Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Verkeer en Waterstaat

ECMWF SEMINAR 2009
7-10 September

Adjoint diagnostics for the atmosphere and ocean

Jan Barkmeijer
KNMI

OUTLINE

- Why do we need an adjoint model and what is it?
- Easy non-trivial example
- Difficulties in developing adjoint models
- Applications for atmospheric/ocean models
  - ‘sensitivity calculation’
  - singular vectors
  - other use, i.e. not ‘initial condition’ related

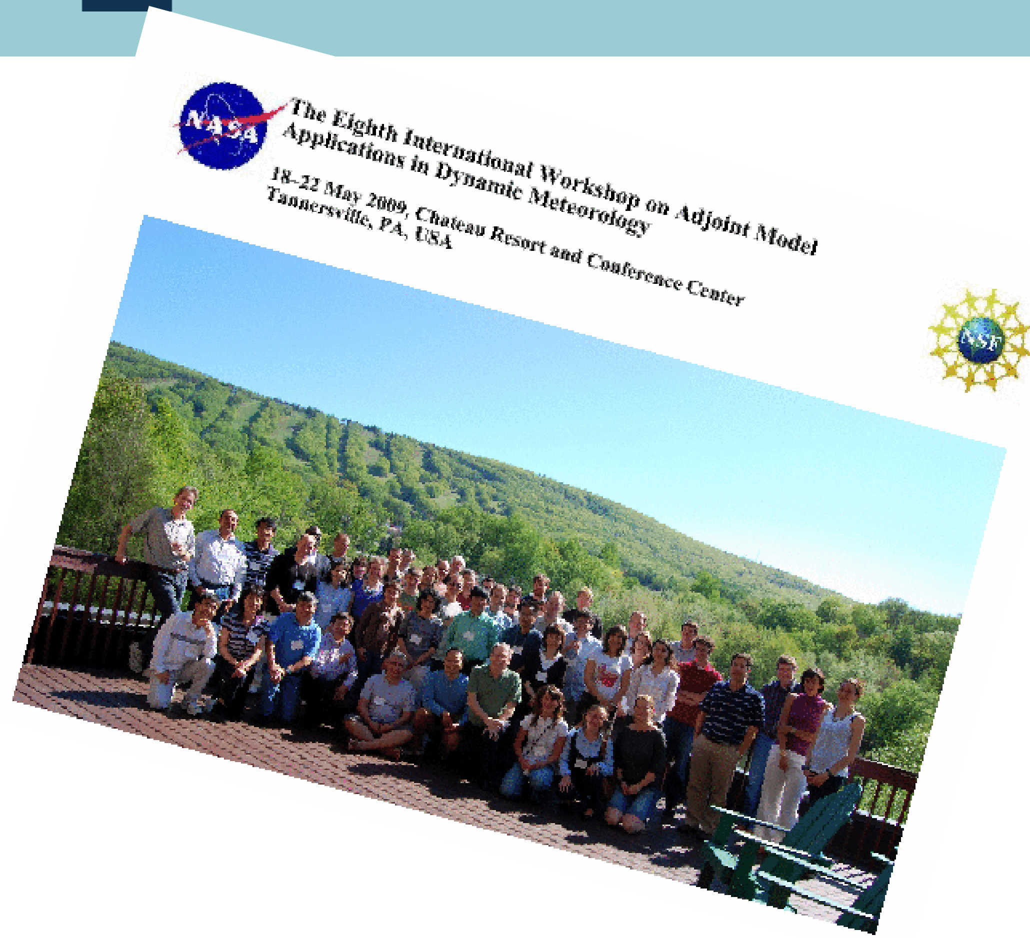
Contributions

Many thanks to:

- Bernard Bilodeau
- Ron Errico
- Ron Gelaro
- Thomas Jung
- Simon Lang
- Martin Leutbecher
- Andy Moore
- Frank Selten
- Florian Sevellec
- Gerard van der Schrier

Workshop on Adjoint Applications in Dynamic Meteorology
Asilomar Conference Center, Pacific Grove, CA 23-28 August 1992

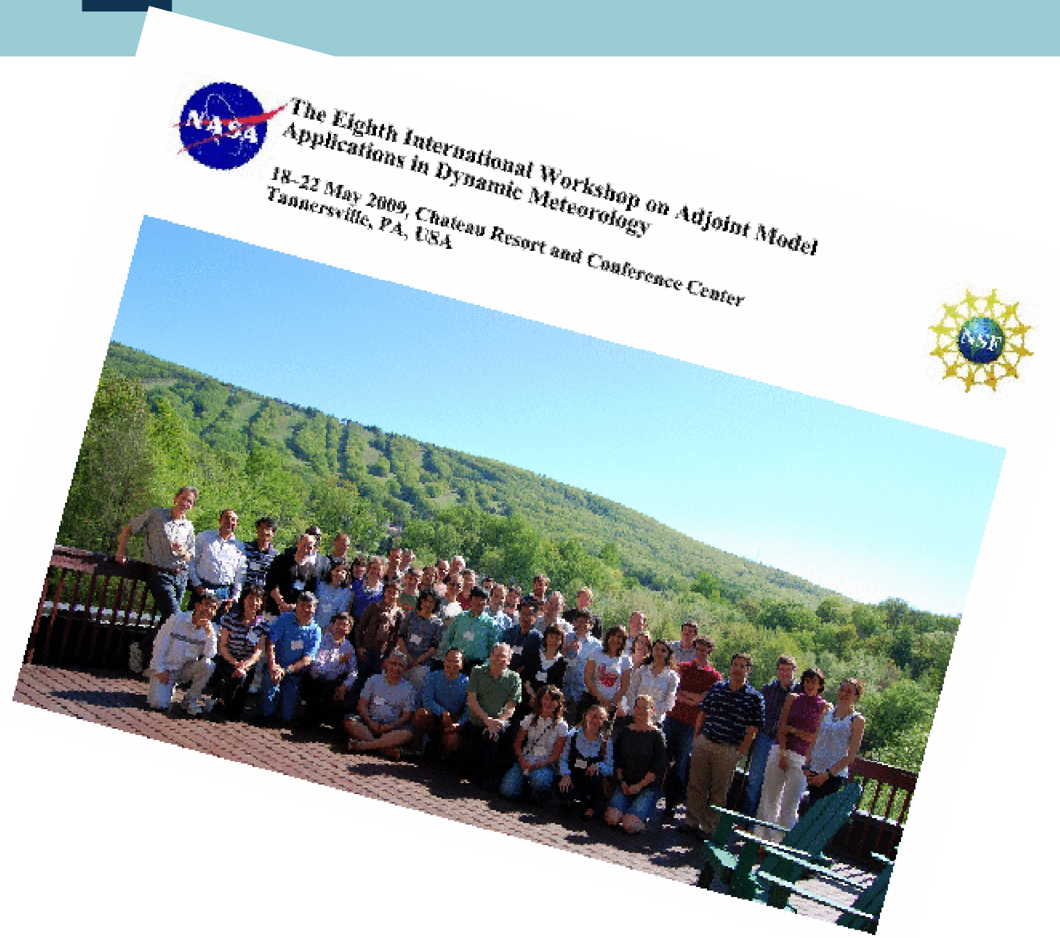
ORGANIZERS
philippe courtier (ECMWF)
john derber (NMC)
ronald m. errico (NCAR)
jean-francois louis (AER)
tomislava vukicevic (NCAR)



Special guests: E. Lorenz (2002) and G.I. Marchuk (2004)

Workshop on Adjoint Applications in Dynamic Meteorology
Asilomar Conference Center, Pacific Grove, CA 23-28 August 1992

ORGANIZERS
philippe courtier (ECMWF)
john derber (NMC)
ronald m. errico (NCAR)
jean-francois louis (AER)
tomislava vukicevic (NCAR)



Special guests: E. Lorenz (2002) and G.I. Marchuk (2004)

→ *Application of adjoint equations to virus infection modelling* (Marchuk *et al.*, 2005).

Why is an adjoint model useful?

Suppose we are dealing with a nonlinear model $\boldsymbol{M}$ of the form:

$$y = \boldsymbol{M}(x)$$

and a differentiable scalar J defined for model output fields y :

$$J = J(y) = J(\boldsymbol{M}(x))$$

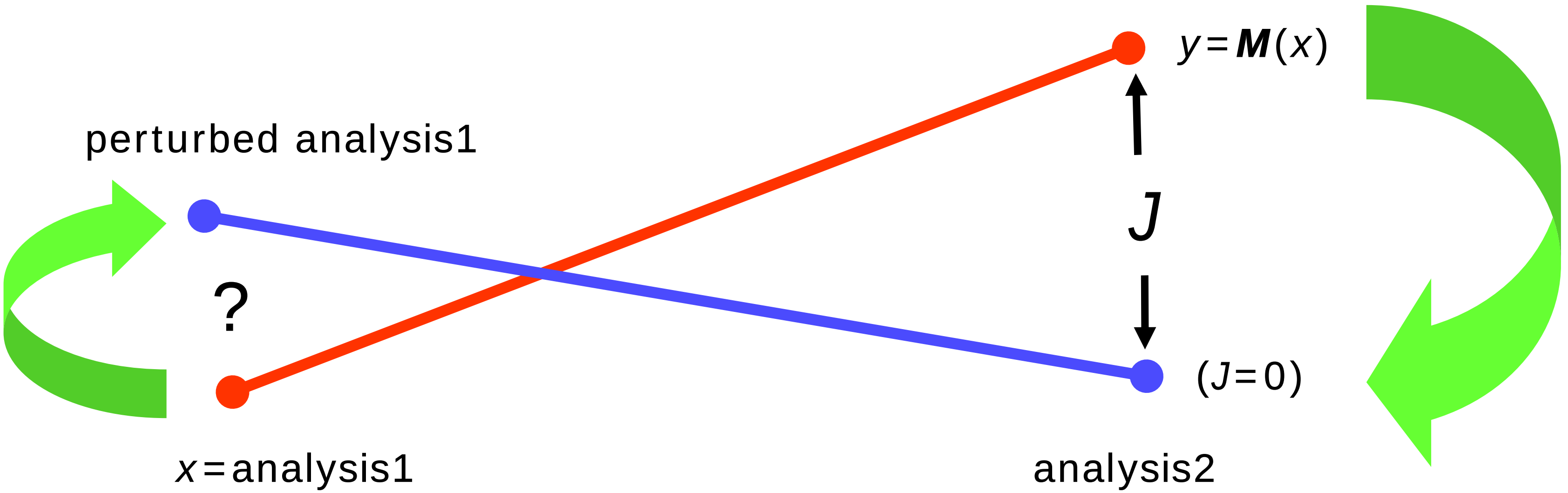
Dependence of J on y is often straightforward,

but determining $\partial J / \partial x$ seems impossible for high-dimensional models.

It would require perturbed model runs for every ($\sim 10^8$) entry of x.

Example 0: Sensitivity calculation

(method to improve a forecast retrospectively)



How to minimize *J* ?

$J(x)$=[y-analysis2,y-analysis2], with [.,.] a suitable inner product

See: Rabier *et al.* (1996), Klinker *et al.* (1998), ...
,..., Isaksen *et al.* (2005), Caron *et al*. (2006),...

Application of the chain rule learns that

$$\frac{\partial J}{\partial x_j} = \sum_{k=1}^{M} \frac{\partial J}{\partial y_k}\frac{\partial y_k}{\partial x_j}$$

Assume that a small perturbation δy_j of y_j is associated to a small perturbation δx_k of x_k through:

$$\delta y_j = \sum_k \frac{\partial y_j}{\partial x_k}\delta x_k =_{def} (\mathbf{M}\delta \mathrm{x})_j$$

and consequently

$$\nabla_x J = \mathbf{M}^T \nabla_y J$$

How to determine M^T?

Assume that the linear model describing the evolution of initial time perturbations has the form

$$(1) \qquad d\varepsilon/dt = \mathbf{L}\varepsilon$$

with propagator **M**: $\varepsilon(t_2) = \mathbf{M}(t_1,t_2)\,\varepsilon(t_1)$

Define **the adjoint model** by

$$(2) \qquad d\varepsilon/dt = -\mathbf{L}^T\varepsilon, \text{ with } [\mathbf{L}a,b] = [a,\mathbf{L}^Tb],$$

with propagator **S** and where [.,.] is a suitable inner product.

N.B. Adjoint model depends on chosen inner product [.,.].

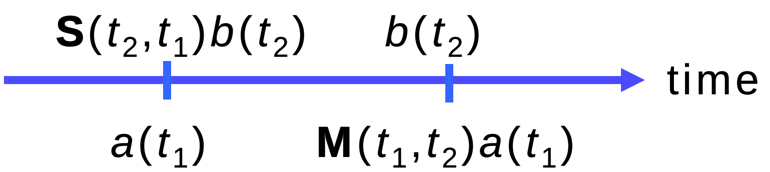
How to determine $\mathbf{M}^T$? (2)

Solutions $a(t)$ and $b(t)$ of (1) and (2) respectively satisfy the property:

$$\frac{d}{dt}\,[a(t),b(t)] = [\mathbf{L}a(t),b(t)] + [a(t),-\mathbf{L}^T b(t)] = 0$$

and consequently

$$[\mathbf{M}(t_1,t_2)a(t_1),b(t_2)] = [a(t_1),\mathbf{S}(t_2,t_1)b(t_2)]$$



$$\mathbf{M}(t_1,t_2)^T = \mathbf{S}(t_2,t_1)$$

YES!

Gradient J can be determined efficiently by running the adjoint model (2) backwards in time!

