A primer on ensemble (weather) prediction

Tom Hamill

NOAA Earth System Research Lab, Physical Sciences Division

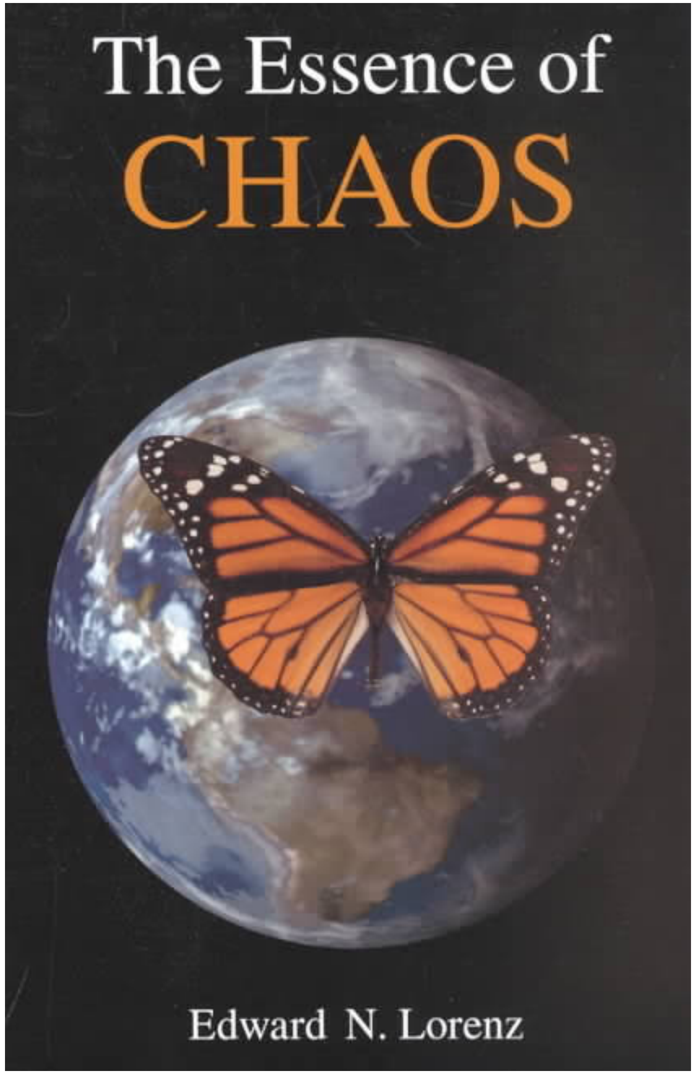
tom.hamill@noaa.gov

I'm not a WRF expert.

Topics today

- Chaos and predictability fundamentals as applied to weather forecasting: why are weather forecasts uncertain?
- Probabilistic forecasting: given this uncertainty, what are we trying to achieve?
- Major sources of forecast uncertainty in regional WRF simulations.
  - Growth of initial-condition uncertainties.
  - Model error
  - Lateral boundary conditions
- Visualization of ensemble forecasts (brief).
- Verification of ensemble forecasts (brief).

Chaos and predictability fundamentals



see this very readable book for more.

Chaos and state-dependent error growth

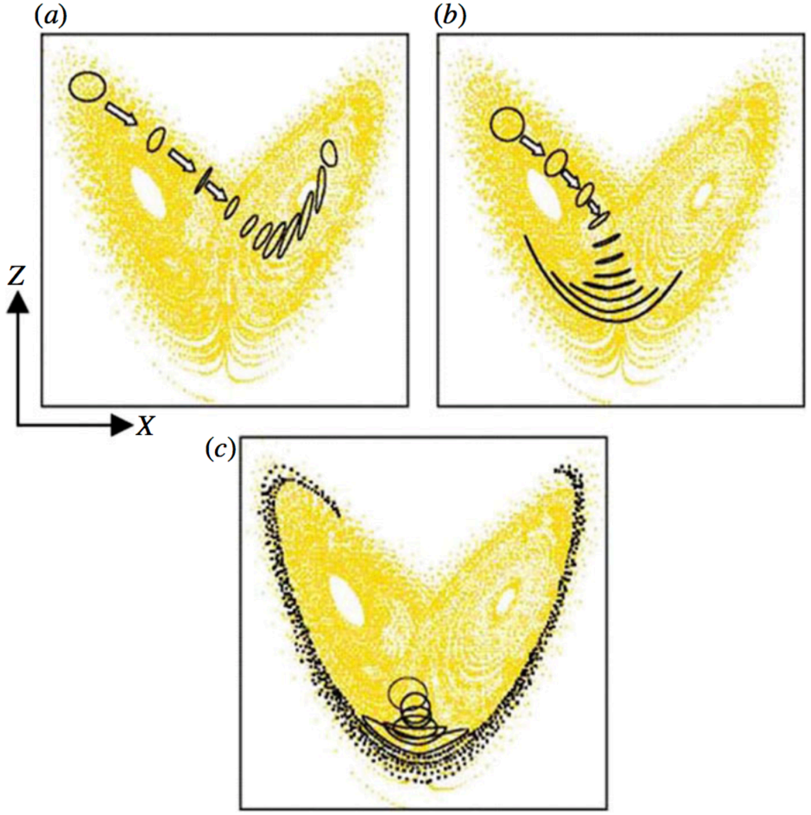


Figure 1. Examples of finite-time error growth on the Lorenz attractor for three probabilistic predictions starting from different points on the attractor. (*a*) High predictability and therefore a high level of confidence in the transition to a different 'weather' regime. (*b*) A high level of predictability in the near term but then increasing uncertainty later in the forecast with a modest probability of a transition to a different 'weather' regime. (*c*) A forecast starting near the transition point between regimes is highly uncertain.

The Lorenz (1963) dynamical system

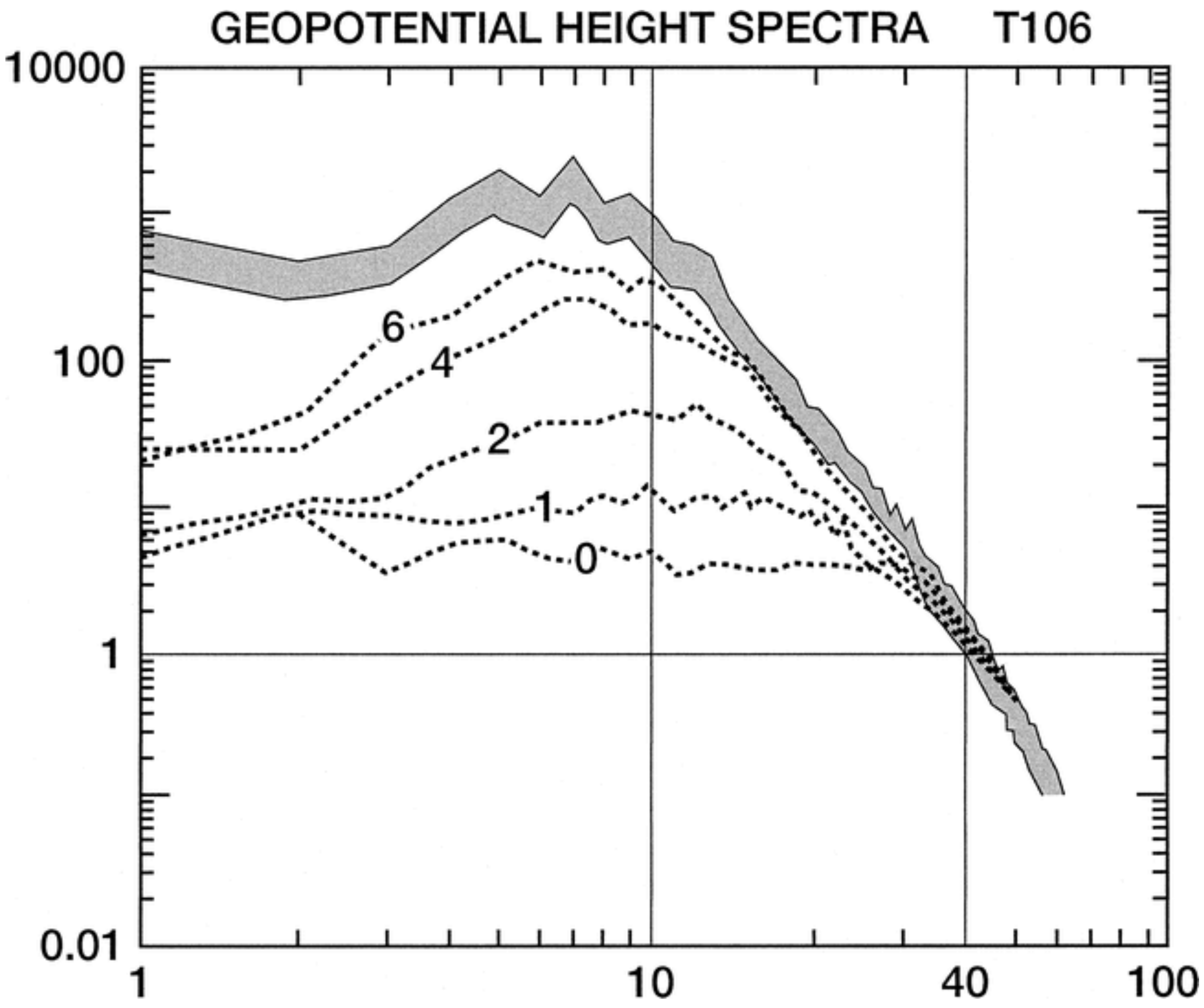
$$\frac{\mathrm{d}x}{\mathrm{d}t} = \sigma(y - x),$$
$$\frac{\mathrm{d}y}{\mathrm{d}t} = x(\rho - z) - y,$$
$$\frac{\mathrm{d}z}{\mathrm{d}t} = xy - \beta z.$$

x, y, z are state variables,
σ, ρ, and β are parameters

$\sigma = 10$, $\beta = 8/3$ and $\rho = 28$

Picture from Slingo and Palmer 2011, Phil. Trans. Royal. Soc A., **369**, 4751-4767

The rate of chaotic error growth and the timescale where errors saturate is a function of the spatial scale.



Ref: Tribbia and Baumhefner, March 2004 *MWR*

1. Typically, the large scales are proportionately well analyzed (smaller error than signal).

2. The smaller scales of motion are less well analyzed, and even at short leads error can be as large as the signal.

3. Errors in small scales grow very rapidly and grow upscale, until they project on synoptic scales. Thereafter, slower, more modal growth.

4. To the extent that there is predictive skill in small-scale features (e.g., thunderstorms) at long leads, it's largely because the synoptic scale determines the general areas where convection is possible.

5. Not shown: predictability time scale of small-scale phenomena even shorter in weakly forced flow (see supplementary slides).

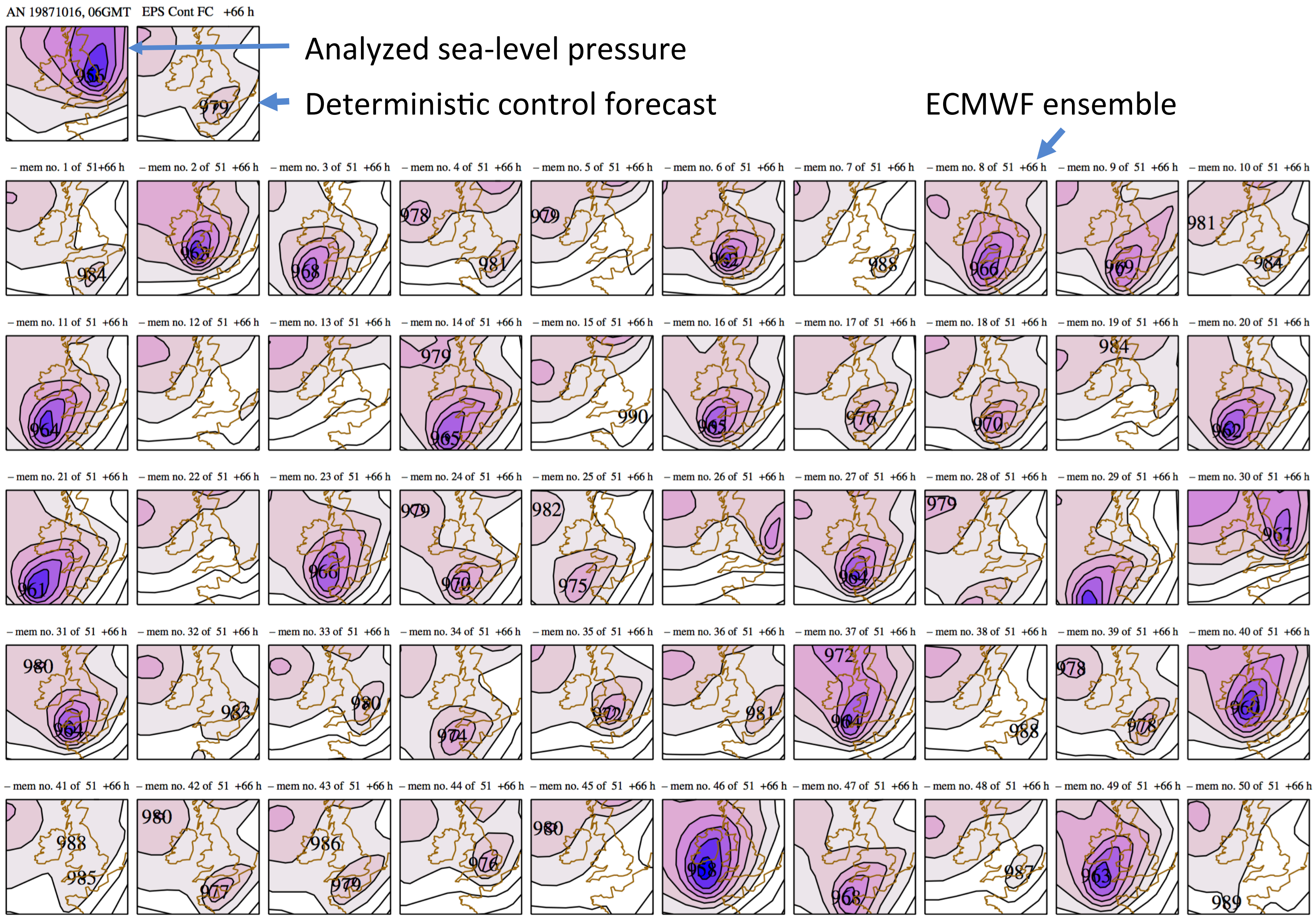


Figure 3. Example of 66 h probabilistic forecast for 15–16 October 1987. Top left shows the analysed deep depression with damaging winds on its southern flank. Top right shows the deterministic forecast, and the remaining 50 panels show other possible outcomes based on perturbations to the initial conditions. A substantial fraction of the ensemble indicates the development of a deep depression.

Ref: Slingo and Palmer 2011, Phil. Trans. Royal. Soc A., **369**, 4751-4767

Does chaos matter for short-range weather? Example: initial-condition uncertainty grows very rapidly, making even short-range deterministic forecasts untenable.

Note that though many simulations forecast a deep low-pressure system, those that do forecast this are commonly mis-forecasting the position, with member 30 being the exception.

Suggestive that model error plays a role even in these short-lead forecasts.

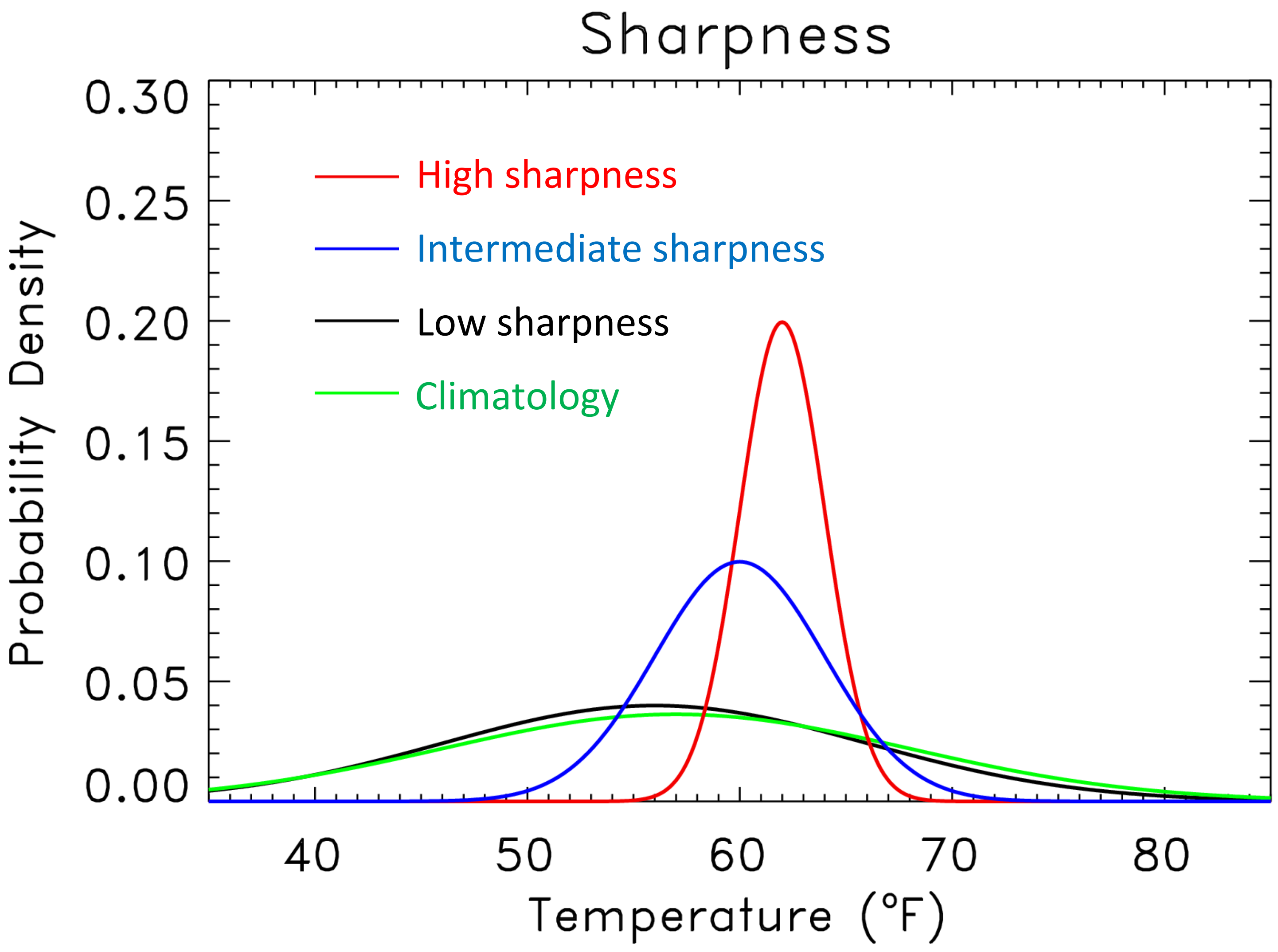
Probabilistic forecasting: given forecast uncertainty, what are we trying to achieve?

