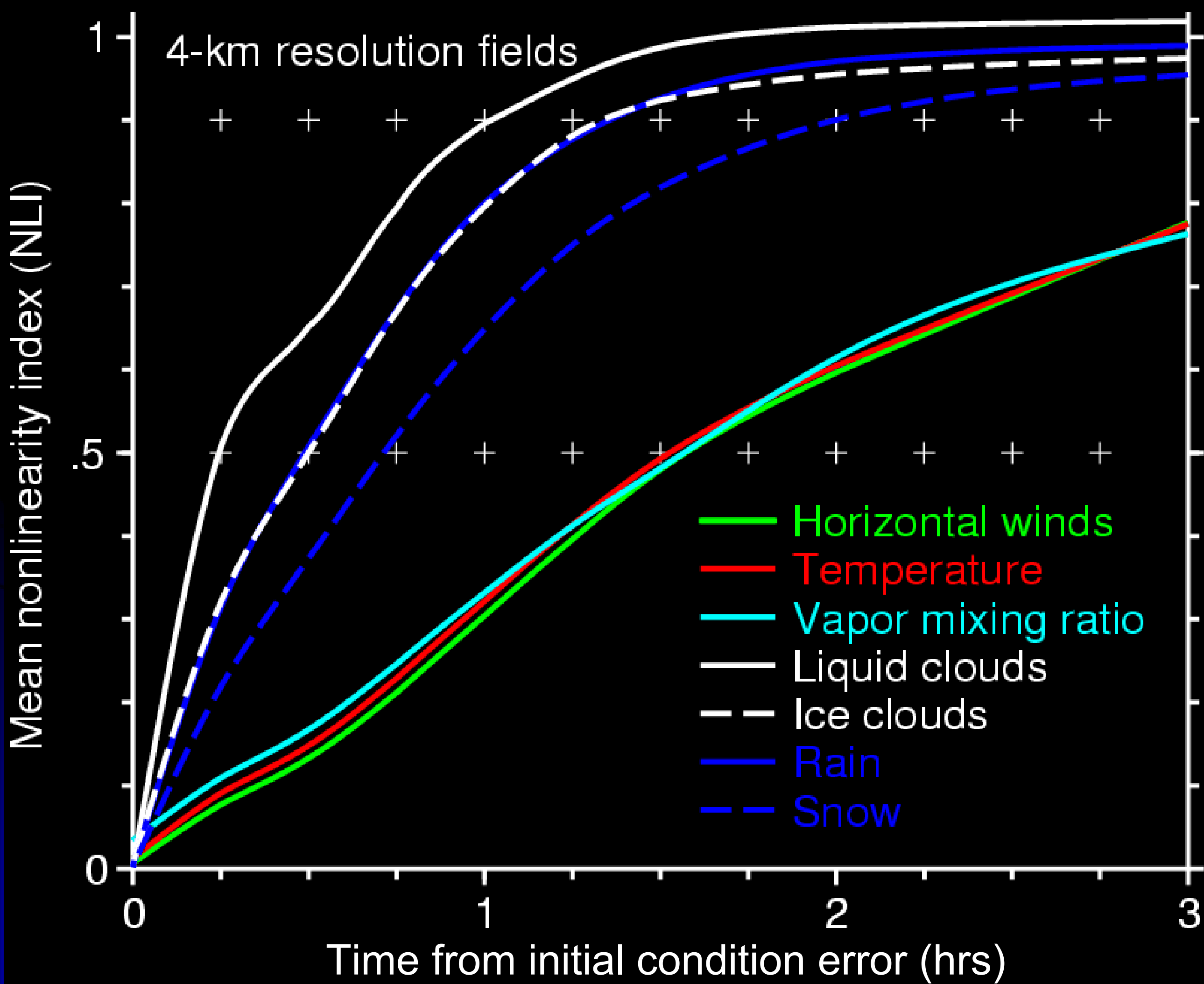

CII: Contradictory information index

$$= \frac{\#\ \text{contradicting pts}}{\#\ \text{non-contradicting}}$$

Results are similar to nonlinear index; (but here rain is worse than clouds)

Note: At CII = .11, one (perfect) data point in ten is worsening the initial state of the model.