Adjoint of barotropic vorticity equation (BVE)

$$\partial\zeta/\partial t = \mathrm{Jac}(\zeta + f, \Delta^{-1}\zeta)$$, with $$\mathrm{Jac}(g,h) = \frac{\partial g}{\partial\lambda}\frac{\partial h}{\partial\mu} - \frac{\partial g}{\partial\mu}\frac{\partial h}{\partial\lambda}$$

One of the components of the linearised BVE reads as:

$$\mathbf{L}\varepsilon = \mathrm{Jac}(\Delta\varepsilon, \psi)$$

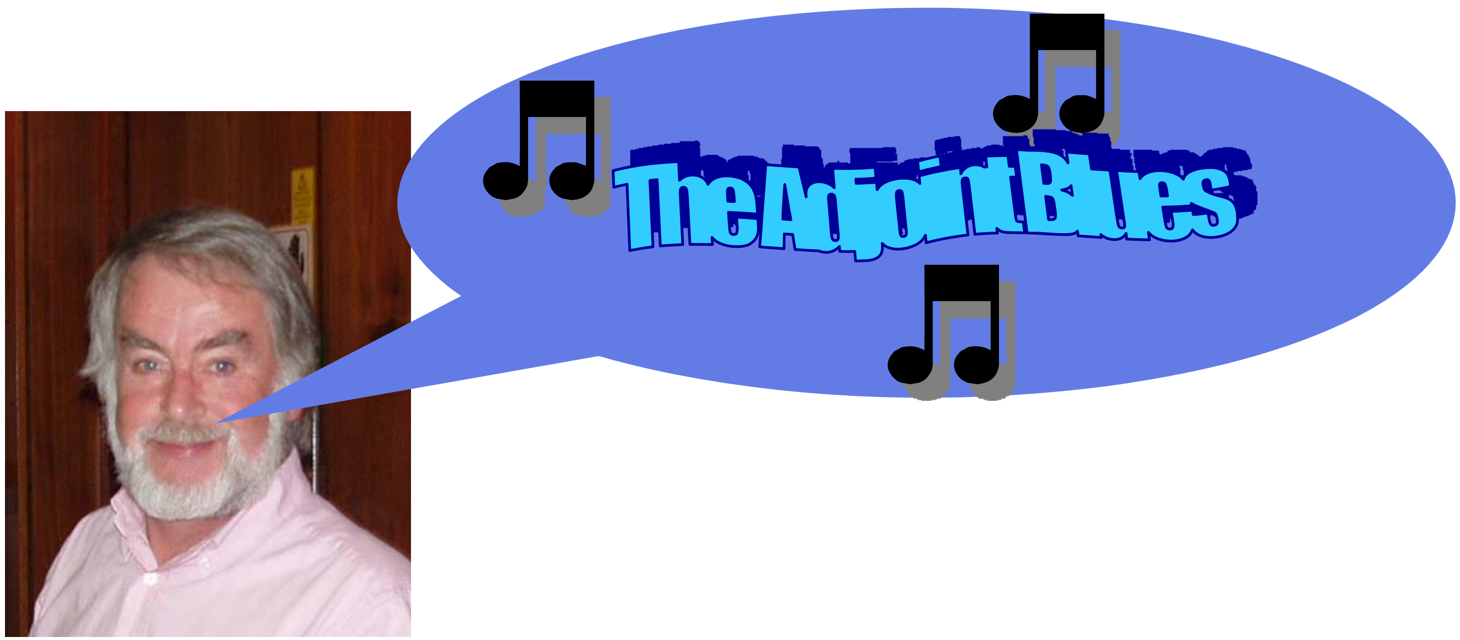
Use inner product defined by: $(a,b) = \iint a.b\,\mathrm{d}\Sigma$

$$(\mathbf{L}a,b) = \iint \mathrm{Jac}(\Delta a,\psi).b\,\mathrm{d}\Sigma \overset{!}{=} \iint \Delta a.\mathrm{Jac}(\psi,b)\mathrm{d}\Sigma =$$

$$-\iint \nabla a.\nabla\mathrm{Jac}(\psi,b)\mathrm{d}\Sigma = \iint a.\Delta\mathrm{Jac}(\psi,b)\mathrm{d}\Sigma = (a,\mathbf{L}^{*}b)$$

and thus $\mathbf{L}^{*}\tilde{\varepsilon} = -\Delta\mathrm{Jac}(\tilde{\varepsilon},\psi)$ (N.B. $\mathbf{L}^{*} \neq \mathbf{L}^{-1}$)

While developing the SL2TL adjoint .....



Linear vs. nonlinear model

- More elaborate physics in linear models, as nonlinear model is growing increasingly complex/realistic.
  (see Marta Janiskova's contribution to 2003 ECMWF seminar)

  - ensemble forecasting (e.g. in the tropics, tropical cyclones)
  - investigate physics driven instability mechanisms
  - data-assimilation of variety of data types (e.g. radar)

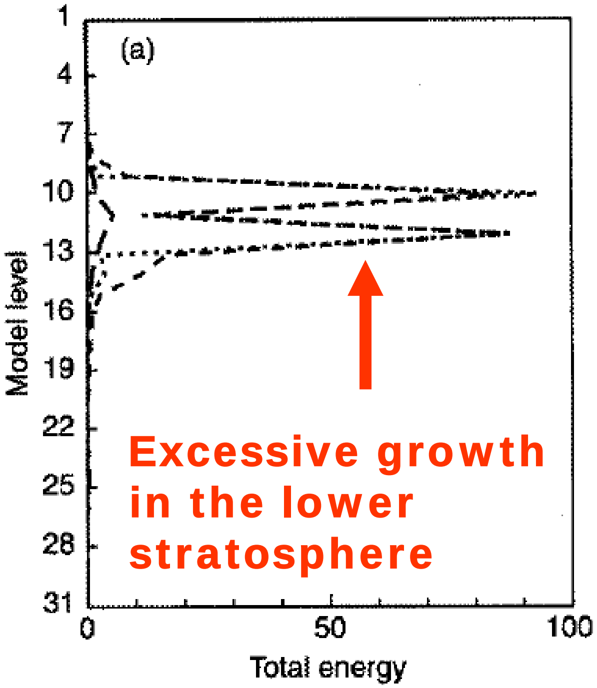
- Not straightforward to check the correctness of linear models.
  (e.g., by comparing difference between two nonlinear runs and the outcome of a linear run)

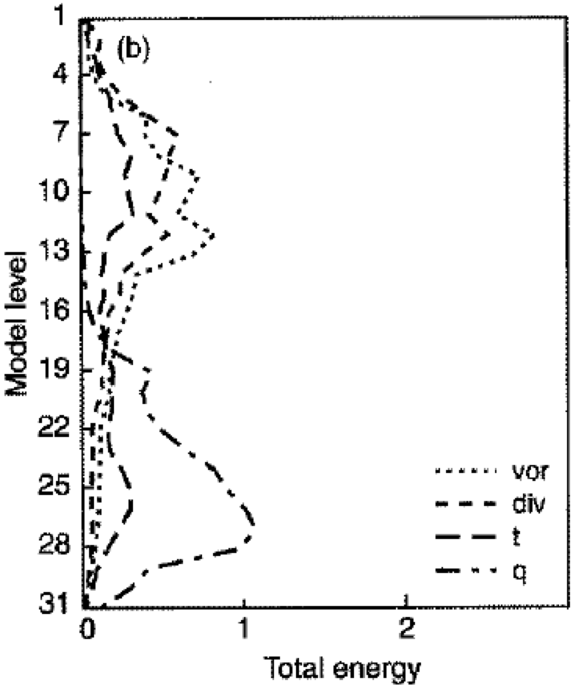
- Possibility to trigger unwanted instability mechanisms.

Spurious perturbation growth in the tropics

linear run

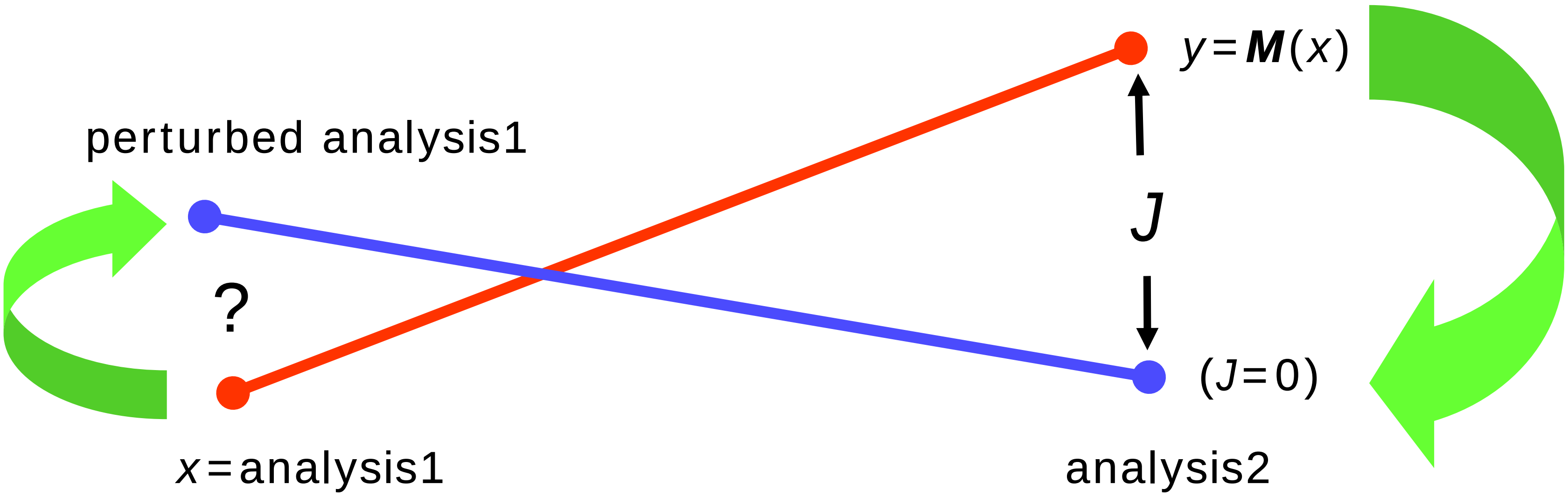
nonlinear (difference of 2 runs)





Example 0: Sensitivity calculation

(how to improve a forecast retrospectively)



Minimize cost function J :

$J(x) = [y\text{-analysis2}, y\text{-analysis2}]$, with [.,.] a suitable inner product

Example 0: Sensitivity calculation (contnd)

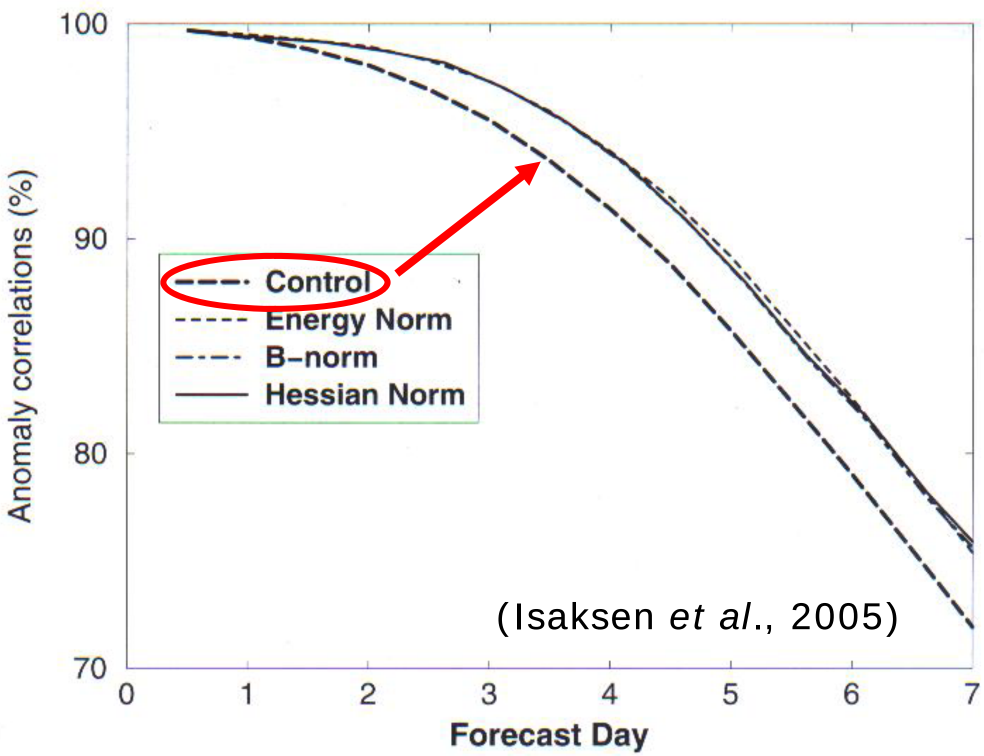


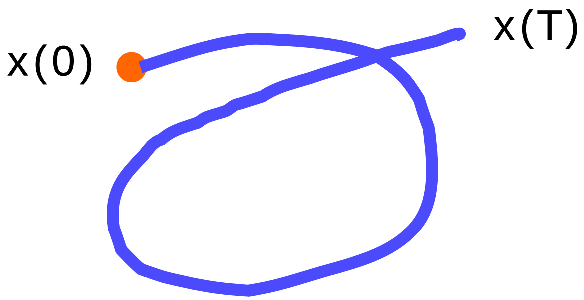
Figure 2. Anomaly correlations of forecast error in terms of 500 hPa geopotential height (m), for the northern hemisphere extratropics averaged over 29 days in December 2001. Forecasts from analyses perturbed by the 'Key Analysis Errors' (dashed curve: energy norm, dot-dashed curve: B-norm, solid curve: Hessian norm) clearly outperform forecasts from the control analyses (thick long dashed) irrespective of which of the three norms is used.

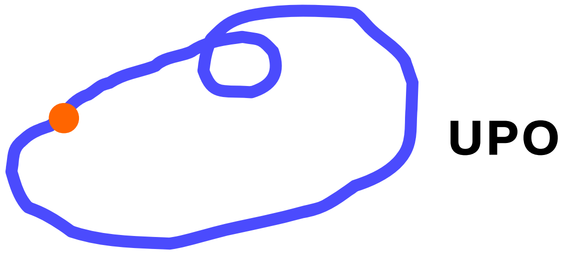
Be careful with the interpretation of 'key analysis errors'

For example, see ECMWF 2003 Seminar

Isaksen: Realism of sensitivity patterns.

Example 1: Periodic weather





Unstable Periodic Orbits (UPO) can be used to describe certain characteristics of the 'climate attractor' efficiently.

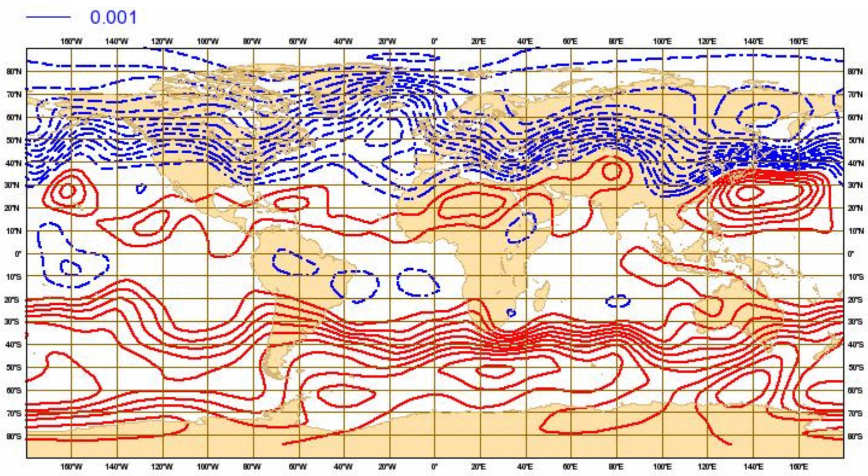
Define a cost function J by: $J(x(0)) = 1/2 \, || x(0) - x(T) ||^2$, with ||.|| the Euclidean norm.