"*Sharpness subject to reliability*."

But what does this mean?

See Gneiting et al. 2007, http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9868.2007.00587.x/abstract

Sharpness of a pdf.



Reminder: appropriately estimated *situational sharpness* is what we strongly desire.

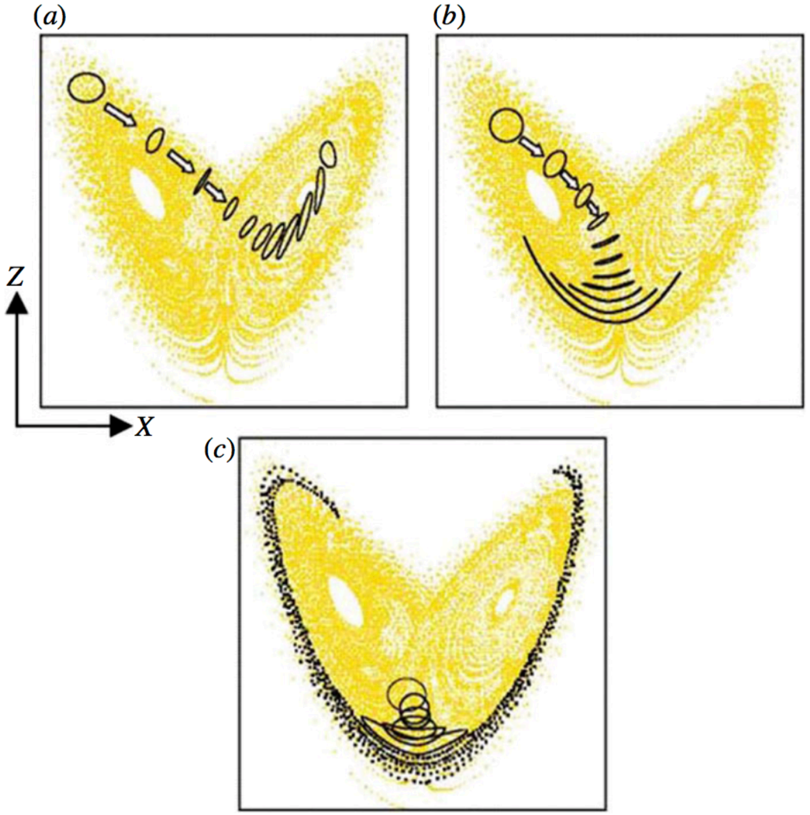
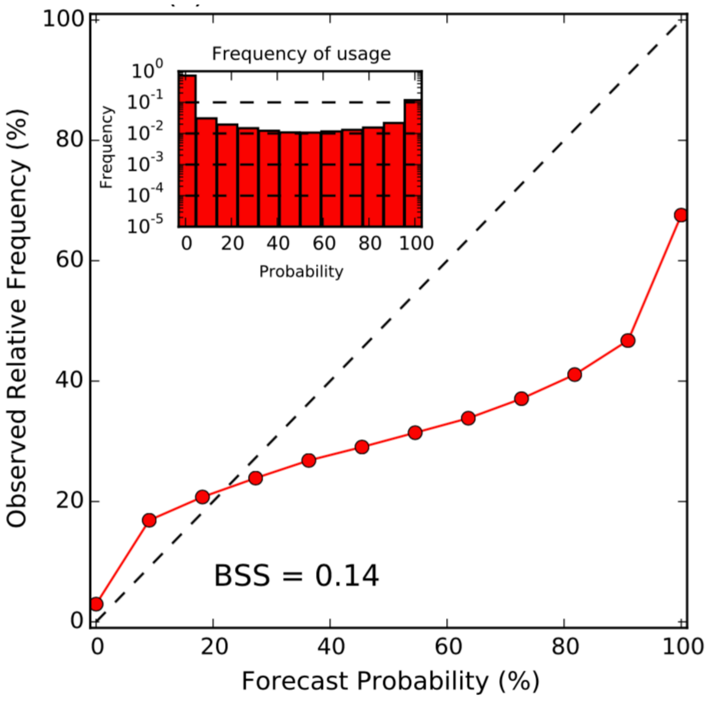


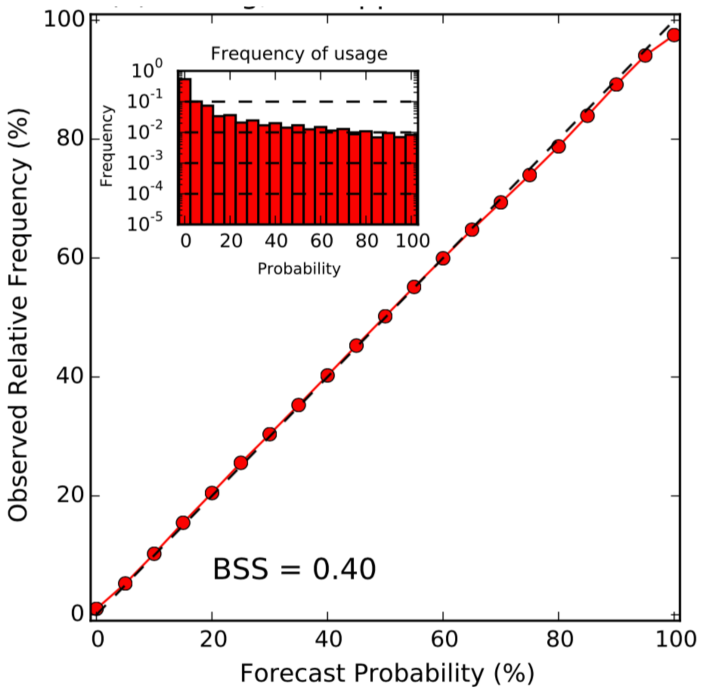
Figure 1. Examples of finite-time error growth on the Lorenz attractor for three probabilistic predictions starting from different points on the attractor. (*a*) High predictability and therefore a high level of confidence in the transition to a different 'weather' regime. (*b*) A high level of predictability in the near term but then increasing uncertainty later in the forecast with a modest probability of a transition to a different 'weather' regime. (*c*) A forecast starting near the transition point between regimes is highly uncertain.

Reliability and sharpness

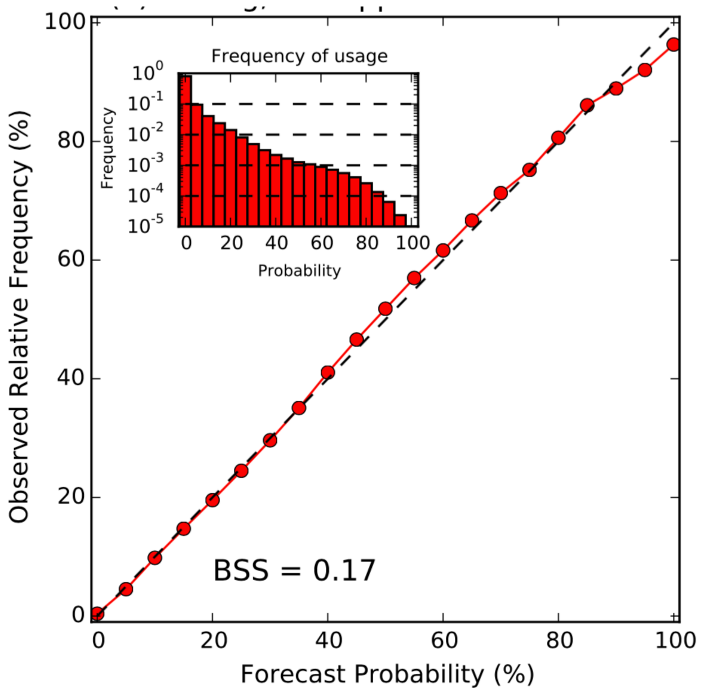
A sharper forecast, more binary, but less reliable



A reliable forecast, but less sharp



A mostly reliable forecast, much less sharp



We'd like high confidence in event probabilities (forecasts are 1.0 or 0.0) and for the events to always happen when 1.0 is forecast, and never happen when 0.0 is forecast. If this is unobtainable (it is), we want to maximize the sharpness (the specificity) subject to retaining reliability.

An ensemble can be “perfect” but not reliable?

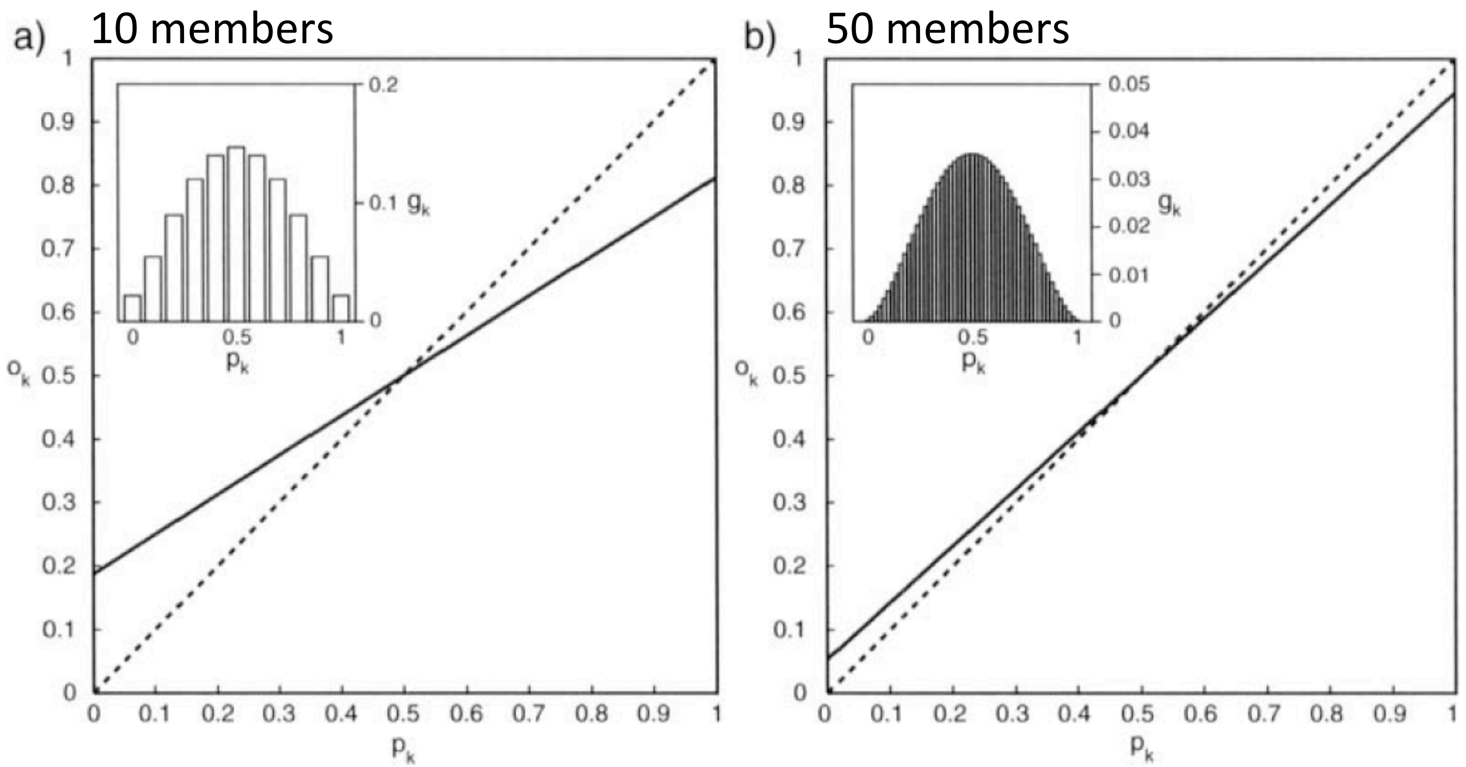
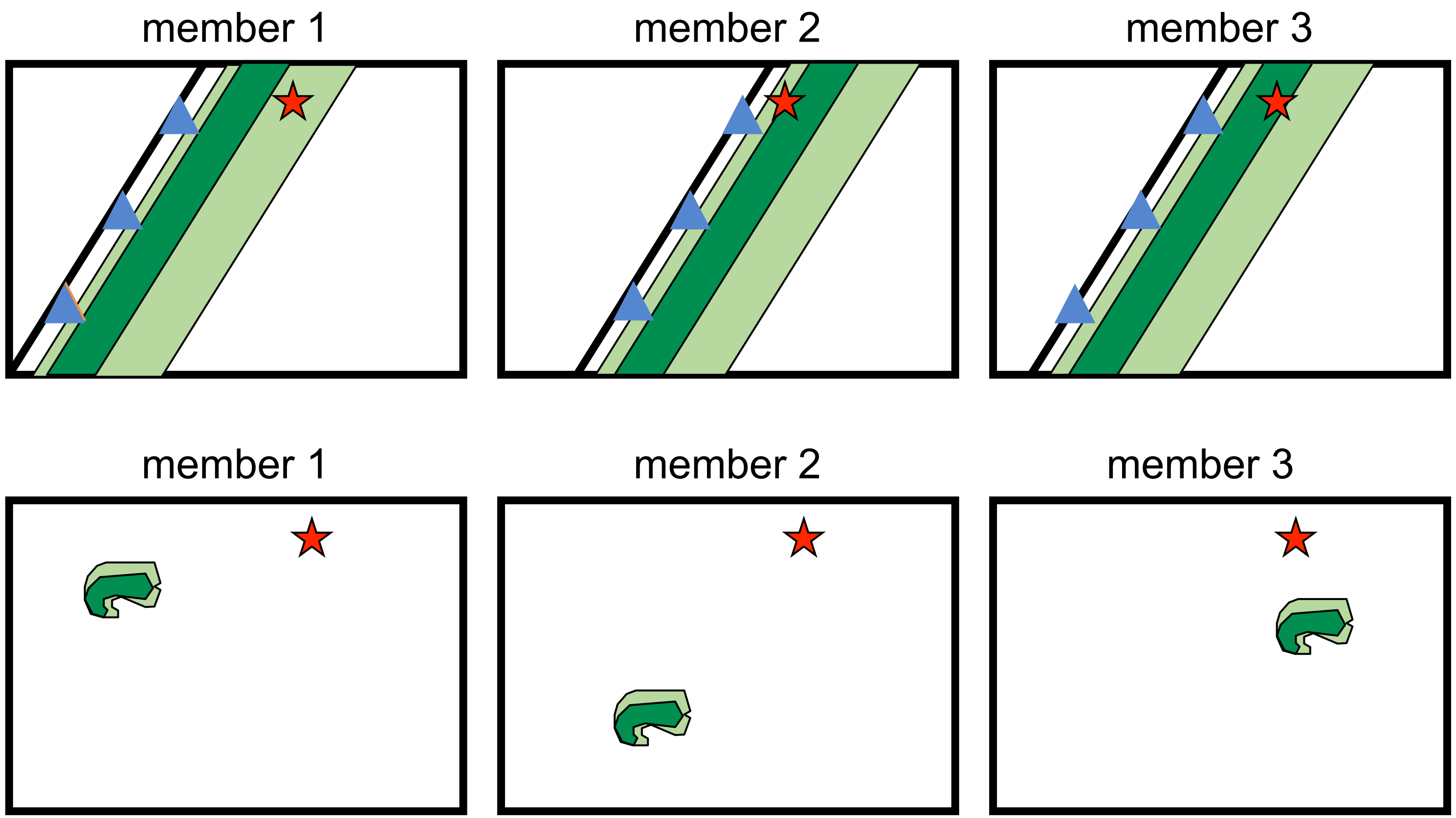


Figure 3. Reliability diagrams for theoretical ensemble forecasts for (a) a 10-member ensemble prediction system (EPS) and (b) a 50-member EPS. Distribution of underlying forecast probabilities is completely reliable and specified by a beta distribution with $r = s = 3$. See text for details and explanation of symbols.

Ensemble members and truth sampled from the same underlying distribution, i.e., prediction system is perfect. And yet with a finite number of members, the resulting probabilities are unreliable. This is because of sampling error.

ref: Richardson 2001 QJRMS

Sample-size issues with ensemble prediction and their relationship to the scale of the phenomenon



- Size of feature approximately the same as spatial variations between members. Moderately sized ensemble adequate for estimating probabilities.
- Size of feature is much smaller than spatial variations between members. Moderately sized ensemble inadequate for directly estimating probabilities. Either large ensemble, or statistical adjustment necessary.

The number of ensemble members you'd need (in the absence of some statistical postprocessing) to achieve reliable forecasts can vary with the phenomenon you are trying to predict.

Major sources of forecast uncertainty in regional simulations, and how we simulate them.

