Parametrizations in Data Assimilation

ECMWF Training Course
19-21 May 2015

Philippe Lopez

Physical Aspects Section, Research Department, ECMWF
(Room 002)

Parametrizations in Data Assimilation

Lecture 1

- **Introduction**
- **An example of physical initialization**
- **A very simple variational assimilation problem**

Lecture 2

- **3D-Var assimilation**
- **The concept of adjoint**
- **4D-Var assimilation**
- **Tangent-linear and adjoint coding**

Lecture 3

- **Issues related to physical parametrizations in assimilation**
- **Physical parametrizations in ECMWF's current 4D-Var system**
- **Examples of applications involving linearized physical parametrizations**
- **Summary and conclusions**

Why do we need data assimilation?

• By construction, **numerical weather forecasts are imperfect**:

← **discrete** representation of the atmosphere in space and time (horizontal and vertical grids, spectral truncation, time step)

← **subgrid-scale processes** (e.g. turbulence, convective activity) **need to be parametrized** as functions of the resolved-scale variables.

← **errors in the initial conditions.**

• **Physical parametrizations** used in NWP models are constantly being improved:

→ more and more prognostic variables (cloud variables, precipitation, aerosols),
→ more and more processes accounted for (e.g. detailed microphysics).

• However, they remain **approximate representations of the true atmospheric behaviour**.

• Another way to improve forecasts is to **improve the initial state.**

• The goal of **data assimilation** is **to periodically constrain the initial conditions of the forecast using a set of accurate observations** that provide our best estimate of the local true atmospheric state.

General features of data assimilation

- **Goal:** to produce an accurate four dimensional representation of the atmospheric state to initialize numerical weather prediction models.

- This is achieved by combining in an optimal statistical way all the information on the atmosphere, available over a selected time window (usually 6 or 12h):

  - ✓ **Observations** with their accuracies (error statistics),
  - ✓ **Short-range model forecast** (background) with associated error statistics,
  - ✓ **Atmospheric equilibria** (e.g. geostrophic balance),
  - ✓ **Physical laws** (e.g. perfect gas law, condensation)

- The optimal atmospheric state found is called the **analysis**.

Which observations are assimilated?

Operationally assimilated since many years ago:

- * **Surface measurements** (SYNOP, SHIPS, DRIBU,…),
- * **Vertical soundings** (TEMP, PILOT, AIREP, wind profilers,…),
- * **Geostationary satellites** (METEOSAT, GOES,…)
  Polar orbiting satellites (NOAA, SSM/I, AIRS, AQUA, QuikSCAT,…):
  - radiances (infrared & passive microwave in clear-sky conditions),
  - products (motion vectors, total column water vapour, ozone,…).

More recently:

- * **Satellite radiances/retrievals in cloudy and rainy regions** (SSM/I, TMI,…),
- * Precipitation measurements from **ground-based radars** and **rain gauges.**

Still experimental:

- * Satellite cloud/precipitation radar reflectivities/products (TRMM, CloudSat),
- * Lidar backscattering/products (wind vectors, water vapour) (CALIPSO),
- * GPS water vapour retrievals,
- * Satellite measurements of aerosols, trace gases,....
- * Lightning data (TRMM-LIS).

Why physical parametrizations in data assimilation?

- In current operational systems, most used observations are directly or indirectly related to **temperature, wind, surface pressure** and **humidity** outside cloudy and precipitation areas (~ 10 million observations assimilated in ECMWF 4D-Var every 12 hours).

- Physical parametrizations are used during the assimilation to link the model's prognostic variables (typically: T, u, v, q_v and P_s) to the observed quantities (e.g. radiances, reflectivities,…).

- Observations related to **clouds** and **precipitation** are starting to be routinely assimilated,

→ but how to convert such information into proper corrections of the model's initial state (prognostic variables T, u, v, q_v and, P_s) is not so straightforward.

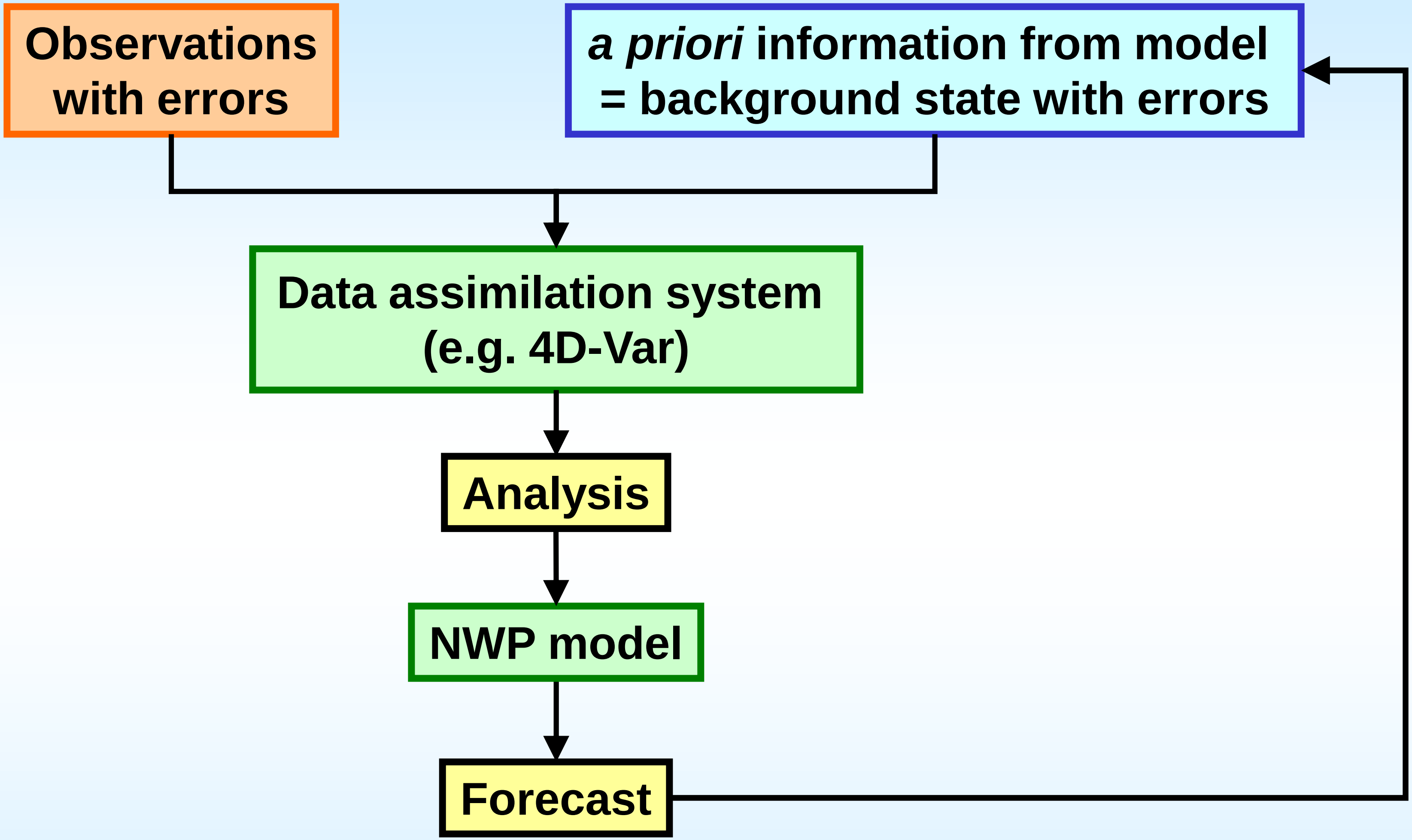
For instance, problems in the assimilation can arise from the discontinuous or non-linear nature of moist processes.

Improvements are still needed…

- More observations are needed to improve the analysis and forecasts of:
  - Mesoscale phenomena (convection, frontal regions),
  - Vertical and horizontal distribution of clouds and precipitation,
  - Planetary boundary layer processes (stratocumulus/cumulus clouds),
  - Surface processes (soil moisture),
  - The tropical circulation (monsoons, squall lines, tropical cyclones).

