Introduction to Adjoint Models

Ronald M. Errico

Goddard Earth Sciences Technology And Research Center
Morgan State University
Global Modeling and Assimilation Office,
NASA Goddard Space Flight Center

Outline

1. Sensitivity analysis
2. Examples of adjoint-derived sensitivity
3. Development of adjoint model software
4. Validation of an adjoint model
5. Use in optimization problems
6. Misunderstandings
7. References

Sensitivity Analysis:
The basis for adjoint model applications

Adjoints in simple terms

Adjoint Sensitivity Analysis for a Discrete Model

The Problem to Consider:

A possibly nonlinear model:

$$\mathbf{y} = \mathbf{m}(\mathbf{x}) \tag{1}$$

A differentiable scalar measure of model output fields:

$$J = J(\mathbf{y}) \tag{2}$$

The result of input perturbations

$$\Delta J = J(\mathbf{x} + \mathbf{x}') - J(\mathbf{x}) \tag{3}$$

A 1st–order Taylor series approximation to ΔJ

$$J' = \sum_i \frac{\partial J}{\partial x_i} \, x_i' \tag{4}$$

The goal is to efficiently determine $\frac{\partial J}{\partial x_i}$ for all i

Adjoint Sensitivity Analysis for a Discrete Model

The Tangent Linear Model (TLM)

Apply a 1st–order Taylor series to approximate the model output

$$y_i' = \sum_j \frac{\partial y_i}{\partial x_j} x_j'$$

$\partial y_i / \partial x_j$ is called either the **Resolvant** matrix of the TLM or the **Jacobian** of the nonlinear model.

Approximate ΔJ by a 1st–order Taylor series in $\mathbf{y}'$

$$J' = \sum_i \frac{\partial J}{\partial y_i} y_i'$$

Adjoint Sensitivity Analysis for a Discrete Model

The Adjoint Model
(Adjoint of the TLM or adjoint of the nonlinear model)

Application of the "chain rule" yields

$$\frac{\partial J}{\partial x_i} = \sum_j \frac{\partial y_j}{\partial x_i} \frac{\partial J}{\partial y_j} \tag{9}$$

Contrast with the TLM

$$y'_i = \sum_j \frac{\partial y_i}{\partial x_j} x'_j \tag{10}$$

A. The variables are different in the two equations
B. The order of applications of the variables related to x and y differ
C. The indices i and j in the matrix operator are reversed

Adjoint Sensitivity Analysis for a Discrete Model

Example Equations

Nonlinear model:

$$\frac{\partial u}{\partial t} = -u \ \frac{\partial u}{\partial x}$$

Discrete NLM (superscript t index, subscript x index)

$$u_i^{n+1} \ = \ u_i^n \ - \ (\Delta t) u_i^n \frac{u_{i+1}^n - u_{i-1}^n}{2(\Delta x)}$$

TLM (linearized about time and space varying solution $\tilde{u}$)

$$u_i'^{n+1} = u_i'^n \ - \ \frac{\Delta t}{2\Delta x} \left[u_i'^n (\tilde{u}_{i+1}^n - \tilde{u}_{i-1}^n) + \tilde{u}_i^n (u_{i+1}'^n - u_{i-1}'^n) \right]$$

Adjoint model:

$$\hat{u}_i^n = \hat{u}_i^{n+1} - \frac{(\Delta t)}{2(\Delta x)} \left[(\tilde{u}_{i+1}^n - \tilde{u}_{i-1}^n) \hat{u}_i^{n+1} + \tilde{u}_{i-1}^n \hat{u}_{i-1}^{n+1} - \tilde{u}_{i+1}^n \hat{u}_{i+1}^{n+1} \right]$$

Adjoint Sensitivity Analysis for a Discrete Model

Additional Notes

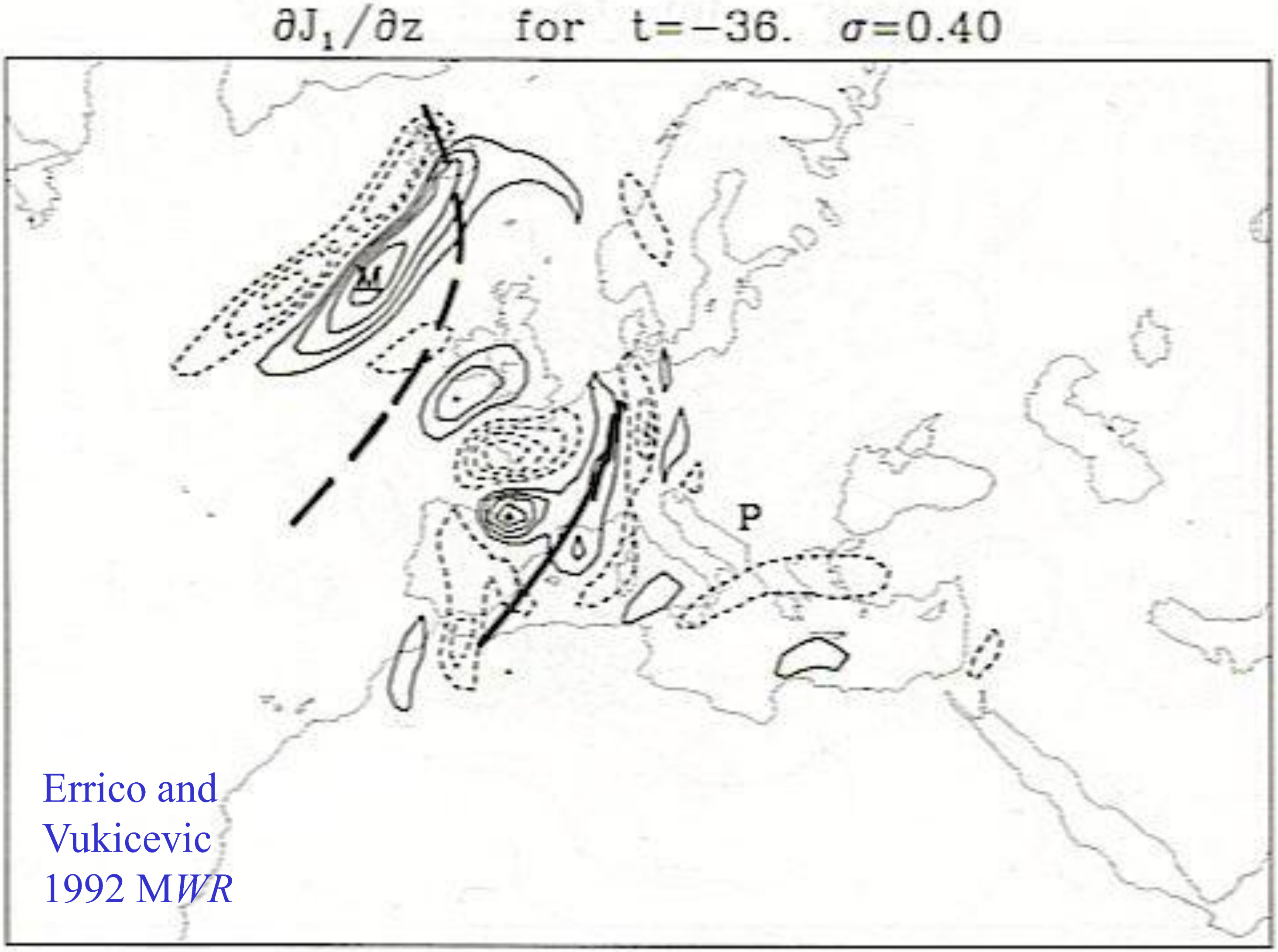
1. Mathematically, the field or vector $\partial J/\partial \mathbf{x}$ is said to reside in the dual space of $\mathbf{x}$.
2. With the simplified notation $\hat{\mathbf{x}} = \partial J/\partial \mathbf{x}$ and $\mathbf{M} = \partial \mathbf{y}/\partial \mathbf{x}$, etc.:

$$J' = \hat{\mathbf{y}}^T \mathbf{y}' = \hat{\mathbf{y}}^T (\mathbf{M}\mathbf{x}') = (\hat{\mathbf{y}}^T \mathbf{M})\mathbf{x}' = (\mathbf{M}^T \hat{\mathbf{y}})^T \mathbf{x}' = \hat{\mathbf{x}}^T \mathbf{x}' \ .$$

3. The adjoint is not generally the inverse: in non-trivial atmospheric models, $\mathbf{M}^T \neq \mathbf{M}^{-1}$.
4. This discrete representation of an adjoint model neglects an important aspect if the sensitivity fields are to be physically interpreted (as will be shown later).