# Linearity of Perturbations: Results

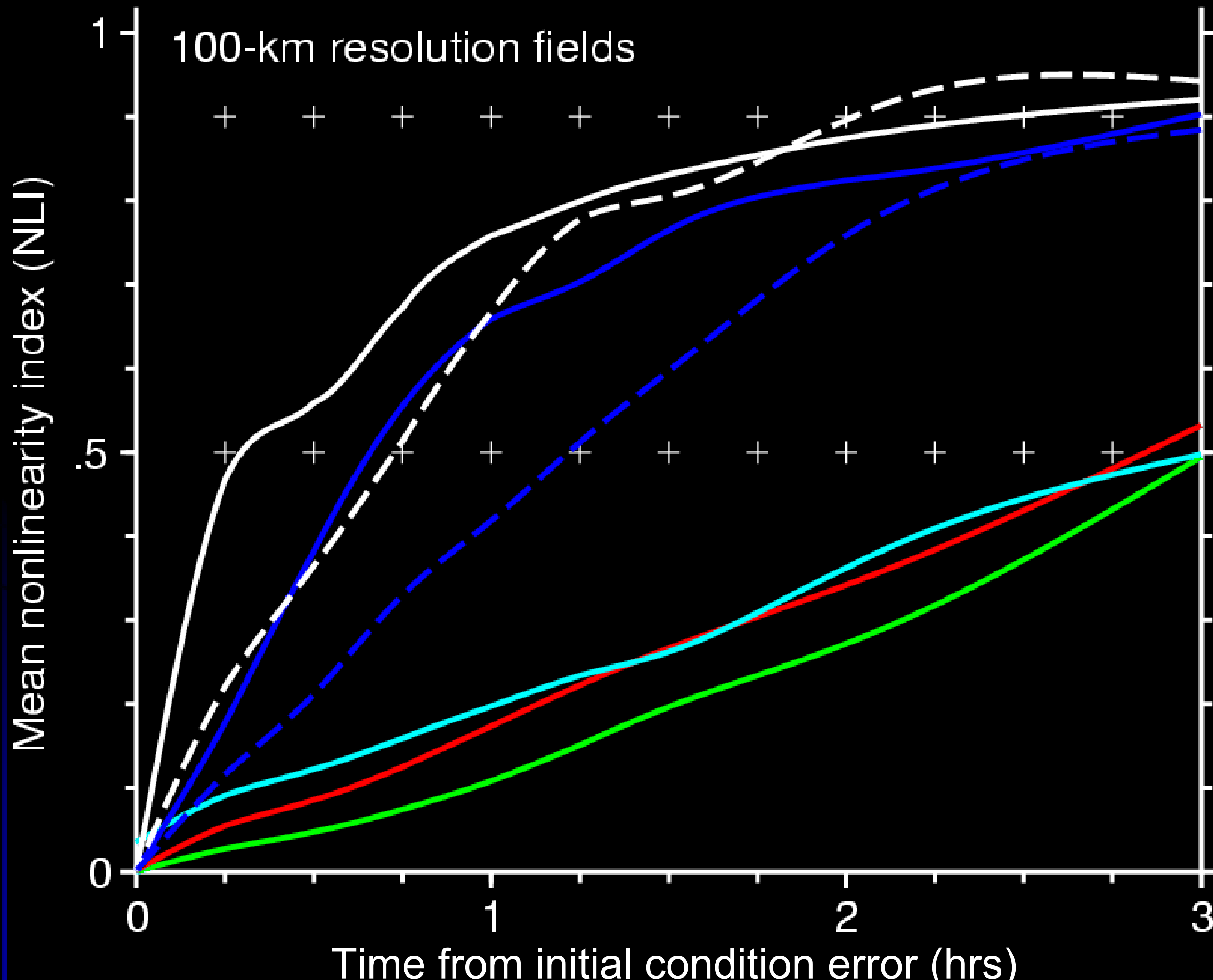



If assimilation is possible only when NLI<0.5 (threshold to be determined), one can assimilate at 4-km resolution:

- 90 minutes of winds, temperature and humidity data;
- 30-45 minutes of precipitation data;
- 15-30 minutes of cloud data.

Note on nonlinearities and assimilation: both an assimilation and a modeling challenge

# Linearity of Perturbations: Results




Nonlinearities are a function of scale → Smoothed fields should be more linear than higher resolution fields.