- Recent developments and improvements have been achieved in:
  - **Data assimilation techniques** (OI → 3D-Var → 4D-Var → Ensemble DA),
  - **Physical parametrizations** in NWP models (prognostic schemes, detailed convection and large-scale condensation processes),
  - **Radiative transfer models** (infrared and microwave frequencies),
  - **Horizontal and vertical resolutions** of NWP models (currently at ECMWF: T1279 ~ 15 km, 137 levels),
  - **New satellite instruments** (incl. microwave imagers/sounders, precipitation/cloud radars, lidars,…).

To summarize…



<u>Physical parametrizations are needed in data assimilation:</u>

- to link the model variables to the observed quantities,

- to evolve the model state in time during the assimilation (e.g. 4D-Var).

Empirical initialization

<u>Example from Ducrocq *et al.* (2000), Météo-France</u>:

- Using the mesoscale research model Méso-NH (prognostic clouds and precipitation).

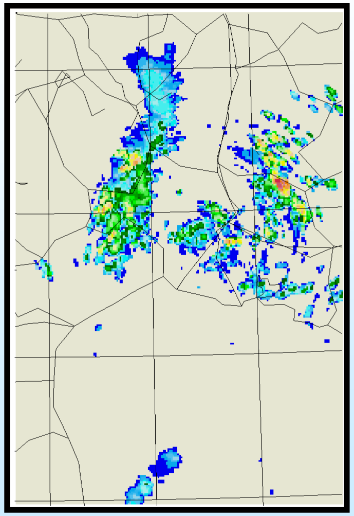
- Particular focus on strong convective events.

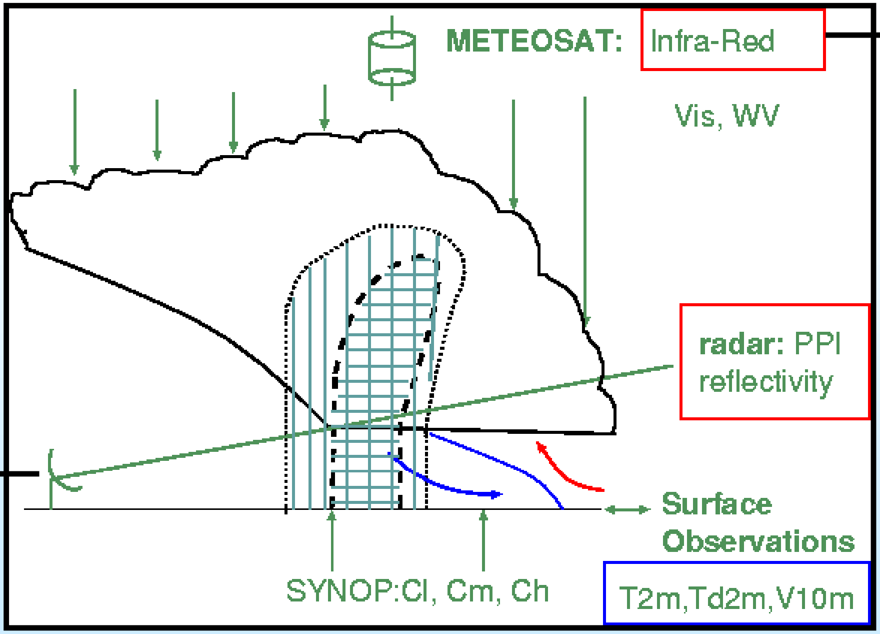
- **<u>Method:</u>** Before running the forecast:

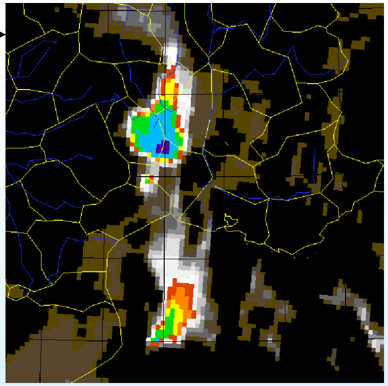
  1) A mesoscale surface analysis is performed (esp. to identify convective cold pools)

  2) the model humidity, cloud and precipitation fields are **empirically adjusted** to match ground-based precipitation radar observations and METEOSAT infrared brightness temperatures.