The gradient of J is given by:

$$\nabla J = [\mathbf{M}^{\mathrm{T}} - \mathbf{Id}](x(\mathrm{T}) - x(0))$$

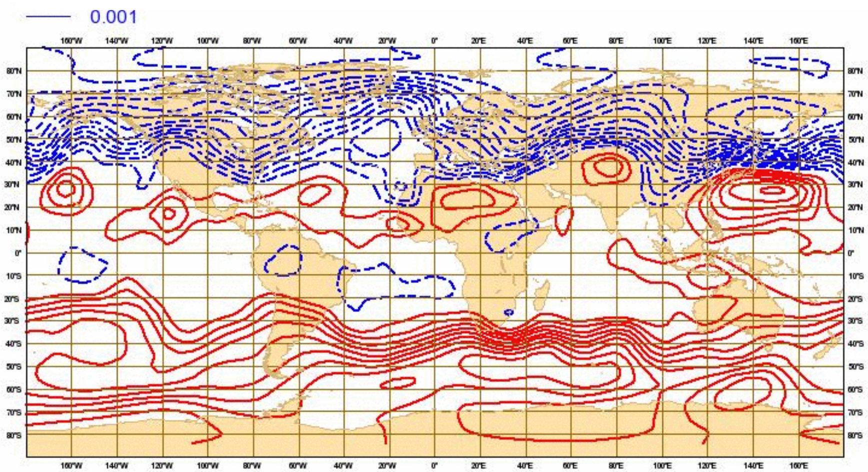
Streamfunction at 500 hPa
in a T21L3 Quasi-Geostrophic model

original initial condition

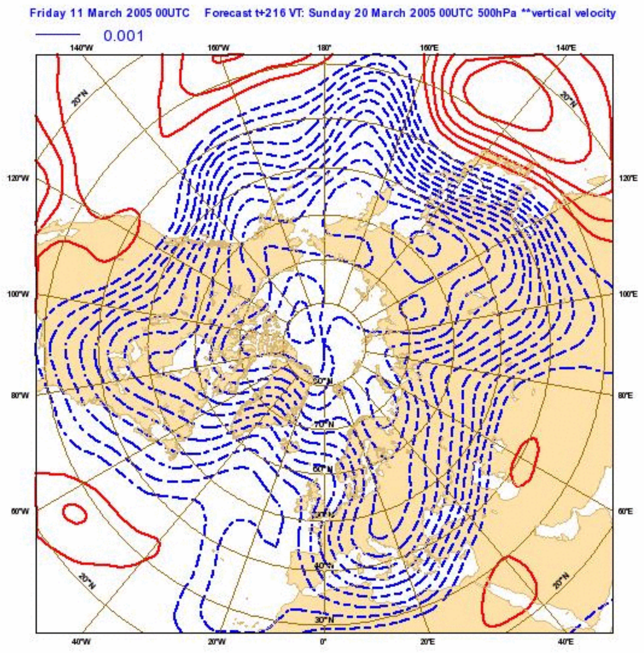


and after minimization

head and tail of UPO
Unstable **P**eriodic **O**rbit
(period is 10 days)



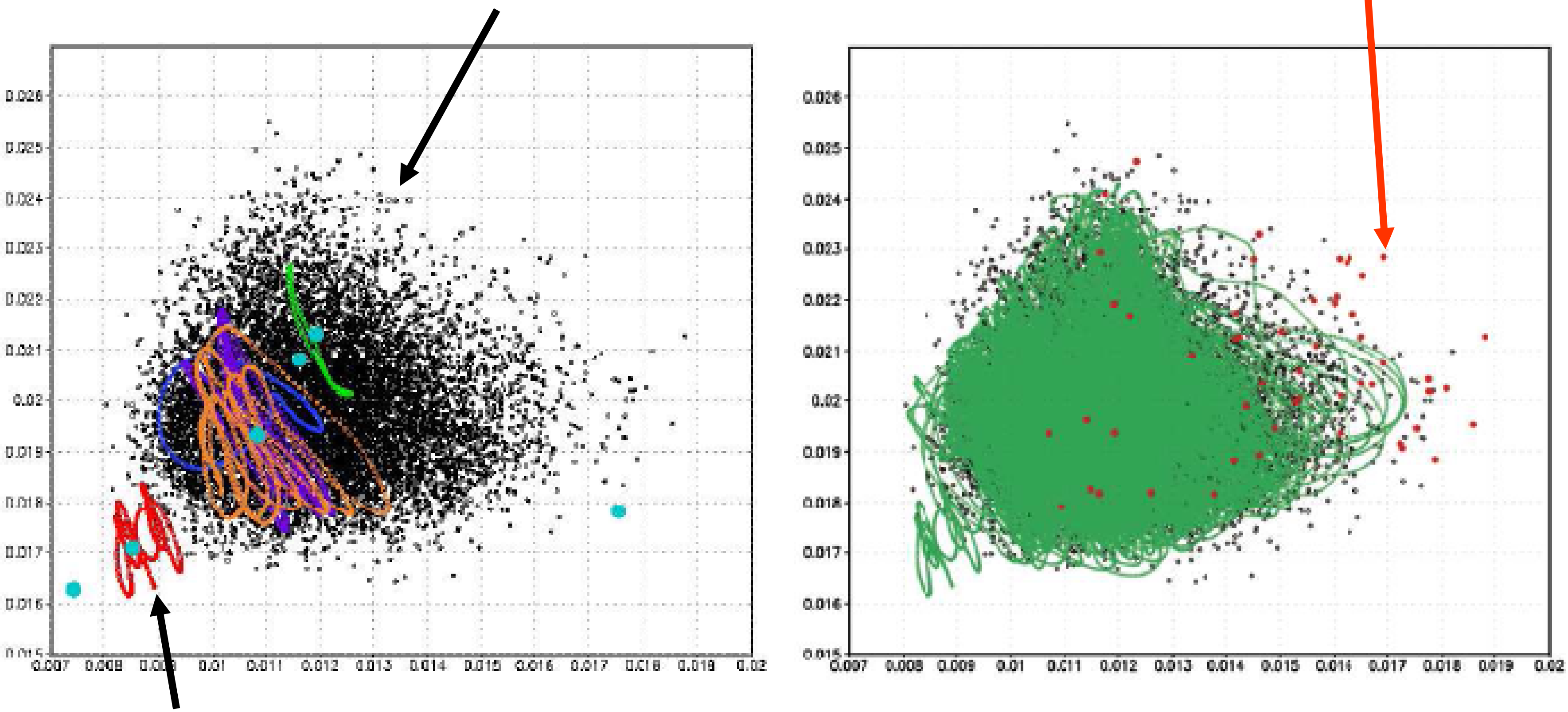
Periodic weather



Almost all data points lie close to a small set of UPO’s

Points on the attractor

not close to a UPO



UPO

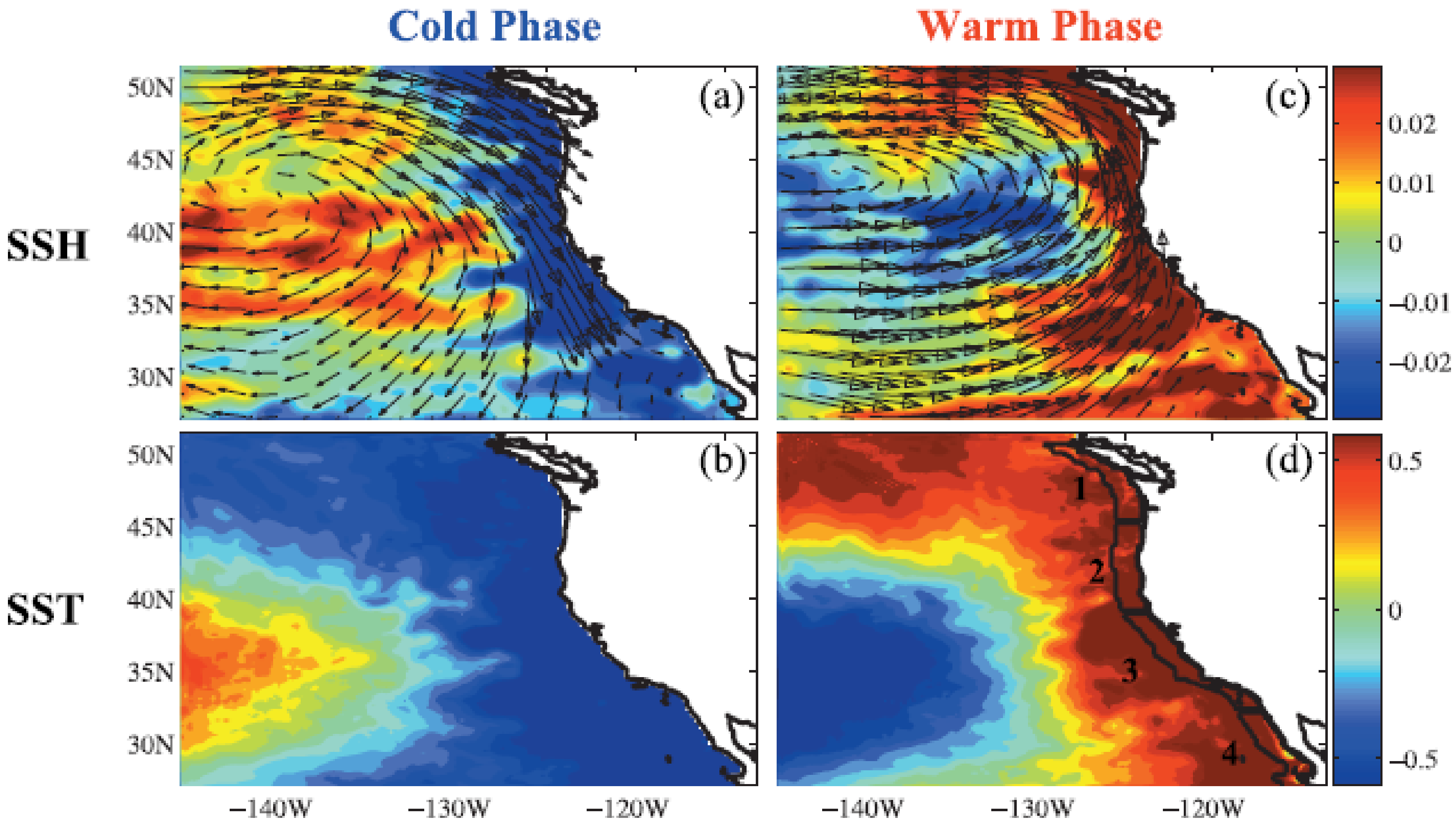
(Gritsun and Branstator, 2007)

Example 2: Upwelling

- Decadal variations in California Current Upwelling Cells (Chhak & Di Lorenzo, 2007)

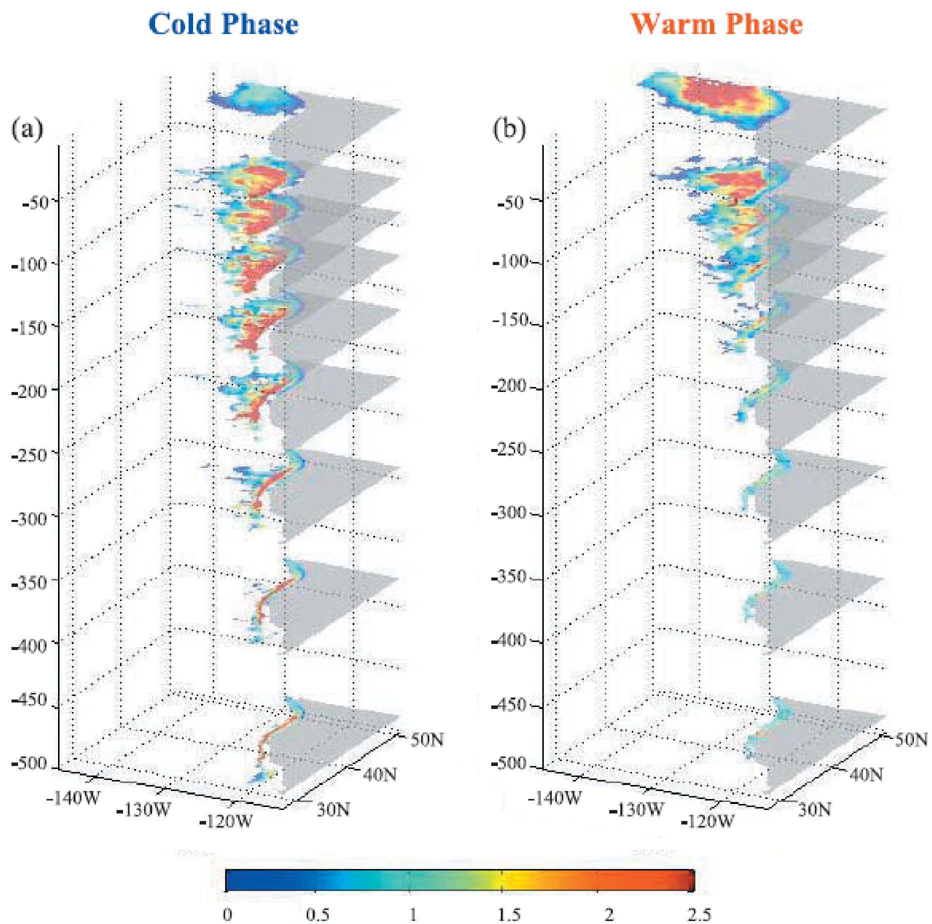
- Model: Regional Ocean Modeling System (ROMS) (Moore et *al.*, 2004)

- Potential mechanisms for the sharp decline in zooplankton biomass off the coast of California after the mid ‘70s



Pacific Decadal Oscillation (PDO)

Passive tracer introduced mid-April each year (55 yrs); adjoint run for 1 yr



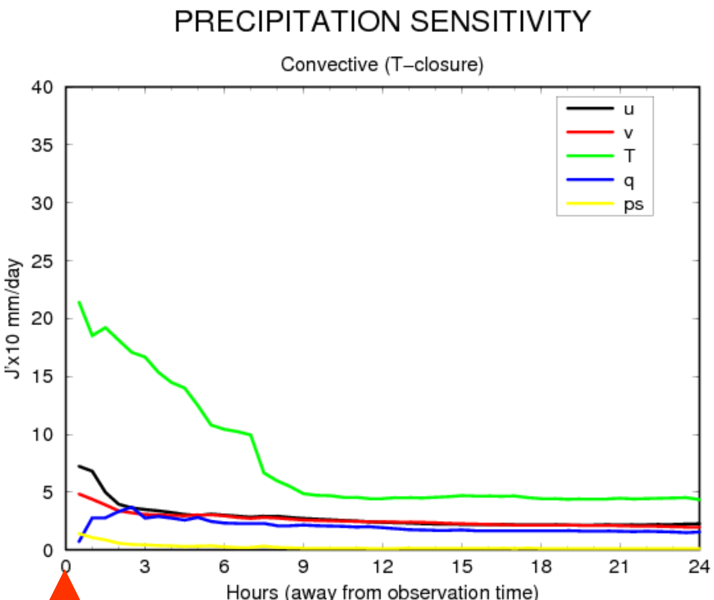
Cold phase: deeper source of upwelled, nutrient rich water.

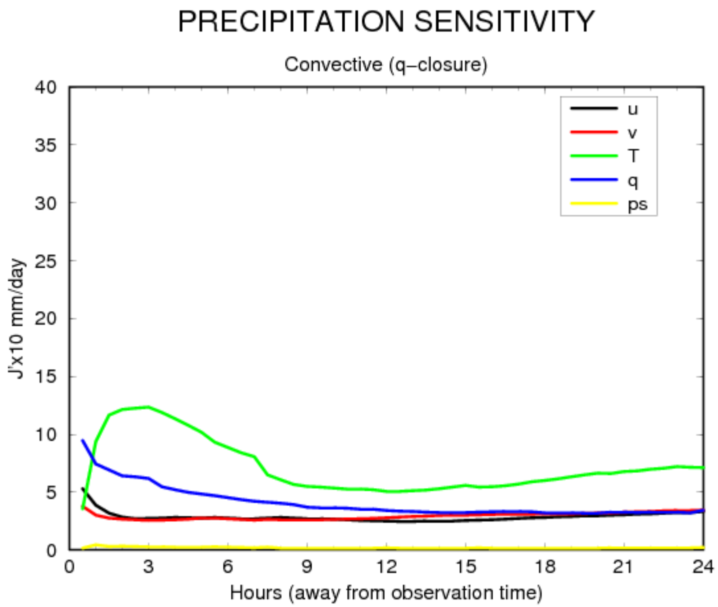
Supported by observations of phytoplankton & zooplankton

Origin of upwelling water (%) 1 yr prior to following year upwelling max.