Examples of Adjoint-Derived Sensitivities

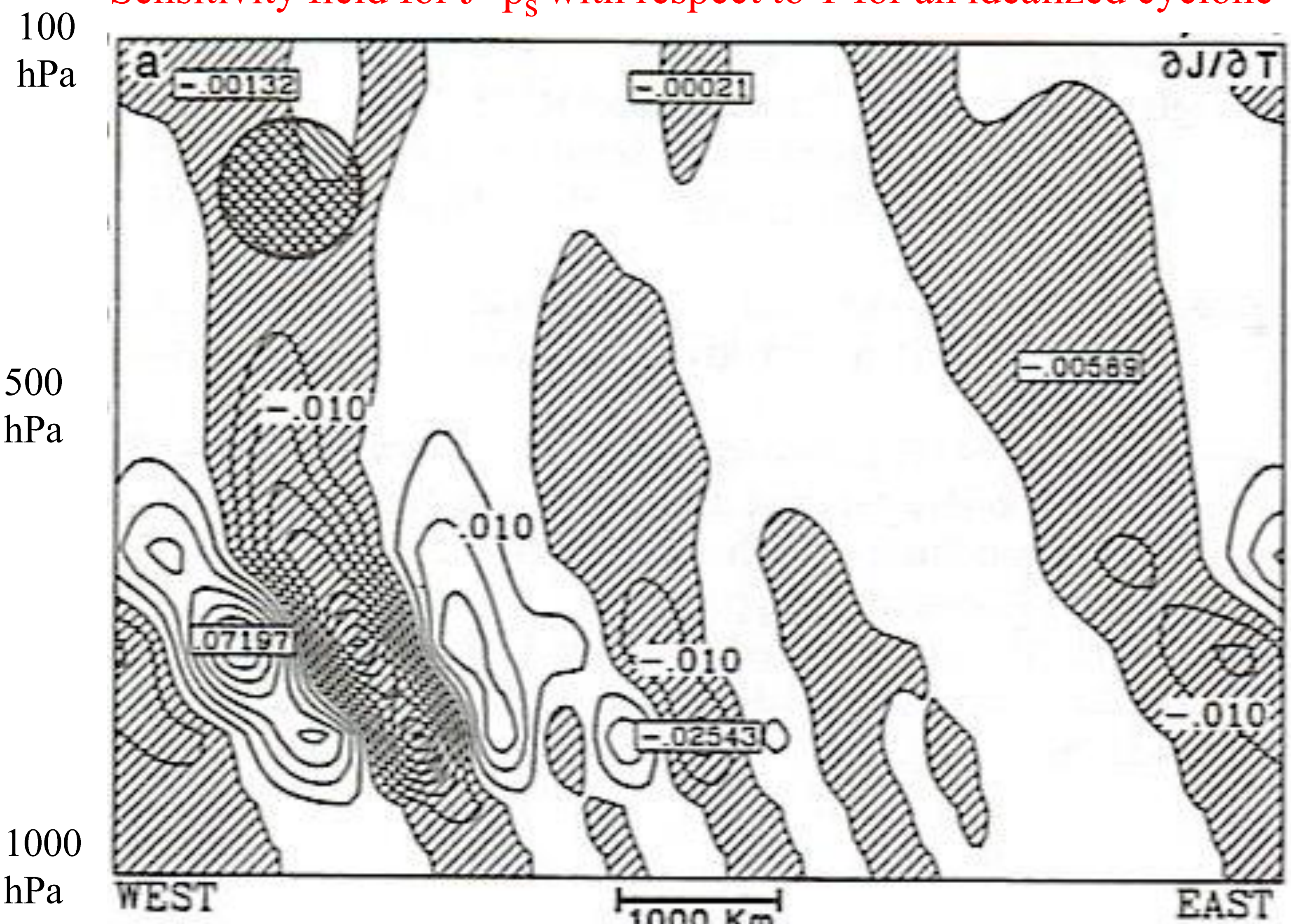
Example Sensitivity Field



Errico and Vukicevic 1992 M*WR*

Contour interval 0.02 Pa/m M=0.1 Pa/m

Sensitivity field for $J=p_s$ with respect to T for an idealized cyclone



From Langland and Errico 1996 *MWR*

Adjoint Sensitivity Analysis

Impacts vs. Sensitivities

A single impact study yields exact response measures (J) for **all** forecast aspects with respect to the **particular** perturbation investigated.

$$\Delta J = J(\mathbf{x} + \mathbf{x}') - J(\mathbf{x})$$

A single adjoint-derived sensitivity yields linearized estimates of the **particular** measure (J) investigated with respect to **all** possible perturbations.

$$J' = \sum_i \frac{\partial J}{\partial x_i} \, x_i'$$

Development of Adjoint Model Software

First consider deriving the TLM and its adjoint model codes directly from the NLM code

Why consider development from code?

1. Eventually, a TLM and adjoint code will be necessary.

2. The code itself is the most accurate description of the model algorithm.

3. If the model algorithm creates different dynamics than the original equations being modeled, for most applications it is the former that are desirable and only the former that can be validated.

Development of Adjoint Model From Line by Line Analysis of Computer Code

Automatic Differentiation

TAMC	Ralf Giering (superceded by TAF)
TAF	FastOpt.com
ADIFOR	Rice University
TAPENADE	INRIA, Nice
OPENAD	Argonne
Others	www.autodiff.org

Development of Adjoint Model From Line by Line Analysis of Computer Code

1. TLM and Adjoint models are straight-forward (although tedious) to derive from NLM code, and actually simpler to develop.
2. Intelligent approximations can be made to improve efficiency.
3. TLM and (especially) Adjoint codes are simple to test rigorously.
4. Some outstanding errors and problems in the NLM are typically revealed when the TLM and Adjoint are developed from it.
5. Some approximations to the NLM physics considered are generally necessary.
6. It is best to start from clean NLM code.
7. The TLM and Adjoint can be formally correct but useless!