Radar

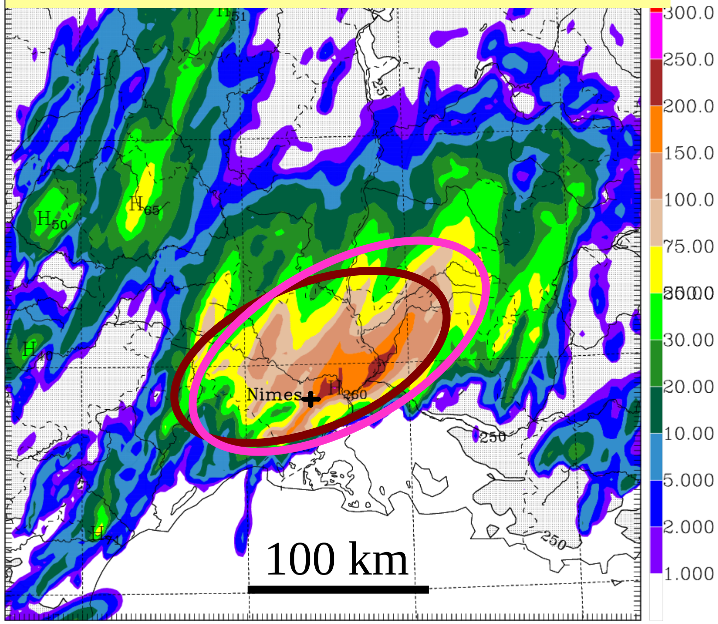




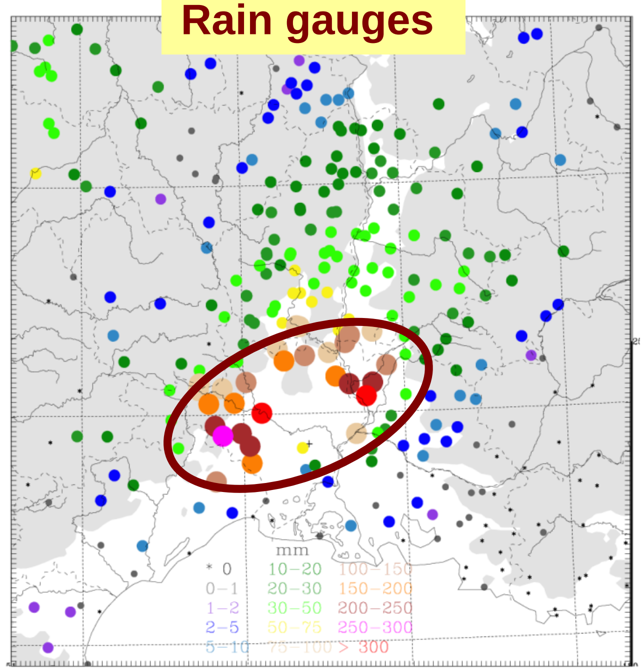


METEOSAT

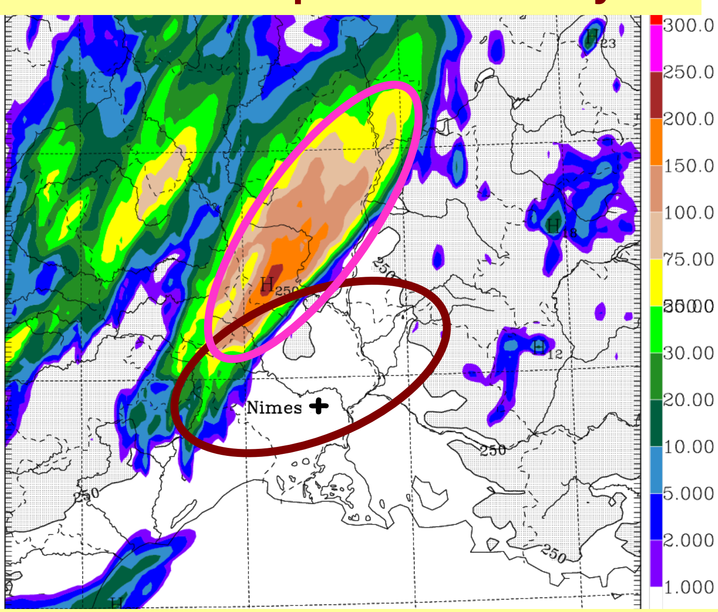
12h FC from **modified analysis**



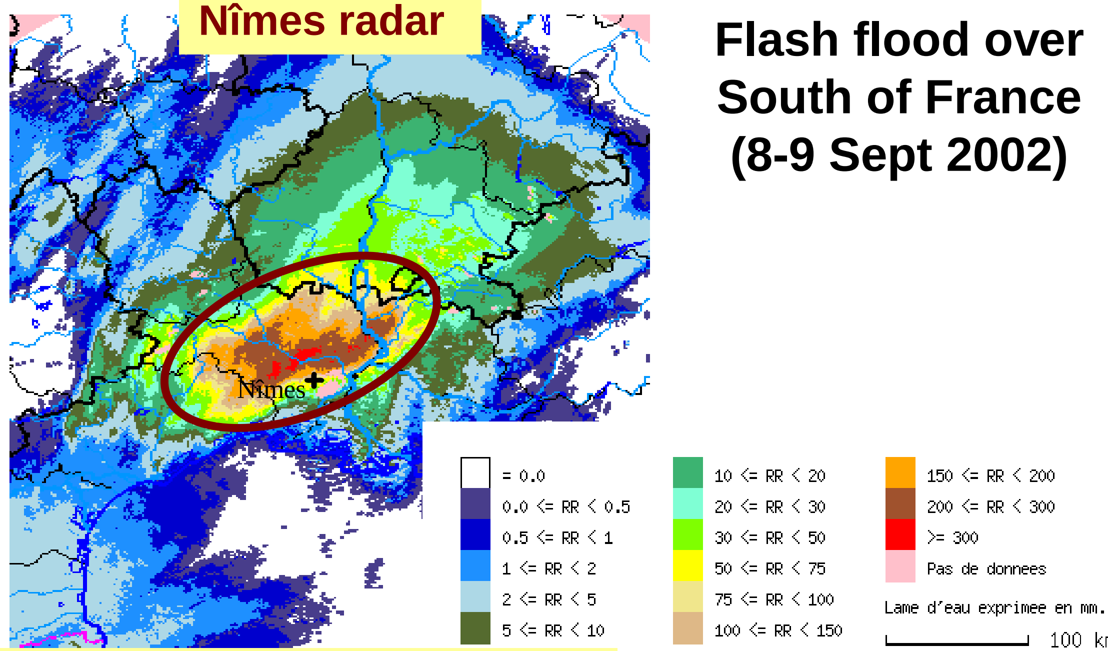
Rain gauges



12h FC from **operational analysis**



Nîmes radar



Ducrocq et al. (2004)

2.5-km resolution model Méso-NH

Flash flood over South of France (8-9 Sept 2002)

12h accumulated precipitation: 8 Sept 12 UTC → 9 Sept 2002 00 UTC

A very simple example of variational data assimilation

- Short-range forecast (background) of 2m temperature from model: x_b with error σ_b.
- Simultaneous observation of 2m temperature: y_o with error σ_o.

The best estimate of 2m temperature (x_a=analysis) minimizes the following cost function:

$$J(x) = \underbrace{\frac{1}{2}\left(\frac{x - x_b}{\sigma_b}\right)^2}_{J_b} + \underbrace{\frac{1}{2}\left(\frac{x - y_o}{\sigma_o}\right)^2}_{J_o}$$

= quadratic distance to background and obs

(weighted by their errors)

In other words:

$$\left(\frac{dJ}{dx}\right)_{x=x_a} = \frac{(x_a - x_b)}{\sigma_b^{\ 2}} + \frac{(x_a - y_o)}{\sigma_o^{\ 2}} = 0 \quad \Leftrightarrow \quad x_a = x_b + \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}(y_o - x_b)$$

And the analysis error, σ_a, verifies:

$$\frac{1}{\sigma_a^2} = \frac{1}{\sigma_b^2} + \frac{1}{\sigma_o^2} \quad \Rightarrow \quad \sigma_a^2 \leq \min(\sigma_b^2, \sigma_o^2)$$

The analysis is a linear combination of the model background and the observation weighted by their respective error statistics.

3D-Var assimilation

$$\mathcal{J}(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + \frac{1}{2}(H(\mathbf{x})-\mathbf{y}_o)^T\mathbf{R}^{-1}(H(\mathbf{x})-\mathbf{y}_o)$$

$\mathbf{B}$ is the background error covariance matrix, $\mathbf{R}$ is the observation error covariance matrix, H is the observation operator (used for converting model state vector $\mathbf{x} = (T, q_v, u, v)$ into observation space).

0D-Var

$$J = \frac{1}{2}\left(\frac{x - x_b}{\sigma_b}\right)^2 + \frac{1}{2}\left(\frac{x - y_o}{\sigma_o}\right)^2$$

3D-Var

$$J = \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + \frac{1}{2}\left(H(\mathbf{x}) - \mathbf{y}_o\right)^T \mathbf{R}^{-1}\left(H(\mathbf{x}) - \mathbf{y}_o\right)$$

3D-Var assimilation

$$\mathcal{J}(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + \frac{1}{2}(H(\mathbf{x})-\mathbf{y}_o)^T\mathbf{R}^{-1}(H(\mathbf{x})-\mathbf{y}_o)$$

$\mathbf{B}$ is the background error covariance matrix, $\mathbf{R}$ is the observation error covariance matrix, H is the observation operator (used for converting model state vector $\mathbf{x} = (T, q_v, u, v)$ into observation space).

The minimization of $\mathcal{J}$ can be performed if its gradient with respect to the atmospheric state $\mathbf{x}$ is known:

$$\nabla_{\mathbf{x}}\mathcal{J}(\mathbf{x}) = \mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + \mathbf{H}^T\mathbf{R}^{-1}(H(\mathbf{x})-\mathbf{y}_o)$$

where $\mathbf{H}^T$ is the transpose of the tangent linear operator derived from the non-linear observation operator H.

Important remarks on variational data assimilation

- Minimizing the cost function J is equivalent to finding the so-called *Best Linear Unbiased Estimator* (BLUE) if one can assume that:

  - **Model background and observation errors are unbiased and uncorrelated**,

  - **their statistical distributions are Gaussian**.
  (then, the final analysis is the maximum likelihood estimator of the true state).

- The analysis is obtained by adding corrections to the background which depend *linearly* on background-observations departures.