Effect of smoothing:

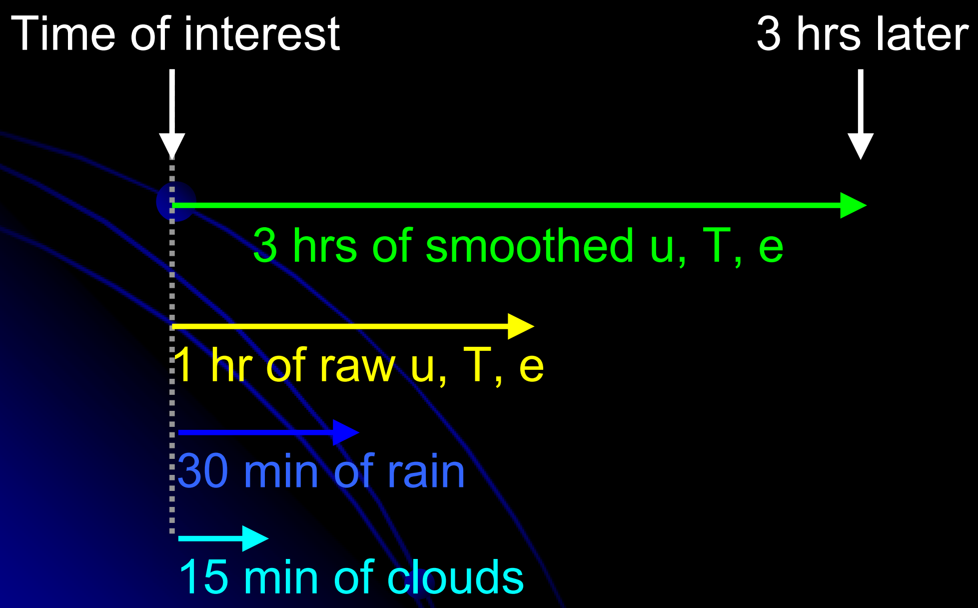
- Significant (+100%) predictability gains for winds, humidity, and temperature;
- Some gains for precipitation; limited gains for clouds

# Assimilation and Predictability: Partial Conclusions

- Clouds and precipitation patterns evolve highly nonlinearly. → Using them to retrieve larger-scale patterns of other variables will be challenging; sadly, our best mesoscale sensors target these variables!
- Ideal assimilation strategy at the mesoscale given perfect measurements of model parameters:
  1) smoothed **u**, *T*, *e* for a few hours; 2) higher-resolution **u**, *T*, *e* for an hour; 3) *R* for 30 min; *LWC* for 15 min. Smoothed (**u** or *T*) and *e* may be just sufficient to constrain the largest sub-synoptic patterns;
  [Instrument-related issues still have to be considered]

# Assimilation and Predictability: Partial Conclusions

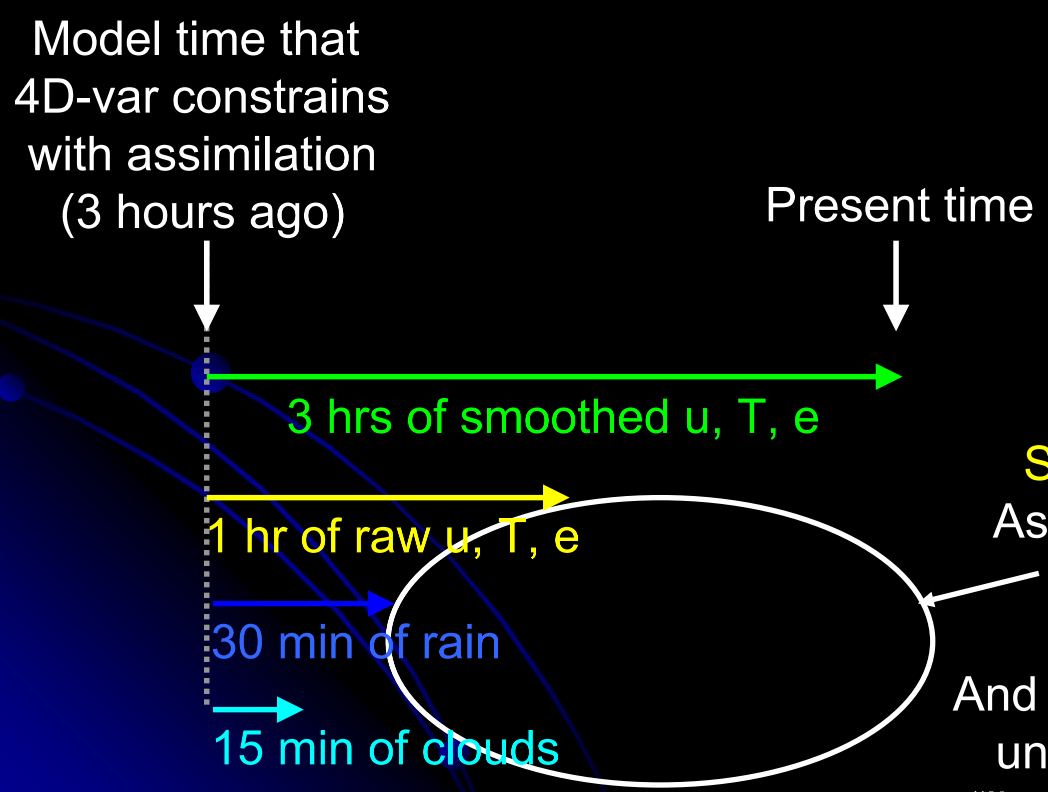
- Assimilating variables with different predictability over different periods may require more than plain 4D-Var.