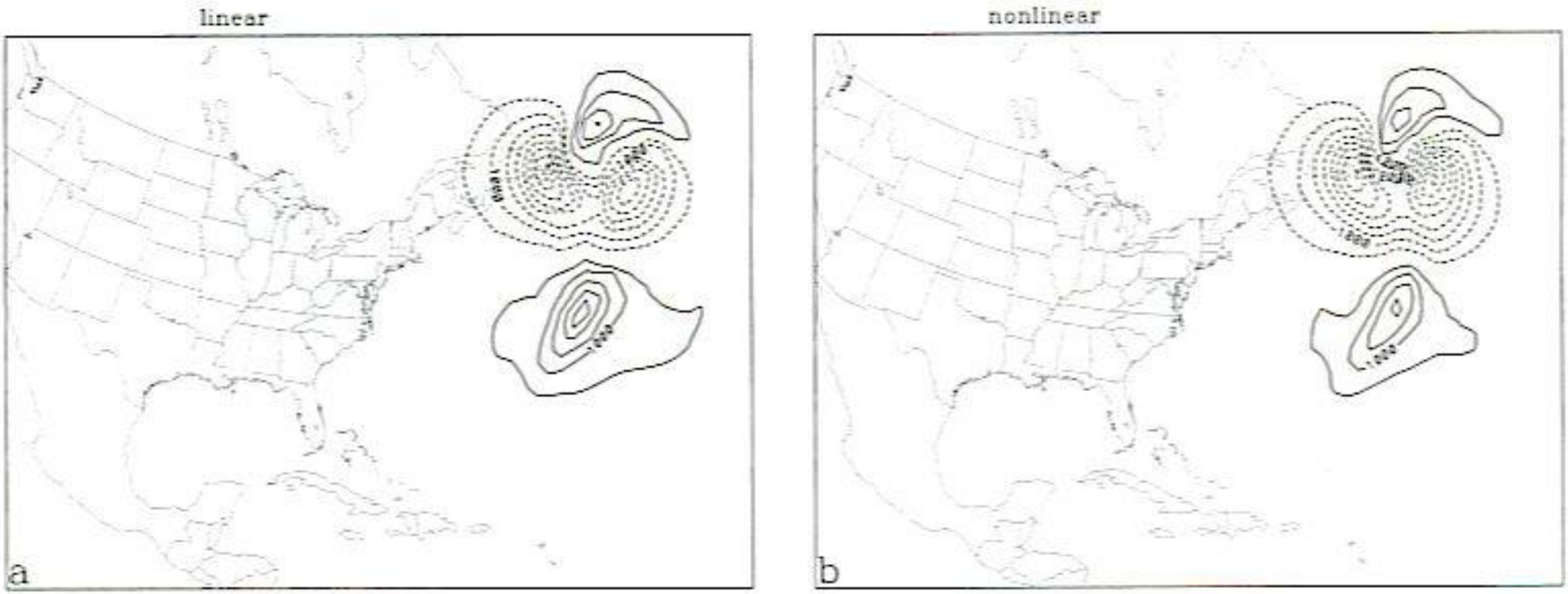
Nonlinear Validation

Does the TLM or Adjoint model tell us anything about the behavior of meaningful perturbations in the nonlinear model that may be of interest?

Linear vs. Nonlinear Results in Moist Model

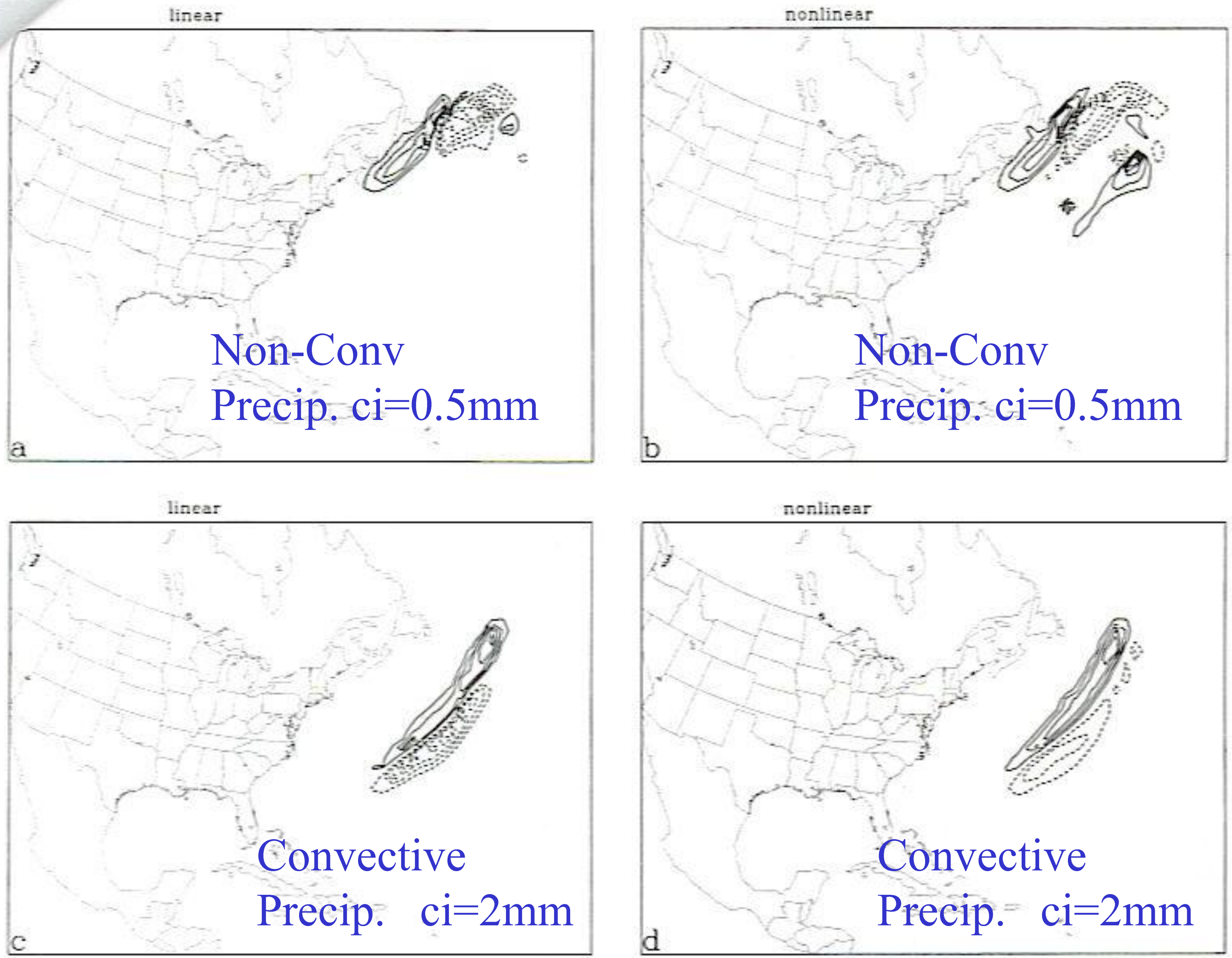
24-hour SV1 from case W1
Initialized with T' =1K
Final ps field shown

Errico and Raeder
1999 *QJRMS*



Contour interval 0.5 hPa

Linear vs. Nonlinear Results in Moist Model



Linear vs. Nonlinear Results

In general, agreement between TLM and NLM results will depend on:

1. Amplitude of perturbations
2. Stability properties of the reference state
3. Structure of perturbations
4. Physics involved
5. Time period over which perturbation evolves
6. Measure of agreement

The agreement of the TLM and NLM is exactly that of the Adjoint and NLM if the Adjoint is exact with respect to the TLM.