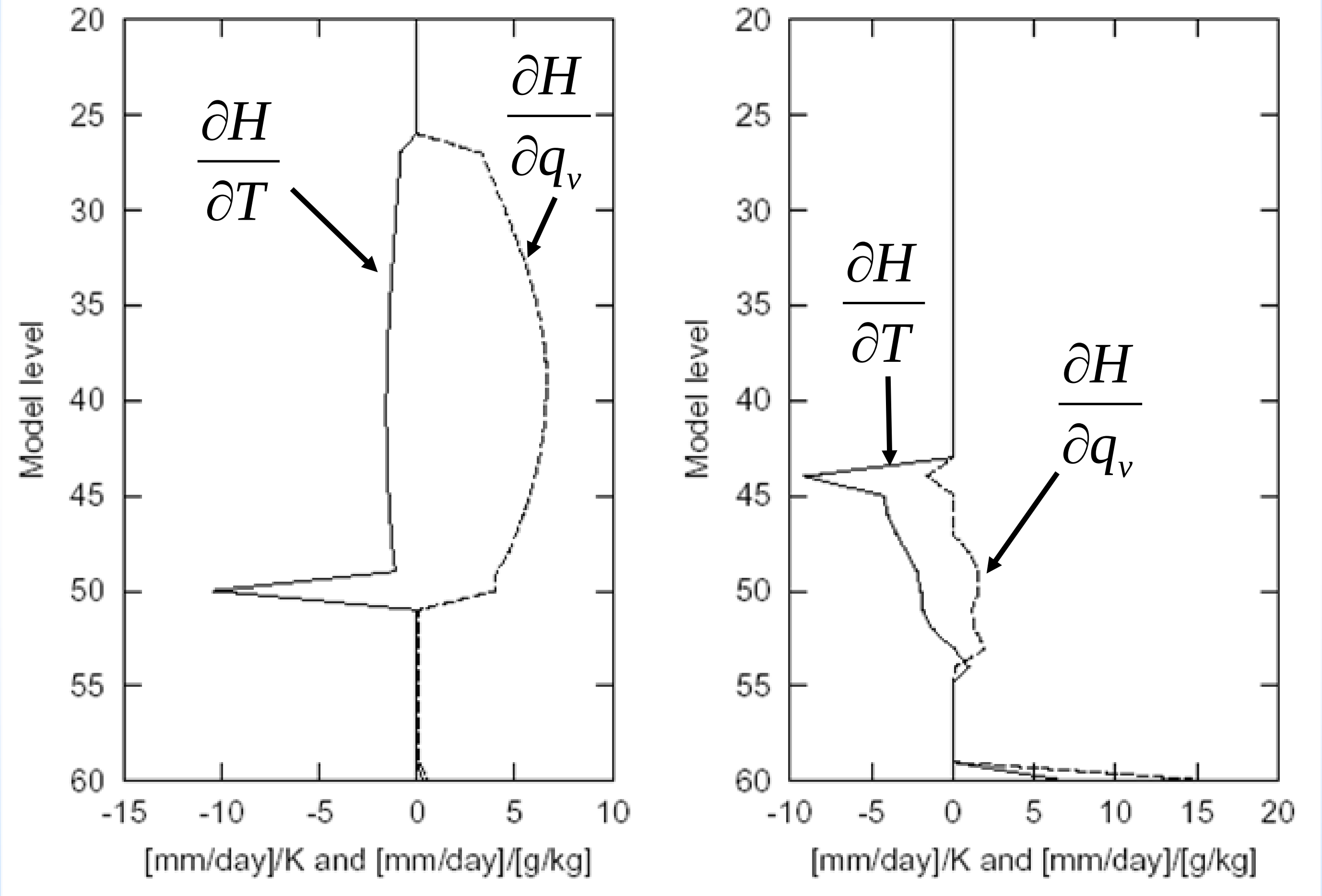
- In this linear context, **the observation operator** (to go from model space to observation space) **must not be too non-linear** in the vicinity of the model state, else the result of the analysis procedure is not optimal.

- The result of the minimization depends on the background and observation error statistics (matrices $\mathbf{B}$ and $\mathbf{R}$) but also on the Jacobian matrix ($\mathbf{H}$) of the observation operator (H).

An example of observation operator

H: input = model state (T, q_v) → output = surface convective rainfall rate

Jacobians of surface rainfall rate w.r.t. T and q_v



Marécal and Mahfouf (2002)

Betts-Miller (adjustment scheme)

Tiedtke (ECMWF's oper mass-flux scheme)

Adjoint technique

- **Non-linear observation operator:**

$$\mathbf{y} = H(\mathbf{x})$$

- **Tangent linear operator:**

$$\delta\mathbf{y} = \mathbf{H}(\delta\mathbf{x})$$

- **$\mathbf{H}$ is the Jacobian matrix derived from H:**

$$\mathbf{H}_{ij} = \frac{\partial y_i}{\partial x_j}$$

$$\delta y_i = \sum_{j=1}^{N} \frac{\partial y_i}{\partial x_j}\, \delta x_j$$

Adjoint technique

- **Observation term of the cost-function:**

$$\mathcal{J}_o = \frac{1}{2}(\mathbf{y} - \mathbf{y}_o)^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{y}_o)$$

- **Gradient with respect to y:**

$$\nabla_y \mathcal{J}_o = \mathbf{R}^{-1}(\mathbf{y} - \mathbf{y}_o)$$

- **Gradient with respect to x:**

$$\frac{\partial \mathcal{J}_o}{\partial x_i} = \sum_{j=1}^{M} \frac{\partial \mathcal{J}_o}{\partial y_j} \underbrace{\frac{\partial y_j}{\partial x_i}}_{\mathbf{H}^T_{ij}}$$

which involves the adjoint (transpose) of the tangent-linear operator.

- **Finally:**

$$\nabla_x \mathcal{J}_o = \mathbf{H}^T (\nabla_y \mathcal{J}_o) = \mathbf{H}^T \mathbf{R}^{-1} (H(\mathbf{x}) - \mathbf{y}_o)$$

Solution of 3D-Var assimilation

- **3D-Var solution in the linear case:**

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{y}_o - H(\mathbf{x}_b))$$

- **with the analysis error covariance matrix A such as:**

$$\mathbf{A}^{-1} = \mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$$

- **3D-Var solution in the non-linear case:**

$$\mathbf{x}^{n+1} = \mathbf{x}^n - \rho_n \, \nabla_x \mathcal{J}(\mathbf{x}^n)$$

which requires an iterative minimization algorithm (e.g. M1QN3, conjugate gradient)

0D-Var

$$J = \frac{1}{2}\left(\frac{x - x_b}{\sigma_b}\right)^2 + \frac{1}{2}\left(\frac{x - y_o}{\sigma_o}\right)^2$$

$$x_a = x_b + \boxed{\frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}}\left(y_o - x_b\right) \qquad \frac{1}{\sigma_a^2} = \frac{1}{\sigma_b^2} + \frac{1}{\sigma_o^2}$$