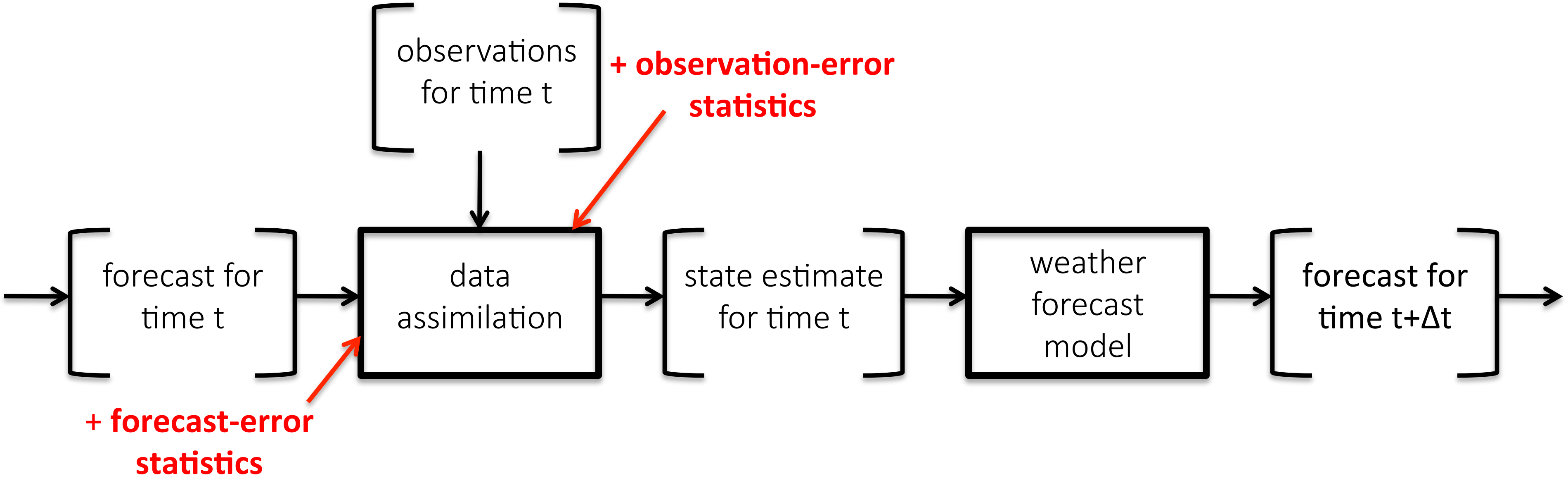
- Initial condition uncertainty.
- Model uncertainty.
- Lateral boundary condition uncertainty.

Initial-condition uncertainty.

There are many ways of producing realistic ensembles of plausible initial conditions.

We'll focus on a common one in WRF, the "Ensemble Kalman Filter" or "EnKF."

State estimation ("data assimilation")



The forecast is uncertain, and the observations are imperfect and scattered. Hence we should expect that the analysis too will be uncertain, so in ensemble prediction we generate multiple realizations of the possible analysis state. How?

From first principles: Bayesian data assimilation

$\mathbf{x}_t$ = (unknown) true model state

$\psi_t = [\mathbf{y}_t, \psi_{t-1}]$ = observations = [today's, all previous]

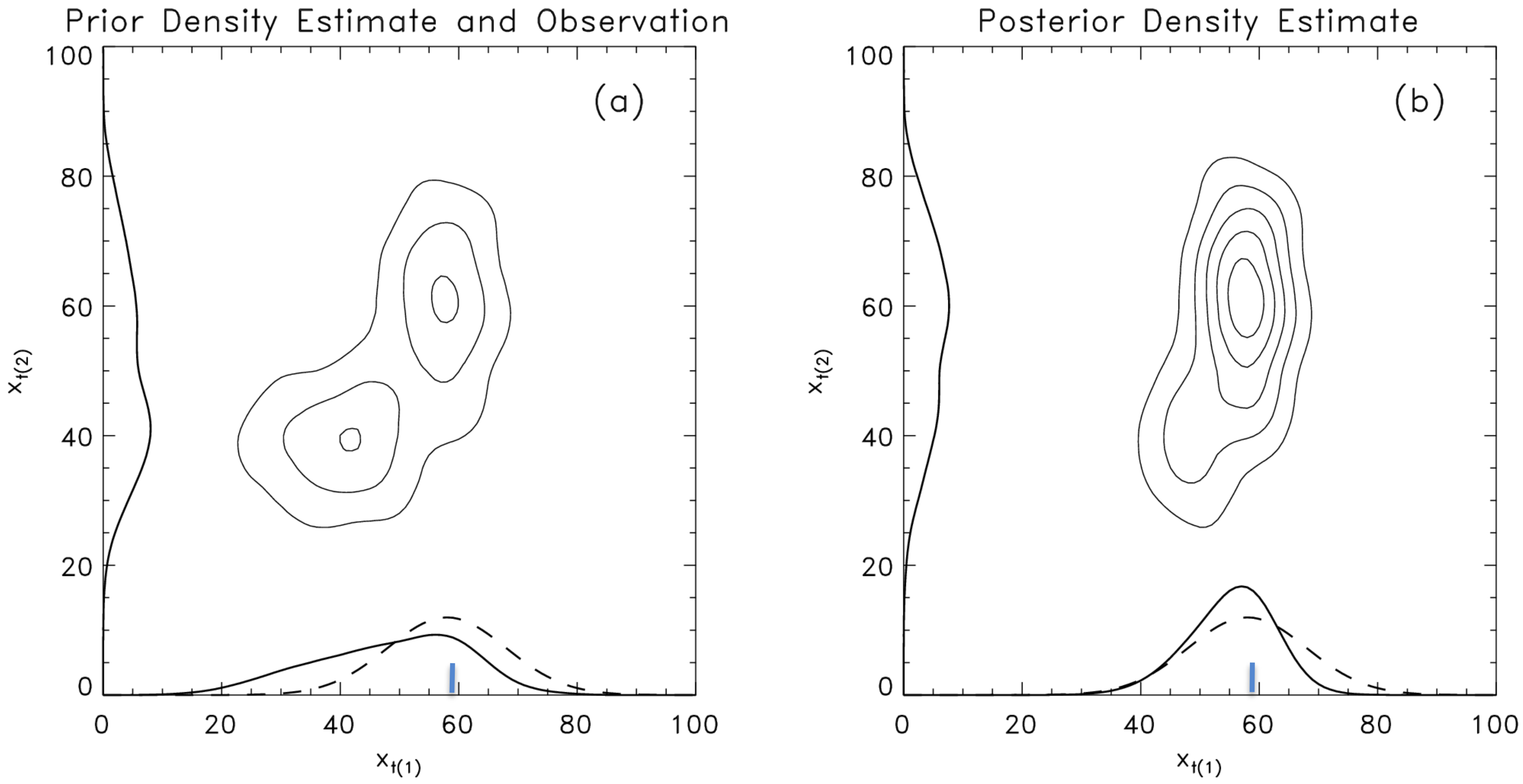
$$P(\mathbf{x}_t | \psi_t) \propto P(\mathbf{y}_t | \mathbf{x}_t) P(\mathbf{x}_t | \psi_{t-1})$$

"posterior" "prior"

A manipulation of Baye's Rule, assuming observation errors are independent in time.

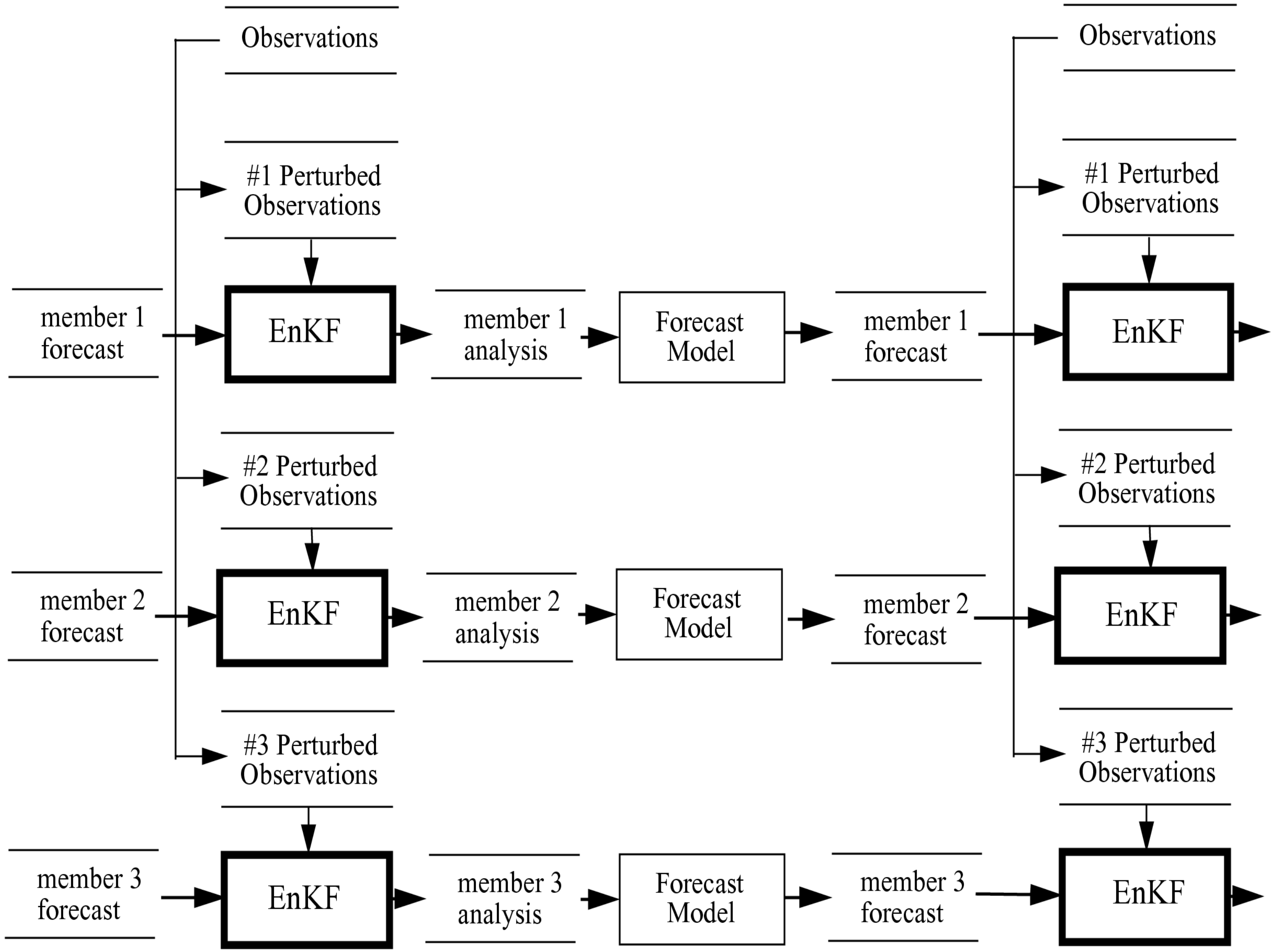
[derivation]

Bayesian data assimilation: a 2-D example



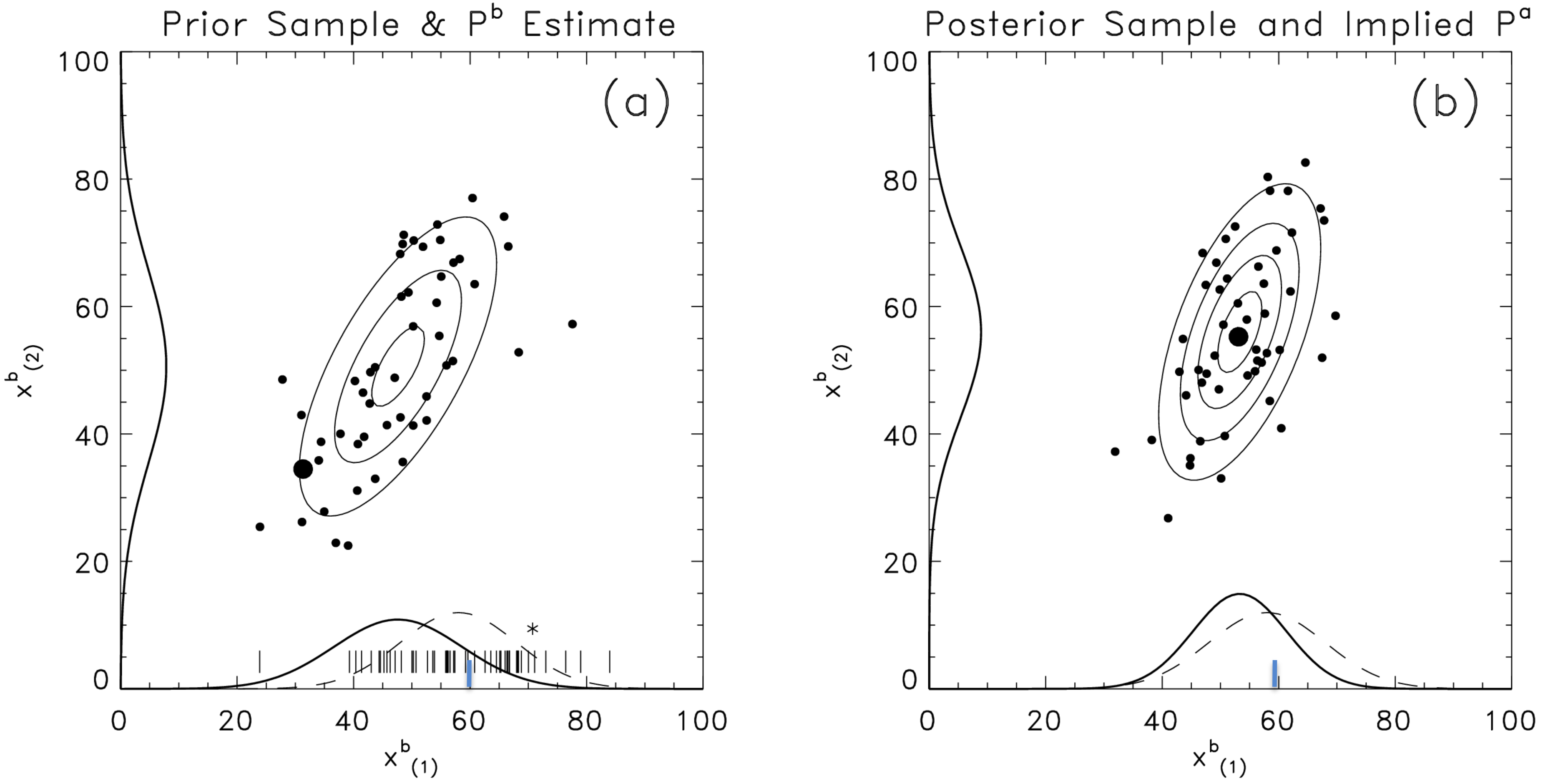
Computationally expensive when highly dimensional! Here, probabilities explicitly updated on 100x100 grid; costs multiply geometrically with the number of dimensions of model state. Also: “curse of dimensionality”

The ensemble Kalman filter: a schematic



(This schematic is a bit of an inappropriate simplification, for EnKF uses every member to estimate background-error covariances)

How the EnKF works: a 2-D example



Start with a random sample from bimodal distribution used in previous Bayesian data assimilation example. Contours reflect the Gaussian distribution fitted to ensemble data. Assumptions are made in EnKF that errors are Gaussian-distributed.

More on the ensemble Kalman filter and ensemble initialization.

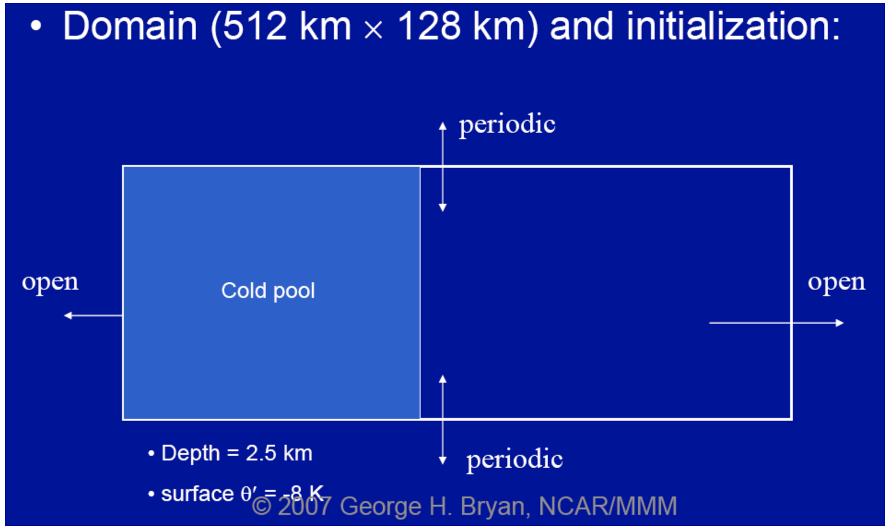
- Hamill, T. M., 2006: Ensemble-based atmospheric data assimilation Chapter 6 of *Predictability of Weather and Climate*, Cambridge Press, 124-156.
- Ehrendorfer and Tribbia, JAS, 1997: Optimal Prediction of Forecast Error Covariances through Singular Vectors
    - Provides a sound theoretical understanding of what we're trying to do to initialize a limited-size ensemble and span the forecast errors as best as possible.

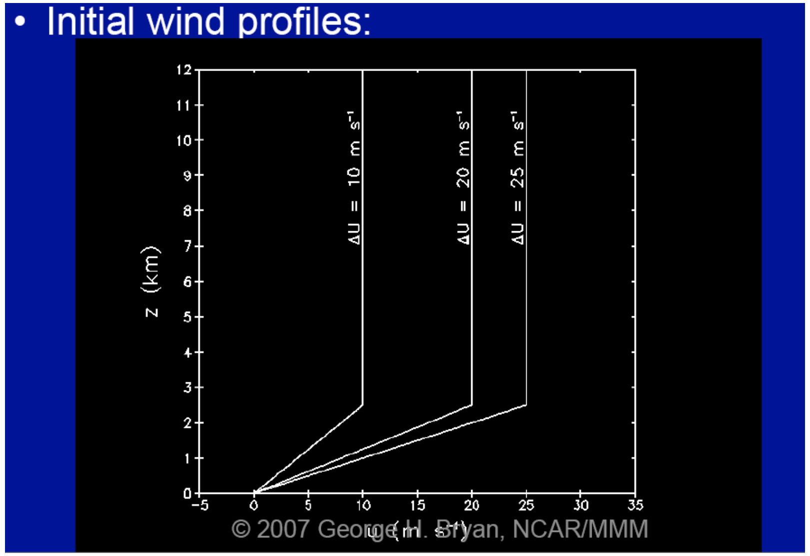
Model uncertainty in many manifestations.

