Issues Regarding the Assimilation of Precipitation Observations

Dr. Ronald M. Errico

Goddard Earth Sciences and Technology Center (UMBC)
Global Modeling and Assimilation Office (NASA)

Outline

1. A general approach
2. Issues peculiar to precip. assim.
3. Examples of some issues
4. Examples of neglected fundamentals
5. Summary and recommendations

A Bayesian Approach

Information from observations: $\rho_o(\mathbf{y}^o|\mathbf{y}^d)$

Information from models $\rho_m(\mathbf{y}^d|\mathbf{H}(\mathbf{x}))$

Information from prior $\rho_p(\mathbf{x}|\mathbf{x}^b)$

$$\rho_a(\mathbf{x}|\mathbf{x}^b, \mathbf{y}^o, \mathbf{H}) = \text{const} \times \rho_p(\mathbf{x}|\mathbf{x}^b) \int_Y \rho_o(\mathbf{y}^o|\mathbf{y}^d)\, \rho_m(\mathbf{y}^d|\mathbf{H}(\mathbf{x}))\, d\mathbf{y}^d$$

If Gaussian input statistics, then Bayesian result is:

$$\rho_a(\mathbf{x}|\mathbf{x}^b, \mathbf{y}^o, \mathbf{H}) = \text{const} \times \exp\left[-\frac{1}{2} J(\mathbf{x})\right]$$

where

$$J(\mathbf{x}) = \left[\mathbf{x} - \mathbf{x}^b\right]^T \mathbf{B}^{-1} \left[\mathbf{x} - \mathbf{x}^b\right] + \left[\mathbf{H}(\mathbf{x}) - \mathbf{y}^o)\right]^T (\mathbf{E} + \mathbf{F})^{-1} \left[\mathbf{H}(\mathbf{x}) - \mathbf{y}^o\right]$$

Implications of the Bayesian Approach

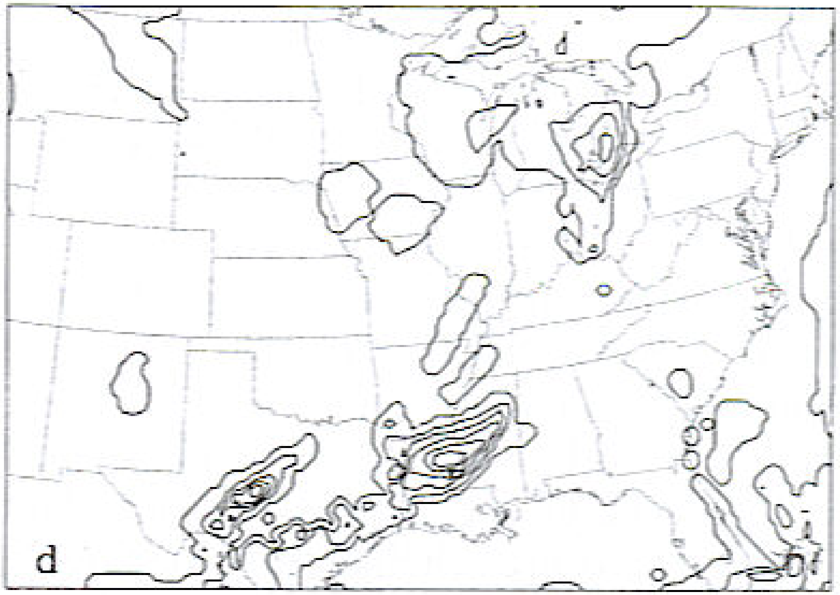
1. Unless the underlying distributions are simple, the problem is computationally intractable for large problems.
2. We see how the different information should be optimally combined.
3. We see what statistical knowledge is required as input.
4. We see that **E** and **F** may be equally important.
5. Results may depend on shapes of distributions, not only their means and variances.
6. We see that selection of a “best” analysis can be somewhat ambiguous.
7. Multi-modality of the PDF can occur, particularly due to model non-linearity.
8. While an explicit Bayesian approach may be impractical, the Bayesian implications of other techniques should be considered.

Some issues peculiar to precipitation assimilation

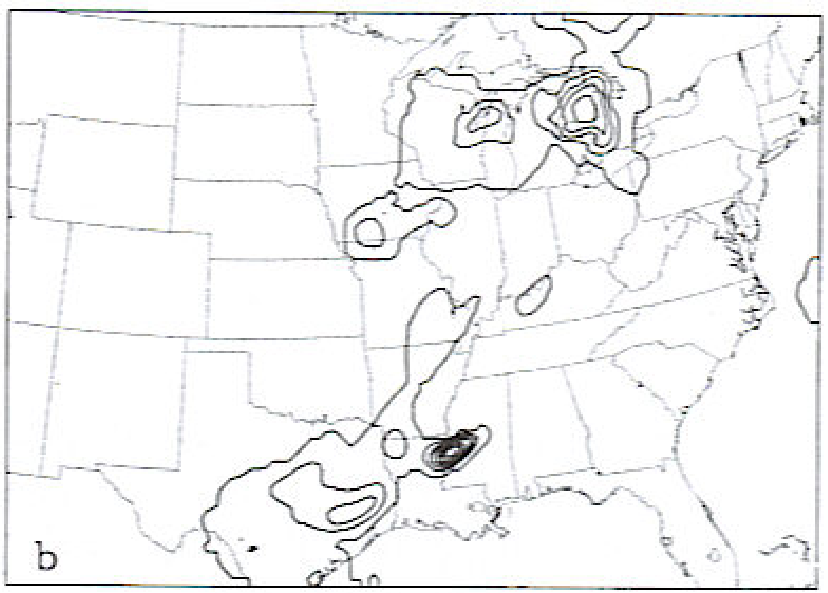
1. Distribution of precipitation errors is likely non-Gaussian.
2. Precipitation forward model error is likely non-negligible.
3. Multi-modal cost functions are likely.
4. Minimization of cost function may be poor objective.
5. Results may be very sensitive to prior statistics.
6. Straight-forward adjoint models of convection may be useless.
7. Common descent algorithms may be useless.
8. Results may be sensitive to precipitation type.
9. Possible incompatibility with gravity wave constraints.

Example of Model Error:
Errico et al. *QJRMS* 2001

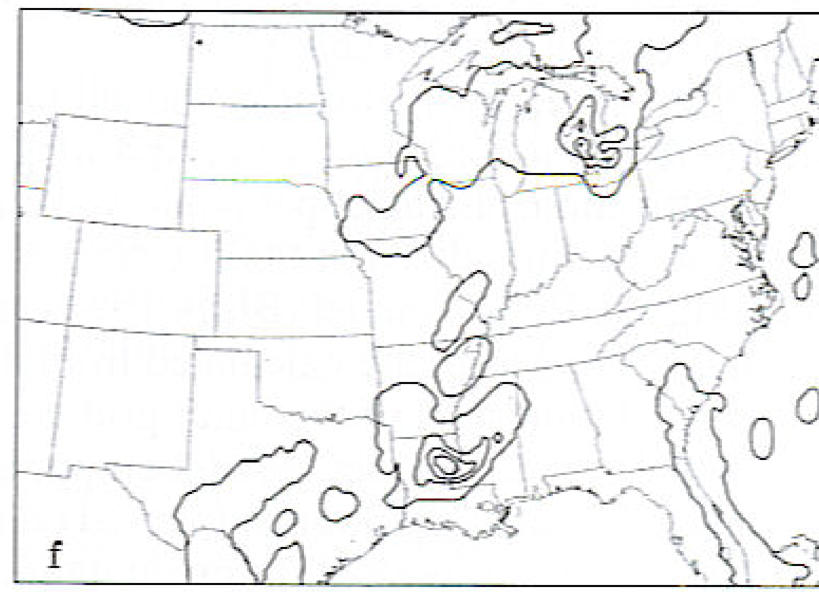
6-hour accumulated precip.
With 3 versions of MM5
Contour interval 1/3 cm