3D-Var

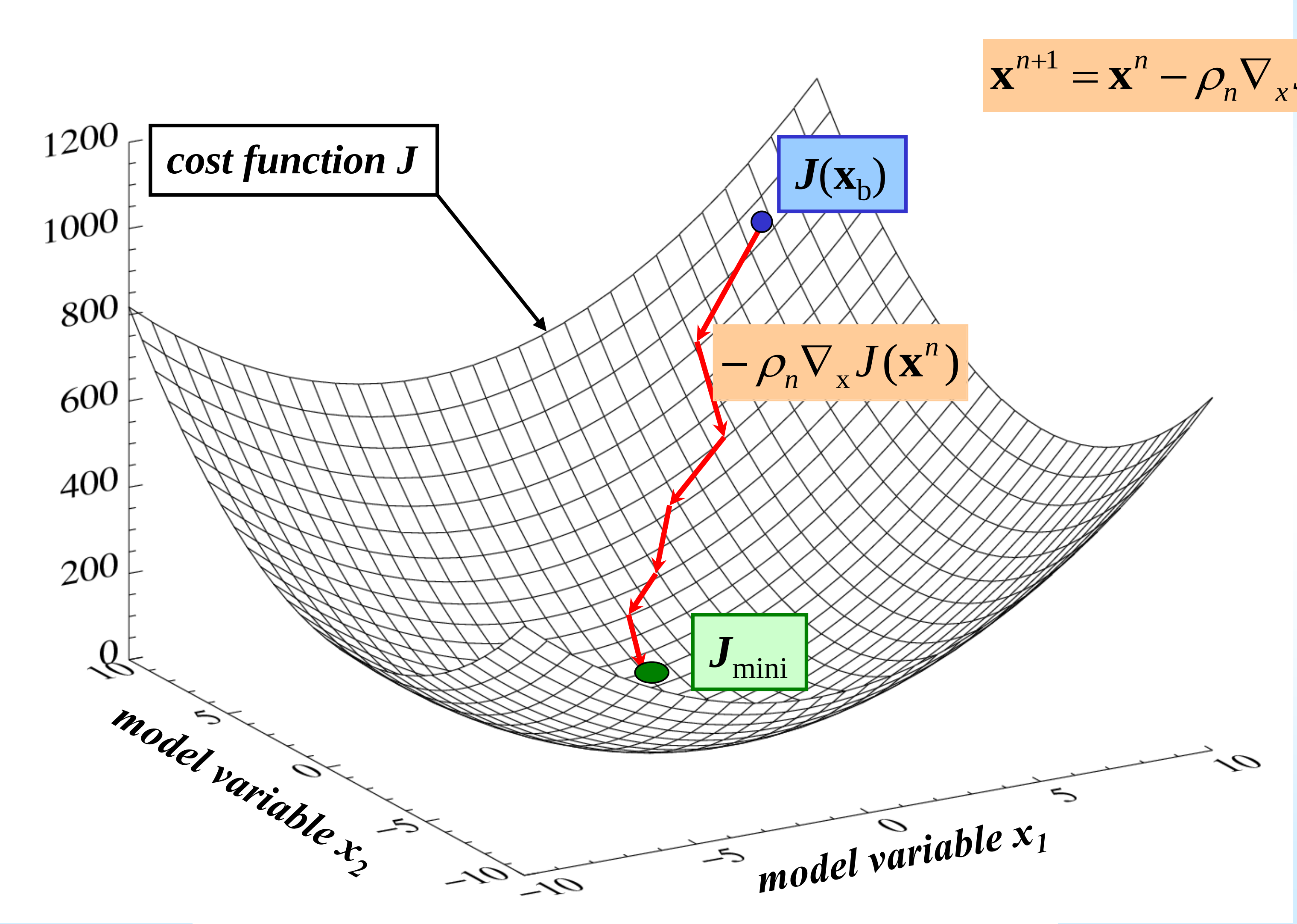
$$J = \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + \frac{1}{2}\left(H(\mathbf{x}) - \mathbf{y}_o\right)^T \mathbf{R}^{-1}\left(H(\mathbf{x}) - \mathbf{y}_o\right)$$

$$\mathbf{x}_a = \mathbf{x}_b + \boxed{\mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}}\left(\mathbf{y}_o - H(\mathbf{x}_b)\right)$$

$$\mathbf{A}^{-1} = \mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$$

The minimization of the cost function J is usually performed using an iterative minimization procedure

$$\mathbf{x}^{n+1} = \mathbf{x}^{n} - \rho_n \nabla_x J(\mathbf{x}^n)$$



Example with control vector $\mathbf{x} = (x_1, x_2)$

4D-Var assimilation

$$\mathcal{J}(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_b) + \frac{1}{2}\sum_{i=0}^{n}(H_i(\mathbf{x}_i) - \mathbf{y_o}_i)^T \mathbf{R}_i^{-1}(H_i(\mathbf{x}_i) - \mathbf{y_o}_i)$$

where $\mathbf{x}_i$ is the model state at time step t_i such as:

$$\mathbf{x}_i = M(t_0, t_i)[\mathbf{x}_0]$$

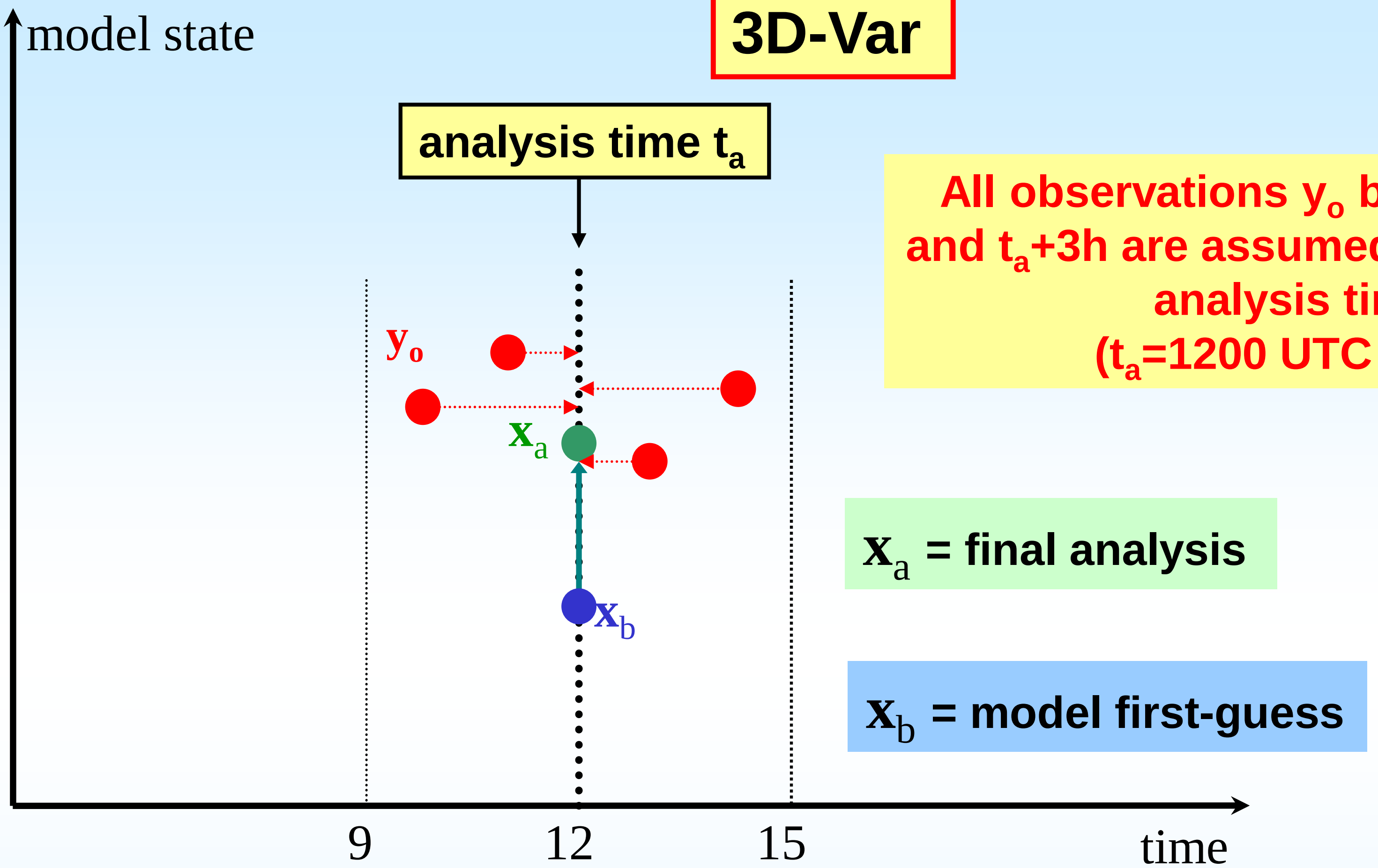
M is the non-linear forecast model integrated between time t_0 and time t_i.

The gradient of the cost function with respect to the initial state $\mathbf{x}_0$ writes:

$$\nabla_{\mathbf{x}_0}\mathcal{J} = \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_b) + \sum_{i=0}^{n}\mathbf{M}^T(t_i, t_0)\mathbf{H}_i^T \mathbf{R}_i^{-1}(H_i(\mathbf{x}_i) - \mathbf{y_o}_i)$$

where $\mathbf{M}^T$ is the adjoint of the forecast model integrated between time t_i and time t_0.