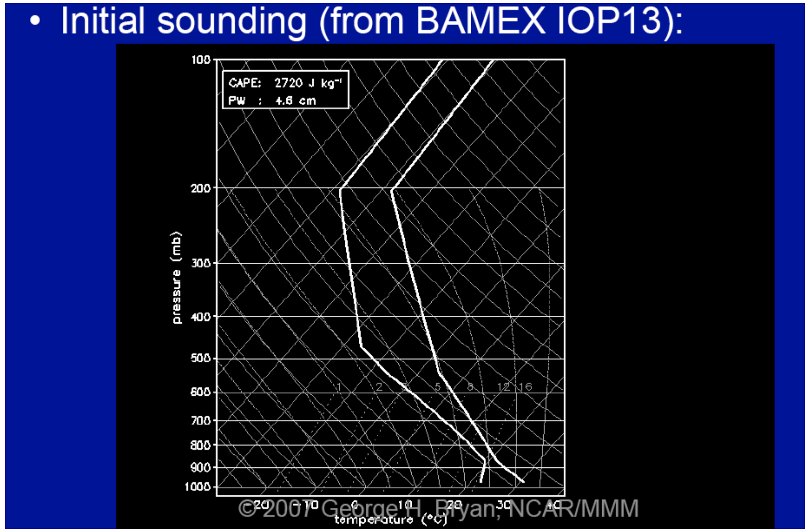
(see Judith Berner's presentation for some methods to deal with this).

Model error at mesoscale:
(1) errors from too coarse a grid spacing

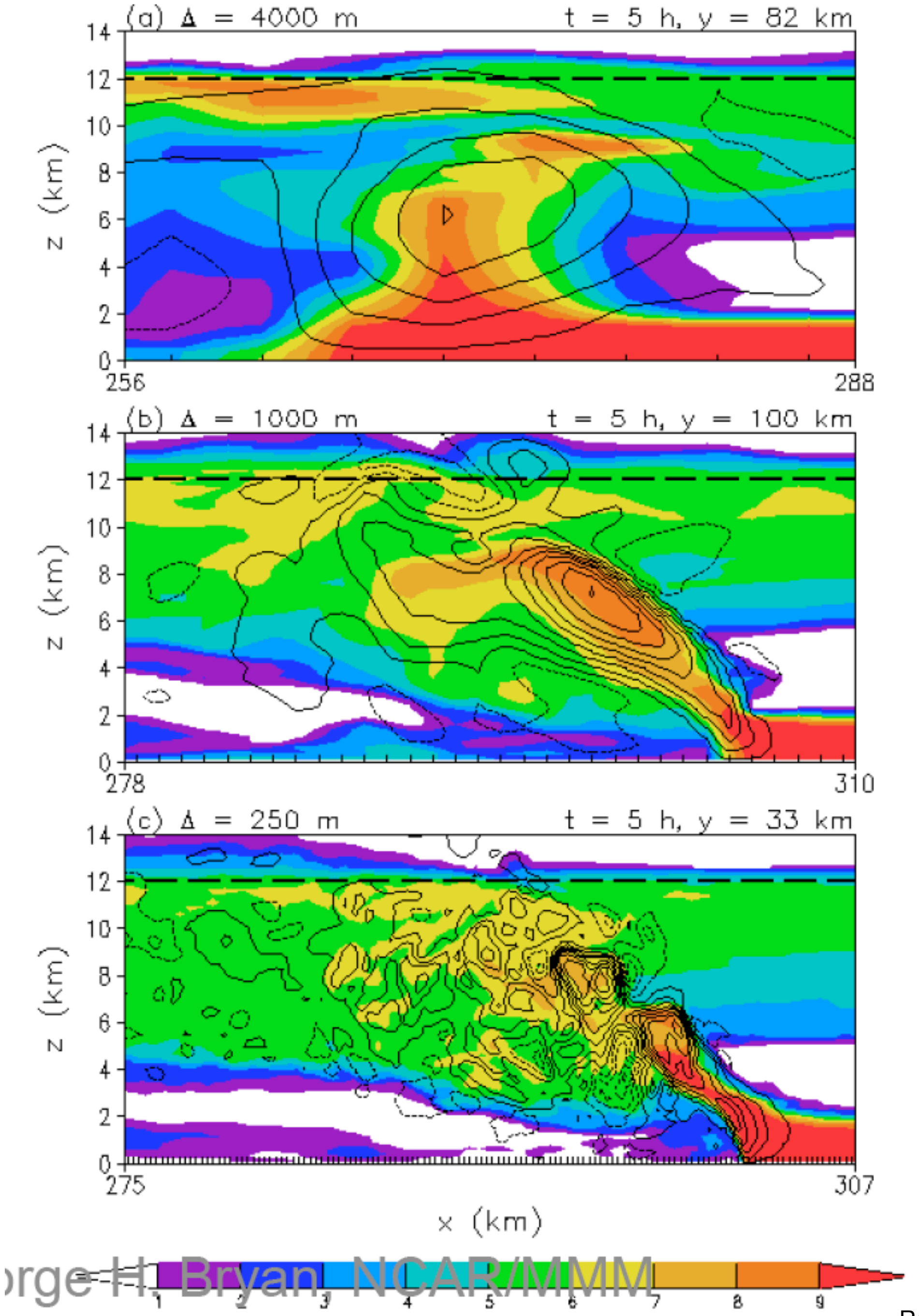
- George Bryan (NCAR) tested convection in simple models with grid spacings from 8 km to 125 m







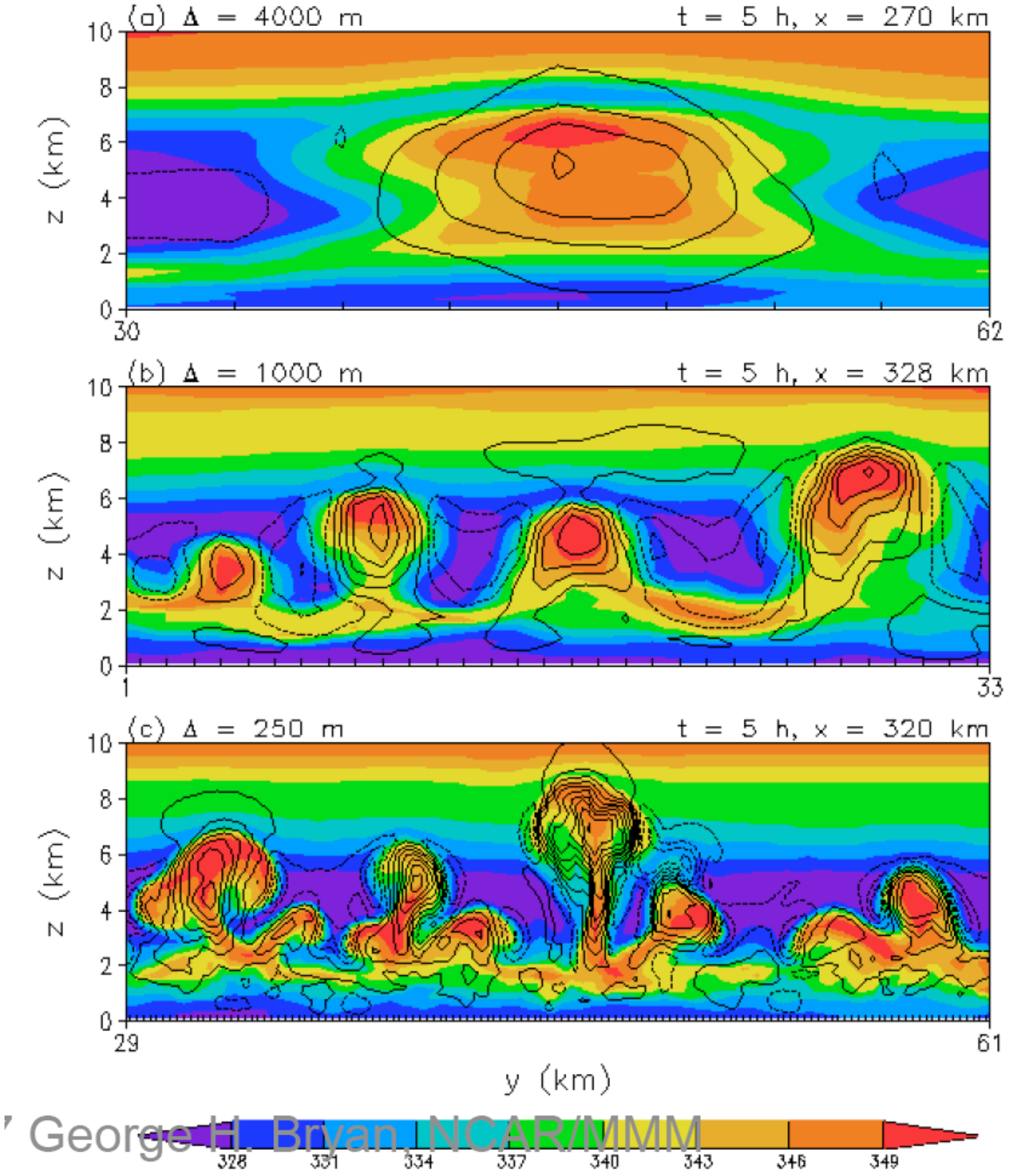
Ref: http://www.mmm.ucar.edu/people/bryan/Presentations/bryan_2007_nssl_resolution.pdf



4 km, 1 km, 0.25 km

- *Across* the squall line vertical cross section for 25 ms^{-1} wind shear. Shading: mixing ratio (g kg^{-1}); contours (vertical velocity (every 4 ms^{-1}).
- Dramatic changes in structure of squall line, updraft, positioning of cold pool.

Ref: http://www.mmm.ucar.edu/people/bryan/Presentations/bryan_2007_nssl_resolution.pdf

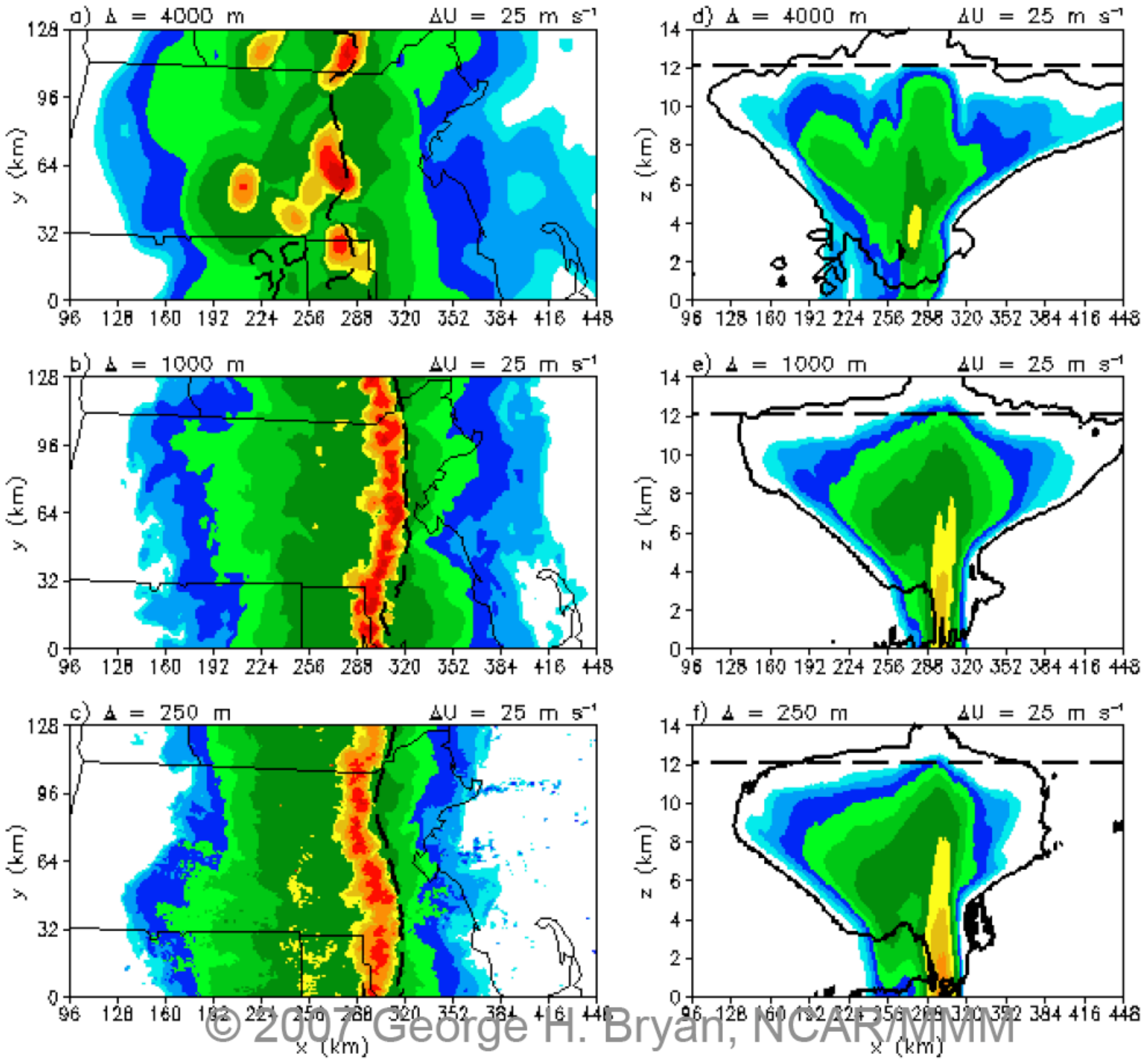


4 km, 1 km, 0.25 km

- *Along* the squall line vertical cross section for 20 ms^{-1} wind shear. Shading: mixing ratio (g kg^{-1}); contours (vertical velocity (every 4 ms^{-1}).
- Updrafts increase in number and intensity with increasing resolution, decrease in size.

Ref: http://www.mmm.ucar.edu/people/bryan/Presentations/bryan_2007_nssl_resolution.pdf

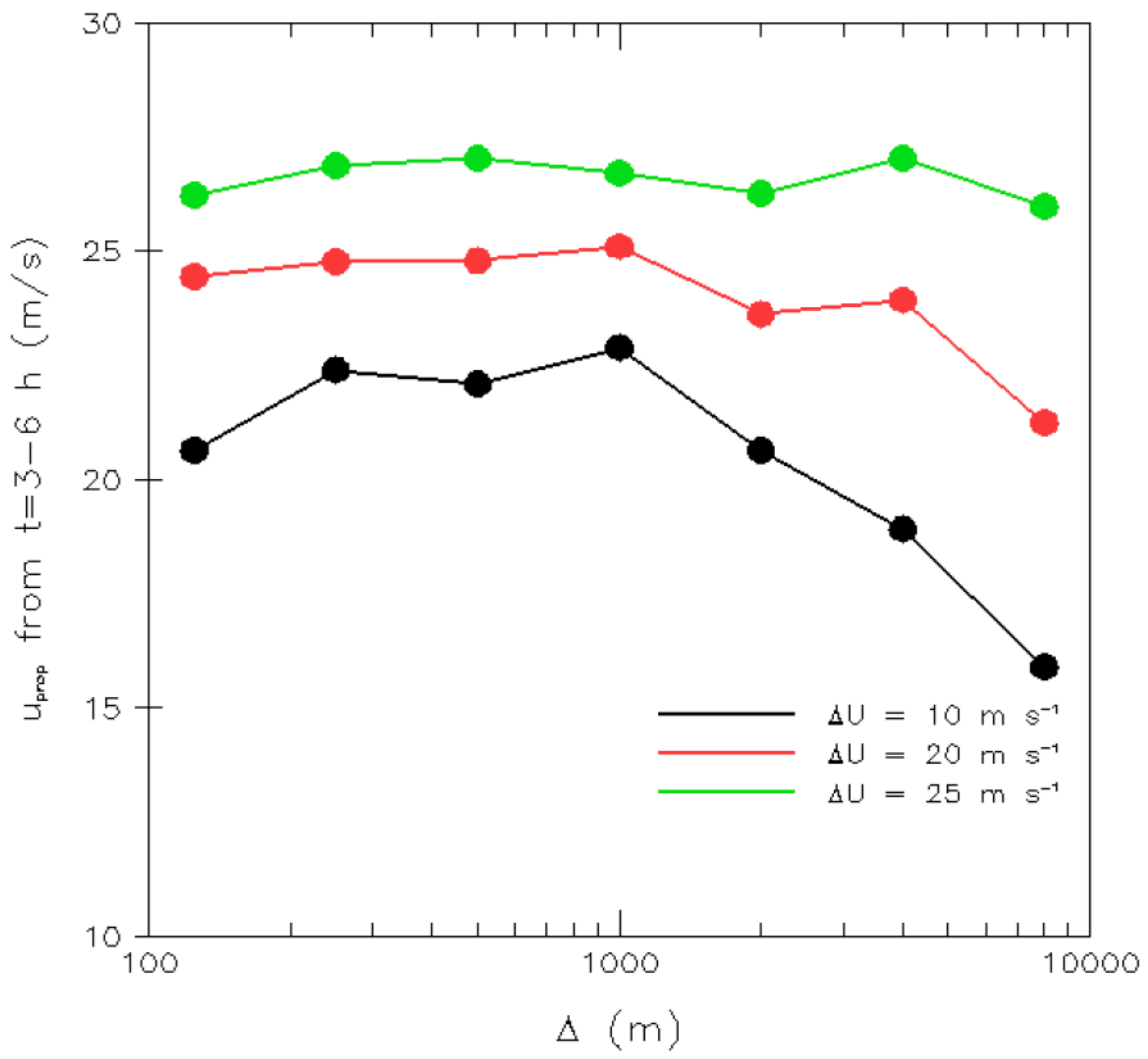
4 km, 1 km, 0.25 km



- Plan view and N-S integrated vertical cross section for 25 ms^{-1} wind shear. Shading: mixing ratio (g kg^{-1}); contours (vertical velocity (every 4 ms^{-1}).
- Here, 1 km and 4 km differences aren't as noticeable.