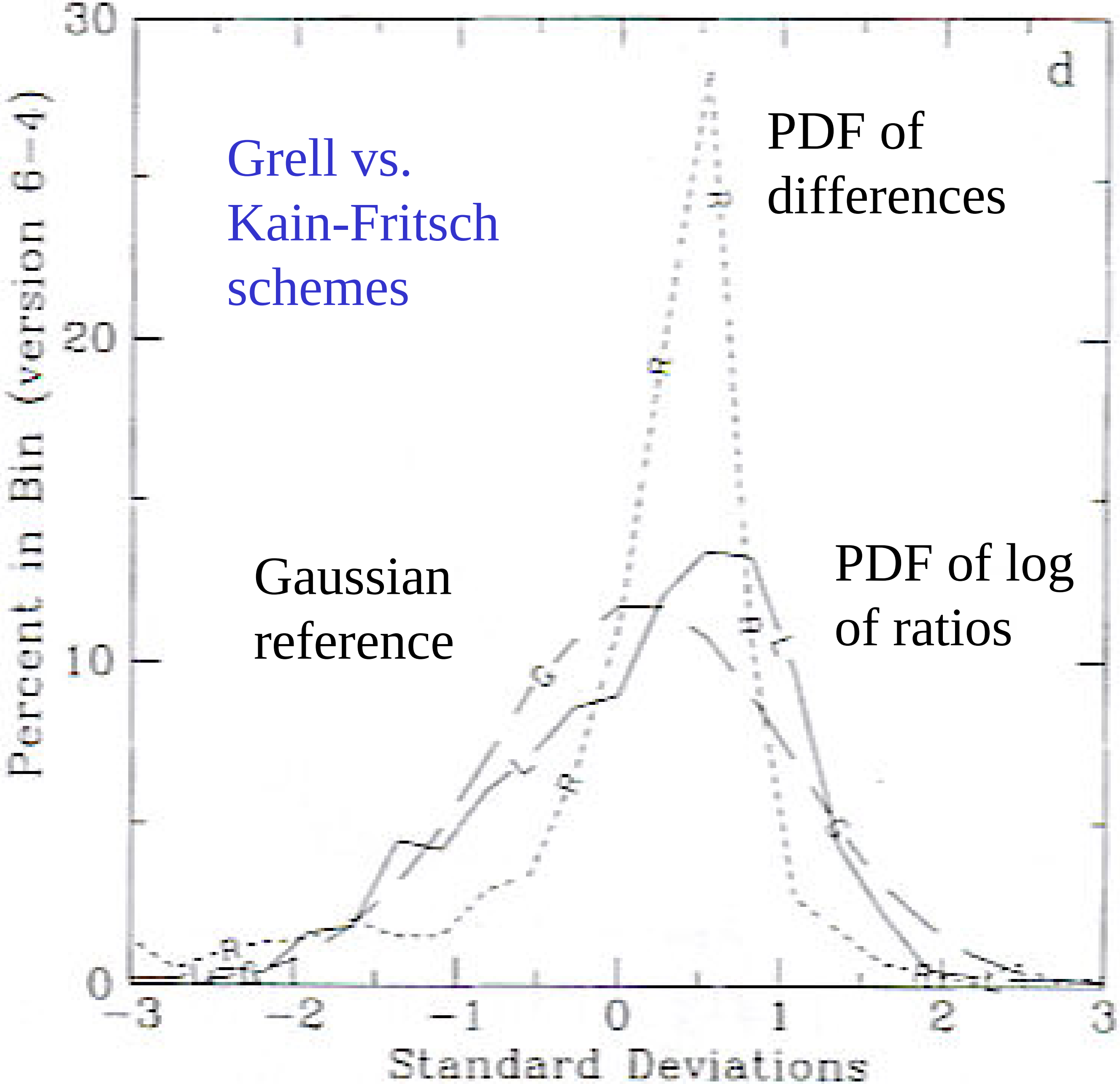
Kain - Fritsch



Betts - Miller



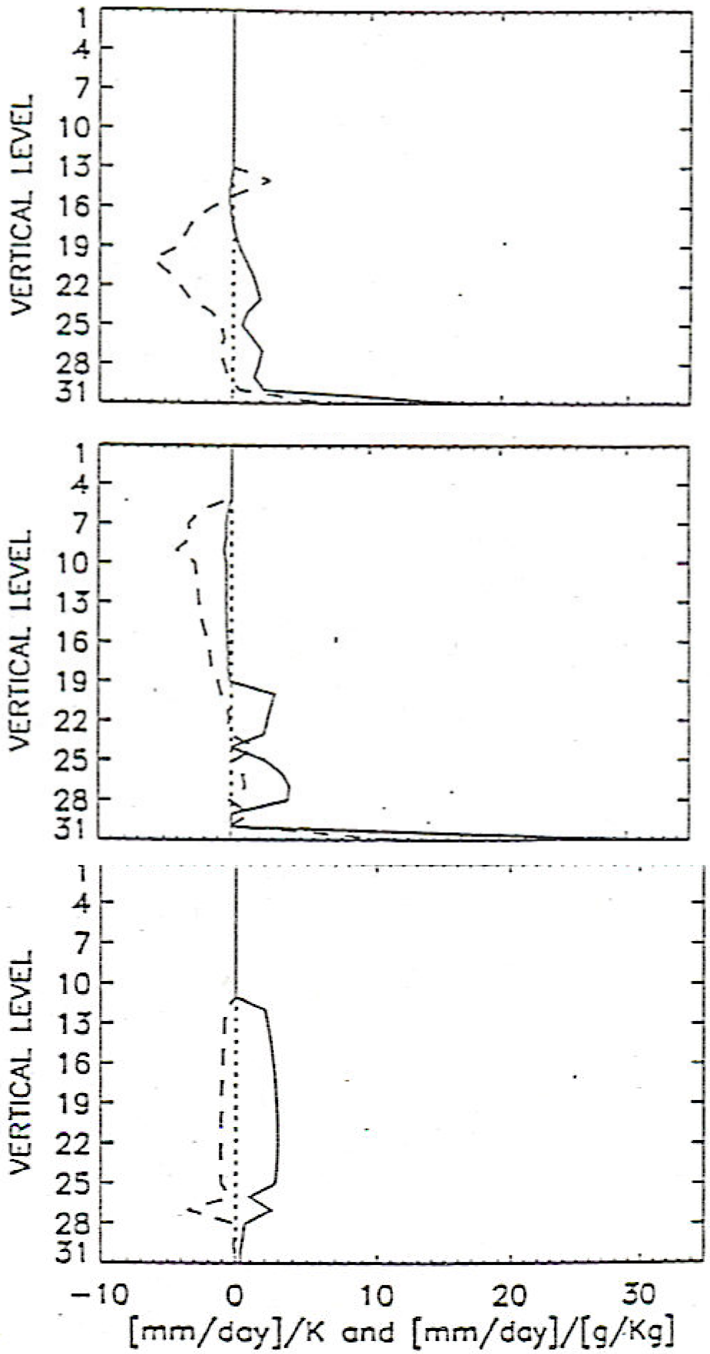
Grell



PDFs of
model
“errors”