3D-Var

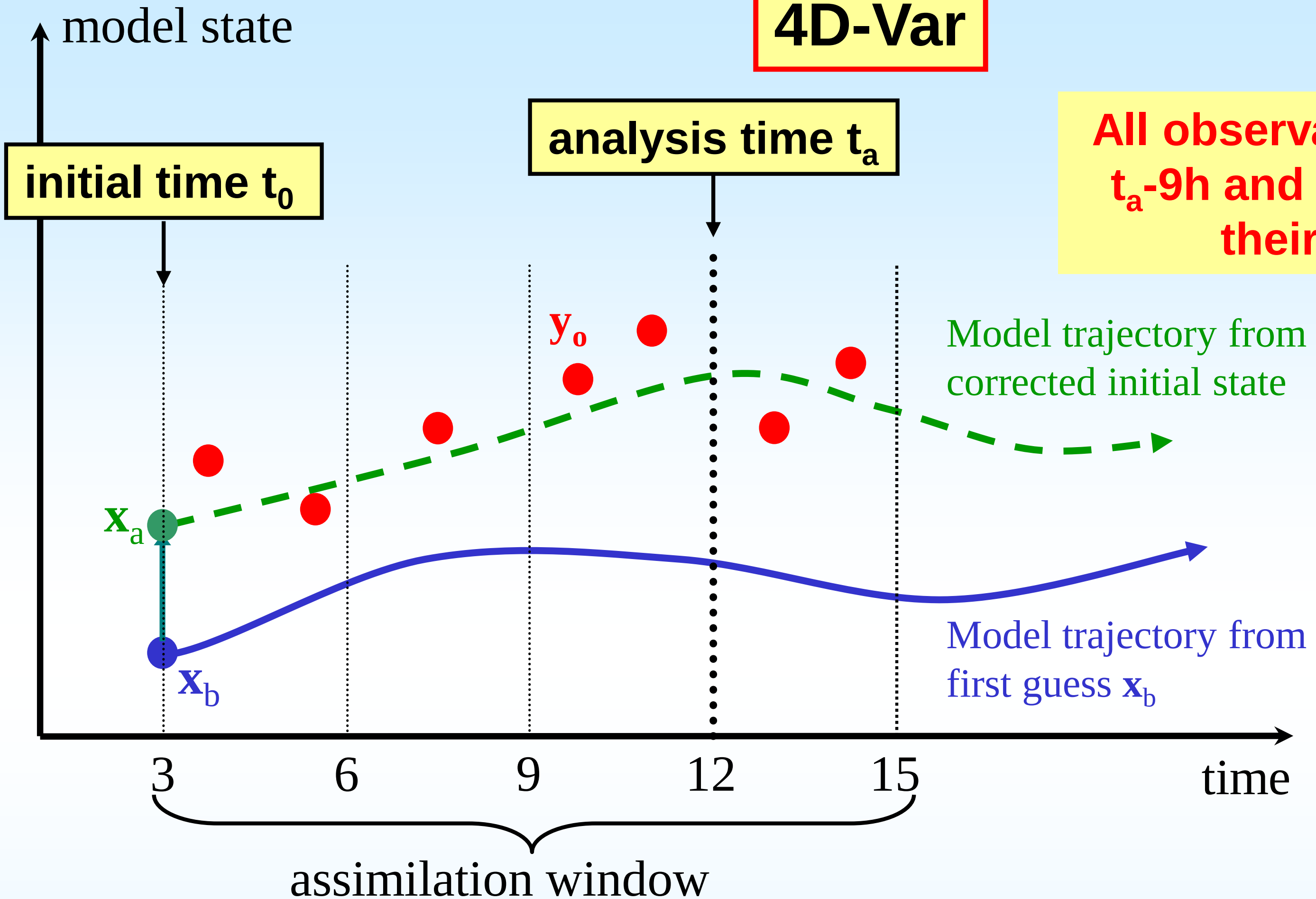


All observations y_o between t_a-3h and t_a+3h are assumed to be valid at analysis time (t_a=1200 UTC here)

$$\min J = \frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + \frac{1}{2}\left(H(\mathbf{x})-\mathbf{y}_o\right)^T \mathbf{R}^{-1}\left(H(\mathbf{x})-\mathbf{y}_o\right)$$

$$\Leftrightarrow \nabla_{\mathbf{x}} J = \mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + \mathbf{H}^T \mathbf{R}^{-1}\left(H(\mathbf{x})-\mathbf{y}_o\right) = 0$$

4D-Var



All observations y_o between t_a-9h and t_a+3h are valid at their actual time

Forecast model is involved in minimization

$$\min J = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_b)^T \ldots_{oi})^T \mathbf{R}_i^{-1}\left(H_i(M[\mathbf{x}_0]) - \mathbf{y}_{oi}\right)$$

Adjoint of forecast model with simplified linearized physics

$$\Leftrightarrow \nabla_{\mathbf{x}} J = \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_b) + \sum_{i=0}^{n} \mathbf{M}^T[t_i, t_0]\mathbf{H}_i^T \mathbf{R}_i^{-1}\left(H_i(M[\mathbf{x}_0]) - \mathbf{y}_{oi}\right) = 0$$

Incremental 4D-Var

- **Model initial state:** $\mathbf{x}_0 = \mathbf{x}_{0b} + \delta\mathbf{x}_0$

- **Observation operator at time t_i:** $H_i(\mathbf{x}_i) = H_i(\mathbf{x}_{ib}) + \mathbf{H}_i\delta\mathbf{x}_i$

  where $\delta\mathbf{x}_i = \mathbf{M}_s(\mathbf{x}_0, \mathbf{x}_i)[\delta\mathbf{x}_0]$

- **The cost function is minimized in terms of increments:**

$$\mathcal{J}(\mathbf{x}_0) = \frac{1}{2}(\delta\mathbf{x}_0)^T\mathbf{B}^{-1}\delta\mathbf{x}_0 + \frac{1}{2}\sum_{i=0}^{n}(\mathbf{H}_i(\delta\mathbf{x}_i) - \mathbf{d}_i)^T\mathbf{R}_i^{-1}(\mathbf{H}_i(\delta\mathbf{x}_i) - \mathbf{d}_i)$$

  where $\mathbf{d}_i = \mathbf{y_o}_i - H_i(\mathbf{x}_{ib})$ is the innovation vector.

- **The gradient of the cost function then writes:**

$$\nabla_{\delta\mathbf{x}_0}\mathcal{J} = \mathbf{B}^{-1}\delta\mathbf{x}_0 + \sum_{i=0}^{n}\mathbf{M}_s^T(t_i, t_0)\mathbf{H}_i^T\mathbf{R}_i^{-1}(\mathbf{H}_i(\delta\mathbf{x}_i) - \mathbf{d}_i)$$

- **The analysis is obtained by adding the optimal $\delta\mathbf{x}_a$ to the model background: $\mathbf{x}_a = \mathbf{x}_b + \delta\mathbf{x}_a$**

- **To account for non-linearities, the trajectory around which the model is linearized can be updated several times (using $\mathbf{x}_a$ as a new $\mathbf{x}_b$).**

- **In operational practice:**

  – $\mathbf{d}_i$ are computed with the non-linear model M at high resolution (T1279 L137) with full physics.

  – $\delta\mathbf{x}_i$ are computed with the tangent linear model $\mathbf{M}_s$ at low resolution (T255 L137) with simplified physics.

  – $\nabla\mathcal{J}$ is computed with the adjoint model $\mathbf{M}_s^T$ at low resolution (T255 L137) with simplified physics.

  – The trajectory at high resolution is updated twice and around 30 iterations are needed in each minimization.

TL AND AD MODELS

- **TANGENT LINEAR MODEL**

  If M is a model such as:

$$\mathbf{x}(t_{i+1}) = M[\mathbf{x}(t_i)]$$

  then the tangent linear model of M, called M', is:

$$\delta\mathbf{x}(t_{i+1}) = M'[\mathbf{x}(t_i)]\delta\mathbf{x}(t_i) = \frac{\partial M[\mathbf{x}(t_i)]}{\partial \mathbf{x}}\delta\mathbf{x}(t_i)$$

- **ADJOINT MODEL**

  The adjoint of a linear operator M' is the linear operator $\mathbf{M}^*$ such that, for the inner product $<,>$,

$$\forall\mathbf{x}, \forall\mathbf{y} \qquad < M'\mathbf{x}, \mathbf{y} > = < \mathbf{x}, \mathbf{M}^*\mathbf{y} >$$

<u>Remarks:</u>

– with the euclidian inner product, $\mathbf{M}^* = {M'}^T$.