Ref: http://www.mmm.ucar.edu/people/bryan/Presentations/bryan_2007_nssl_resolution.pdf

4 km, 1 km, 0.25 km



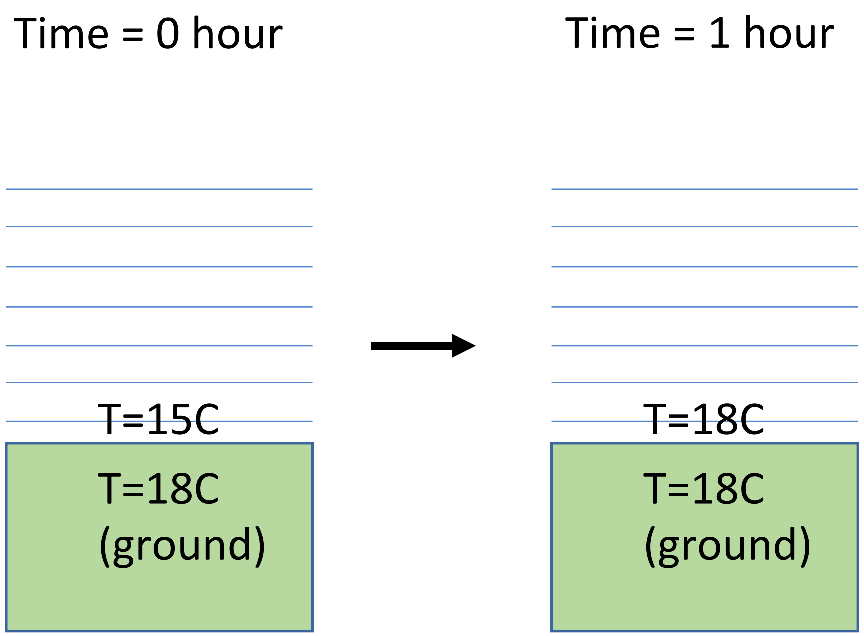
- System propagation speed approximately converged at 1 km for high-shear cases.
- For low-shear environment (more weakly forced) resolutions above 1 km are increasingly inadequate.

Ref: http://www.mmm.ucar.edu/people/bryan/Presentations/bryan_2007_nssl_resolution.pdf

Model errors at mesoscale:
(2) those darn parameterizations! A few examples.

- Land-surface parameterization (and state estimate)
- Deep convective parameterization
- Microphysical parameterization
- etc.

2-m temperature bias can be significant at forecast lead times of *1 hour.*



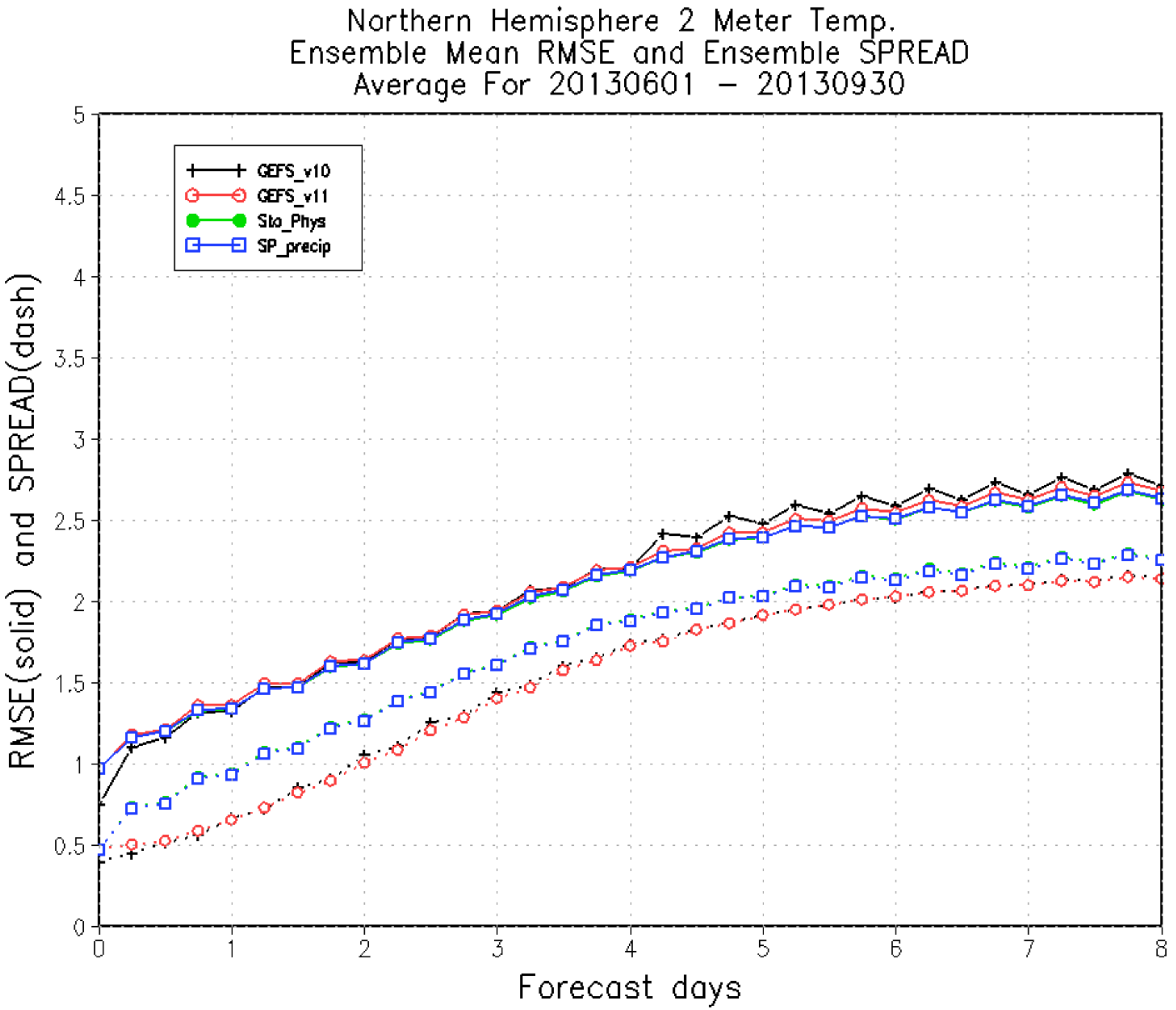
Suppose we have very carefully estimated the thermodynamic profile of the atmospheric state, and the analyzed temperature near the ground is 15C (error ~ 1C).

Suppose the ground temperature is not well quantified, possibly with significant bias, due to lack of observations and significant model error in the land-surface model.

In a very short amount of time, the near-surface atmospheric temperature equilibrates to be more consistent with the (biased) soil temperature, with its much greater thermal energy storage.

This is a large problem for near-surface predictions at short leads. The answer to such problems is probably better addressed more directly (by improving the LSM) than by designing ways to introduce several K of ensemble spread in the first few hours of the forecast. This isn't uncertainty, it's systematic bias, and is better addressed as such.

Recent results with 2-meter temperature spread in NCEP Global Ensemble Forecast System

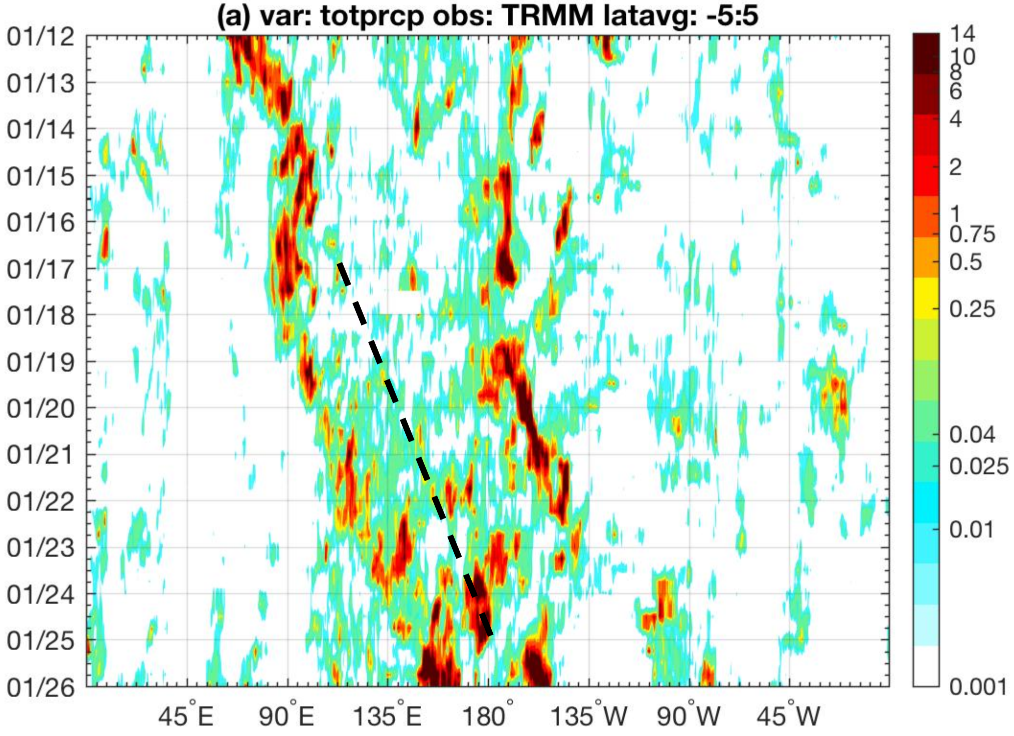


Even with relatively modern atmospheric stochastic parameterization package (blue curve; discussed in greater depth later), 2-meter temperature spread is much smaller than RMS error. Consistency of the two over many samples is necessary but not sufficient for reliability.

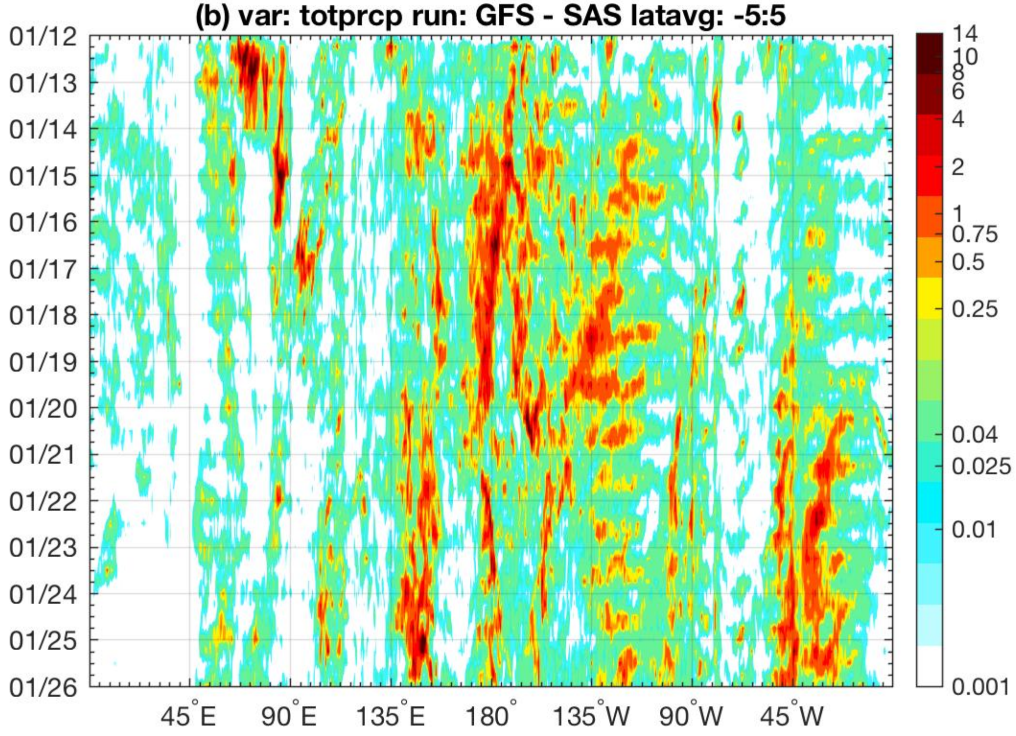
c/o Walt Kolczynski, NCEP/EMC

Error in tropical variability estimates with a current-generation global NWP model.

Analyzed precipitation



GFS forecast precipitation

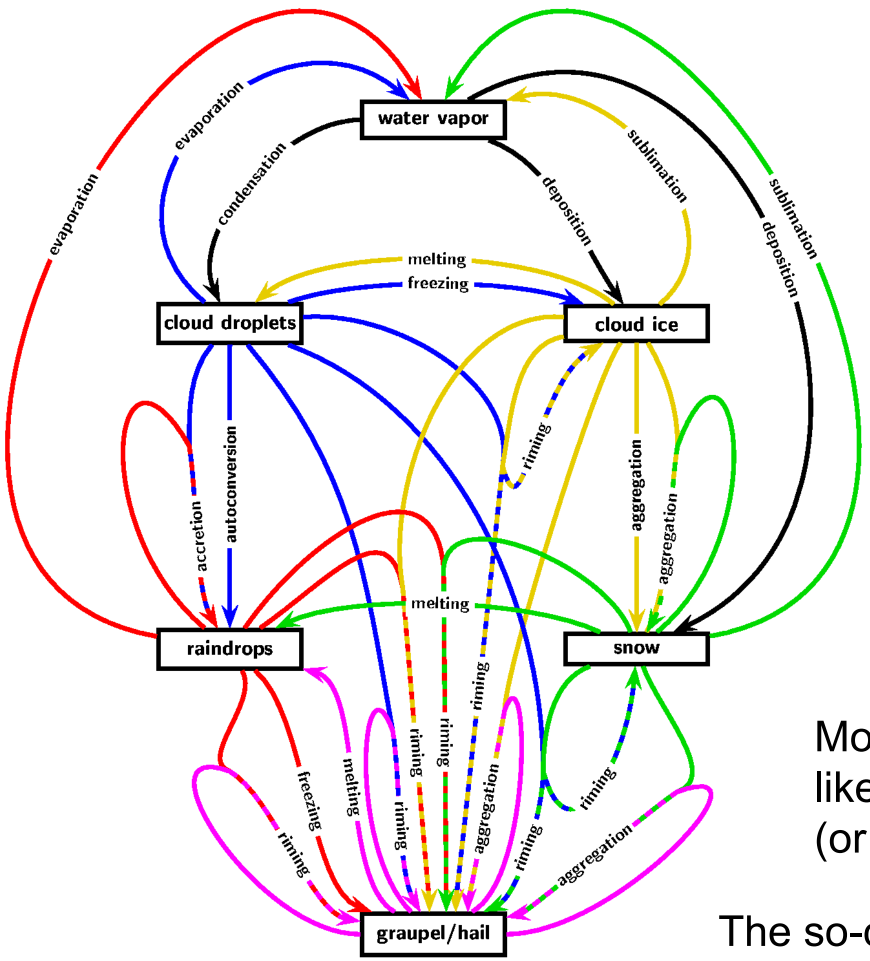


Kelvin waves are not well simulated with current National Weather Service global forecast system (GFS). These tropical convective clusters can force extratropical wave trains. There may be many reasons, including deficiencies in parameterizations of deep convection, diurnal evolution of SST, and more.

c/o Juliana Dias, U. Colorado CIRES

Model error at mesoscale:

Example: cloud microphysical processes



Conversion processes, like snow to graupel conversion by riming, are very difficult to parameterize but very important in convective clouds.

Especially for snow and graupel the particle properties like **particle density** and **fall speeds** are important parameters. The assumption of a constant particle density is questionable.

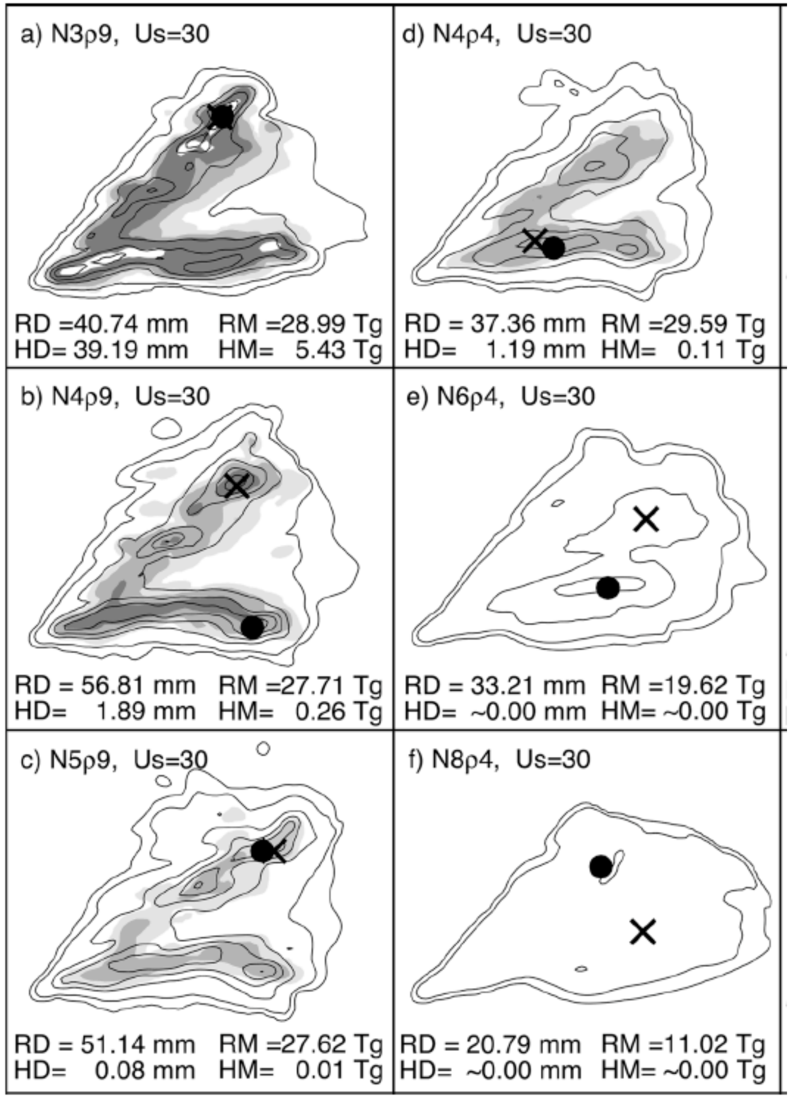
Aggregation processes assume certain collision and sticking efficiencies, which are not well known.