Errico et al.
2001 *QJRMS*

Jacobians of Precipitation



RAS scheme

ECMWF scheme

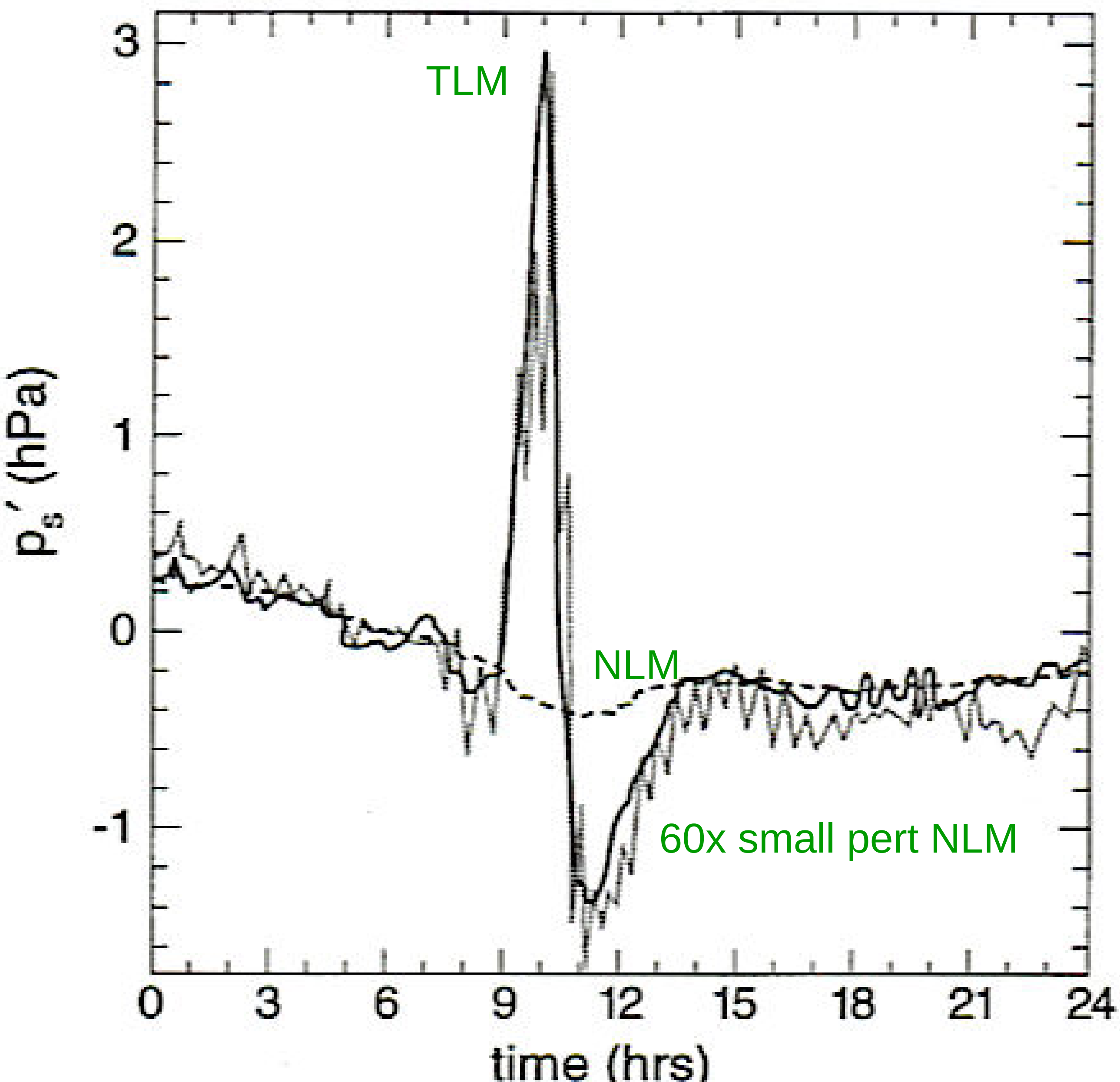
BM scheme

$$\frac{\partial R}{\partial T} \text{ dashed}$$

$$\frac{\partial R}{\partial q} \text{ solid}$$

Fillion and Mahfouf 1999 *MWR*

Tangent linear vs. nonlinear model solutions

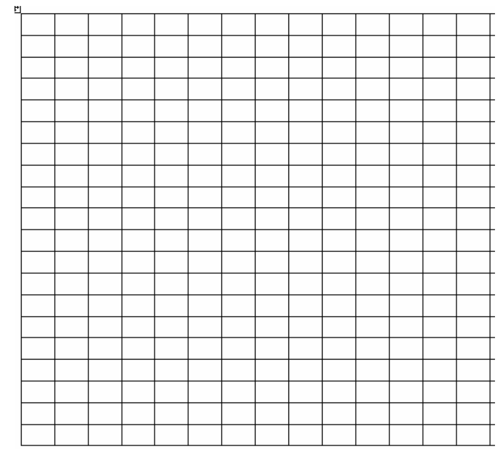


Errico and Raeder 1999 *QJRMS*

Statistically-Based Sub-Grid Parameterization

Model for small volume of mass Δm:

$$r_i = \begin{cases} a_i[q_i - q_s(T_i, p_i)] + \epsilon_i & \text{if } q_i > q_s(T_i, p_i) \\ \varepsilon_i & \text{otherwise} \end{cases}$$



Consider average over large volume V

$$\overline{r} = \frac{1}{I\Delta m}\sum_{i=1}^{I} r_i \Delta m$$

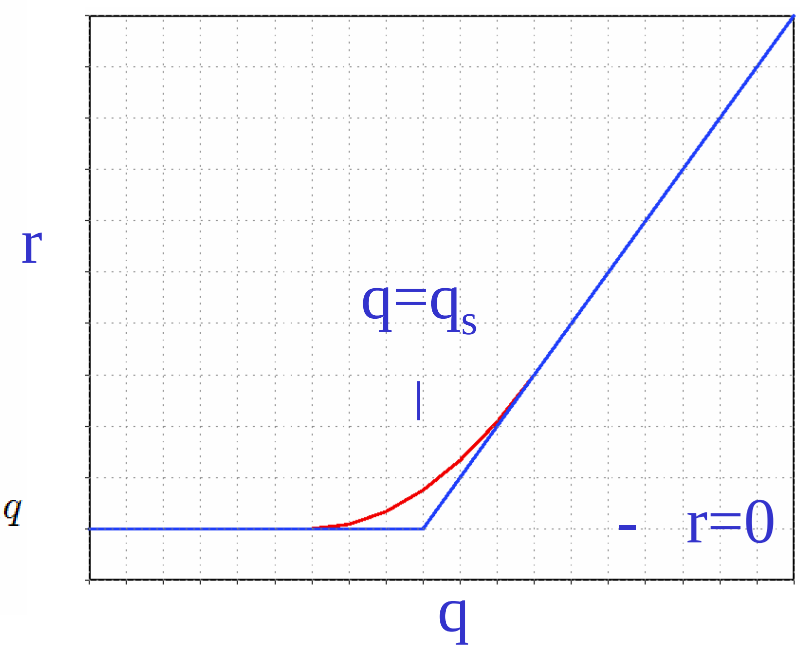
In general, $\overline{r} \neq r_m(\overline{q}, \overline{T}, \overline{p})$

Consider a uniform distribution of $q - q_s$ within V:

$$-\Delta q \leq q_i - q_s(T_i, p_i) \leq \Delta q$$

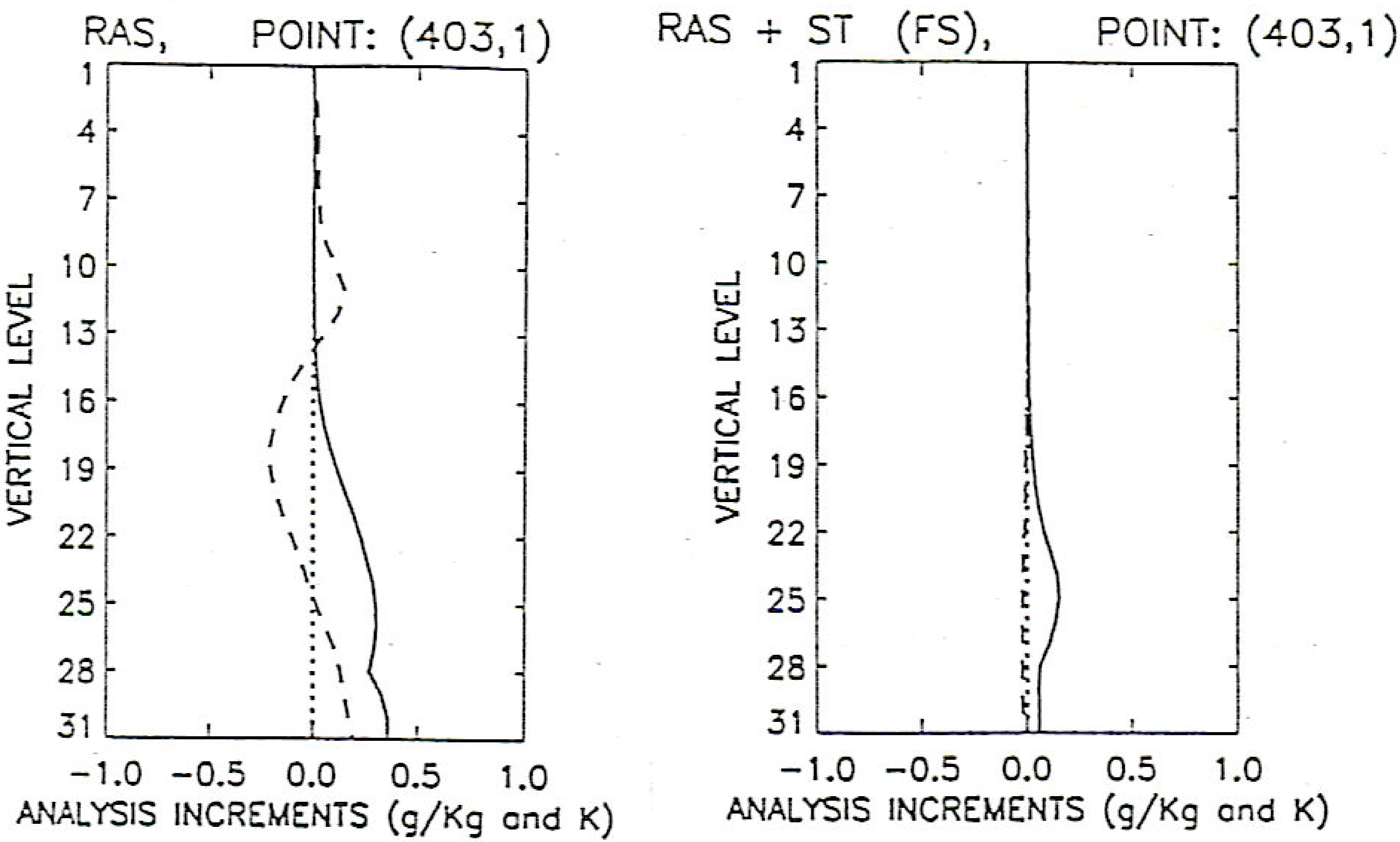
A new model:

$$r_{m*} = \begin{cases} a(\overline{q} - \overline{q_s}) & \text{if } \overline{q} - \overline{q_s} > \Delta q \\ 0 & \text{if } \overline{q} - \overline{q_s} < -\Delta q \\ \frac{a}{4\Delta q}(\overline{q} - \overline{q_s} + \Delta q)^2 & \text{otherwise} \end{cases}$$