Scenario 1: Case reanalysis
No problems here.

# Assimilation and Predictability: Partial Conclusions

- Assimilating variables with different predictability over different periods may require more than plain 4D-Var.




Scenario 2a: Real-time processing

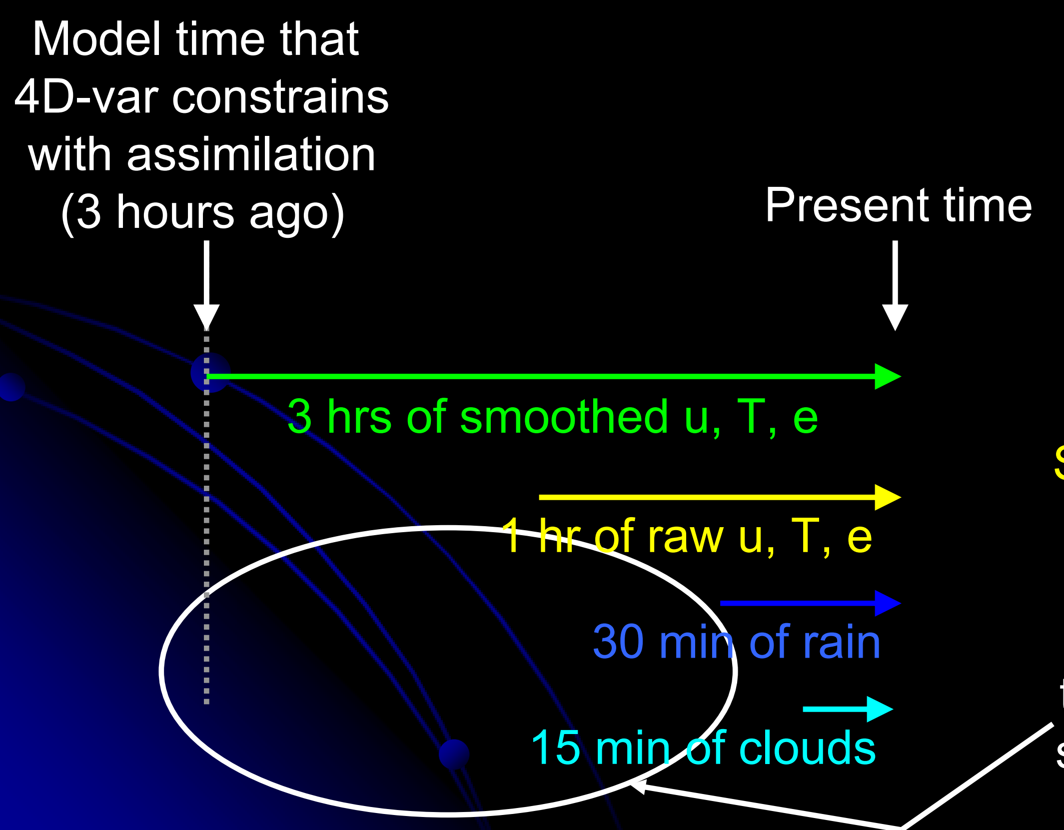
Assimilation does not take advantage of the latest data available.