Most schemes do not include **hail processes** like wet growth, partial melting or shedding (or only very simple parameterizations).

The so-called **ice multiplication** (or Hallet-Mossop process) may be very important, but is still not well understood

from Axel Seifert presentation to NCAR ASP summer colloquium

Sensitivity of deep convective storms to graupel properties in a microphysical parameterization



Effect of assumed graupel density and particle size distribution, i.e. size and fall speed, in a storm split spawning supercells. Contours: rain isohyets: shading: hail/graupel depths greater than .01, 0.1, 1, and 10 mm.

• : location of maximum graupel accumulation.

? : location of maximum hail accumulation.

Main point plausible changes in parameters used in the microphysical parameterizations can cause large changes in precipitation amount, type, and location.

Ref: Gilmore et al., Nov 2004 MWR

Lateral boundary conditions.

Lateral boundary condition issues for limited-area models (and limited-area ensemble forecasts)

- With 1-way LBCs, small scales in domain cannot interact with scales larger than some limit defined by domain size.
- LBCs generally provided by coarser-resolution forecast models, and this "sweeps" in low-resolution information, sweeps out developing high-resolution information.
- Physical process parameterizations for model driving LBCs may be different than for interior. Can cause spurious gradients
- LBC info may introduce erroneous information for other reasons, e.g., model numerics.
- LBC initialization can produce transient gravity-inertia modes.

Ref: Warner et al. review article, *BAMS*, November 1997

Influence of domain size

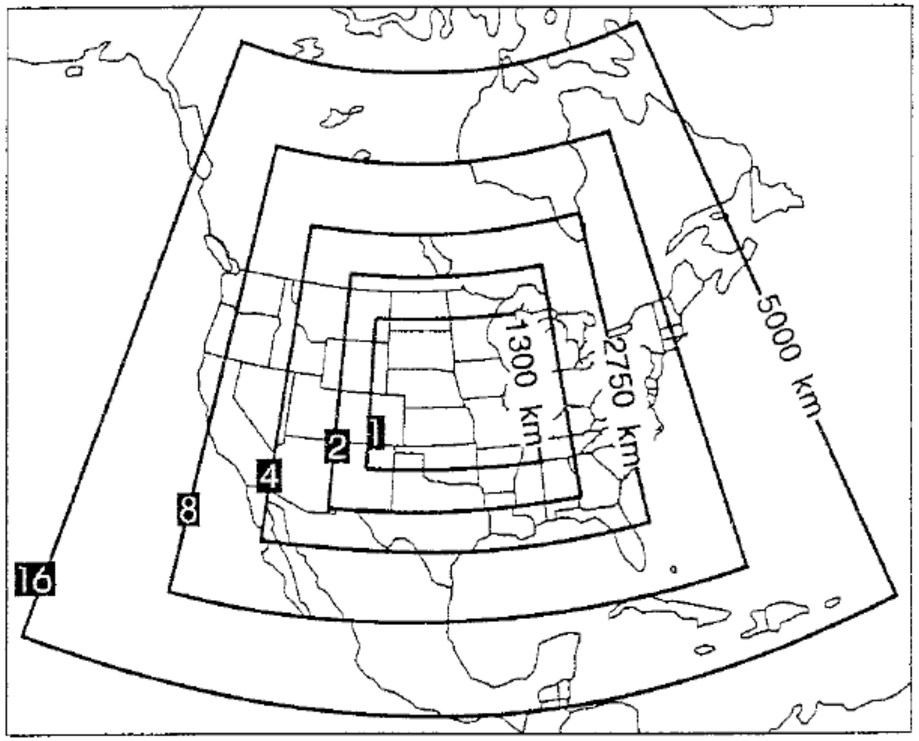


FIG. 3. Five collocated integration domains of the 80-km grid increment Eta Model used in the domain-size sensitivity study. The grid number corresponds to the factor by which the grid is larger than that of the smallest grid. From Treadon and Peterson (1993).

T-126 global model driving lateral boundary conditions for nests with 80-km and 40-km grid spacing of limited-area model.

from Warner et al. Nov 1997 *BAMS*, and Treadon and Peterson (1993), *Preprints, 13th Conf. on Weather Analysis and Forecasting*

Influence of domain size, continued.

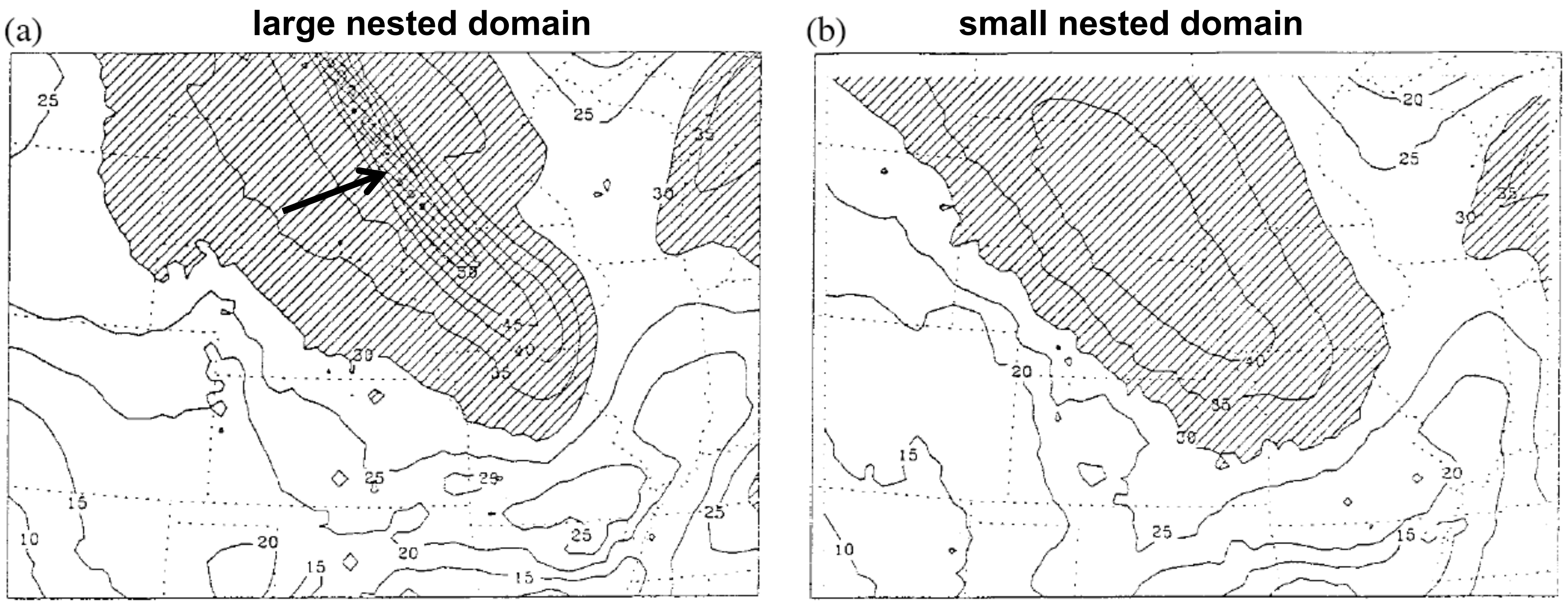
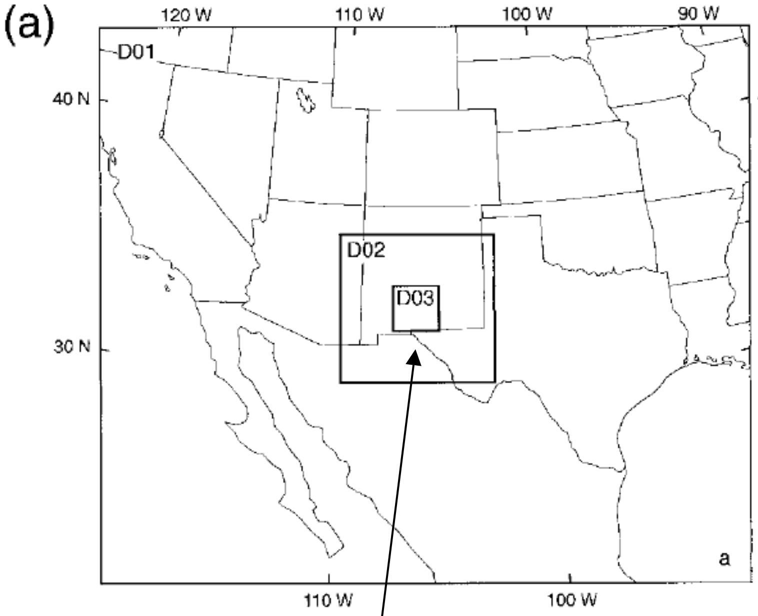


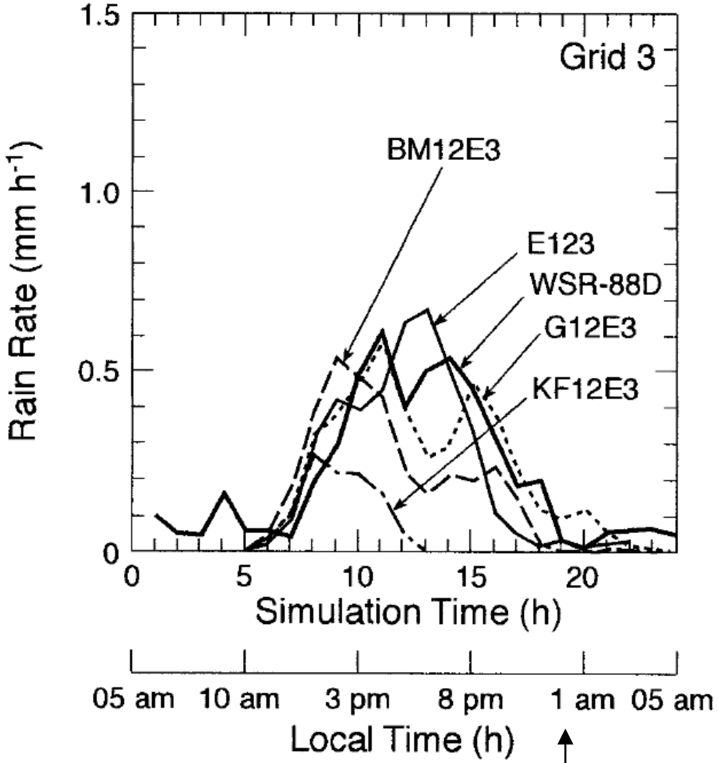
FIG. 6. Simulated 250-hPa isotachs (m s^{-1}) from the 40-km grid increment Eta Model initialized at 1200 UTC 3 August 1992 for the largest computational domain (a) and the smallest (b). The isotach interval is 5 m s^{-1}. From Treadon and Peterson (1993).

40-km nested domain in global model had thin, realistic jet streak using large domain (left) and smeared-out, unrealistic jet streak using small domain (right). High resolution of interior domain not useful here because of sweeping in of low-resolution information.

Ref: ibid

Problems caused by using outer domain convective parameterization with explicit convection in nest



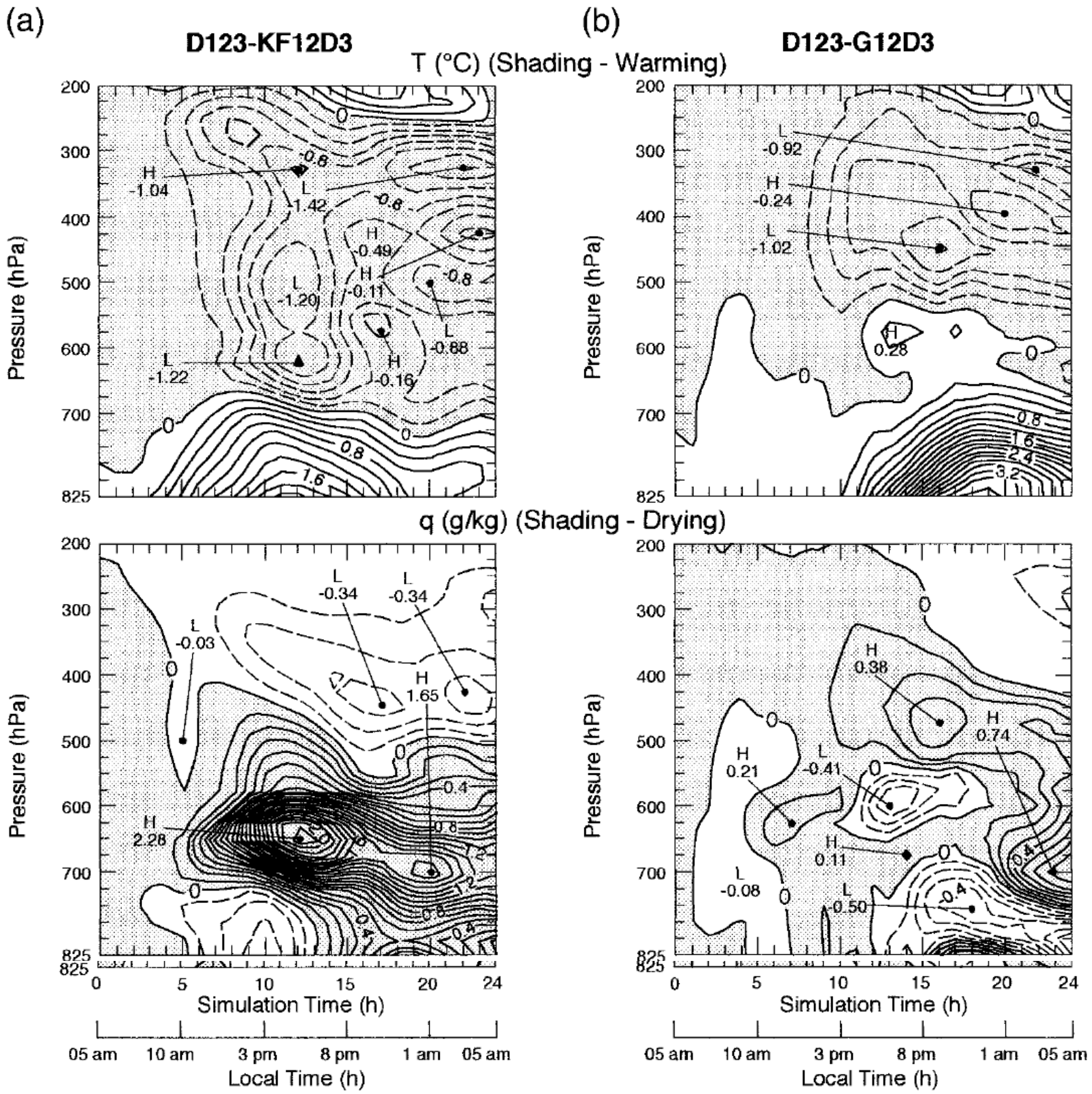


Simulation of nested domains, explicitly resolved convection on inner (3.3 km grid spacing, various parameterized convection on outer (10, 30 km).

Rainfall on inner domain affected by choice of what is done on outer domain. E123 is explicit on each domain, KF12E3 is Kain-Fritsch on 1&2, explicit on 3.

Warner and Hsu, *MWR*, July 2000

Problems caused by using outer domain convective parameterization with explicit convection in nest



In comparison with a dry simulation, the effect of parameterized convection on the thermodynamics of the interior grid is to heat the upper troposphere and dry the middle troposphere.

Lessons:

(1) If possible, use **large** convectively resolving domains.

(2) Develop/utilize convective parameterizations with physically reasonable mass-field responses.

(3) Develop ways of tuning convective parameterizations to minimize nonphysical competition between explicit and parameterized convection.

Verification of ensembles (very briefly).

- Summary measures:
  - Brier scores (BS) and skill scores (BSS)
  - Continuous ranked probability score (CRPS) and skill score (CRPSS).
  - Relative Operating Characteristic (ROC) and ROC skill score (ROCSS).
- Measures of reliability includes:
  - Reliability diagrams (slides 9)
  - Rank histograms (aka “Talagrand diagrams”)
  - ”Spread-skill” diagrams (slide 30)
- Measures of sharpness include:
  - Inset histogram on reliability diagrams (slide 11).
  - Indirectly, higher BSS, CRPSS (or lower BS, CRPS)

Verification resources

- Hamill presentation to NCAR workshop on ensemble verification, here.
- Beth Ebert / WMO web page on verification here.
- Dan Wilks' book chapter in "Statistical Methods in the Atmospheric Sciences"
- Jolliffe and Stephenson book, "Forecast Verification."

Displaying ensemble information: a few resources.