Example 3: Sensitivity in surface precipitation for 2 convection schemes

- costfunction *J* is mean surface precipitation over a selected domain
- relative importance of u,v,T,q and ps depends on convection scheme





start of adjoint model integration

(Mahfouf and Bilodeau, 2007)

SINGULAR VECTORS

If more realistic models with many thousands of variables also have the property that a <u>few of the eigenvalues of A^TA</u> are much larger than the remaining, a study based upon a small ensemble of initial errors should give a reasonable estimate of the growth rate of random errors … It would appear then, that the best use could be made of computational time by choosing only a small number of error fields for superposition upon a particular initial state …(Lorenz 1965)

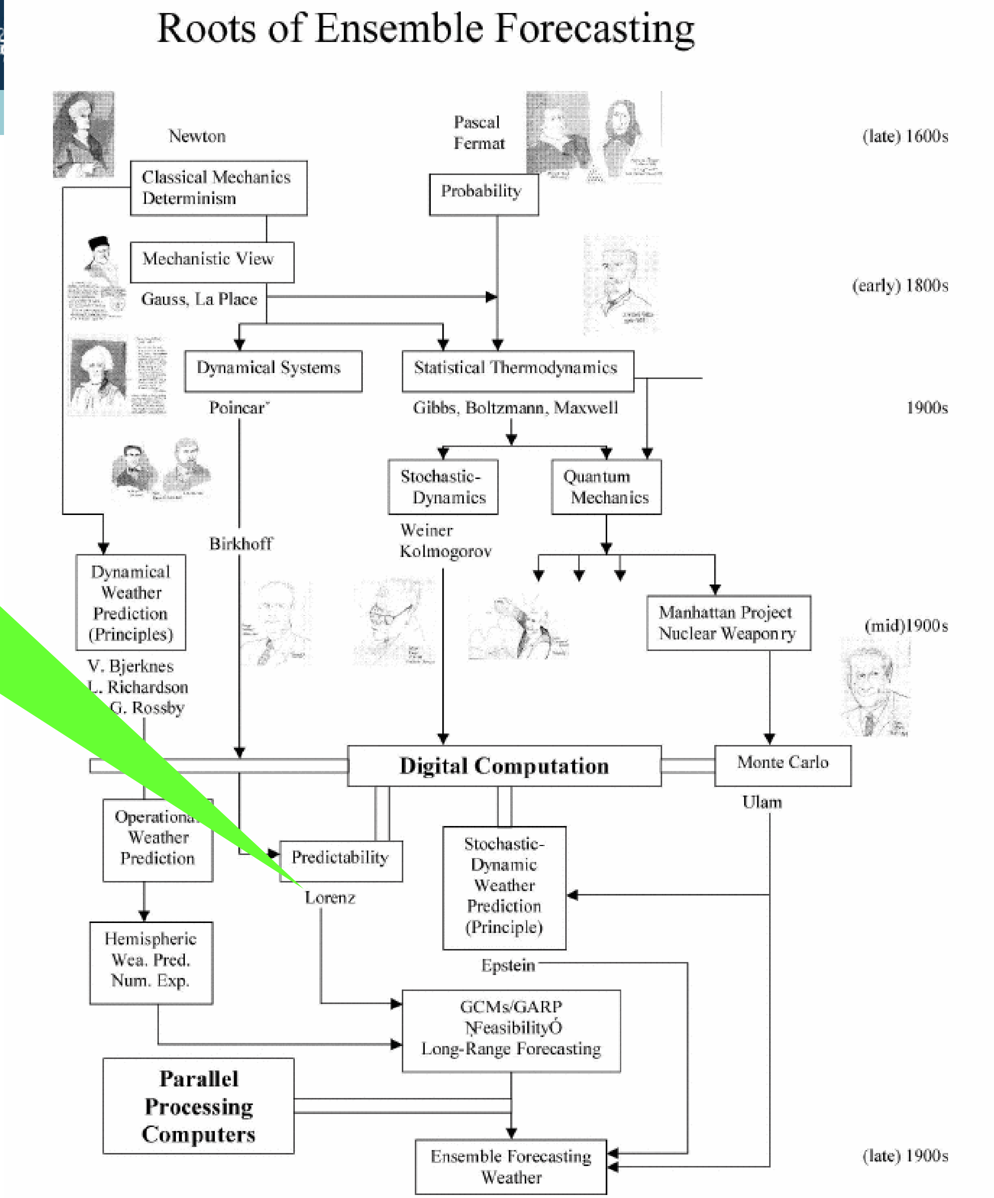


FIG. A1. Schematic diagram: Scientific roots of ensemble forecasting. (Portraits by the author.)

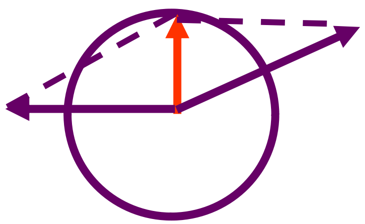
(Lewis, 2005)

Perturbations ε of the initial condition that maximize the ratio

final norm →

begin norm →

$$\frac{(\mathbf{EM}\varepsilon, \mathbf{M}\varepsilon)}{(\mathbf{C}\varepsilon, \varepsilon)} = \frac{(\mathbf{M}^{\mathrm{T}}\mathbf{EM}\varepsilon, \varepsilon)}{(\mathbf{C}\varepsilon, \varepsilon)}$$



Orr mechanism (1907)

where **M** is the propagator of the tangent linear model and **C** and **E** define a perturbation norm at initial and final time respectively.

- Popular choice: C=E= ‘total energy’ norm

$$\boxed{\mathbf{E} = \iiint [u'^2 + v'^2 + \frac{c_p}{T_r} T'^2] \, d\Sigma \frac{\partial p_r}{\partial \eta} d\eta + \iint [R \frac{c_p}{T_r} \ln p_s'^2] \, d\Sigma}$$

- Other choice for C: Hessian norm

$$\mathbf{C} = \nabla\nabla J_{\mathrm{4DVAR}} = \mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H} = \mathbf{A}^{-1}$$

A : estimate of the analysis error covariance matrix
B / R :background/observational error covariance matrix

Equivalently, Hessian singular vectors (HSV) satisfy:

$$\mathbf{M}^{T}\mathbf{E}\mathbf{M}\,\varepsilon = \lambda\,\mathbf{A}^{-1}\varepsilon$$

Solvers exist, even when **M** and **A** are not known explicitly

Note that

$$(\mathbf{E}^{1/2}\mathbf{M}\mathbf{A})\mathbf{M}^{T}\mathbf{E}\mathbf{M}\,\varepsilon = \lambda(\mathbf{E}^{1/2}\,\mathbf{M}\mathbf{A})\mathbf{A}^{-1}\varepsilon$$

or

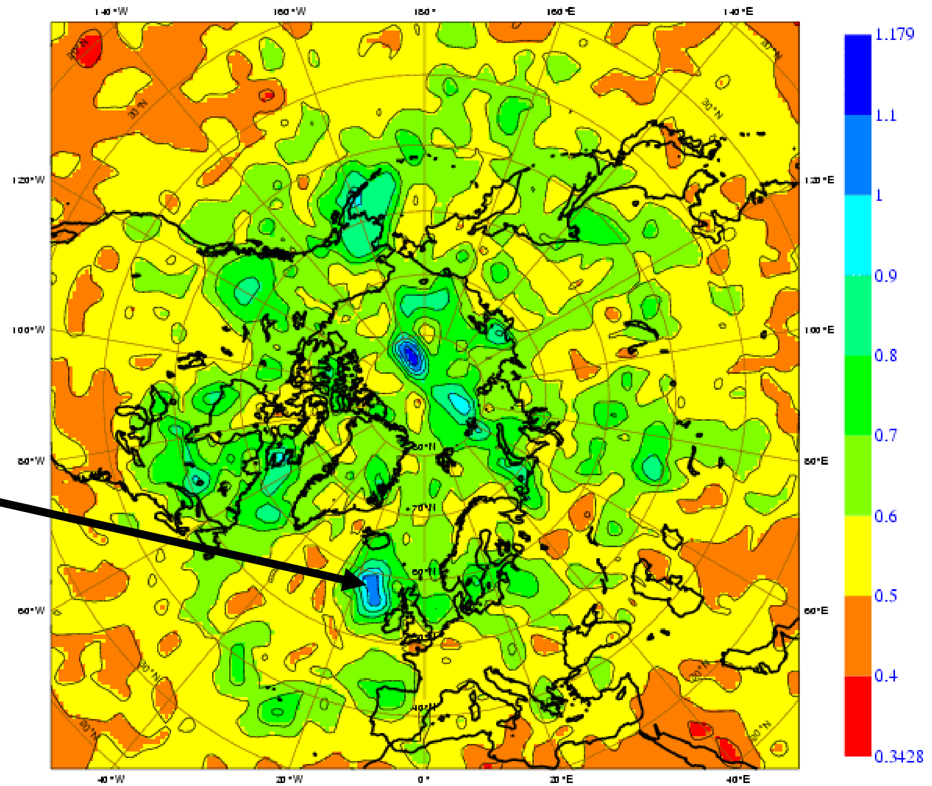
$$\mathbf{E}^{1/2}\,(\mathbf{M}\mathbf{A}\mathbf{M}^{\mathbf{T}})\,\mathbf{E}^{1/2}\,\mathbf{E}^{1/2}\,\mathbf{M}\,\varepsilon = \lambda\mathbf{E}^{1/2}\mathbf{M}\,\varepsilon$$

so

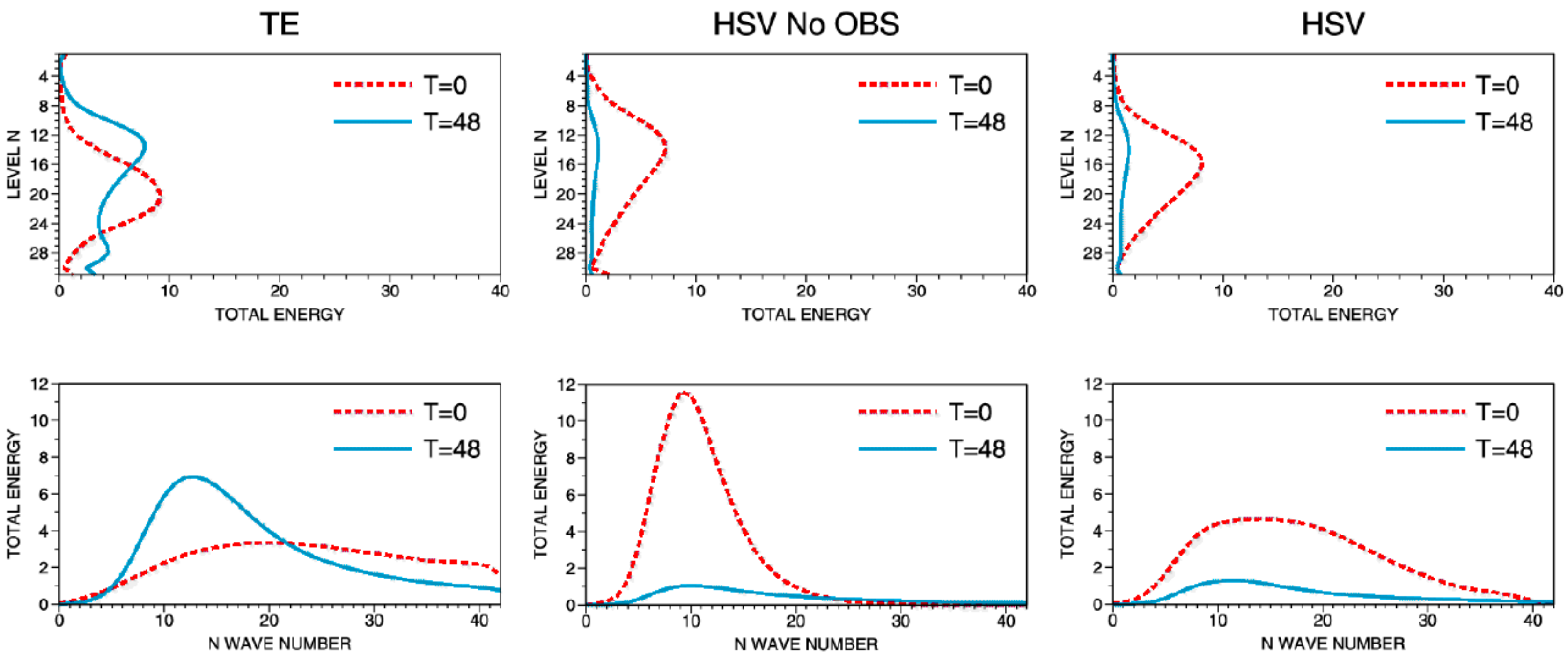
time-evolved HSVs $\mathbf{E}^{1/2}\mathbf{M}\,\varepsilon$ are eigenvectors of $\mathbf{E}^{1/2}(\mathbf{M}\mathbf{A}\mathbf{M}^{\mathbf{T}})\,\mathbf{E}^{1/2}$, which is the forecast error covariance matrix in the E-norm (Observe $\mathbf{M}\mathbf{A}\mathbf{M}^{T} = \mathbf{M}\,\delta\delta^{T}\mathbf{M}^{\mathbf{T}} = \mathbf{M}\delta\,(\mathbf{M}\delta)^{T}$)