→ Far from ideal

And assimilating 3 hrs of everything adds unusable non-linear “noise” to usable “linear” data and hurts more than helps.

# Assimilation and Predictability: Partial Conclusions

- Assimilating variables with different predictability over different periods may require more than plain 4D-Var.

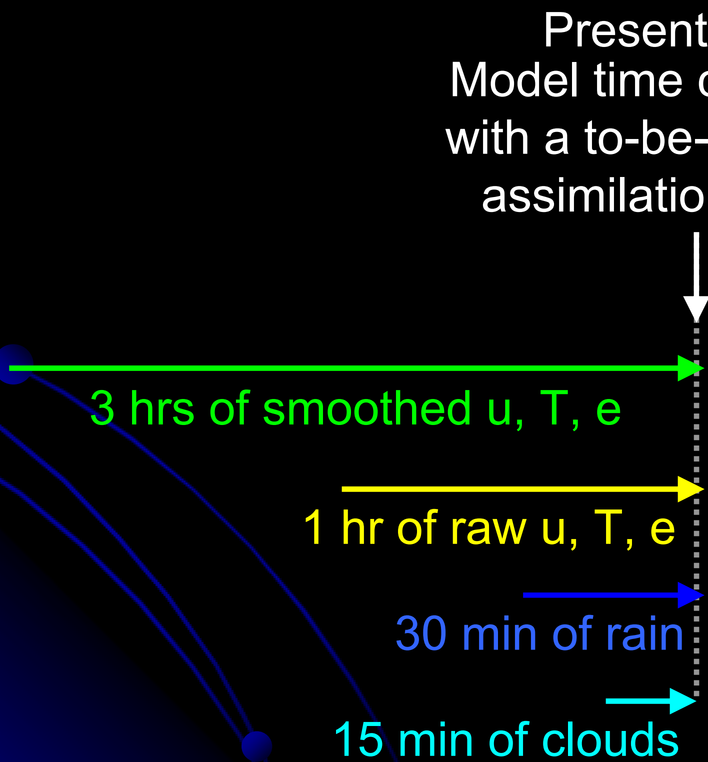



Scenario 2b: Real-time processing

Assimilation of most fields will fail because the data is several "predictability times" away from the time for which the assimilation system is trying to adjust the initial conditions of.

# Assimilation and Predictability: Partial Conclusions

- Assimilating variables with different predictability over different periods may require more than plain 4D-Var.




Scenario 2c: Real-time processing
Ideal solution, but it requires deriving an approach that constrains the <u>end time</u> of the assimilation window, not the beginning time as traditional 4D-var does.

Possible? Or should multiple sliding assimilation windows be used?

# Evaluating Observations

Q: How much signal of the initial perturbation is there in observations?

It depends on how much an observation **y** changes compared with the uncertainty σ(**y**) in that observation and on the number of observations made per unit time per instrument.

Signal strength:

$$S_{dataset}\left(\Delta\mathbf{x},T\right)=\sum_{i}\frac{\left[y_i\left(\mathbf{x}+\Delta\mathbf{x},T\right)-y_i\left(\mathbf{x},T\right)\right]^2}{\sigma\left(y_i\right)^2}$$