Efficient solution of optimization problems

Optimal Perturbations

Type I

Maximize $J' = \sum_i \frac{\partial J}{\partial x_i} x'_i$

Given the constraint: $C = \frac{1}{2} \sum_i w_i x_i'^2$

Solution Method: Minimize the augmented variable

$$I = \sum_i \frac{\partial J}{\partial x_i} x'_i + \lambda \left(C - \frac{1}{2} \sum_i w_i x_i'^2 \right)$$

$$\frac{\partial I}{\partial x'_i} = \frac{\partial J}{\partial x_i} - \lambda w_i x'_i$$

Solution:

$$x'_i(\text{optimal}) = \frac{\lambda}{w_i} \frac{\partial J}{\partial x_i}$$

$$\lambda = \sqrt{2C} \left[\sum_i \frac{1}{w_i} \left(\frac{\partial J}{\partial x_i} \right)^2 \right]^{-1}$$

Optimal Perturbations

Type II

Minimize $C = \frac{1}{2} \sum_i w_i x_i'^2$

Given the constraint: $J' = \sum_i \frac{\partial J}{\partial x_i} x_i'$

Solution Method (as before)

Solution:

$$x_i'(\text{optimal}) = \frac{\lambda}{w_i} \frac{\partial J}{\partial x_i}$$

$$\lambda = J' \left[\sum_i \frac{1}{w_i} \left(\frac{\partial J}{\partial x_i} \right)^2 \right]^{-1}$$

Optimal Perturbations

Singular Vectors

Maximize the L2 norm: $$N = \tfrac{1}{2}\mathbf{y}'^T\mathbf{N}\mathbf{y}' \tag{40}$$

Given the TLM: $$\mathbf{y}' = \mathbf{M}\mathbf{x}' \tag{41}$$

And the constraint: $$1 = C = \tfrac{1}{2}\mathbf{x}'^T\mathbf{C}\mathbf{x}' \tag{42}$$

Solution Method: Minimize the augmented variable $I(\mathbf{x}')$:

$$I = \frac{1}{2}\mathbf{x}'^T\mathbf{M}^T\mathbf{N}\mathbf{M}\mathbf{x}' + \lambda^2\left(C - \frac{1}{2}\mathbf{x}'^T\mathbf{C}\mathbf{x}'\right) \tag{43}$$

$$\frac{\partial I}{\partial \mathbf{x}'} = \mathbf{M}^T\mathbf{N}\mathbf{M}\mathbf{x}' - \lambda^2\mathbf{C}\mathbf{x}' \tag{44}$$

For $\mathbf{z} = \mathbf{C}^{\frac{1}{2}}\mathbf{x}'$, the solution is an eigenvalue problem

$$\lambda^2\mathbf{z} = \mathbf{C}^{-\frac{1}{2}}\mathbf{M}^T\mathbf{N}\mathbf{M}\mathbf{C}^{-\frac{1}{2}}\mathbf{z} \tag{45}$$

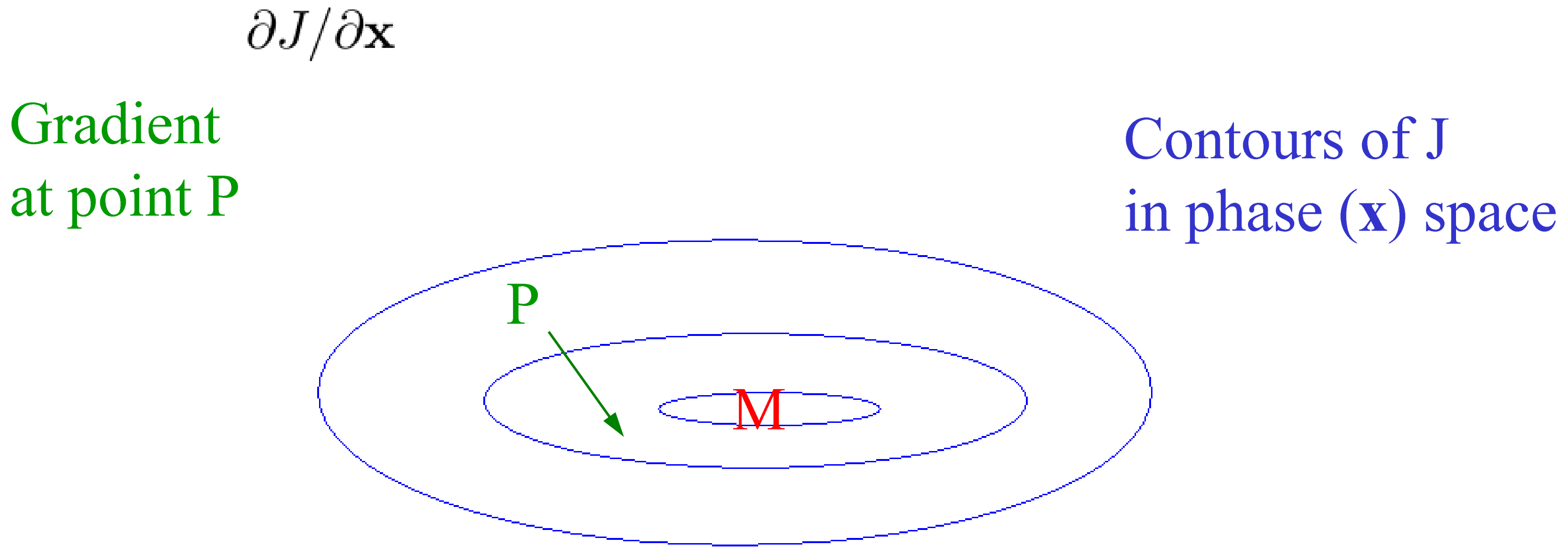
Optimal Perturbations

Additional Notes Regarding SVs

1. λ are the **singular values** of the matrix $\mathbf{N}^{\frac{1}{2}}\mathbf{M}\mathbf{C}^{-\frac{1}{2}}$.
2. The set of $\mathbf{x}'$ form an orthonormal basis with respect to the norm $\mathbf{C}$.
3. If $\mathbf{C}$ and $\mathbf{N}$ are the Euclidean norm $\mathbf{I}$, then $\mathbf{x}' = \mathbf{z}$ are the right (or initial) **singular vectors** (or SVs) of $\mathbf{M}$ and $\mathbf{y}' = \mathbf{M}\mathbf{x}'$ are the left (or final or evolved) singular vectors of $\mathbf{M}$. *The same terminology is used even for more general norms.*
4. $\lambda^2 = N/C$ for each solution.
5. If $\mathbf{C}$ is the inverse of the error covariance matrix, then the evolved SVs are the EOFs (or PCs) of the forecast error covariance, and truncations using the leading SVs maximize the retained error variance. (Ehrendorfer and Tribbia 1997 *JAS*)
6. The SVs and λ^2 depend on the norms used; i.e., on how measurements are made. This dependency is removed only by introducing some other constraint or condition.
7. SVs produced for semi–infinite periods are equivalent to Lyupanov vectors (Legras and Vautard, 1995 *ECMWF Note*).

The more general nonlinear optimization problem

Find the local minima of a scalar nonlinear function J(**x**).