Dependence on precipitation type

Fillion and Mahfouf 1999 *MWR*



Adjoint-derived, optimal perturbations

Errico, Raeder and Fillion, 2003 *Tellus*

Consider $J = J(\mathbf{x})$

Determine initial perturbation $\mathbf{x}'$ that maximizes:

$$J' = \left(\frac{\partial J}{\partial \mathbf{x}}\right)^{\mathrm{T}} \mathbf{x}'$$

Given initial constraint:

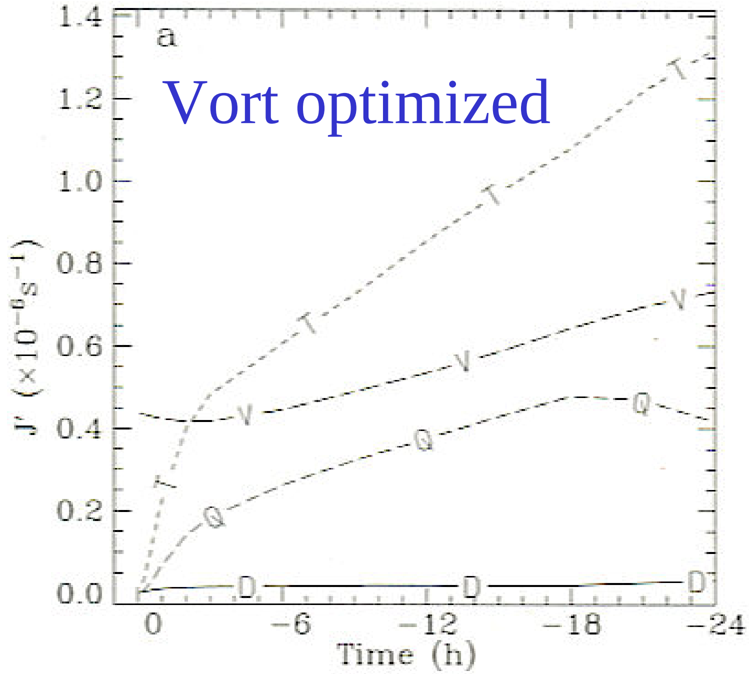
$$C = \frac{1}{2}\mathbf{x}'^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{x}'$$

Solution:

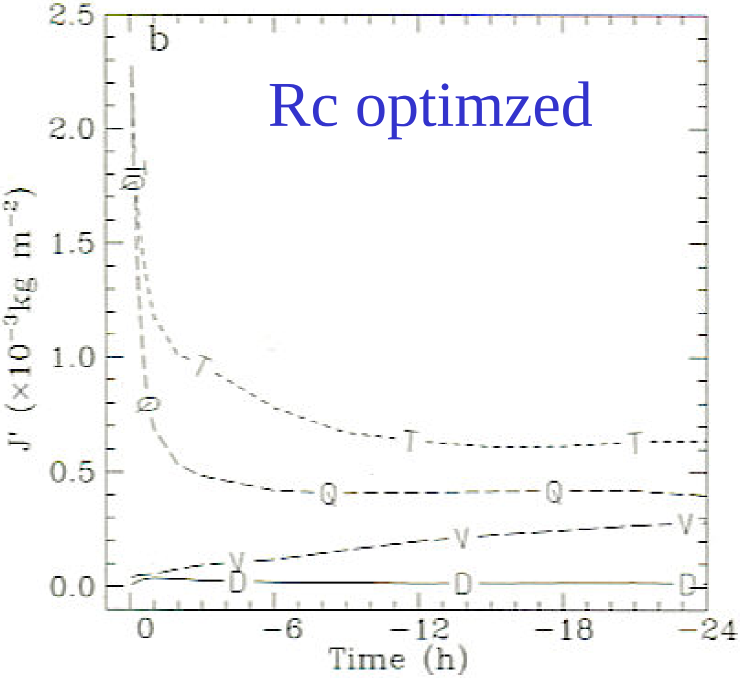
$$\mathbf{x}' = \lambda^{-1}\mathbf{B}\frac{\partial J}{\partial \mathbf{x}}$$

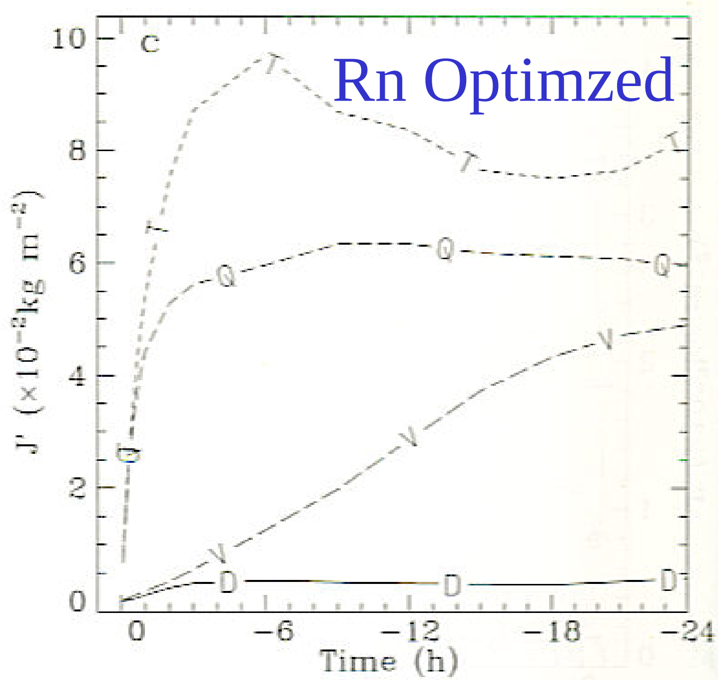
$$\mathrm{Max}(J') = \sqrt{2C\left(\frac{\partial J}{\partial \mathbf{x}}\right)^{\mathrm{T}}\mathbf{B}\frac{\partial J}{\partial \mathbf{x}}}$$

Errico et al. 2003 *Tellus*



Impacts for adjoint-derived optimal perturbations for forecasts starting indicated hours in the past.





The apparent neglect of many fundamentals

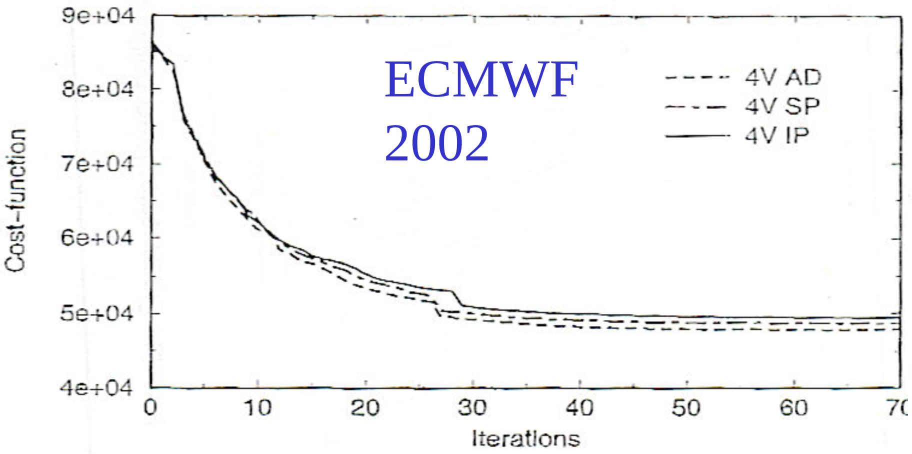
Few statistical considerations
- background estimates ignored
- background error correlations ignored
- observations considered too accurate (and Gaussian)
- representativeness (forward model) error ignored

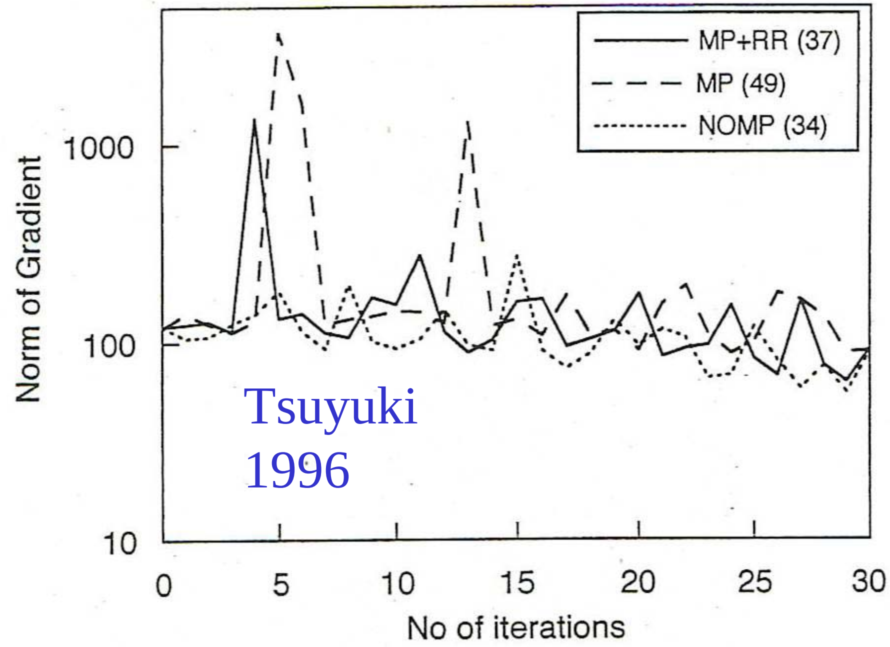
Few balance considerations
- univariate error statistics
- unbalanced reference states