Hessian singular vectors know about this:

higher analysis uncertainty



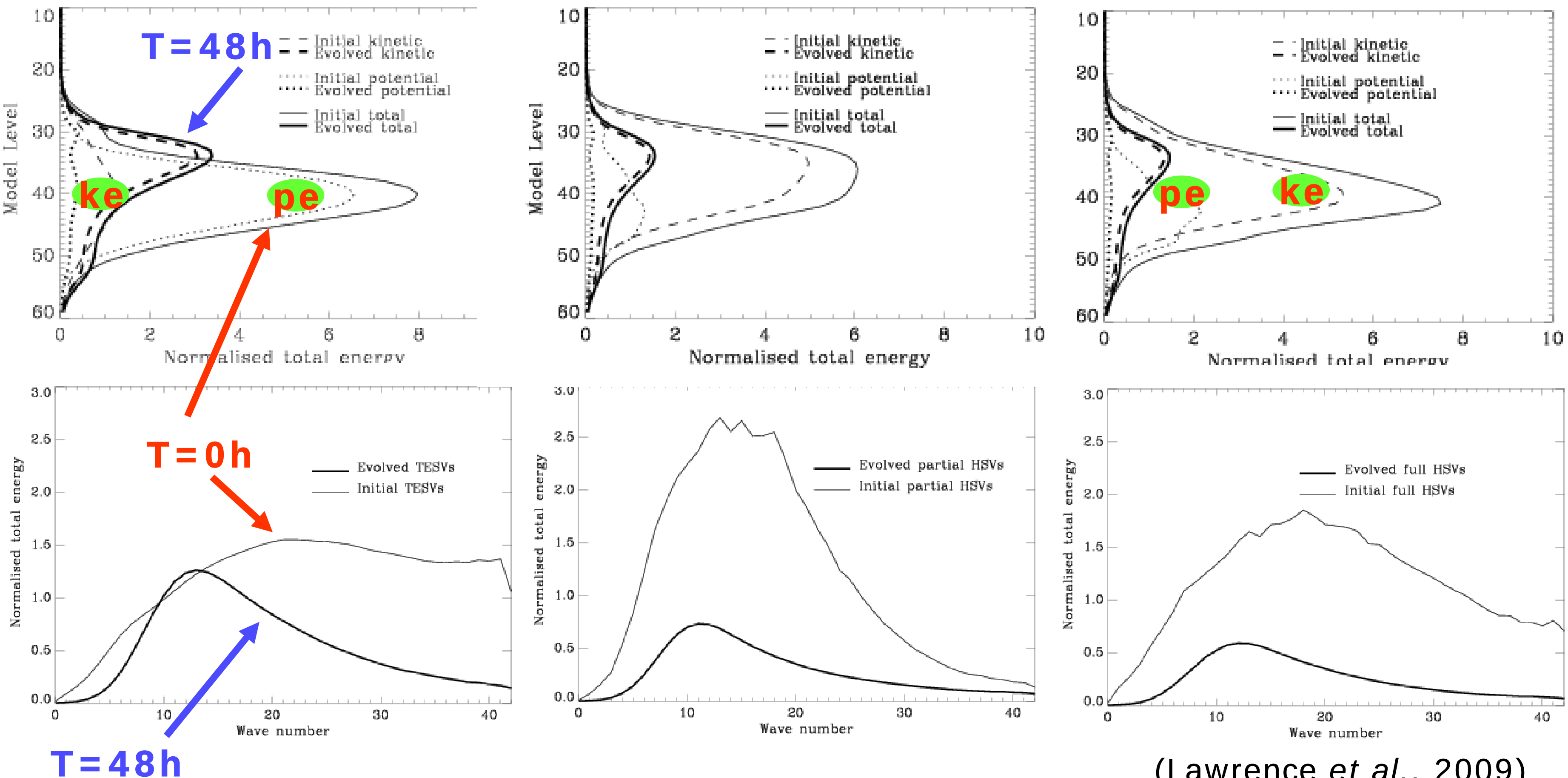
Analysis error variance field for temperature at 500 hPa



- reduced growth (50%) for HSV’s in terms of total energy.
- potential (kinetic) energy dominant for initial TE (Hessian) SVs
- no energy for wave number >25 in case of HSV without observations

(Barkmeijer et *al*., 2000)

TE **HSV no OBS** **HSV**



(Lawrence *et al.*, 2009)

The Observing System Research and Predictability Experiment (THORPEX)

One of the objectives:

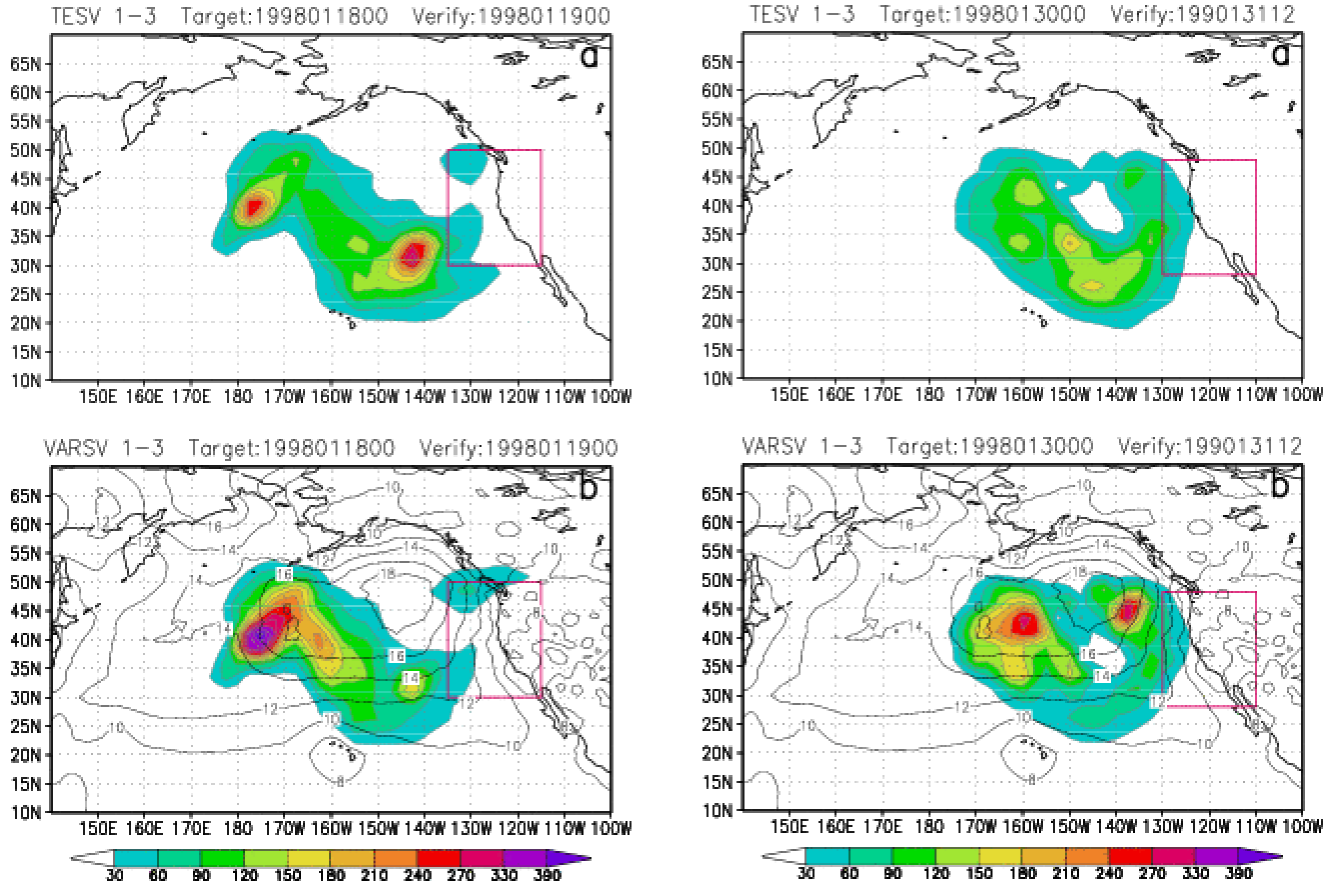
To improve weather forecasts by collecting observations in data-sensitive locations where analysis errors would have the largest impact on the forecast for a specific event or region of interest

data-sensitive areas

case 1 **case 2**

TESV

VARSV



(Gelaro *et al.*, 2002)

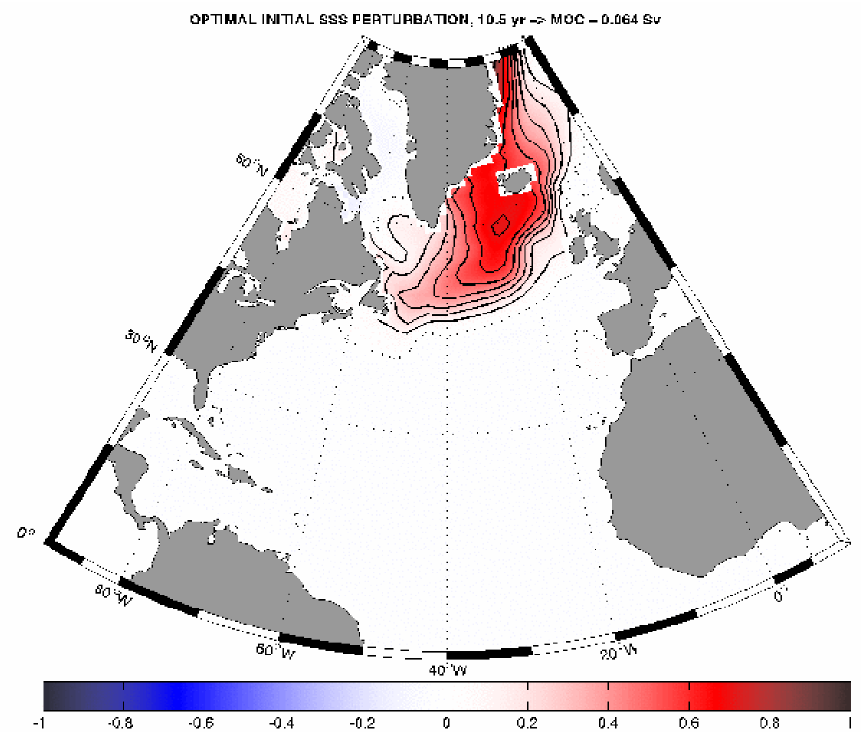
Example 2: Sea Surface Salinity SVs

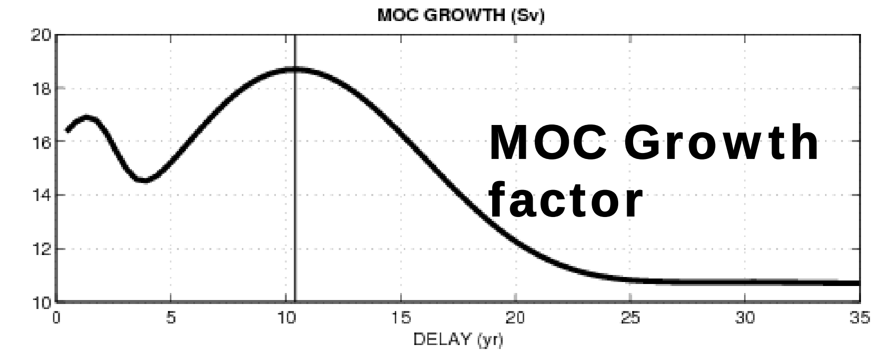
- Optimal surface salinitiy perturbations for the Meridional Overturning Circulation (Sévellec *et al.*, 2008)

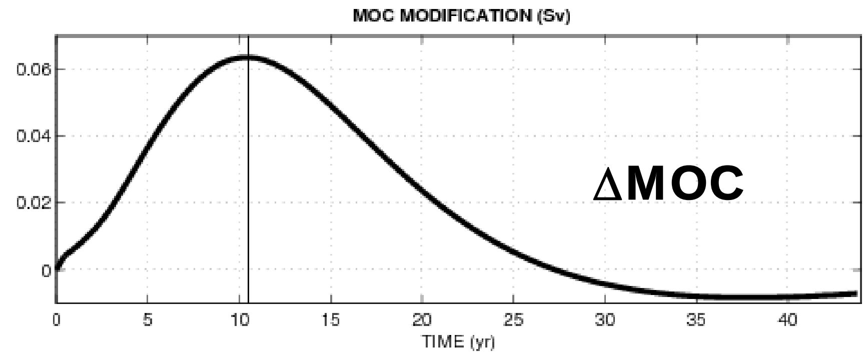
- Final norm measures (northward) mass transport at 48^{o}N in the Atlantic

- Model: OPA and OPA Tangent Adjoint Model (OPATAM, Weaver *et al., 2003)*

- Estimate the influence of sea surface salinity (SSS) perturbations on the North-Atlantic circulation as suggested by observations and modeling studies.







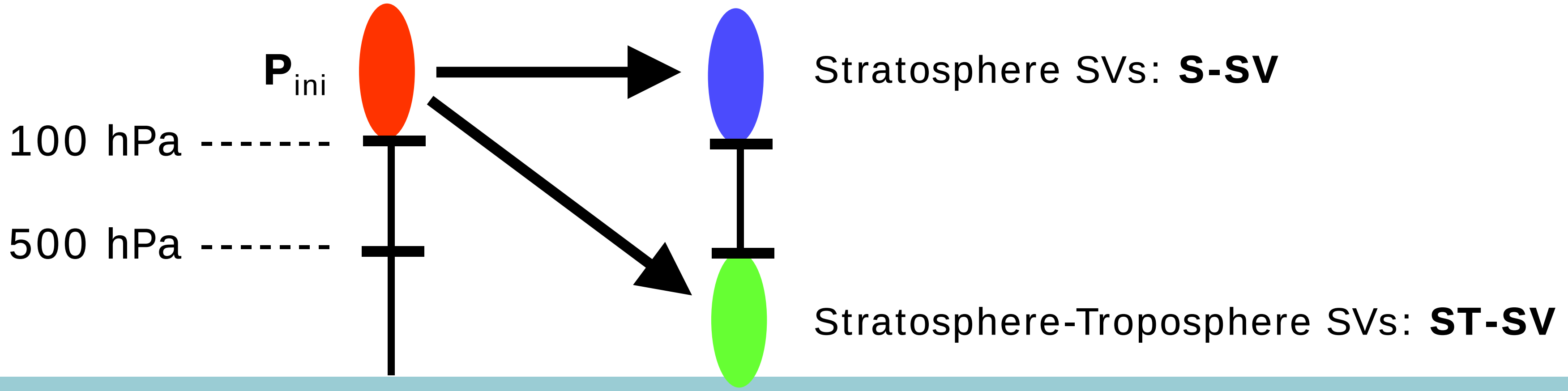
Optimal SSS Perturbation (scaled to 1 psu amp.)

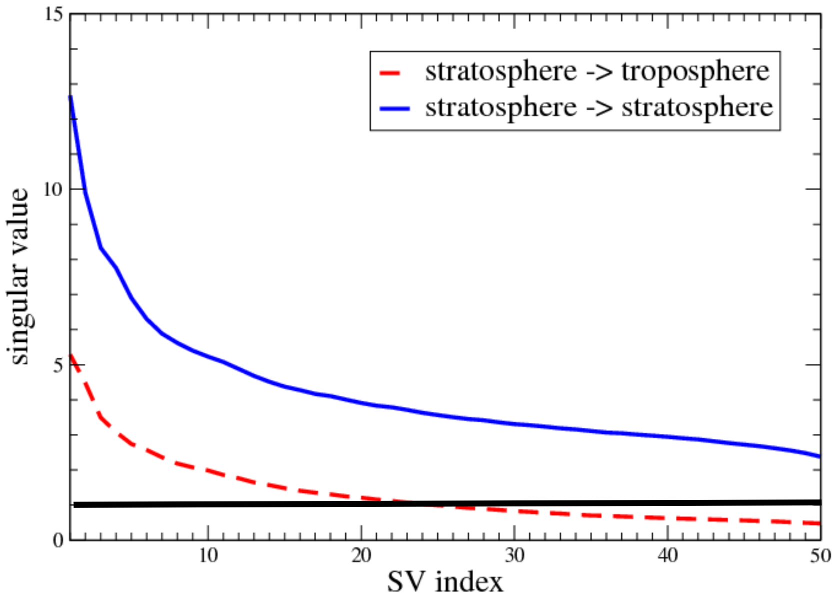
- Upper bounds based on GSA: ~0.8 Sv (11% of mean)
- SST perturbation less optimal: 5^{o}C required to achieve SSS impact