– in variational assimilation, $\nabla_x \mathcal{J} = \mathbf{M}^* \nabla_y \mathcal{J}$, where $\mathcal{J}$ is the cost function.

EXAMPLE OF ADJOINT CODING

- **non-linear statement**

$$x = y + z^2$$

$$\begin{aligned} z &= z \\ y &= y \\ x &= y + z^2 \end{aligned}$$

- **tangent linear statement**

$$\begin{aligned} \delta z &= \delta z \\ \delta y &= \delta y \\ \delta x &= \delta y + 2z\delta z \end{aligned}$$

or in a matrix form:

$$\begin{pmatrix} \delta z \\ \delta y \\ \delta x \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2z & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} \delta z \\ \delta y \\ \delta x \end{pmatrix}$$

EXAMPLE OF ADJOINT CODING

- **adjoint statement**
  - transpose matrix

$$\begin{pmatrix} \delta z^* \\ \delta y^* \\ \delta x^* \end{pmatrix} = \begin{pmatrix} 1 & 0 & 2z \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} \delta z^* \\ \delta y^* \\ \delta x^* \end{pmatrix}$$

or in the form of equation set:

$$\begin{aligned} \delta z^* &= \delta z^* + 2z\delta x^* \\ \delta y^* &= \delta y^* + \delta x^* \\ \delta x^* &= 0 \end{aligned}$$

As an alternative to the matrix method, adjoint coding can be carried out using a **line-by-line** approach (what we do at ECMWF).

Automatic adjoint code generators do exist, but the output code is not optimized and not bug-free.

Basic rules for line-by-line adjoint coding (1)

Adjoint statements are derived from tangent linear ones in a reversed order

Tangent linear code	Adjoint code
$\delta x = 0$	$\delta x^* = 0$
$\delta x = A\ \delta y + B\ \delta z$	$\delta y^* = \delta y^* + A\ \delta x^*$ $\delta z^* = \delta z^* + B\ \delta x^*$ $\delta x^* = 0$
$\delta x = A\ \delta x + B\ \delta z$	$\delta z^* = \delta z^* + B\ \delta x^*$ $\delta x^* = A\ \delta x^*$

do k = 1, N
  $\delta x(k) = A\ \delta x(k-1) + B\ \delta y(k)$
end do
	do k = N, 1, – 1 (Reverse the loop!)
	  $\delta x^*(k-1) = \delta x^*(k-1) + A\ \delta x^*(k)$
	  $\delta y^*(k) = \delta y^*(k) + B\ \delta x^*(k)$
	  $\delta x^*(k) = 0$
	end do

if (*condition*) *tangent linear code*	if (*condition*) *adjoint code*

And do not forget to initialize local adjoint variables to zero !

Basic rules for line-by-line adjoint coding (2)

To save memory, the trajectory can be recomputed just before the adjoint calculations.

<u>The most common sources of error in adjoint coding are:</u>

1) Pure coding errors (often: confusion trajectory/perturbation variables),
2) Forgotten initialization of local adjoint variables to zero,
3) Mismatching trajectories in tangent linear and adjoint (even slightly),
4) Bad identification of trajectory updates:

Tangent linear code	Trajectory and adjoint code

if ($x > x0$) then
  $\delta x = A\, \delta x / x$
  $x = A\, \mathrm{Log}(x)$
end if
	------------- Trajectory ----------------
	$x_{store} = x$ (storage for use in adjoint)
	if ($x > x0$) then
	  $x = A\, \mathrm{Log}(x)$
	end if
	-------------- Adjoint ------------------
	if ($x_{store} > x0$) then
	  $\delta x^* = A\, \delta x^* / x_{store}$
	end if

TEST FOR TANGENT LINEAR MODEL

- **Taylor formula:**

$$\lim_{\lambda \to 0} \frac{M(\mathbf{x} + \lambda\delta\mathbf{x}) - M(\mathbf{x})}{M'(\lambda\delta\mathbf{x})} = 1$$

Perturbation scaling factor

λ	RATIO
0.1E-09	0.9994875881543574E+00
0.1E-08	0.9999477148855701E+00
0.1E-07	0.9999949234236705E+00
0.1E-06	0.9999993501022509E+00
0.1E-05	0.9999999496119013E+00
0.1E-04	0.9999996111338369E+00
0.1E-03	0.9999953179193711E+00
0.1E-02	0.9999724488345042E+00
0.1E-01	0.9998727842790062E+00
0.1E+00	0.9978007454264978E+00
0.1E+01	0.9683066504549524E+00

machine precision reached

TEST FOR ADJOINT MODEL

- **adjoint identity:**

$$\forall \mathbf{x}, \forall \mathbf{y} \quad < M'.\mathbf{x}, \mathbf{y} > = < \mathbf{x}, \mathbf{M}^*.\mathbf{y} >$$

$< F(X) , Y > =$ -.13765102625251640000E-01

$< X , F^*(Y) > =$ -.13765102625251680000E-01

ratio of norms = 1.00000000000000005

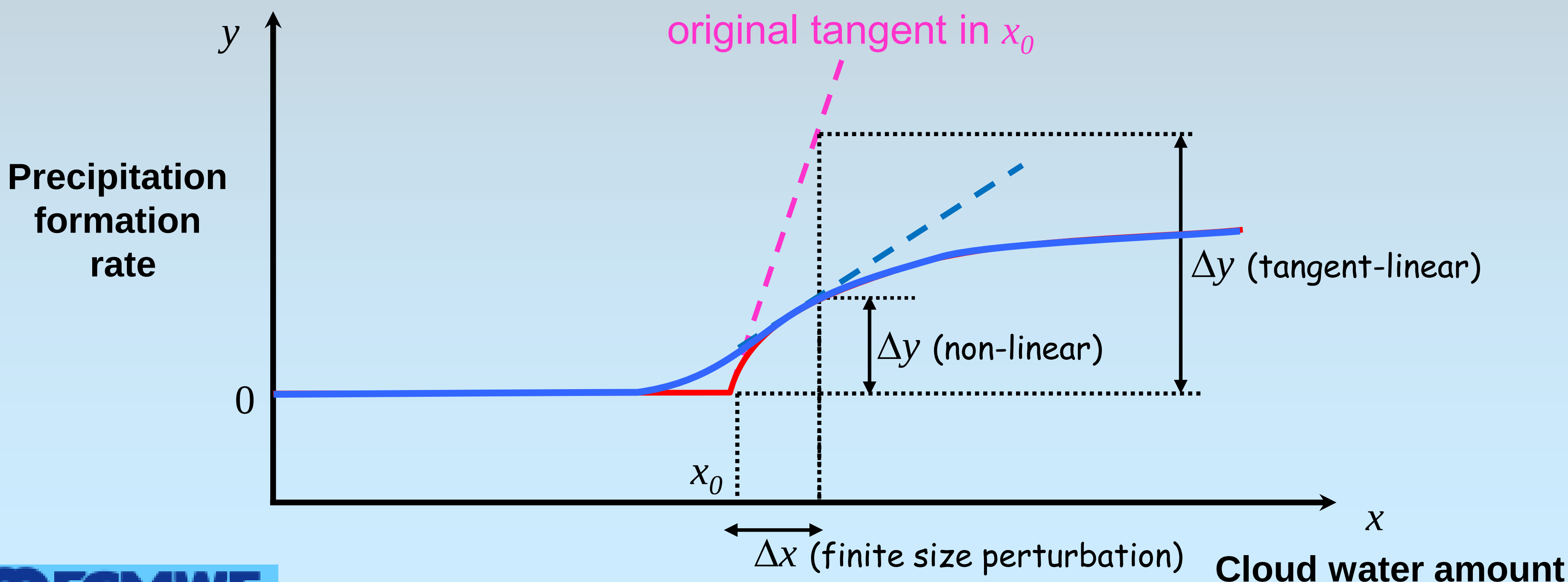
THE DIFFERENCE IS 11.351 TIMES THE ZERO OF THE MACHINE

Linearity assumption

- Variational assimilation is based on the strong assumption that the analysis is performed in a **(quasi-)linear** framework.