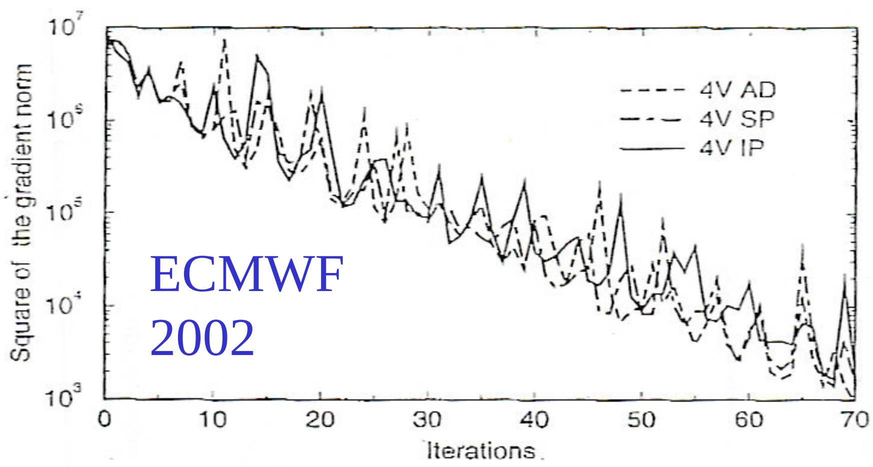
Limited evaluation
- limited cases
- limited measures

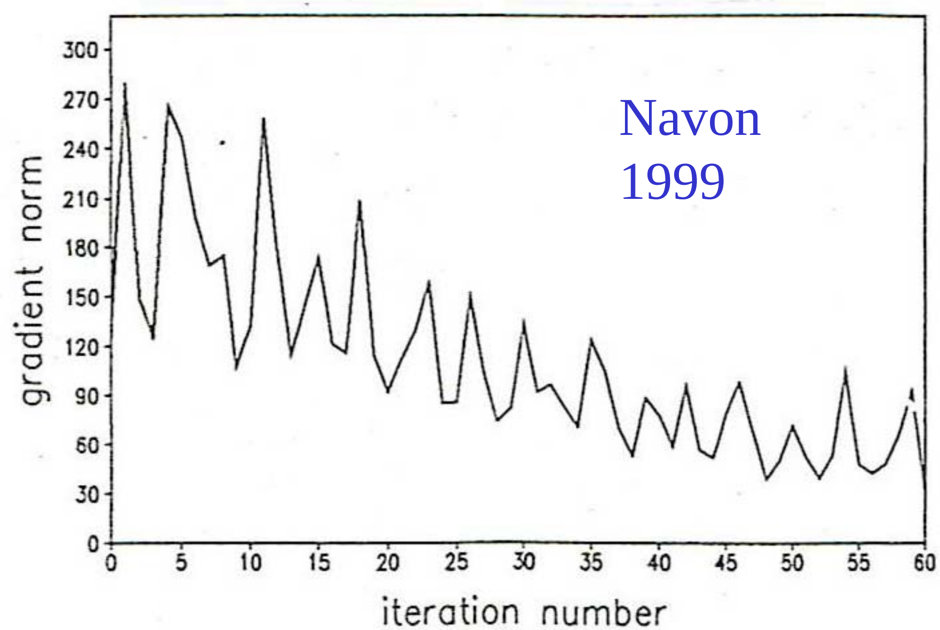
Some strange results
- ultra rapid convergence rates
- mis-characterization of sizes of terms
- little decrease of norm of J-gradient

Convergence of 4DVAR









Summary

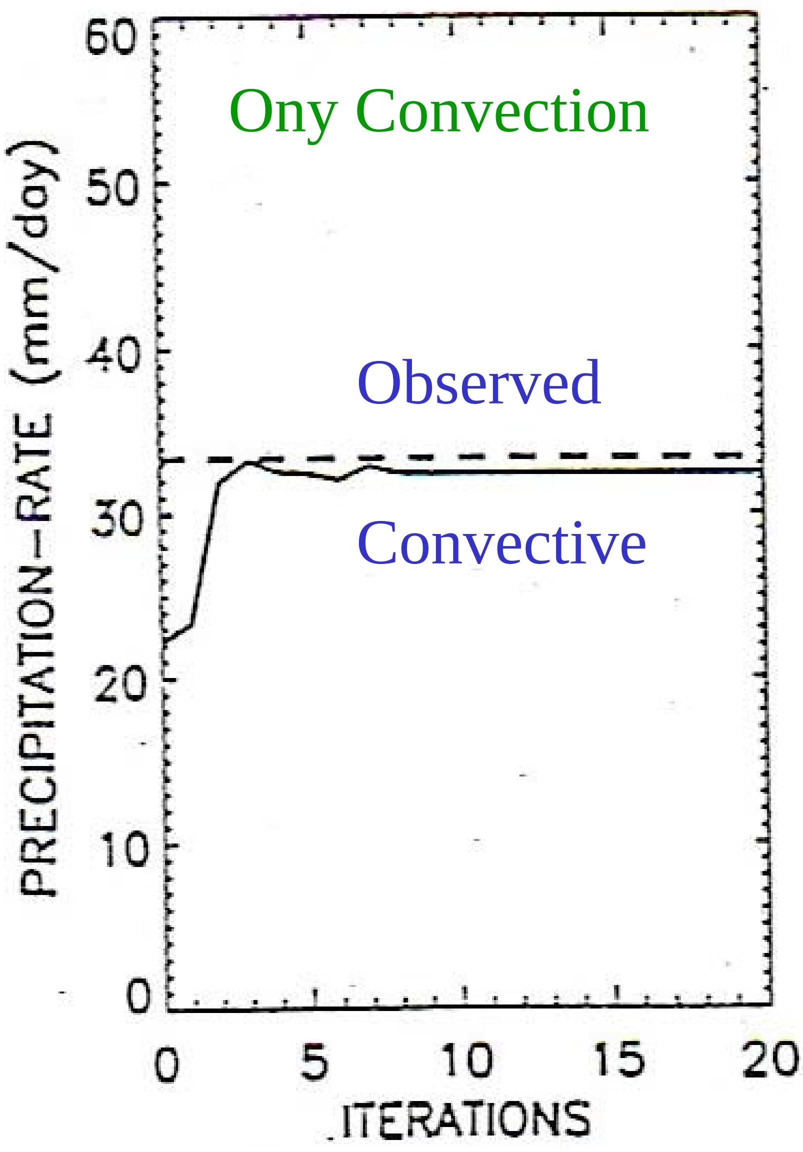
1. I am confused!
2. How can so many apparently fundamental aspects of the problem be neglected, yet such good results be reported?
3. Only in rare cases is enough information provided to help explain question 2.
4. Since only an improvement over some baseline is required, it is not necessary that “correct” procedures are used, just “useful” ones.
5. Successes may reveal more about the baseline results than about the correctness of a new assimilation procedure.
6. With both observation and forward model errors likely very large, what should be a realistic expectation of the usefulness of precipitation information and how can this be realized?