- SREF web page (http://www.spc.noaa.gov/exper/sref/)
- NCAR ensemble (http://ensemble.ucar.edu/index.php?d=2017031312)
- NWS situational awareness tables (http://ssd.wrh.noaa.gov/satable/)

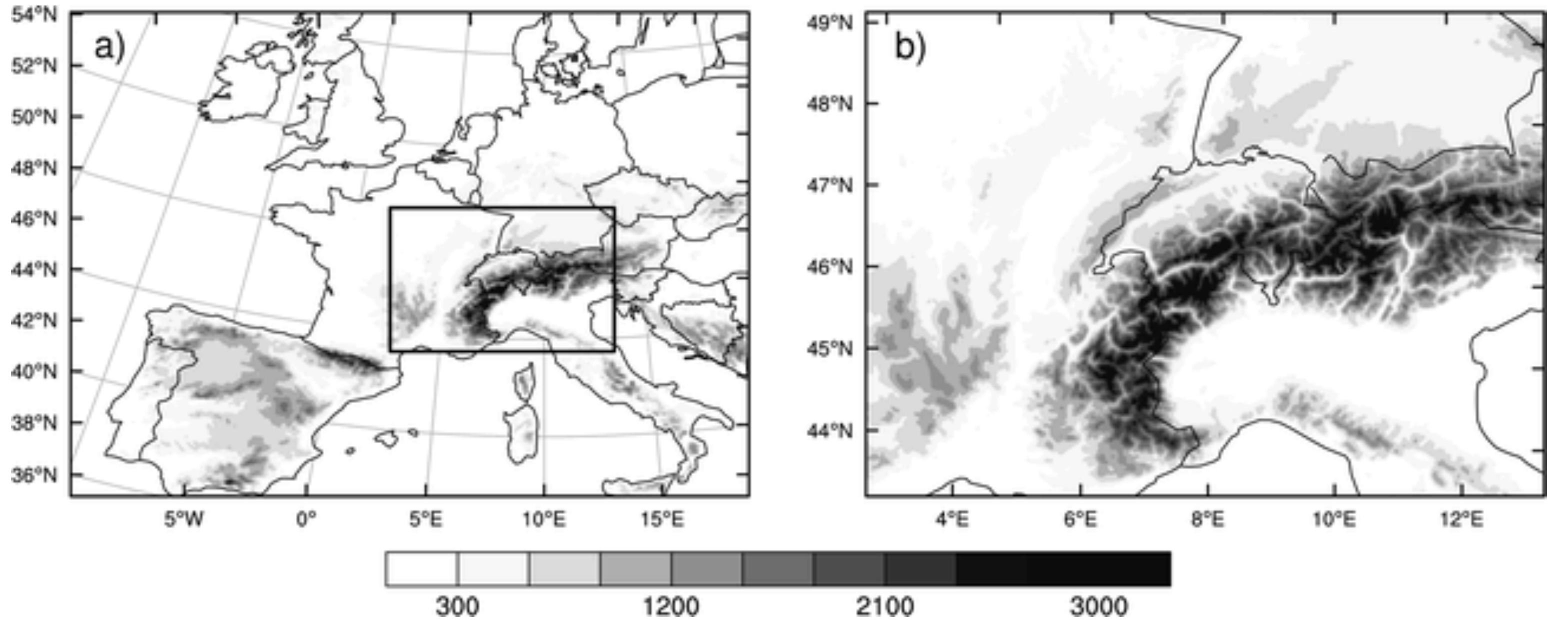
Conclusions

- In many (most) situations, practical weather prediction can be highly uncertain and should lead you to question whether you can get useful guidance from a single deterministic simulation.
- Ensembles are the common method for simulating this uncertainty.
- A properly constructed regional ensemble must deal with uncertainty that includes:
    - Initial-condition uncertainty.
    - Uncertainty of the model forecast itself (hard).
    - Overly confident estimates of forecast uncertainty due to boundary conditions.
- Questions?

Supplementary slides

Why is the predictability of small-scale phenomena poorer in weakly-forced situations?

Understanding predictable and less predictable intense precipitation events in the Alps



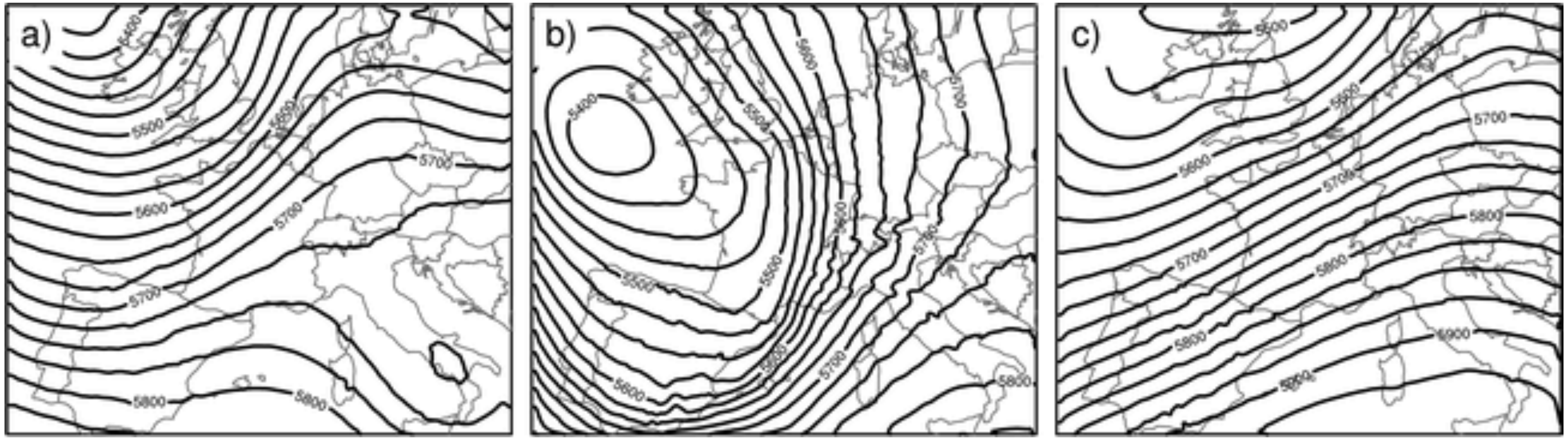
- Integration domains and topography (m) of the (a) 7- and (b) 2.2-km LM simulations. Six-member ensemble in the interior domain using shifting initialization times. LBCs for larger domain from ECMWF forecast.

Ref: Hohnegger et al., August 2006 *MWR*.

Understanding predictable and less predictable intense precipitation events in the Alps

500-hPa initial conditions for 3 cases

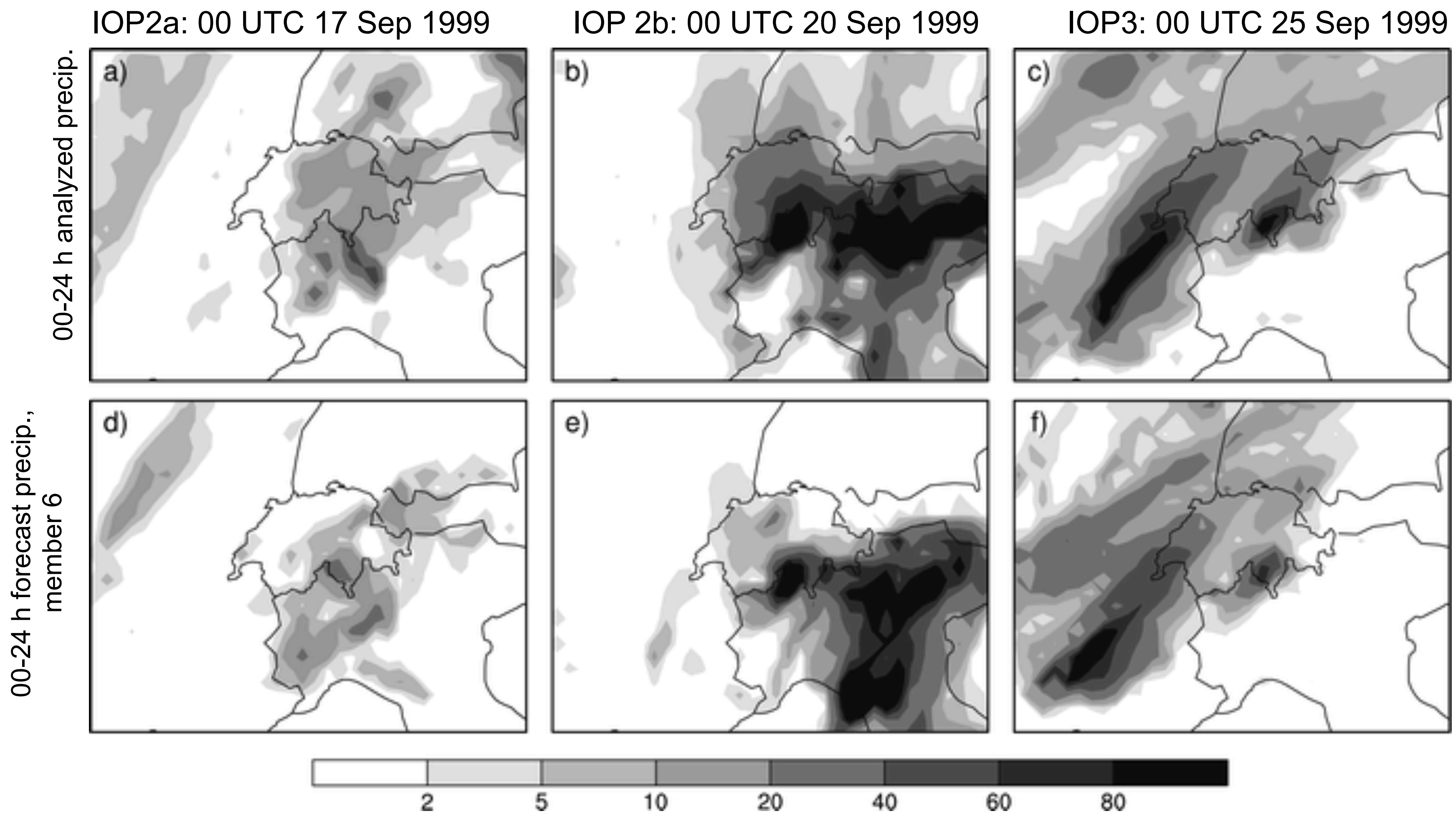
IOP2a: 00 UTC 17 Sep 1999 IOP 2b: 00 UTC 20 Sep 1999 IOP3: 00 UTC 25 Sep 1999



- Data from Mesoscale Alpine Program (MAP), Bougeault et al., *BAMS*, 2001.

Ref: Hohnegger et al., August 2006 *MWR*.

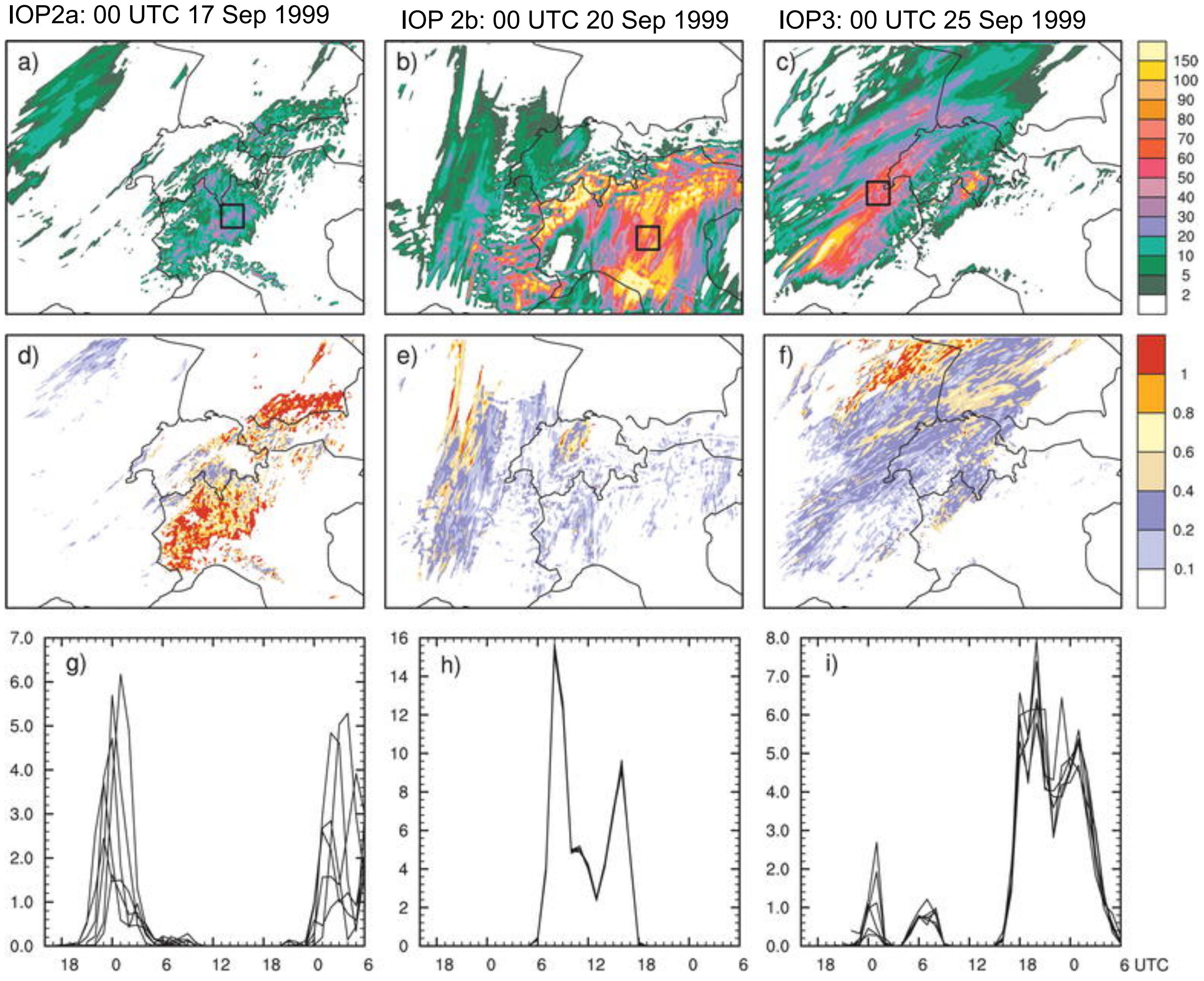
Understanding predictable and less predictable intense precipitation events in the Alps



- Reasonable correspondence between model forecast and analyzed precipitations.

Ref: Hohnegger et al., August 2006 *MWR*.

Understanding predictable and less predictable intense precipitation events in the Alps



30-h accumulated ensemble-mean precipitation (mm)

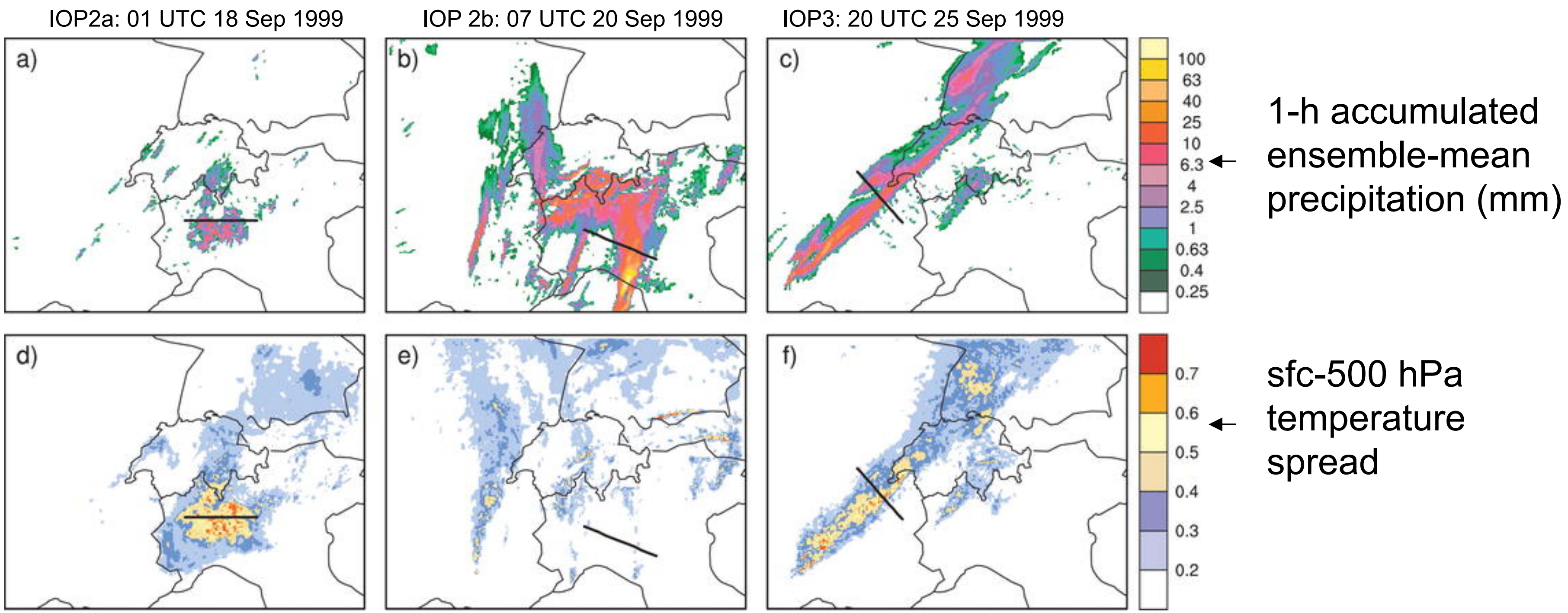
30-h normalized precipitation spread

time series for each member in boxed regions

- Normalized spread: IOP2a > IOP3 >> IOP2b. Why?