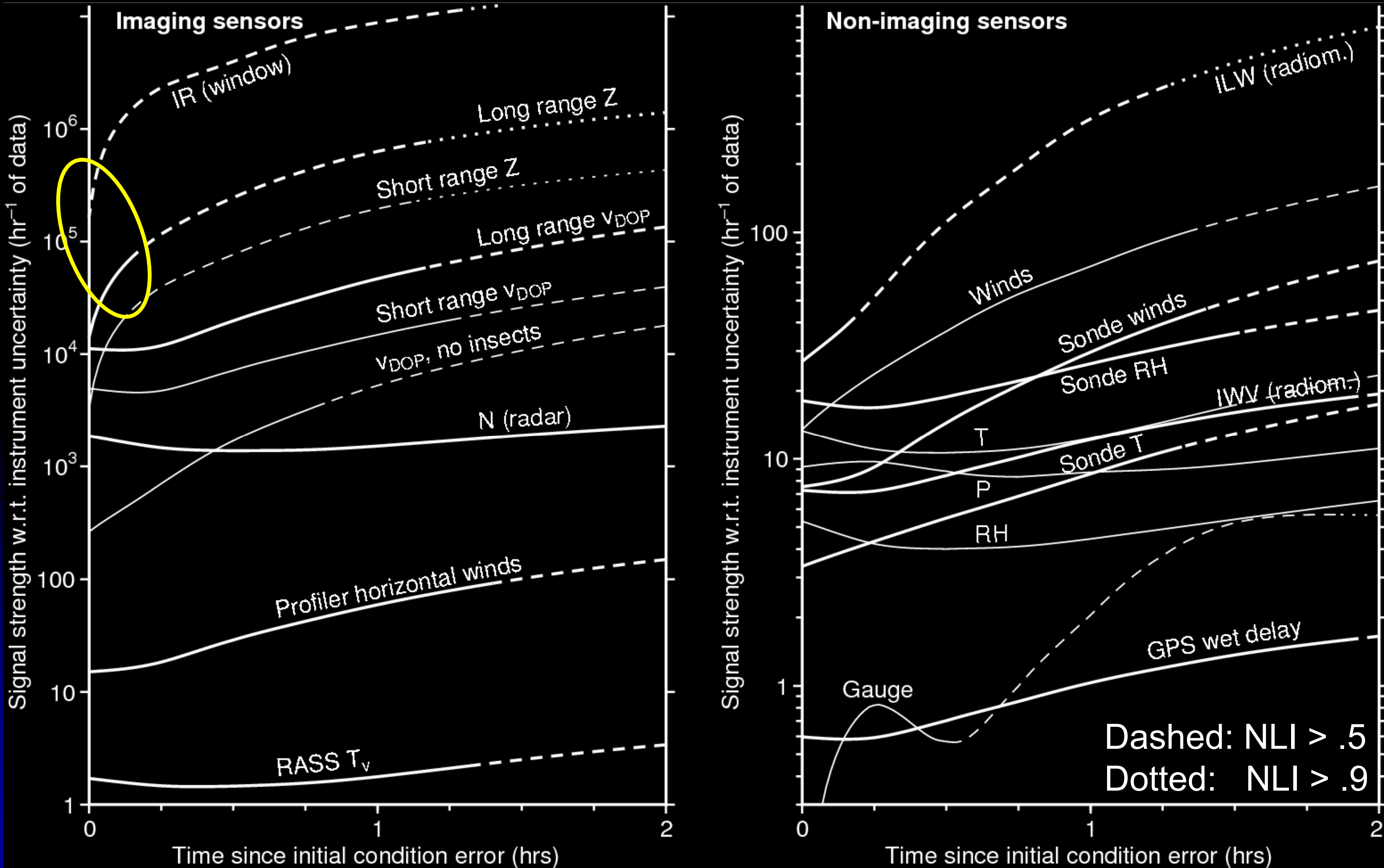
# Evaluating Observations

Many measurements were simulated: surface stations, raingauges, radiosondes, microwave radiometers, ground-based GPS receivers, radars, and satellite-borne IR images.

Signal strength per instrument per hour of data was computed. Note that:

- Signal strength ≠ Usefulness
- Some quantities are harder to assimilate than others: complex link between measurement and atmospheric parameter (e.g., Z → LWC), data difficult to simulate (e.g., surface measurements)…

# Signal Strength Results



# Tentative Recommendations

For instrument developers:

- Aim at midlevel moisture; then low and midlevel temperature.

For data assimilation researchers:

- Ideal assimilation strategy at the mesoscale: 1) smoothed winds, temperature, moisture for a few hours; 2) higher-resolution winds, temperature, moisture for one hour; 3) radar reflectivity for 10 min; 4) one thermal-IR satellite image;
- Different initial condition variables should be constrained with different assimilation windows. New 4D-Var formalism needed?

To be explored (not by me!):

- How tolerant are different assimilation approaches to variables that evolve nonlinearly?

# Questions Unanswered in this Work

- How many perturbations of initial conditions result in similar signals on observations? (beyond the scope of this study)
- How to use and not misuse the results presented here (I still need to think about this)

# THE END