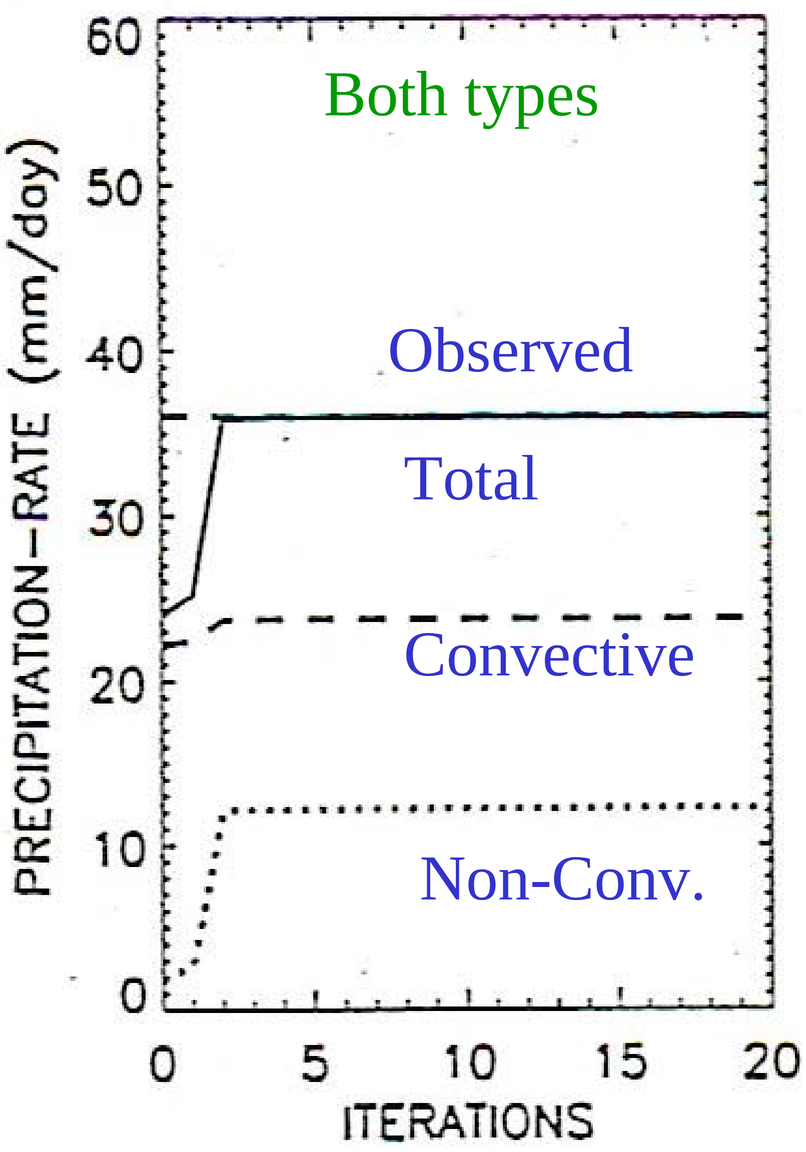
Recommendations

1. Be skeptical.
2. Ask lots of questions.
3. Consider Bayesian implications.
4. Determine reasonable error estimates.
5. Estimate what issues are generally important.
6. Explain results.
7. Encourage research at research institutions.
8. Entrain some interested experts.

Dependence on precipitation type

Fillion and Mahfouf 1999 *MWR*





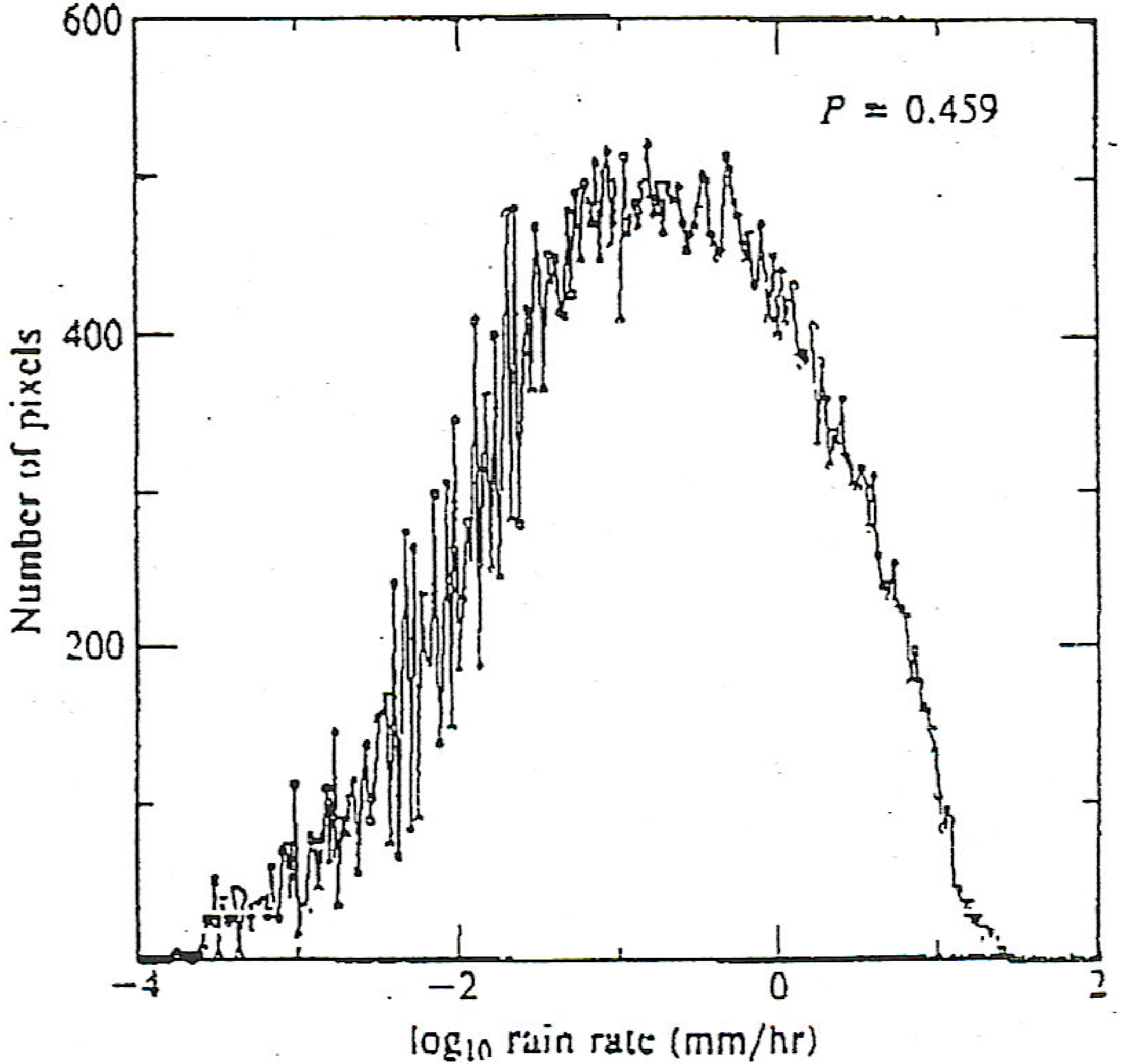


FIG. 2. A histogram of $\log_{10}$ of the rain rate obtained from a large number of 40 × 40 km^2 GATE pixels.

The distribution of precipitation

Kedem and Chiu 1986

Special treatment of 0

From Errico et al. 2000 *QJRMS*

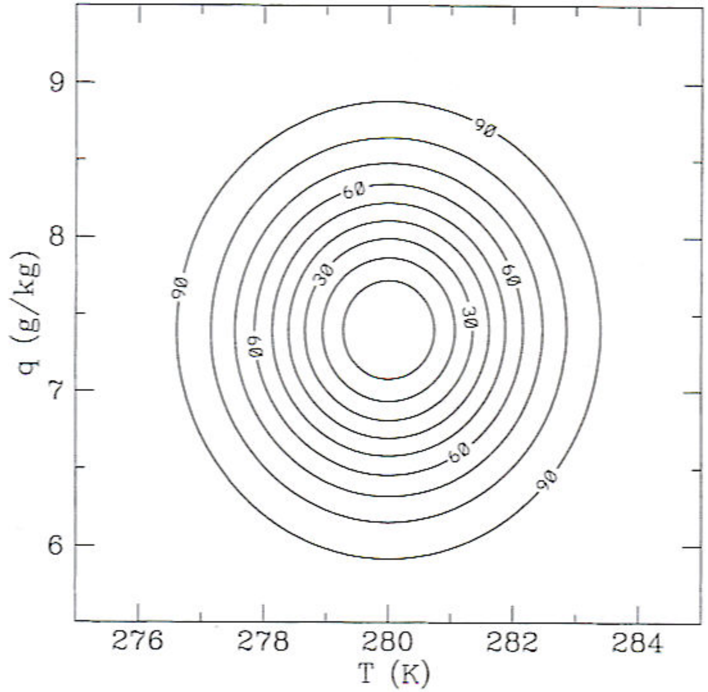
$$P(x = 0|y) = P_0(y)$$

$$P(0 < x \le x_2|y) = [1 - P_0(y)]F(x_2, y)$$

$$P_0(y) = \alpha \exp(-y/\beta)$$

$$F(x_2, y) = \int_0^{x_2} \rho(x|y)\ dx$$

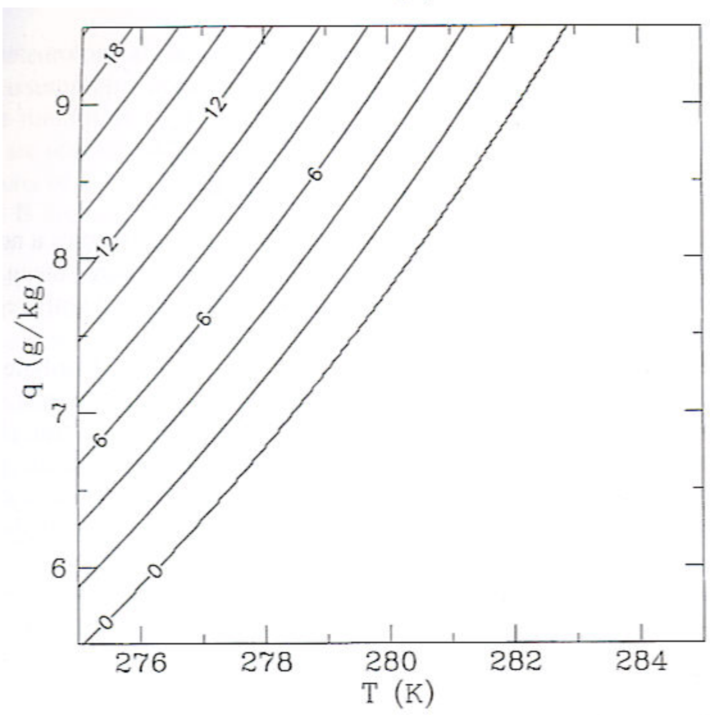
$$\rho(x|y) = \frac{1}{(2\pi)^{\frac{1}{2}}\, sx} \exp\left[-0.5 s^{-2} \left[\frac{\ln(x)}{\ln(\max[y, c])}\right]^2\right]$$