Example 3: Stratosphere-Troposphere interaction

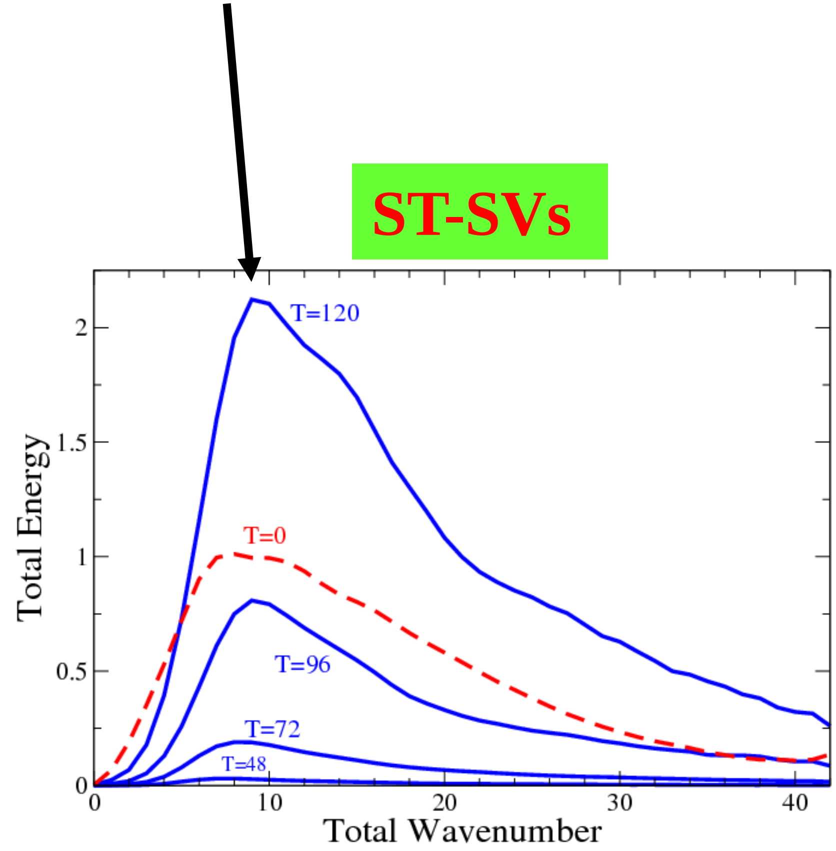
Look for structures ε that originate in the stratosphere and grow in the stratosphere or (lower) troposphere, by using appropriate projection operators $\mathbf{P}_{\mathrm{ini}}$ and $\mathbf{P}_{\mathrm{evo}}$.

$$\frac{<\mathbf{P}_{\mathrm{evo}}\mathbf{M}\mathbf{P}_{\mathrm{ini}}\varepsilon, \mathbf{P}_{\mathrm{evo}}\mathbf{M}\mathbf{P}_{\mathrm{ini}}\varepsilon>}{<\mathbf{P}_{\mathrm{ini}}\varepsilon\ , \mathbf{P}_{\mathrm{ini}}\varepsilon>}$$





Energy distribution below 500 hPa of linearly evolved ST-SV at T=120h

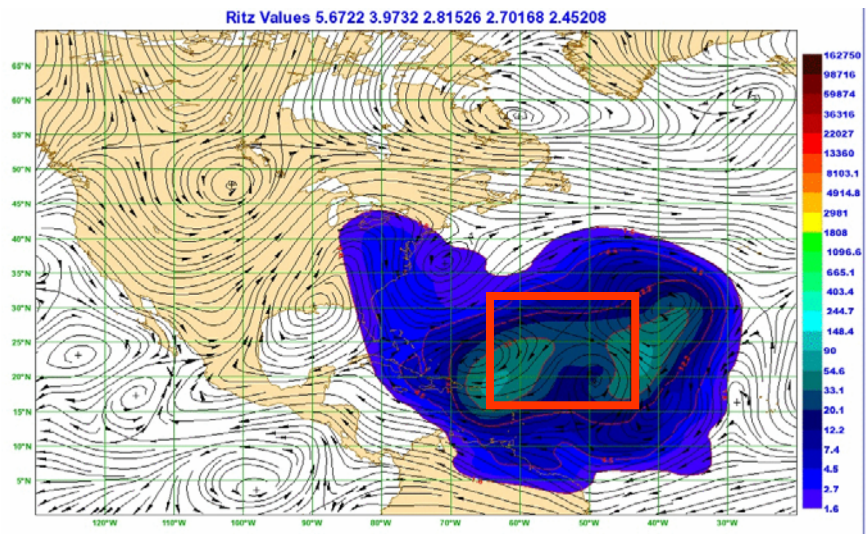


Downward propagating structures are possible even during summer conditions

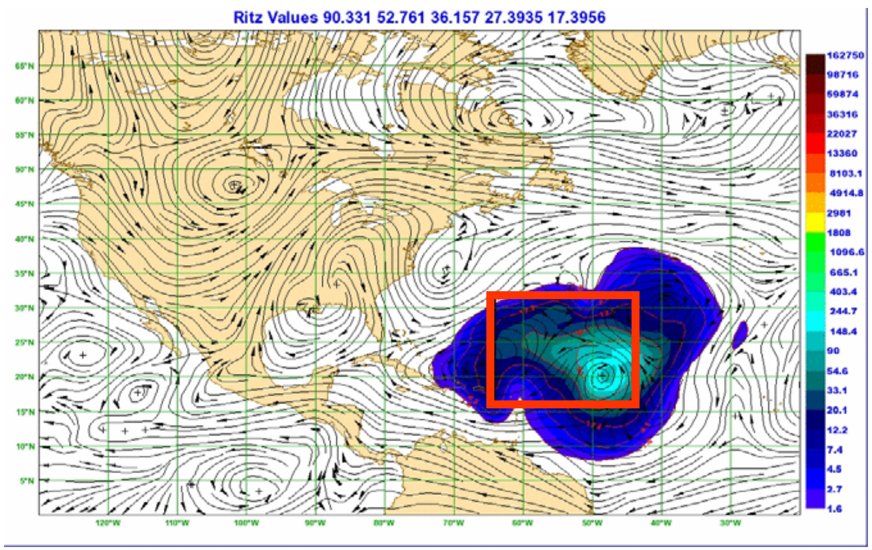
Example 4: Tropical singular vectors

- Case: Tropical cyclone Helene (September 2006) as seen from space shuttle Atlantis

resolution T42



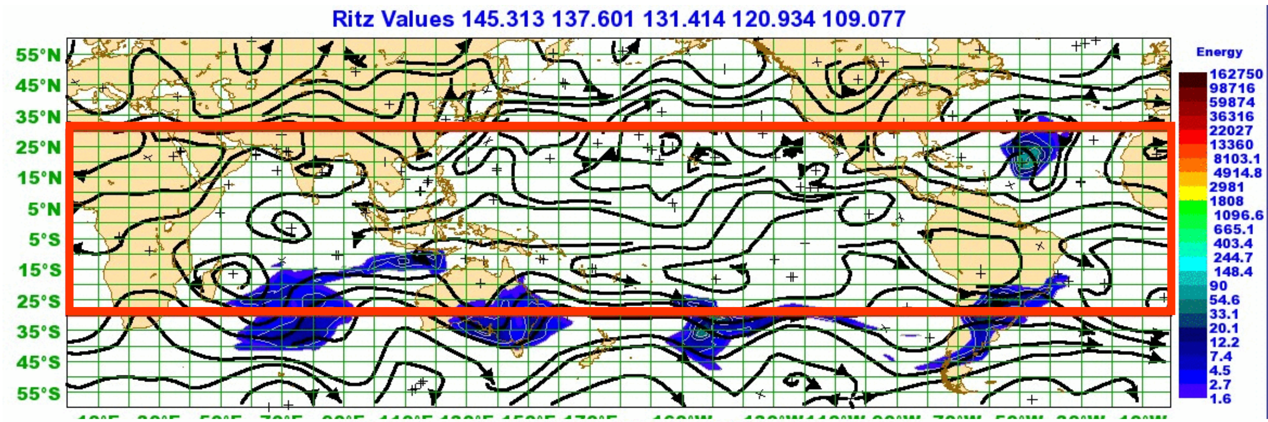
T159



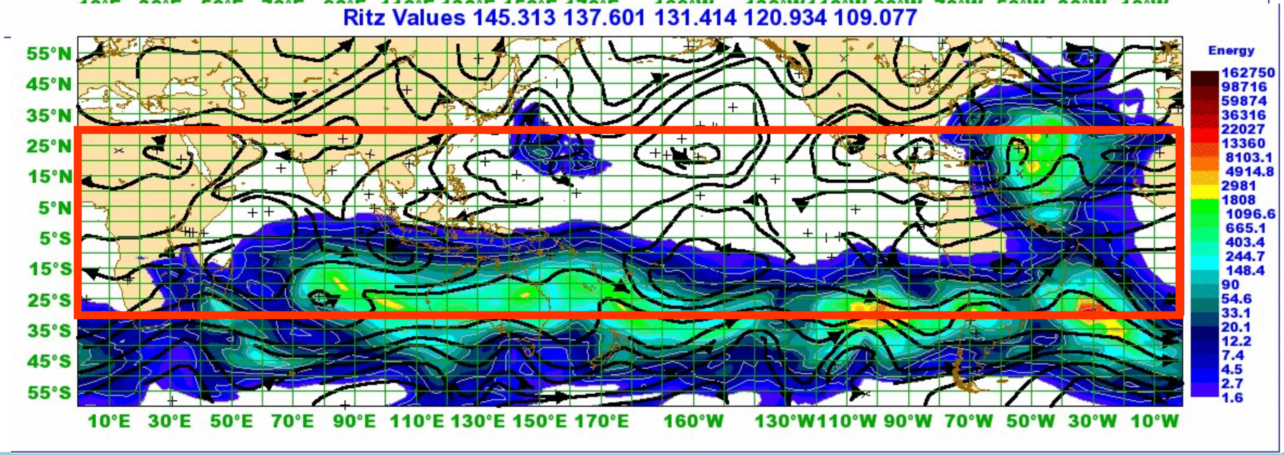
(Courtesy of S. Lang)

- Target area is entire tropical strip 30°S – 30°N !
- Tropical cyclone Helene shows up in the leading SVs

T=0h



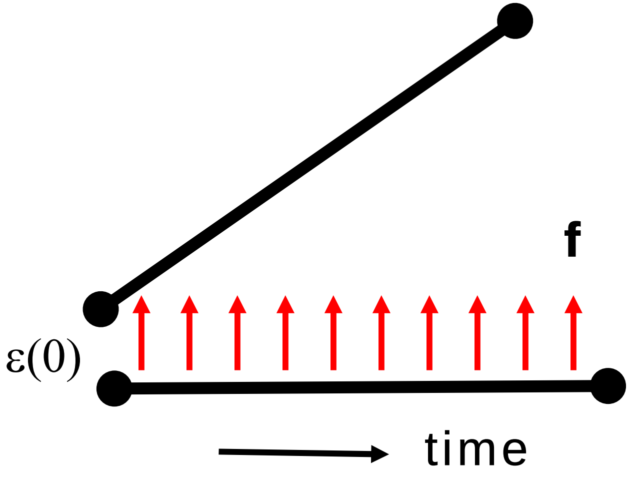
T=48h



Forcing

The regular SV and Sensitivity calculation can be exploited for studying model uncertainty. Assume error evolution is given by

$$\frac{d\varepsilon}{dt} = \mathbf{L}\varepsilon + \mathrm{f}(t)$$



then $\varepsilon(t) = \mathbf{M}\varepsilon(0) + \int_0^{t} \mathbf{M}(s,t)\mathrm{f}(s)\,\mathrm{d}s$

and in case $\varepsilon(0) = 0$

$$\varepsilon(t) = \int_0^{t} \mathbf{M}(s,t)\mathrm{f}(s)\,\mathrm{d}s \; = \mathbf{N}\mathrm{f}$$

Together with the corresponding adjoint $\mathbf{N}^*$
forcing singular vectors/sensitivity can be determined.

The Reynolds system

Assume error dynamics is governed by a stable 2x2-matrix **A**

$$\mathbf{A}=\begin{pmatrix} -1 & 10 \\ 0 & -2 \end{pmatrix}$$

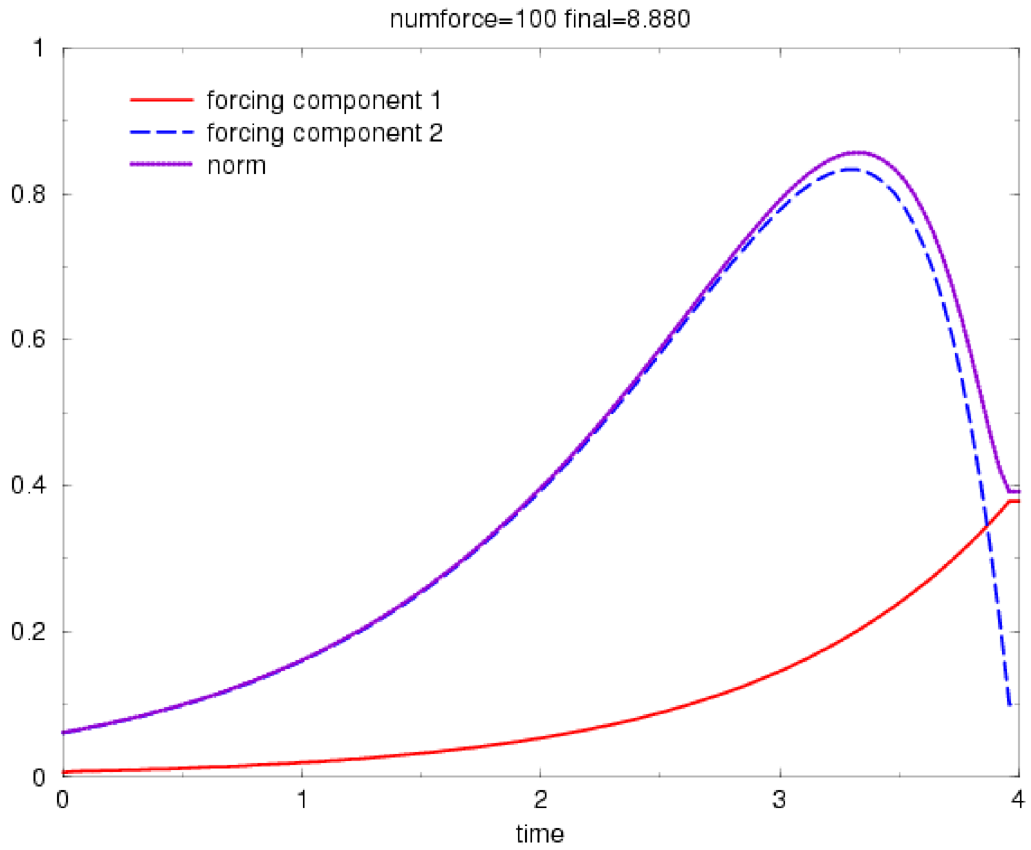
and further subject to a forcing f(s):