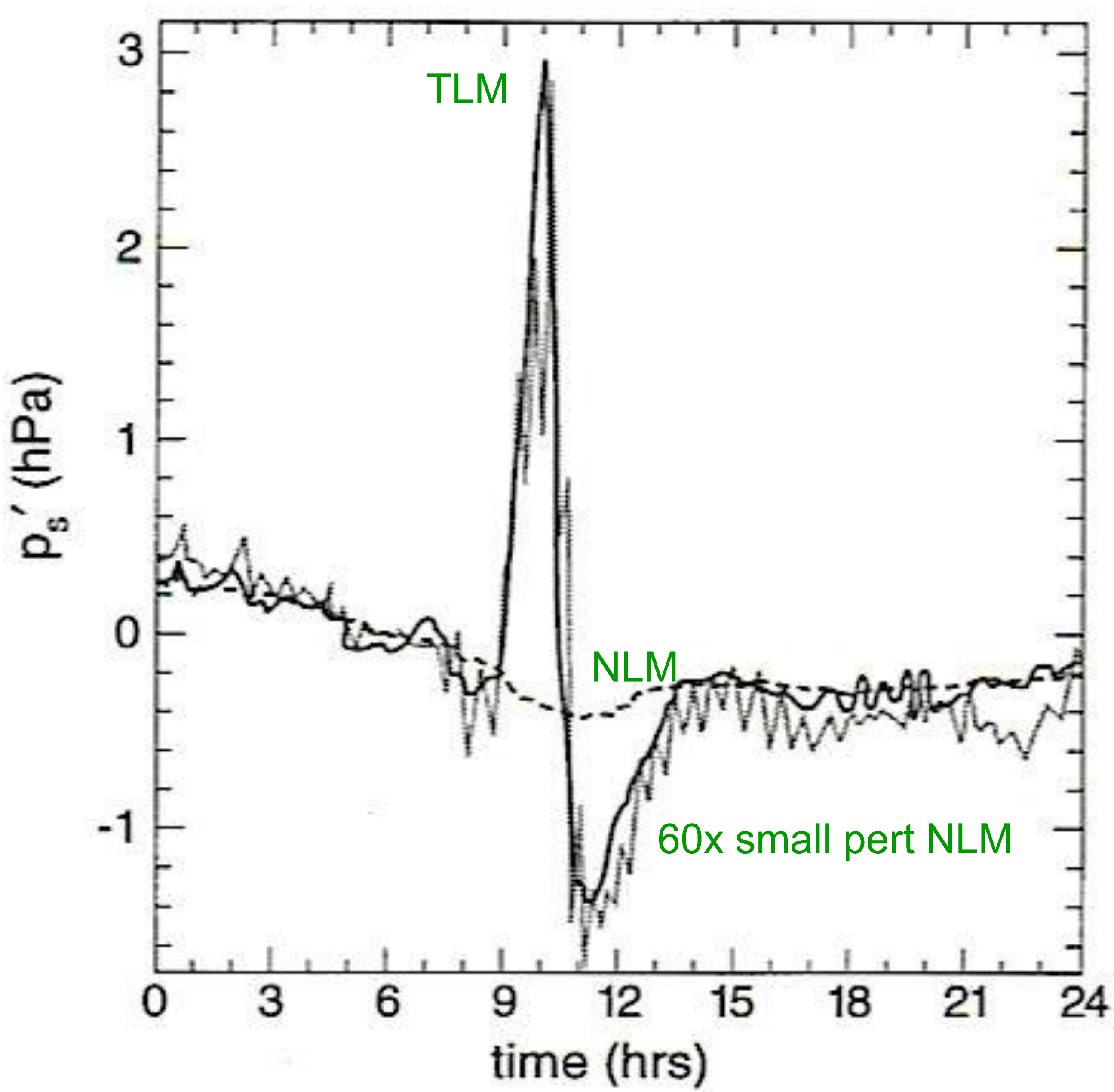
The Energy Norm

$$E = \frac{1}{2A} \int \left[u'^2 + v'^2 + \frac{C_p}{T_r} T'^2 + \frac{RT_r}{p_{sr}^2} p_s'^2 \right] \, dA \, d\sigma$$

Errico, R.M., 2000: Interpretations of the total energy and rotational energy norms applied to determination of singular vectors. *Quart. J. Roy. Meteor. Soc.*, **126A**, 1581–1599.

Problems with Physics

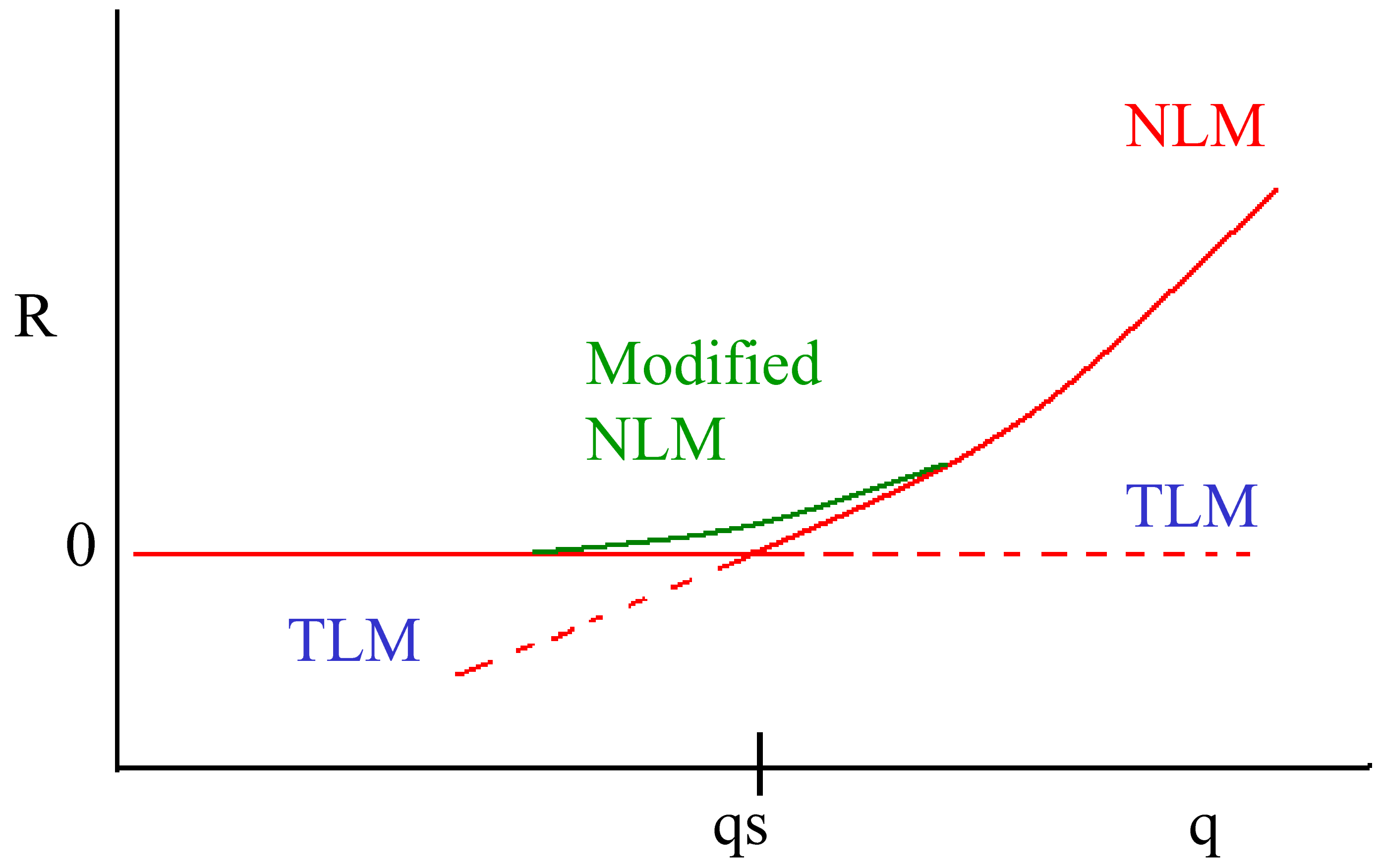
Tangent linear vs. nonlinear model solutions