Ref: Hohnegger et al., August 2006 *MWR*

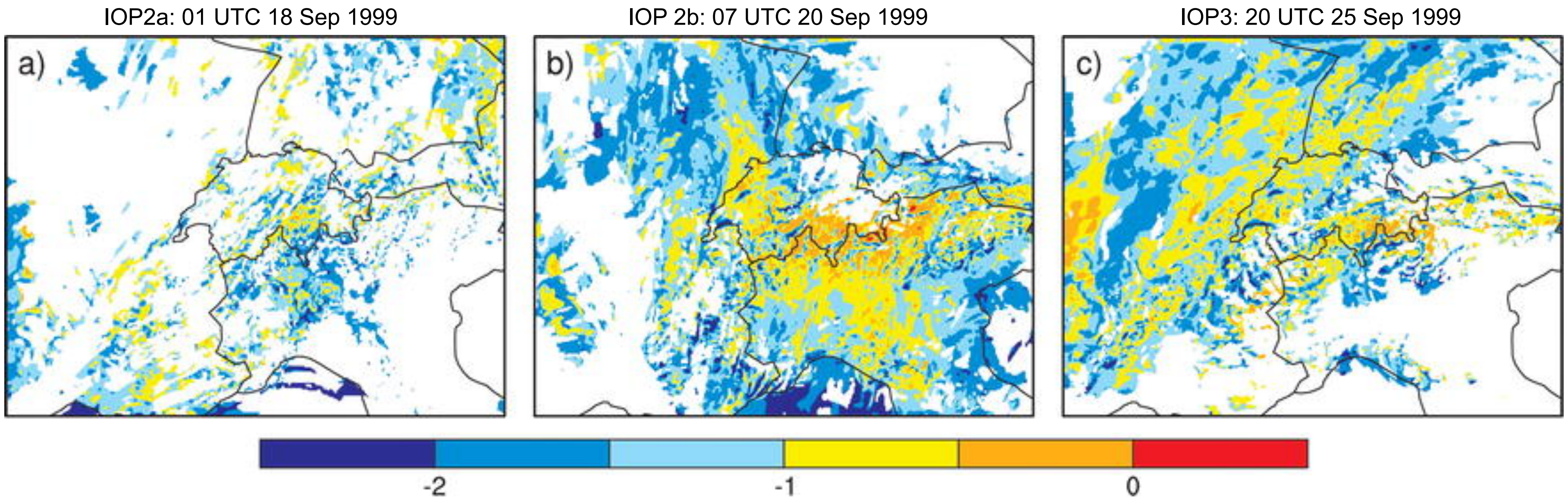
Understanding predictable and less predictable intense precipitation events in the Alps



- Temperature spread particularly small in IOP2b's precipitation region. Why?

Ref: Hohnegger et al., August 2006 *MWR*

Understanding predictable and less predictable intense precipitation events in the Alps

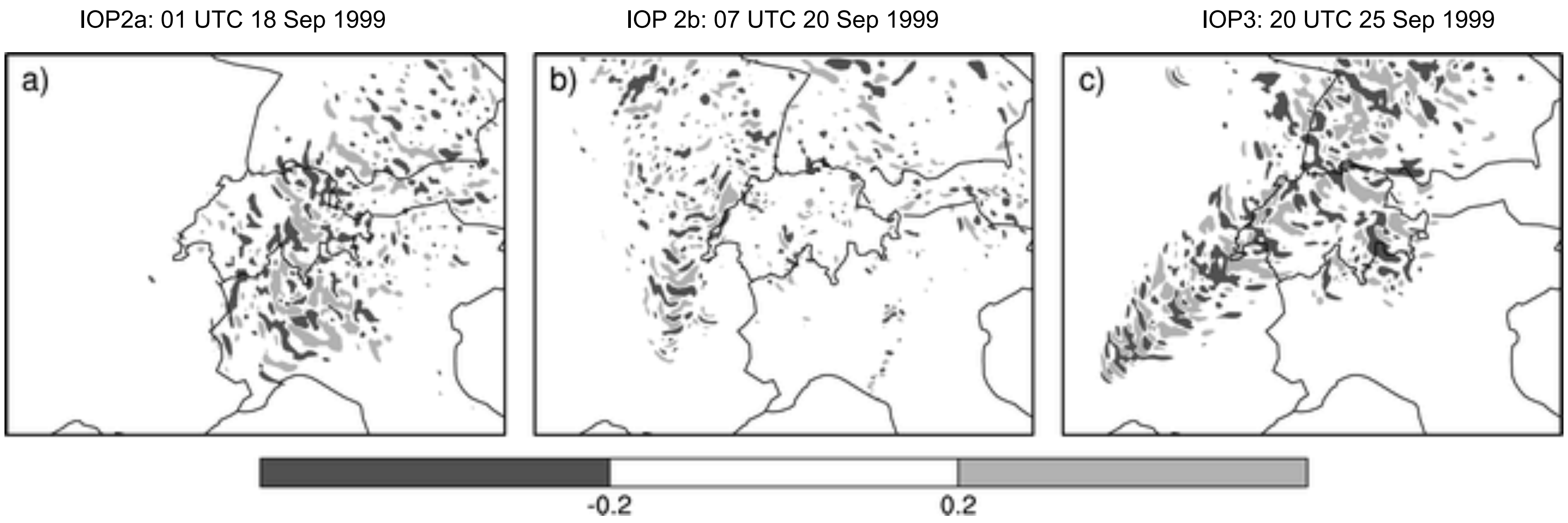


Vertical minimum of the moist Brunt–Väisälä frequency N^2_m (10−4 s−2) derived for ensemble member 6. Cloud-free grid points are masked in white.

- IOP2b has plenty of moist instability relative to the other IOPs, so instability is not the source of unpredictability.

Ref: Hohnegger et al., August 2006 *MWR*

Understanding predictable and less predictable intense precipitation events in the Alps

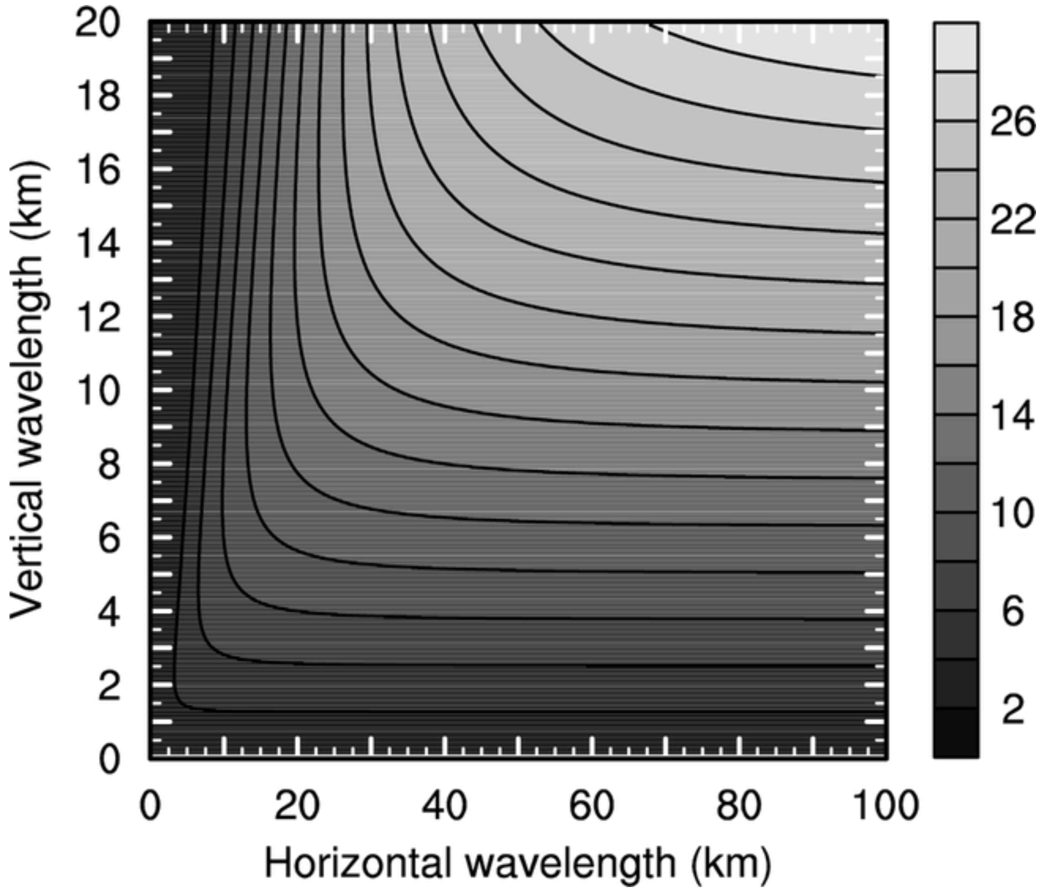


Temperature difference (K) between ensemble members 5 and 6 at a height of 13.6 km.

- Perturbations related to internal gravity wave activity.

Ref: Hohnegger et al., August 2006 *MWR*

Understanding predictable and less predictable intense precipitation events in the Alps



Theoretically derived critical wind speed U_{crit} (m s−1) allowing upstream propagation of energy as a function of horizontal and vertical wavelengths and for N = 0.01 s−1 [see Eq. (4)]

Consider propagation of gravity waves in a dry airstream, uniform stratification and windspeed. Linear analysis as in Holton text (2004, eq. 7.45a)

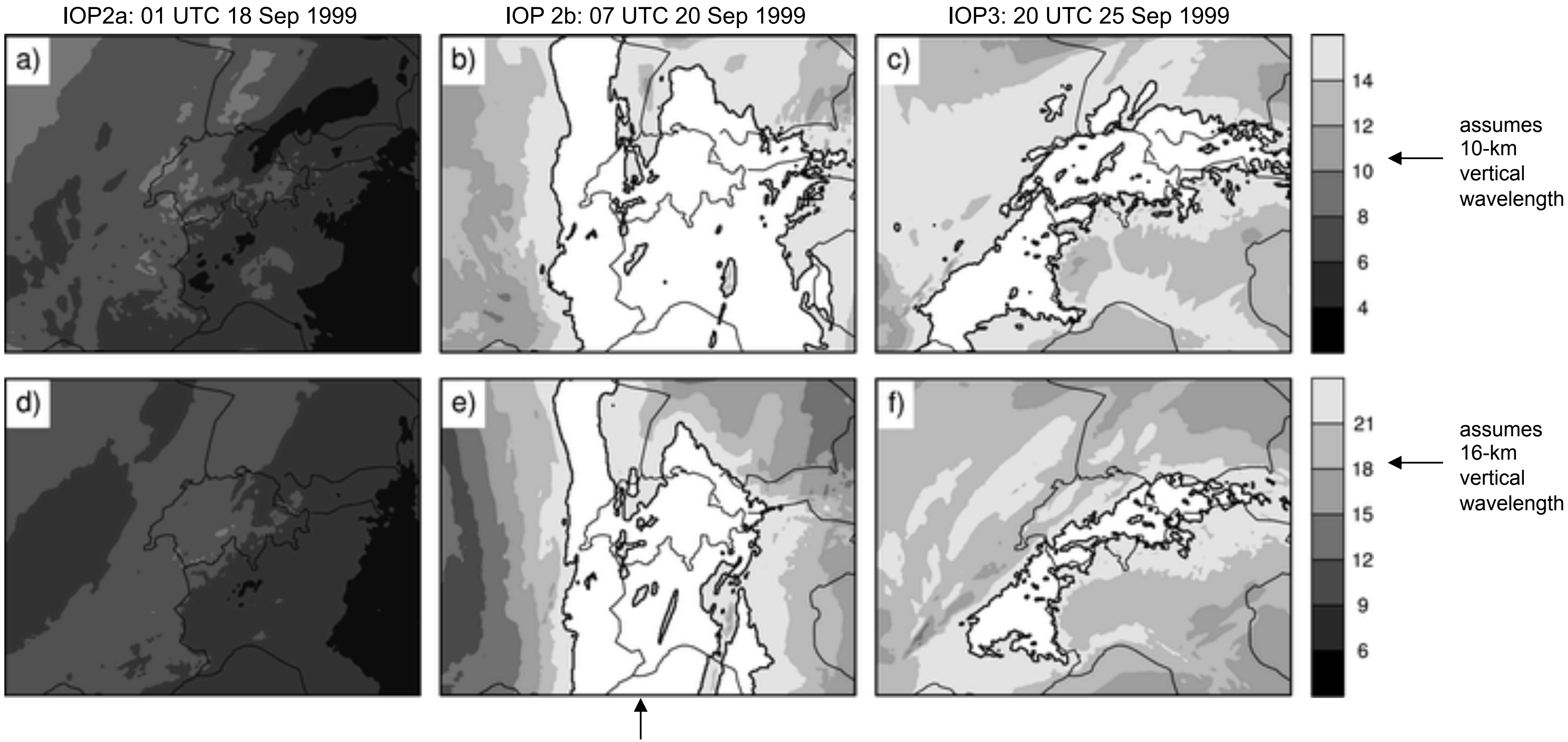
$$U_o \leq \frac{Nm^2}{(m^2 + k^2)\sqrt{m^2 + k^2}} = U_{\text{crit}}. \qquad (4)$$

where *k* and *m* are the vertical and horizontal wavenumber, N^2 is Brunt-Väisälä frequency. When windspeed is less than critical, gravity waves can propagate against mean flow and stay in source region long enough to grow, else they are swept out of growth region. Plot shows that deep gravity waves have higher critical speed threshold and can propagate upstream under broader range of conditions.

Ref: Hohnegger et al., August 2006 *MWR*

Understanding predictable and less predictable intense precipitation events in the Alps

Ensemble mean of the horizontal wind velocity U_o (m s−1). Values larger than U_{crit} inhibiting upstream energy propagation are masked in white. Values for U_o and N have been averaged over half a vertical wavelength.



IOP2b's winds above critical threshold, prohibiting local growth of perturbations from gravity-wave activity.

Ref: Hohnegger et al., August 2006 *MWR*

Synthesizing Hohnegger et al.

- Mesoscale perturbations get stimulated in regions of moist convection[1].
- Perturbations may grow locally if they can remain in a region of moist instability, reducing predictability. High wind speeds tend to sweep the nascent perturbations away from genesis region.[2]
- Reinforces hypothesis that mesoscale predictability is lengthened when large-scale forcing is strong.

[1] See also Zhang et al. 2003 *JAS*, Bei and Zhang, *QJRMS*, 2007.
[2] See also Huerre and Monkewitz 1985 *J. Fluid Mech*, Snyder and Joly 1998 *QJRMS,* and literature on "local baroclinic instability"

Bayesian data assimilation background

$$P(\mathbf{x}_t|\psi_t) \propto P(\psi_t|\mathbf{x}_t)P(\mathbf{x}_t)$$ Bayes rule

$$P(\psi_t|\mathbf{x}_t) = P(\mathbf{y}_t|\mathbf{x}_t)P(\psi_{t-1}|\mathbf{x}_t)$$ assuming independence of errors in time.

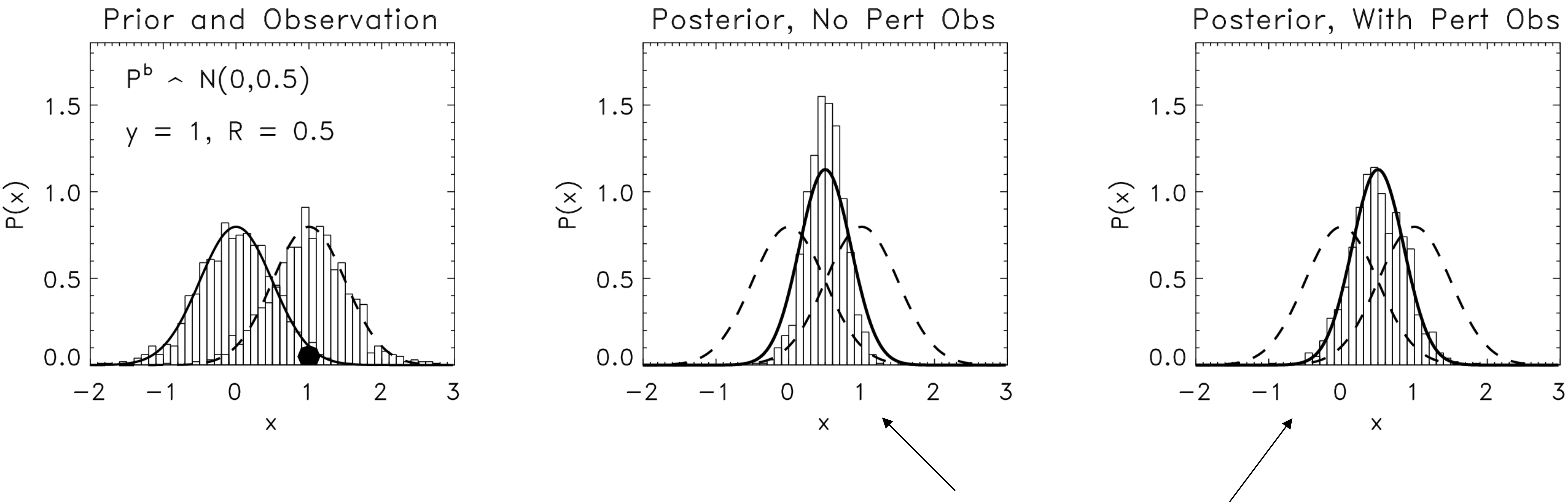
$$P(\mathbf{x}_t|\psi_t) \propto P(\mathbf{y}_t|\mathbf{x}_t)P(\psi_{t-1}|\mathbf{x}_t)P(\mathbf{x}_t)$$

$$P(\psi_{t-1}|\mathbf{x}_t)P(\mathbf{x}_t) = P(\mathbf{x}_t|\psi_{t-1})$$ Bayes rule

$$P(\mathbf{x}_t|\psi_t) \propto P(\mathbf{y}_t|\mathbf{x}_t)P(\mathbf{x}_t|\psi_{t-1})$$

[back]

Why perturb the observations?



histograms denote the ensemble values; heavy black line denotes theoretical expected analysis-error covariance

Model error at mesoscale:

Summary of microphysical issues in convection-resolving NWP

- Many fundamental problems in cloud microphysics are still unsolved.
- The lack of in-situ observations makes any progress very slow and difficult.
- Most of the current parameterization have been designed, operationally applied and tested for stratiform precipitation only.
- Most of the empirical relations used in the parameterizations are based on surface observation or measurements in stratiform cloud (or storm anvils, stratiform regions).
- Many basic parameterization assumptions, like N_0=const., are at least questionable in convective clouds.
- Many processes which are currently neglected, or not well represented, may become important in deep convection (shedding, collisional breakup, ...).
- One-moment schemes might be insufficient to describe the variability of the size distributions in convective clouds.
- Two-moment schemes haven‘t been used long enough to make any conclusions.
- Spectral methods are overwhelmingly complicated and computationally expensive. Nevertheless, they suffer from our lack of understanding of the fundamental processes.

from Axel Seifert presentation to NCAR ASP summer colloquium