$$d\varepsilon/dt = \mathbf{A}\varepsilon + \mathbf{f}(t)$$

Look for Forcing Singular Vectors FSV, which maximize

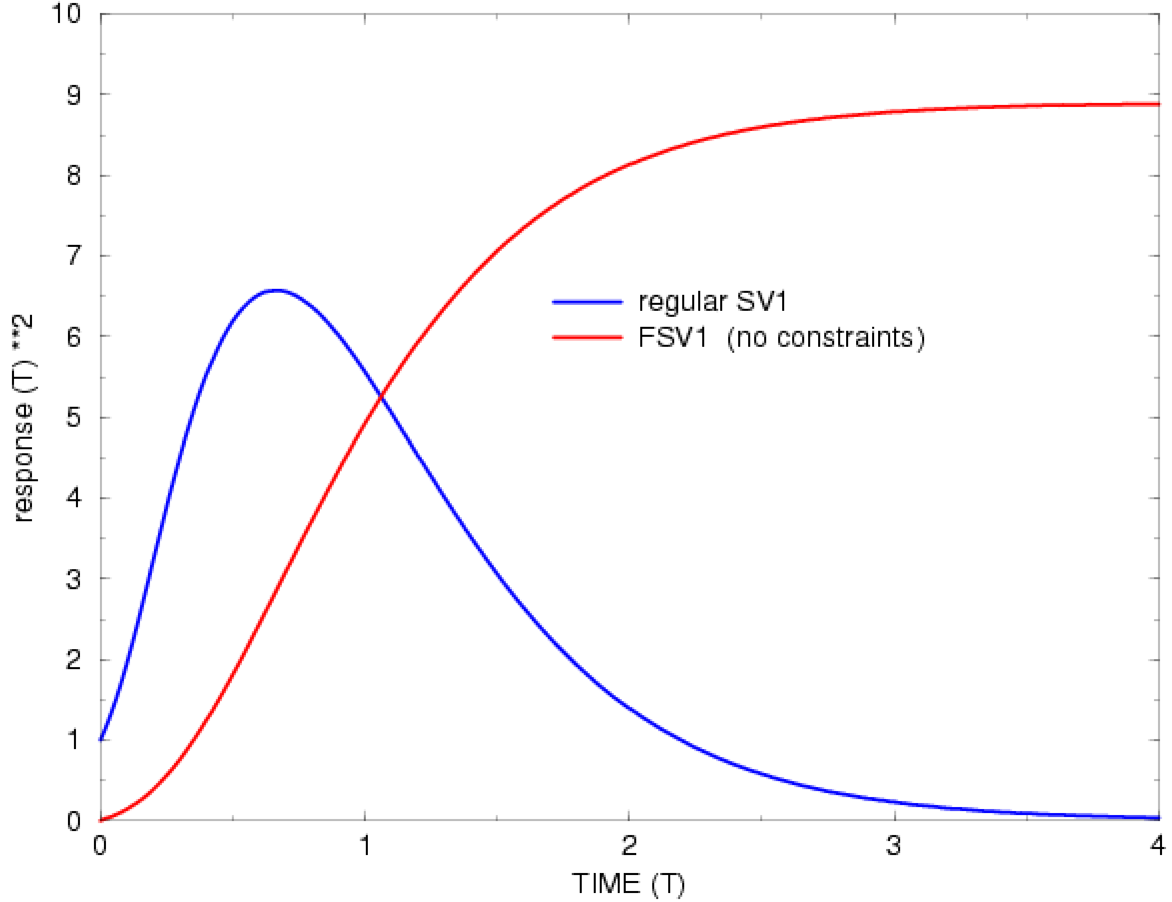
$(\mathbf{N}f , \mathbf{N}f)$ for unit-sized f and $\mathbf{N}f = \int_0^t \mathbf{M}(s,t)\mathrm{f}(s)\,\mathrm{d}s$

Components of the leading FSV for the Reynolds system using an optimization time of 4 units.

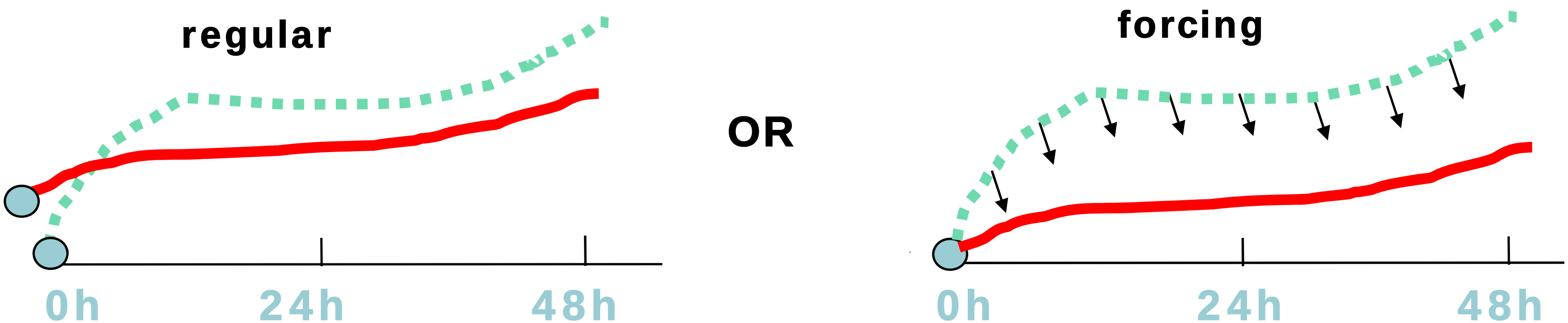


(Farrell and Ioannou, 2005)

Size of the perturbation at optimization time



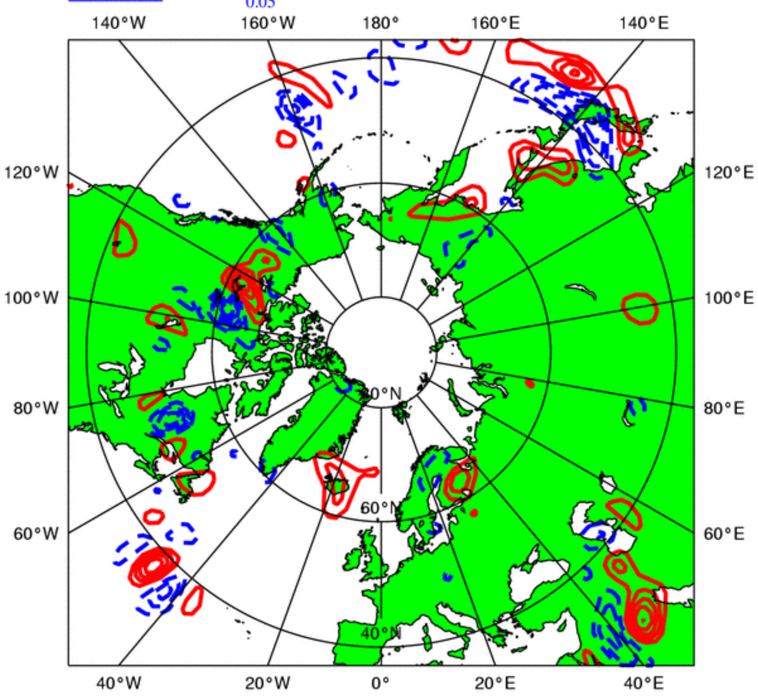
Tendency perturbations in the sensitivity calculation



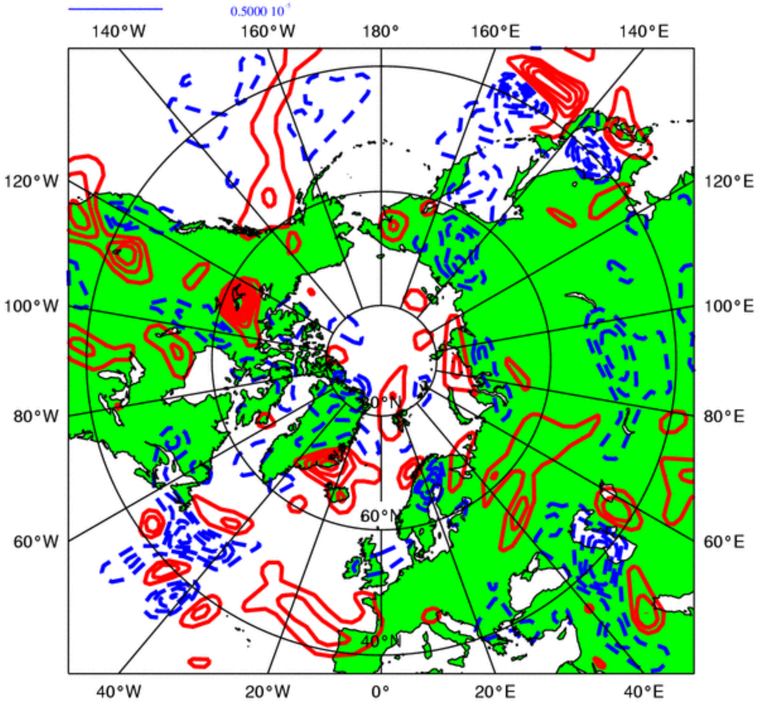
Use **N** and its adjoint **N*** (instead of **M** and **M***) to determine a constant tendency perturbation (forcing) **f,** which decreases the forecast error, or equivalently, minimizes the following cost function:

$$J=\langle fc-an+\mathbf{Nf}\ ,\ fc-an+\mathbf{Nf}\rangle$$

Temperature perturbations at 700 hPa

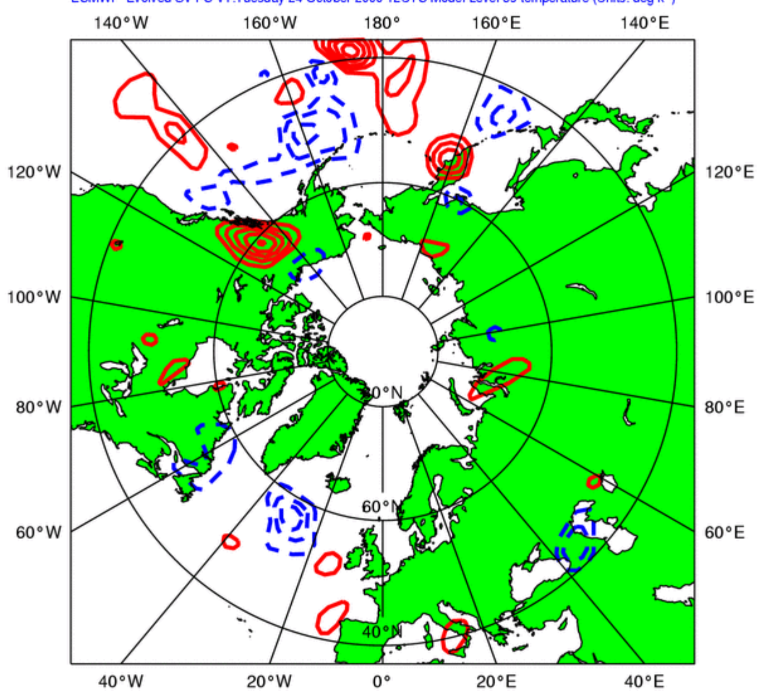


default sensitivity
(T=0h)

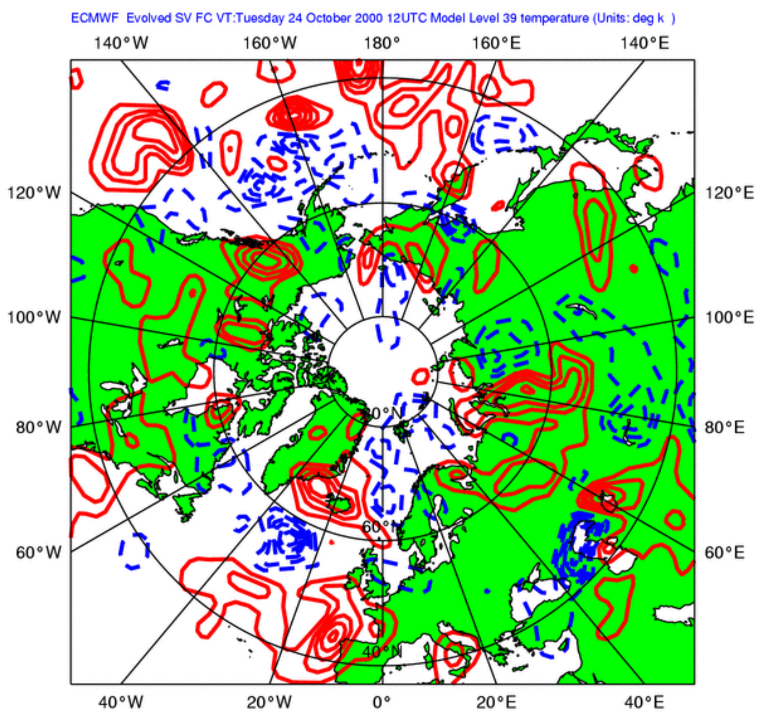


forcing sensitivity
(constant during 48-h forecast)