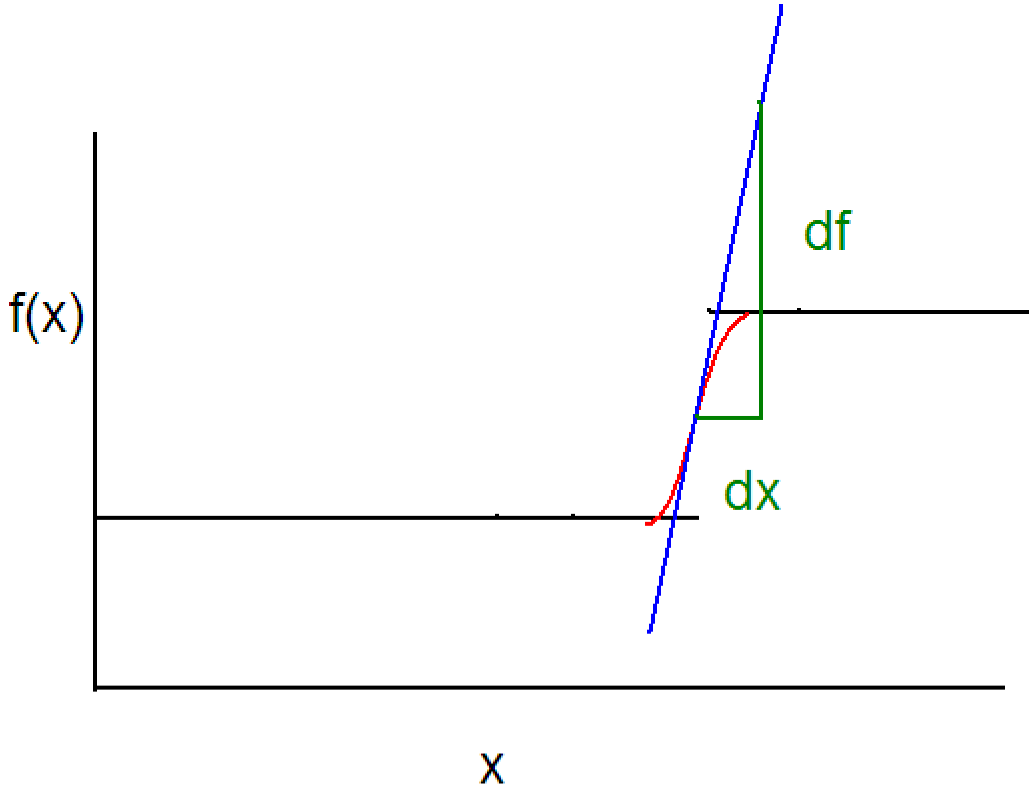
Errico and Raeder 1999 *QJRMS*

Problems with Physics

Consider Parameterization of Stratiform Precipitation



Example of a potentially worse problem introduced by smoothing

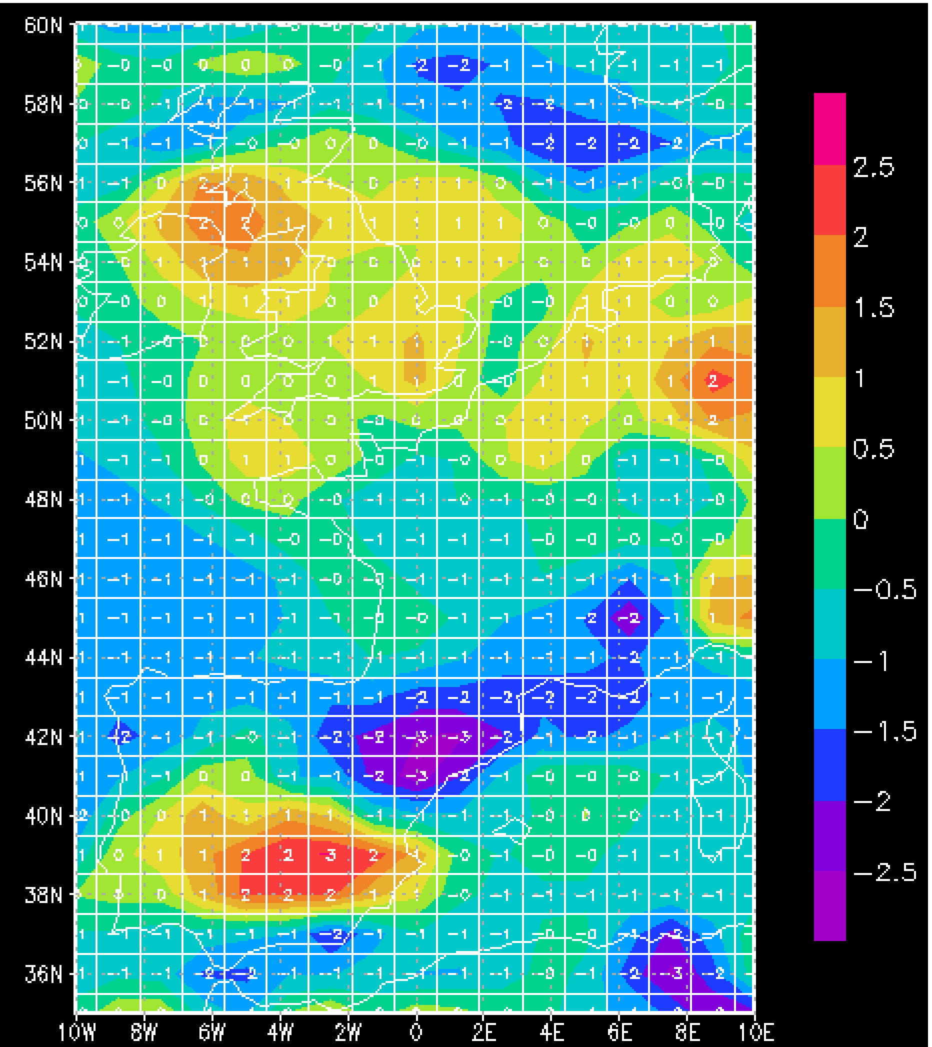


Other Important Considerations

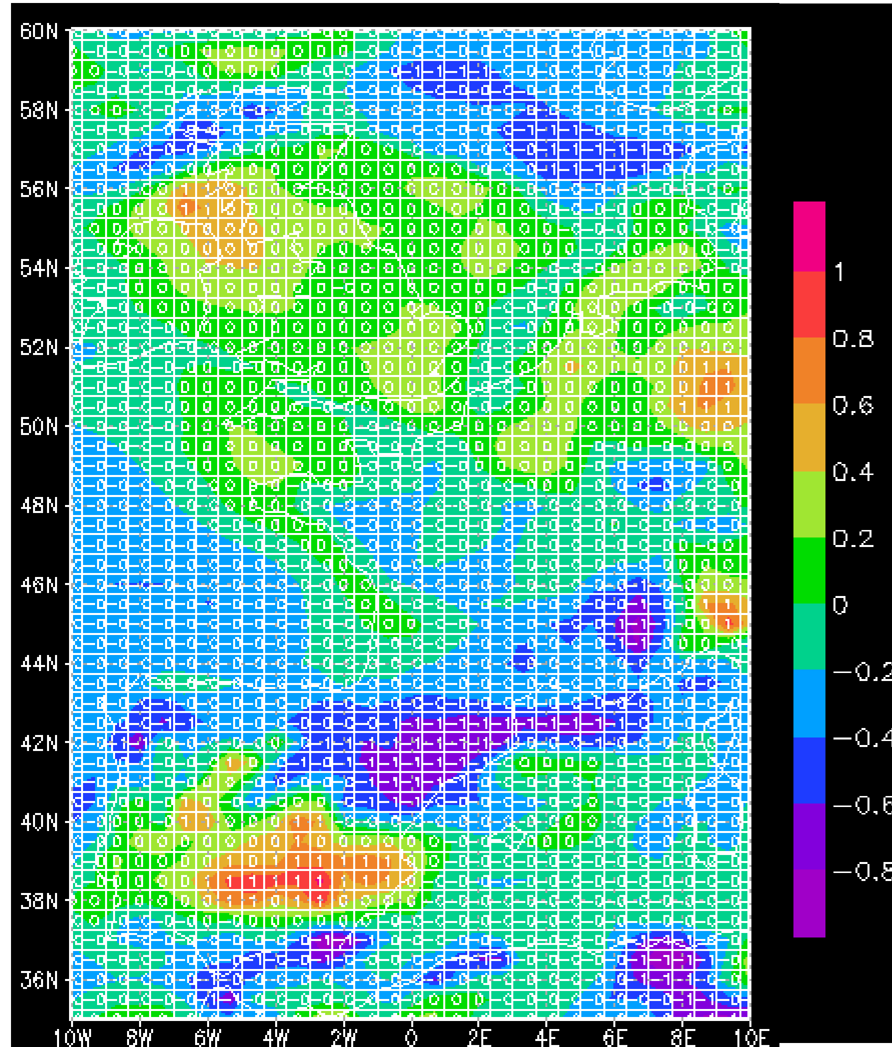
Physically-based norms and the interpretations of sensitivity fields