A Bayesian Example

Errico et al. 2000 *QJRMS*

Pdf of prior information



Forward Model

$$r_m(T,q) = \begin{cases} a(T,p_o)\left[q - q_s(T,p_o)\right] & \text{if } q > q_s(T,p_o) \\ 0 & \text{otherwise} \end{cases}$$

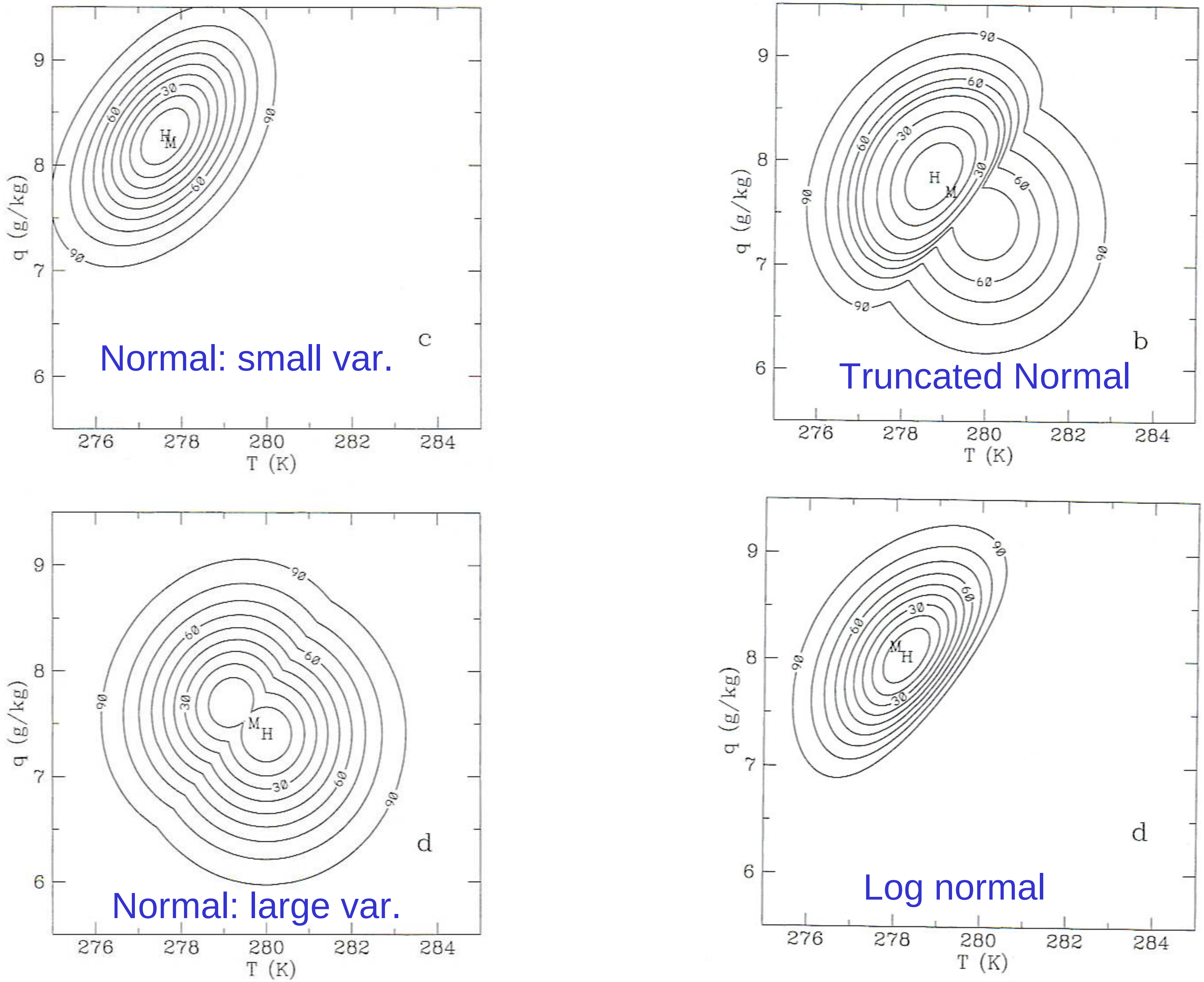
$$q_s(T,p_o) = 0.622\ \frac{v_p(T)}{p_o - v_p(T)}$$

$$v_p(T) = 611.2 \exp\left[17.67 \frac{T - 273.15}{T - 29.65}\right]$$

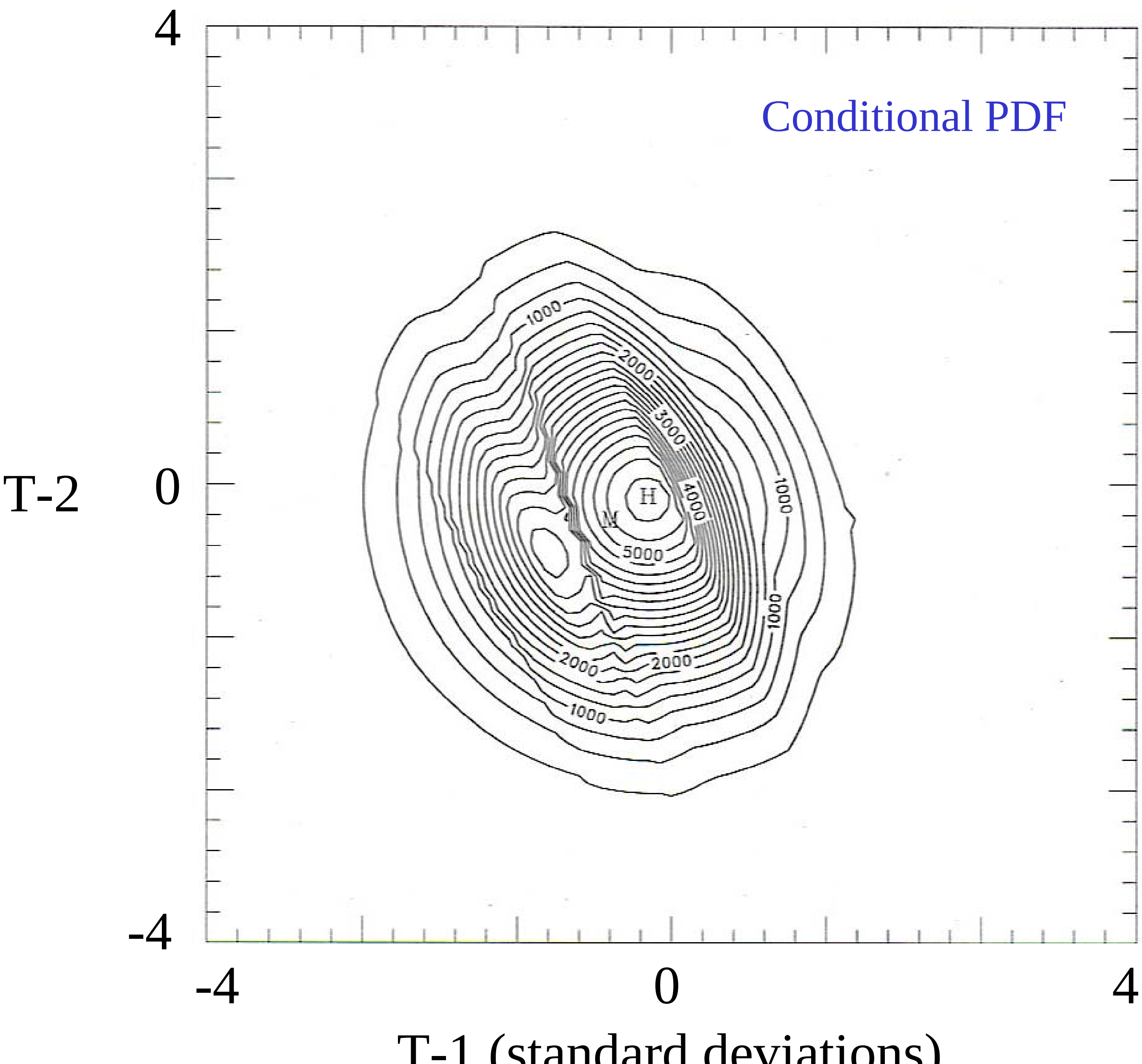
$$a(T,p) = \left(1 + \frac{L}{c_p}\frac{\partial q_s}{\partial T}\right)^{-1}$$