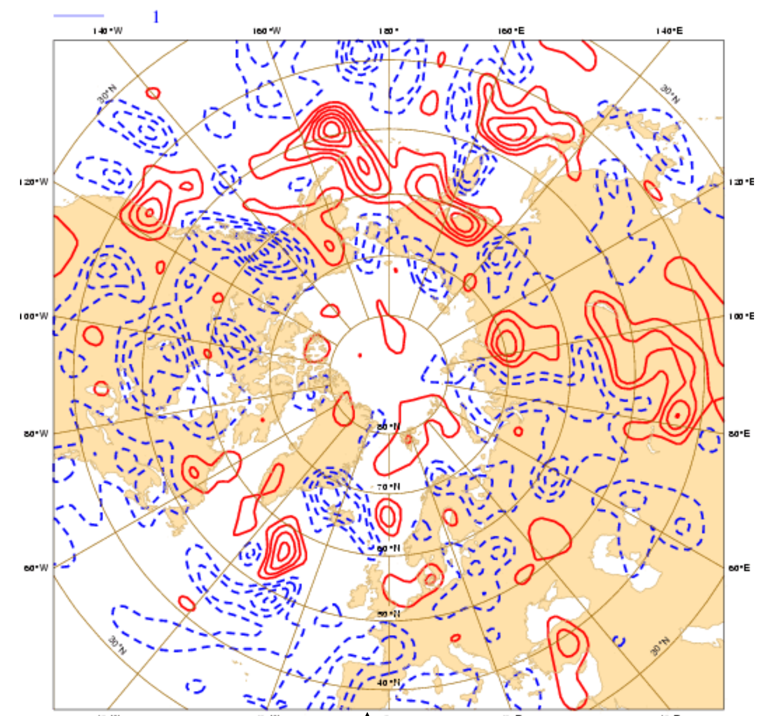
T500 impact at forecast time T=48h



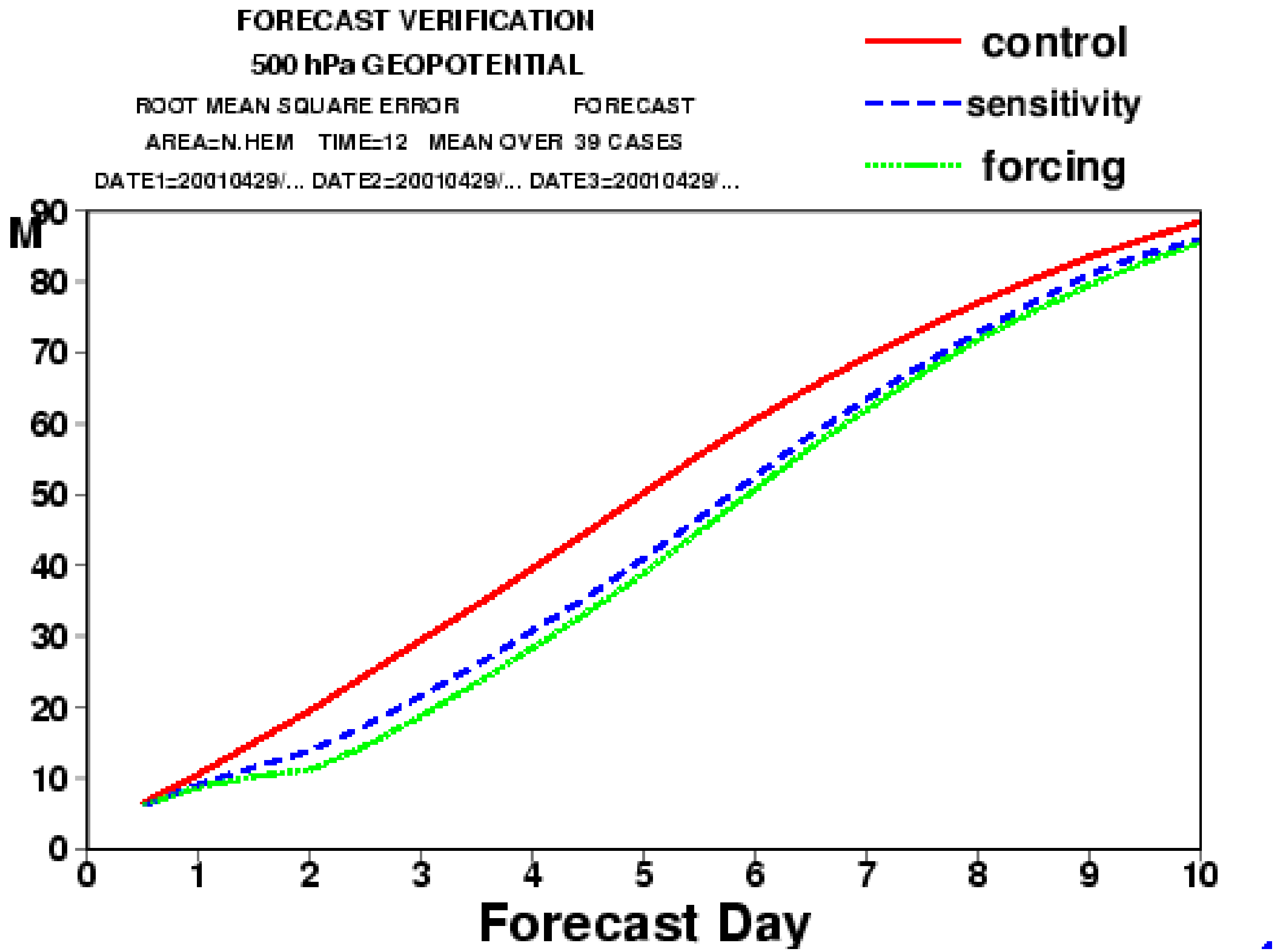
default



forcing

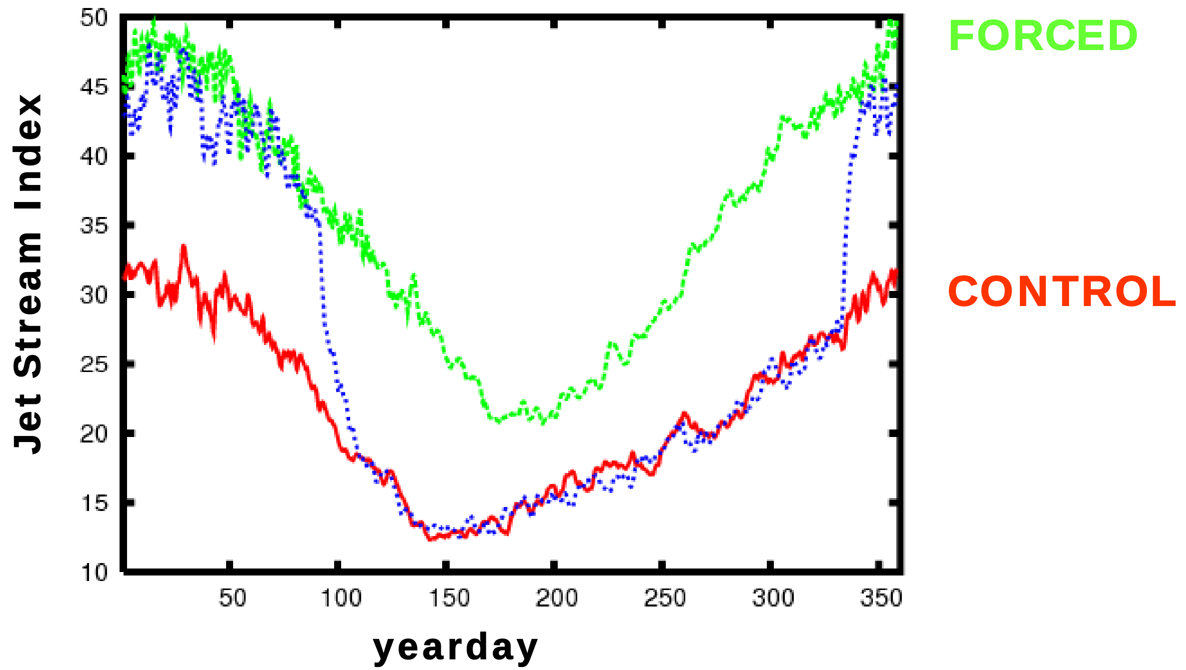


2-day forecast error temperature 500 hPa contour interval 1K



(Barkmeijer *et al.*, 2003)

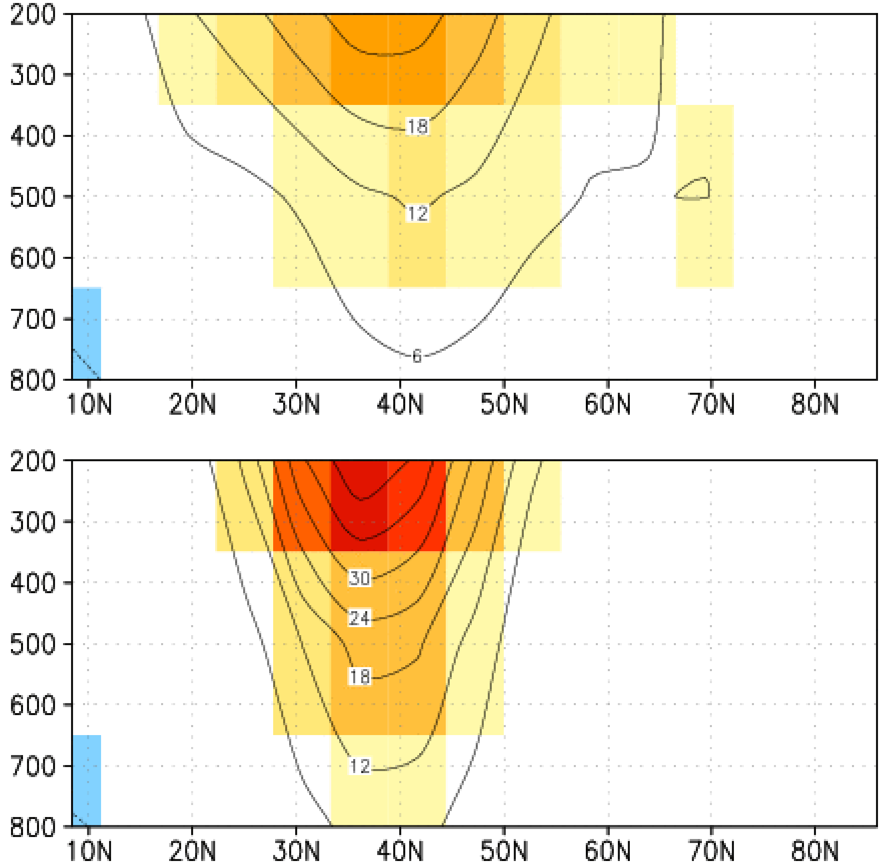
Example 1: Increasing the Atlantic subtropical jet



- Determine every 72 hours an atmospheric forcing, which increases the Atlantic subtropical jet
- Apply this forcing in a coupled atmosphere-ocean model
- Run the coupled model for 10 years

control

forced

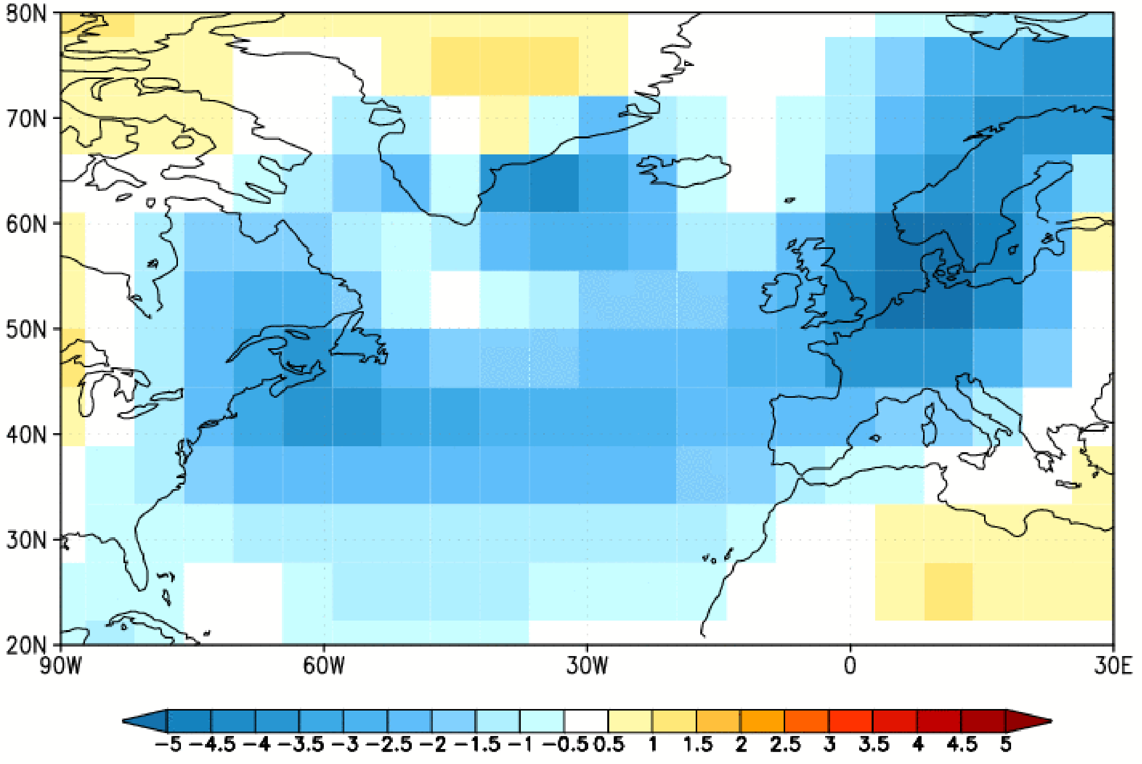


- Subtropical jet more zonal in the forced run

- Atmospheric meridional heat transport over the North Atlantic is reduced

(Van der Schrier et *al.*, 2007)

Resulting in a cooling over the Atlantic in the forced run.



Change in surface temperature for the forced run compared to the control run

Adjoint models have been very useful and instructive in understanding ocean and atmospheric models!

Thank you for your attention

Stratospheric forcing with the ECMWF model

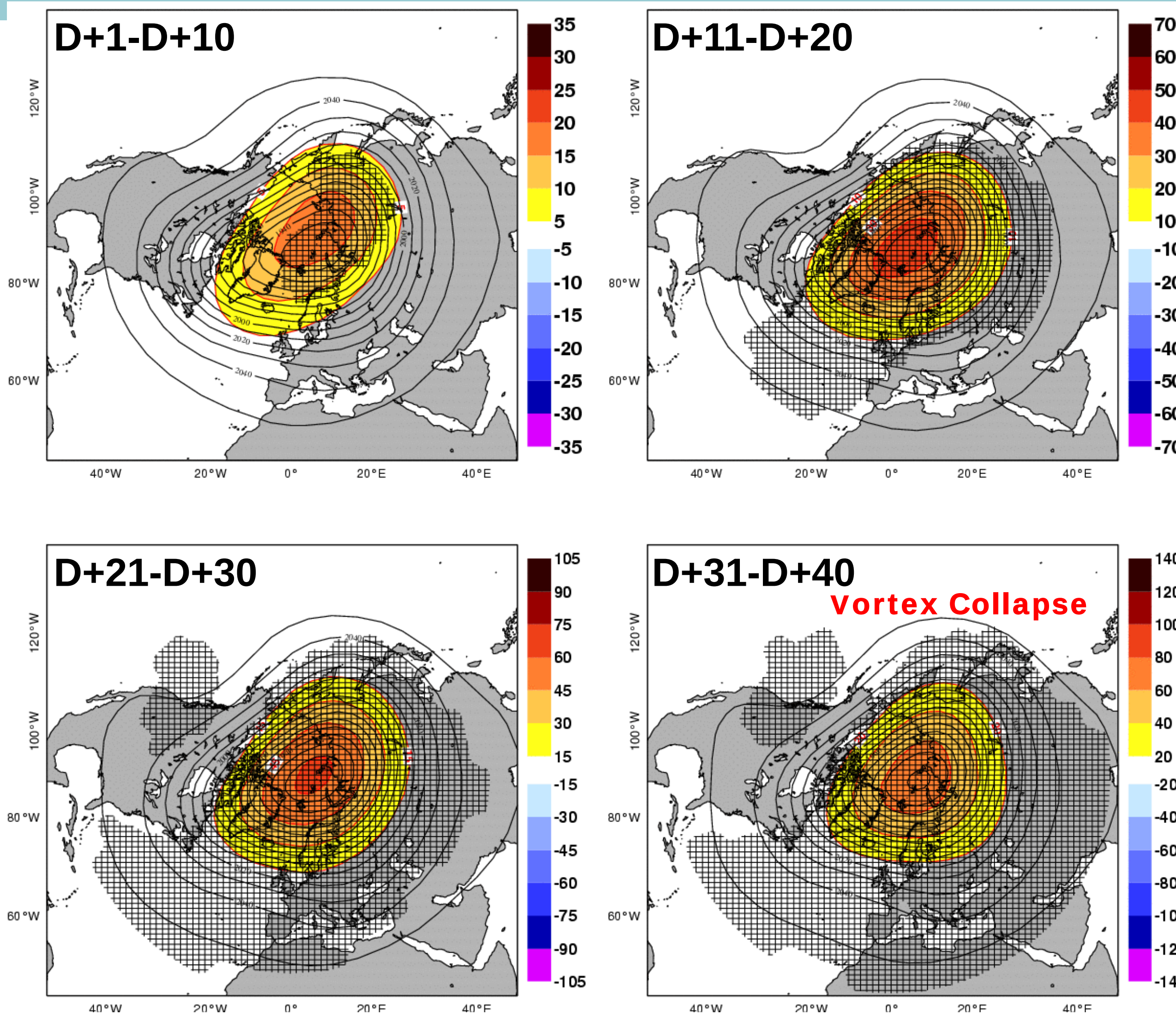
- Apply a forcing **F** to the model tendency
- Forcing **F** is constructed to change the strength of the stratospheric polar vortex (18 sensitivity calculations).
- Perform 60 forty-day T95L60 integrations during DJF 1982-2001 with

$$dx/dt=EC(x),\ dx/dt=EC(x) + \mathbf{F} \text{ and } dx/dt=EC(x) - \mathbf{F}$$

- Forcing **F** is small and zero below 150 hPa
- and **F** is kept constant during the integration

Jung and Barkmeijer (2006)

Stratospheric Response (50hPa) Weak-Ctl



Z1000 Response (Weak-CTL)