∂ (error “energy”) / ∂ (Tv 24-hours earlier)

1 x 1.25 degree lat-lon



0.5 x 0.0625 degree lat-lon



From R. Todling

Sensitivities of continuous fields

Consider $J(\mathbf{f}(\mathbf{x}))$ where J is a scalar function of a set f_i of continuous fields represented by the vector $\mathbf{f}$, each defined within a multi–dimensional space $\mathbf{x}$. Then, the real functional expression

$$\delta J = \langle \frac{\partial J}{\partial \mathbf{f}}, \delta \mathbf{f} \rangle$$

should be interpreted as

$$\sum_i \int_S dS(\mathbf{x}) \; \frac{\partial J}{\partial f_i}(\mathbf{x}) \; \delta f_i(\mathbf{x})$$

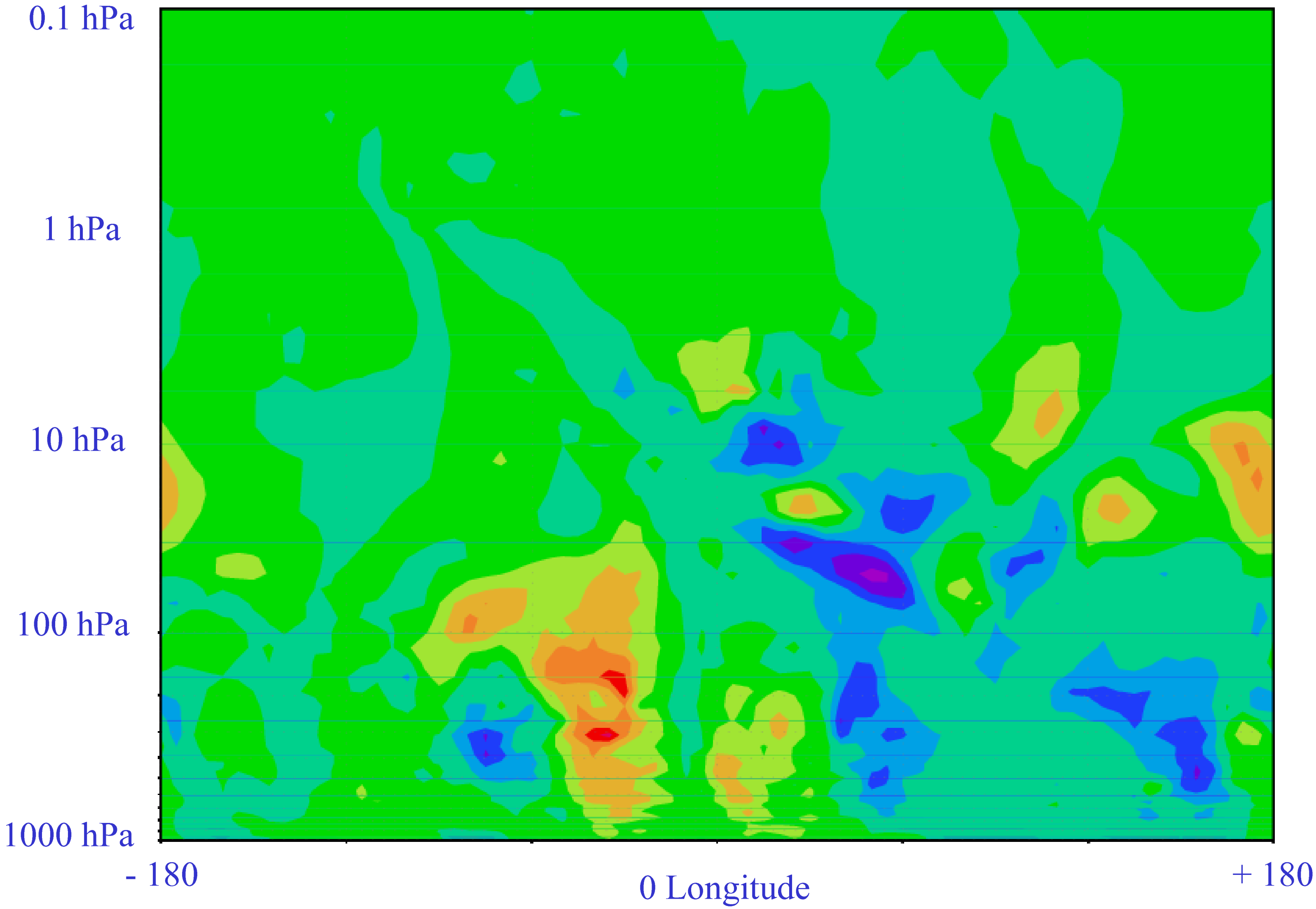
where S is a volume, mass, or other metric. With this interpretation, $\partial J / \partial f_i$ has physical units of $J \times f_i^{-1} \times S^{-1}$; i.e., it is a kind of sensitivity density.

This field of sensitivity density is relatively independent of the grid on which it is represented, but to estimate the change of J due to a perturbation $\delta \mathbf{f}$ applied at grid point $\mathbf{x}_G$, the grid volume dS at this point must be considered; i.e.,

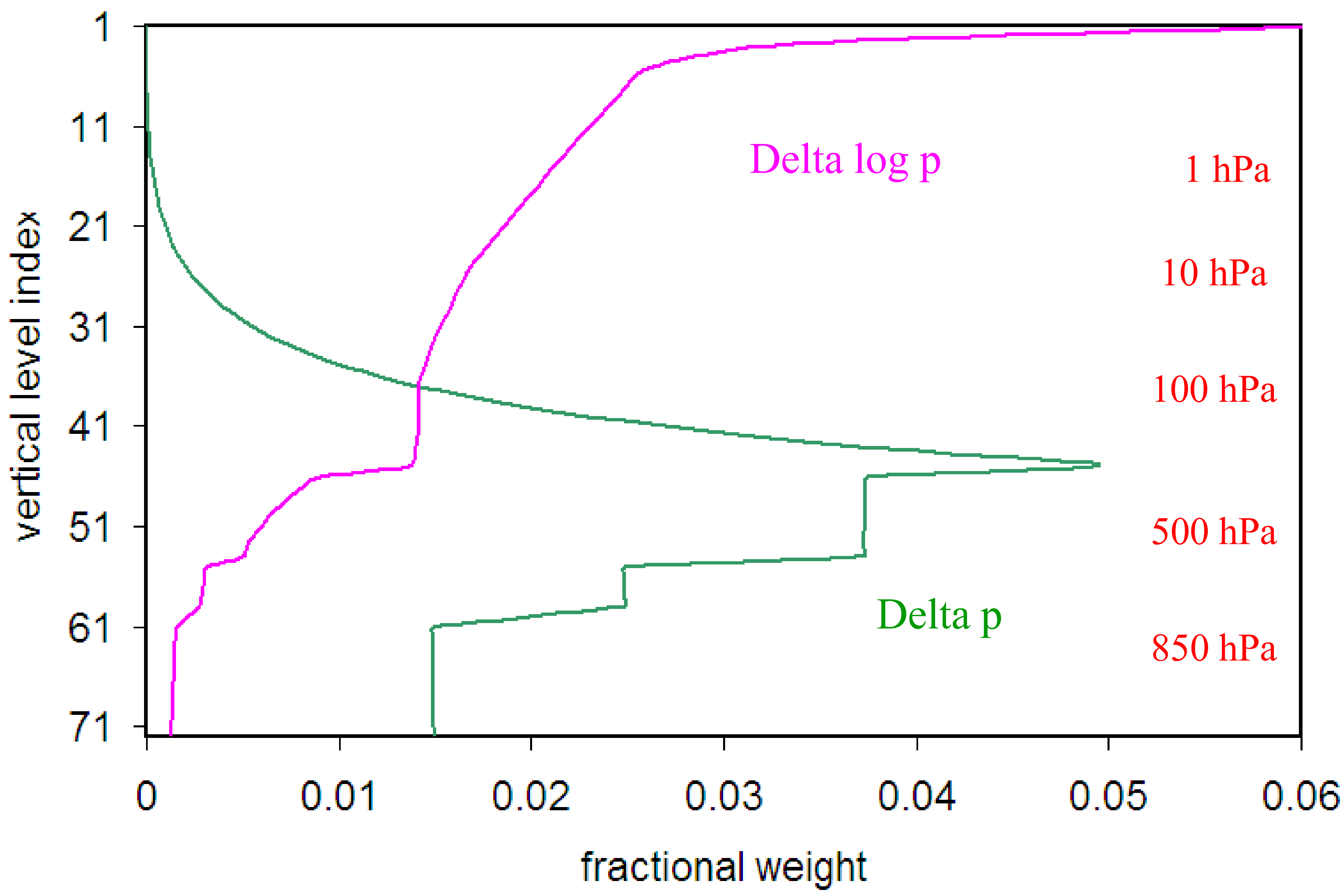
$$\frac{\partial J}{\partial f_i}(\mathbf{x}_G) = \int_{S(\mathbf{x}_G)} dS(\mathbf{x}) \; \frac{\partial J}{\partial f_i}(\mathbf{x})$$