Dependence of posterior PDF on Obs. Error Distribution

Errico et al. 2000 *QJRMS*

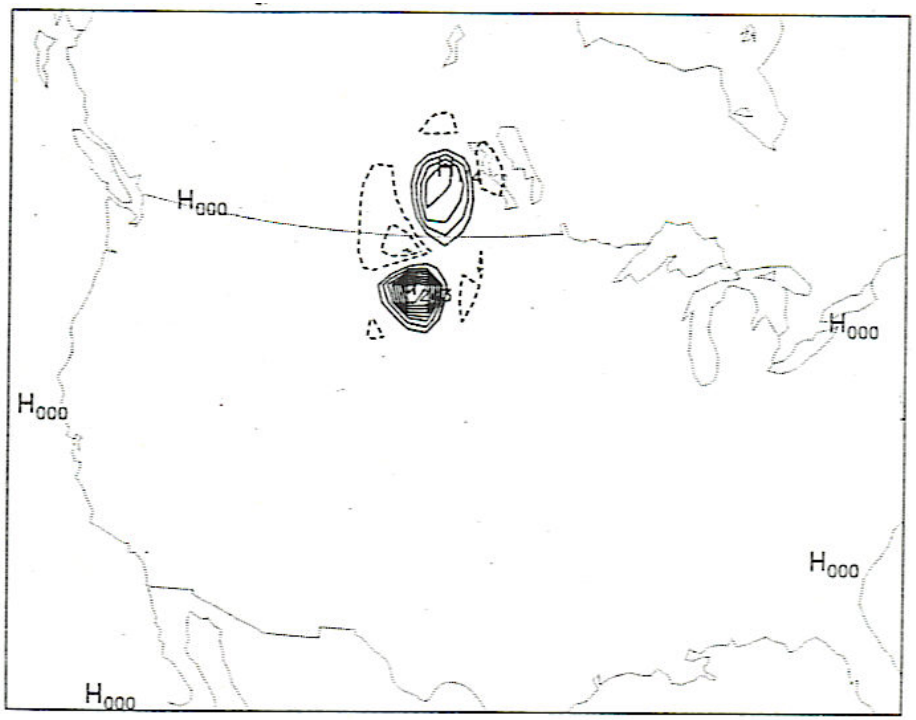
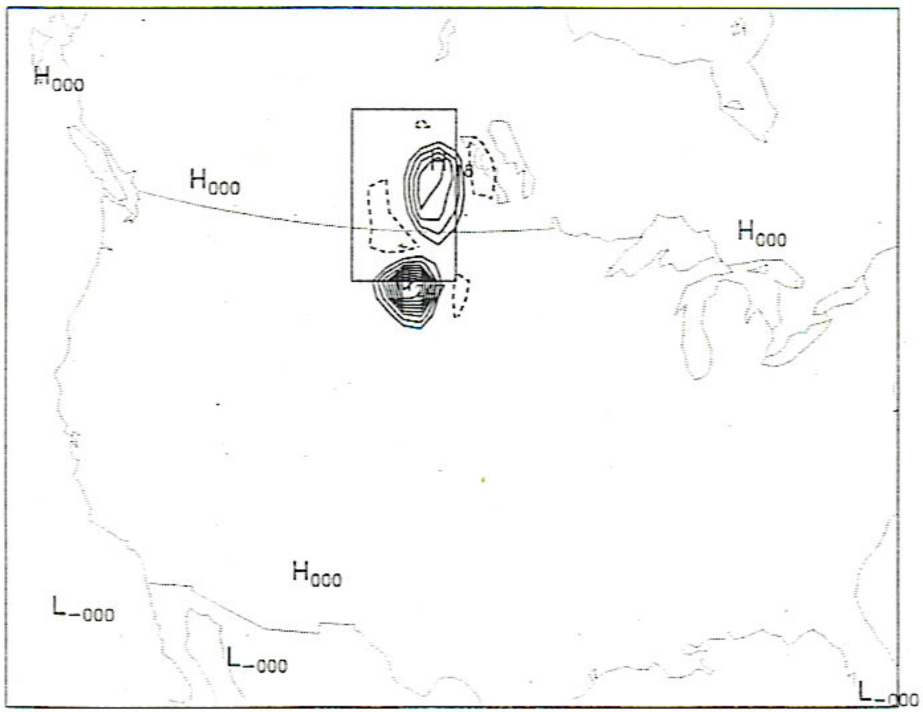


Posterior (analysis) PDF of 1DVAR of Convection



$$\partial(\text{Precipitation})/\partial T(500\text{hPa})$$

6-hour forecast

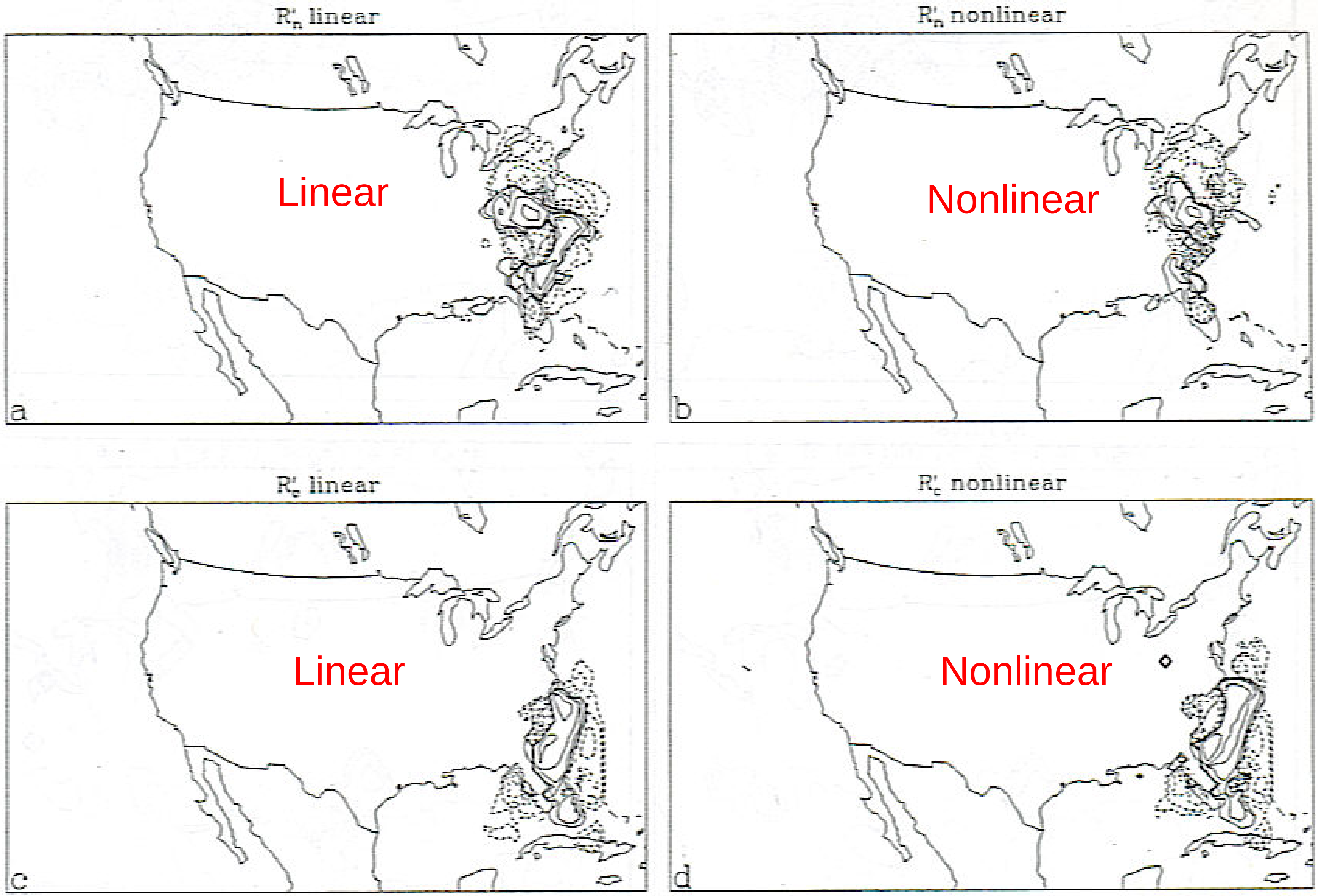
Full Field	Gravitational-mode field
	

Contour interval 0.0025 mm/K

Comparison of TLM and Nonlinearly Produced Precip Rates
12-Hour Forecasts with SV#1

Non-Convective

Convective.



Errico et al.
QJRMS 2004

Contours: 0.1, 0.3, 1., 3., 10. mm/day

Linear vs. Nonlinear Results: 12-hour SV

Linear

Nonlinear

Non-Convective Precip.

Convective Precip.