It is safer to base physical interpretations of sensitivity on its density, but then sensitivities to grid point perturbations become less obvious.

Sensitivity of J with respect to u 5 days earlier at 45^{O}N, where J is the zonal mean of zonal wind within a narrow band centered on 10 hPa and 60^{O}N. (From E. Novakovskaia)

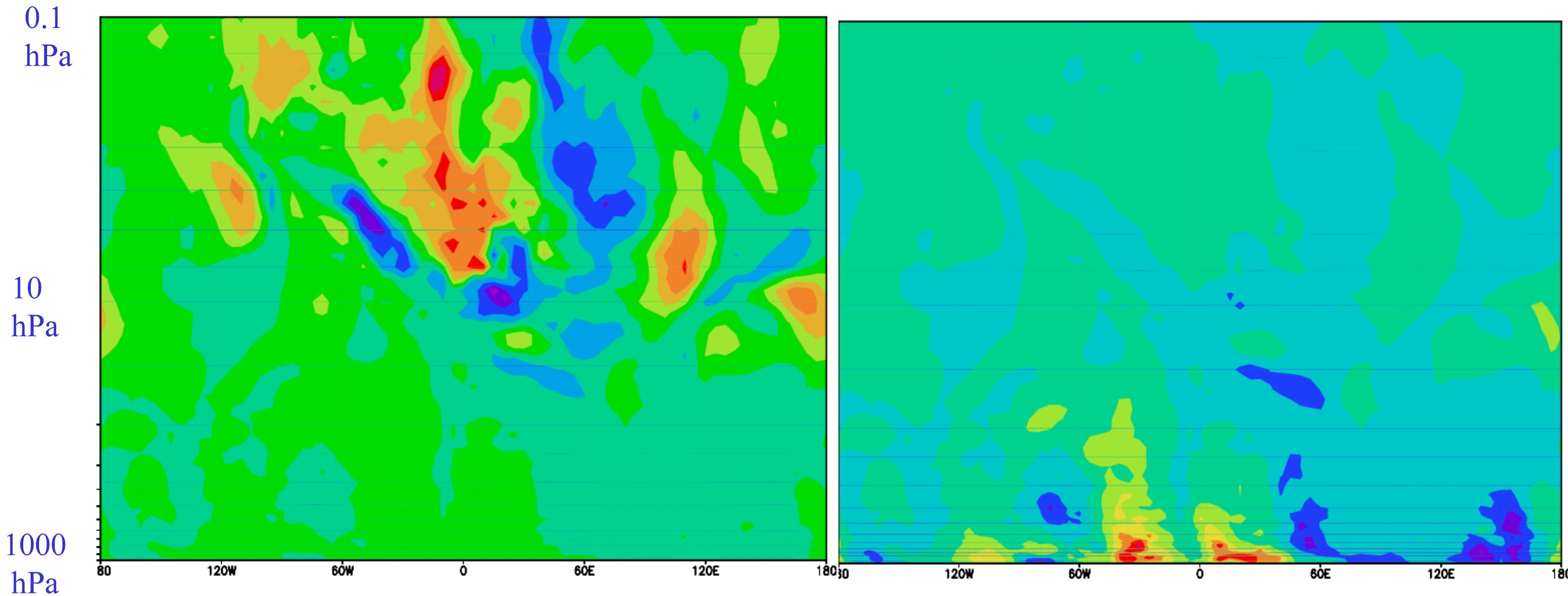


Rescaling options for a vertical grid



2 Re-scalings of the adjoint results

Mass weighting | Volume weighting



From E. Novakovskaia

Summary

Misunderstanding #1

False: Adjoint models are difficult to understand.

True: Understanding of adjoints of numerical models primarily requires concepts taught in early college mathematics.

Misunderstanding #2

False: Adjoint models are difficult to develop.

True: Adjoint models of some component algorithms are simpler to develop than their parent models, and almost trivial to check, but adjoints of some model components or formulations can pose difficult problems.

Misunderstanding #3

False: Automatic adjoint generators easily generate perfect and useful adjoint models.

True: Problems can be encountered with automatically generated adjoint codes that are inherent in the parent model. Do these problems also have a bad effect in the parent model?

Misunderstanding #4

False: An adjoint model is demonstrated useful and correct if it reproduces nonlinear results for ranges of very small perturbations.

True: To be truly useful, adjoint results must yield good approximations to sensitivities with respect to meaningfully large perturbations. This must be part of the validation process.

Misunderstanding #5

False: Adjoints are not needed because the EnKF is better than 4DVAR and adjoint results disagree with our notions of atmospheric behavior.

True: Adjoint models are more useful than just for 4DVAR. Their results are sometimes profound, but usually confirmable, thereby requiring new theories of atmospheric behavior. It is rare that we have a tool that can answer such important questions so directly!

Summary

1. Adjoint models are powerful tools for efficiently estimating sensitivity.

2. Results from an adjoint model can be tested in the parent nonlinear model.

3. Many problems encountered in the development of adjoint models reveal problems within the parent models.

4. Utilization of adjoint models has been slow for a variety of reasons.

5. There remains much to be still be learned.

References

Errico, R.M., 1997: What is an adjoint model. *Bull. Am. Meteor. Soc.*, **78,** 2577-2591.

Errico, R.M. and K.D. Raeder, 1999: An examination of the accuracy of the linearization of a mesoscale model with moist physics. *Quart. J. Roy. Met. Soc.*, **125**, 169-195.

Janisková, M., Mahfouf, J.-F., Morcrette, J.-J. and Chevallier, F., 2002: Linearized radiation and cloud schemes in the ECMWF model: Development and evaluation. *Quart. J. Roy. Meteor. Soc.,* **128,** 1505-1527

Papers by R. Gelaro, C. Reynolds, F. Rabier, M. Ehrendorfer, R. Langland, M. Leutbecher, D. Holdaway

Adjoint Workshop series:
8th http://gmao.gsfc.nasa.gov/events/adjoint_workshop-8/
9th http://gmao.gsfc.nasa.gov/events/adjoint_workshop-9/