- However, in the case of physical processes, strong non-linearities can occur in the presence of discontinuous/non-differentiable processes (e.g. **switches or thresholds** in cloud water and precipitation formation).

→ "Regularization" needs to be applied: **smoothing of functions, reduction of some perturbations.**



Linearity issue

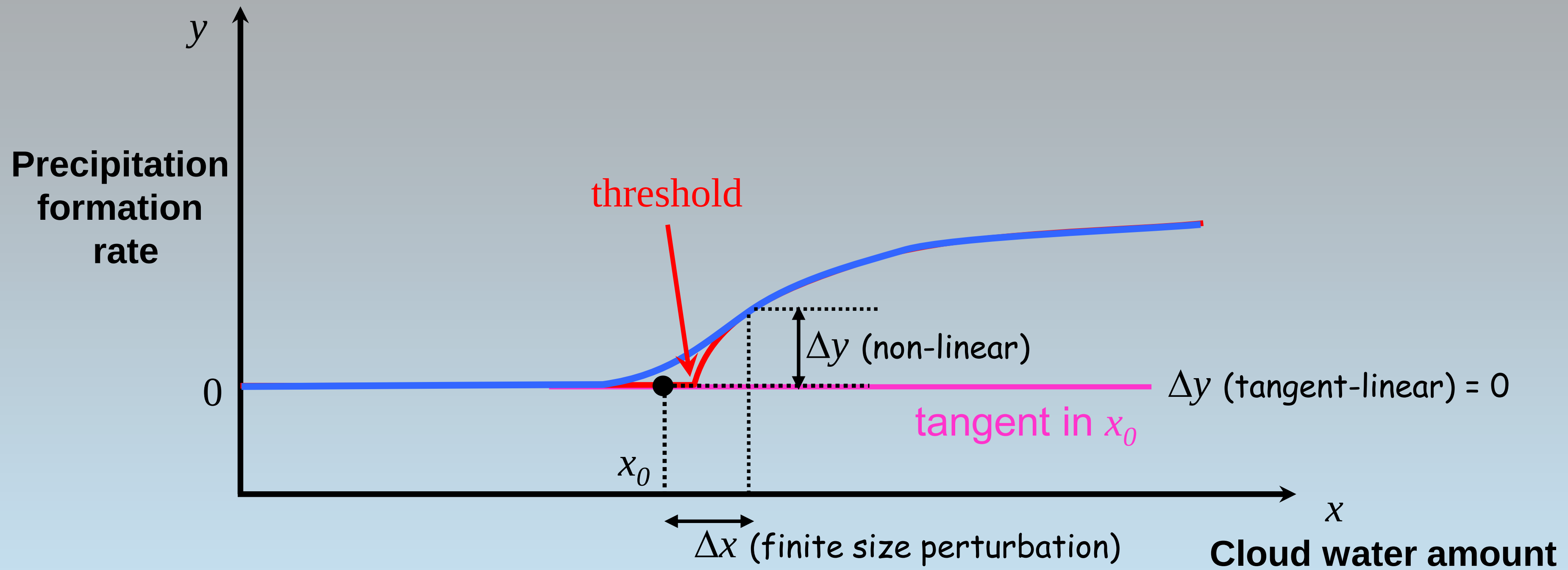


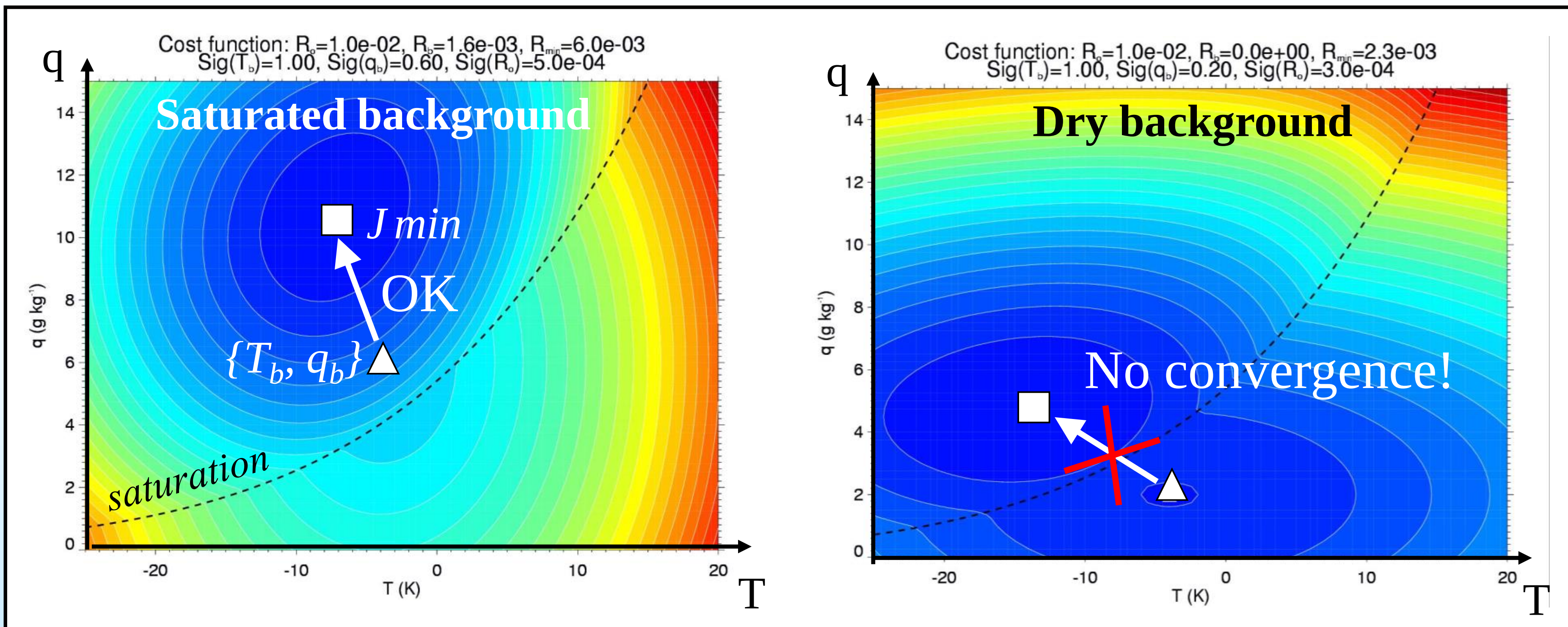
Illustration of discontinuity effect on cost function shape:

Model background = $\{T_b, q_b\}$; Observation = RR_{obs}

Simple parametrization of rain rate:

$$RR = \begin{cases} \alpha\{q - q_{sat}(T)\} & \text{if } q > q_{sat}(T), \\ 0 & \text{otherwise} \end{cases}$$

$$J = \frac{1}{2}\left(\frac{T - T_b}{\sigma_T}\right)^2 + \frac{1}{2}\left(\frac{q - q_b}{\sigma_q}\right)^2 + \frac{1}{2}\left(\frac{\alpha[q - q_{sat}(T)] - RR_{obs}}{\sigma_{RR_{obs}}}\right)^2$$



Single minimum of cost function

Several local minima of cost function

REGULARIZATION OF VERTICAL DIFFUSION SCHEME

- **perturbation of the exchange coefficients is neglected,** $K' = 0$ ***(Mahfouf, 1999)***

  (*exchange coefficient* K *is given by* $K = l^2 \left\| \frac{\partial \mathbf{V}}{\partial z} \right\| f(Ri)$)

- **original computation of the Richardson number** Ri

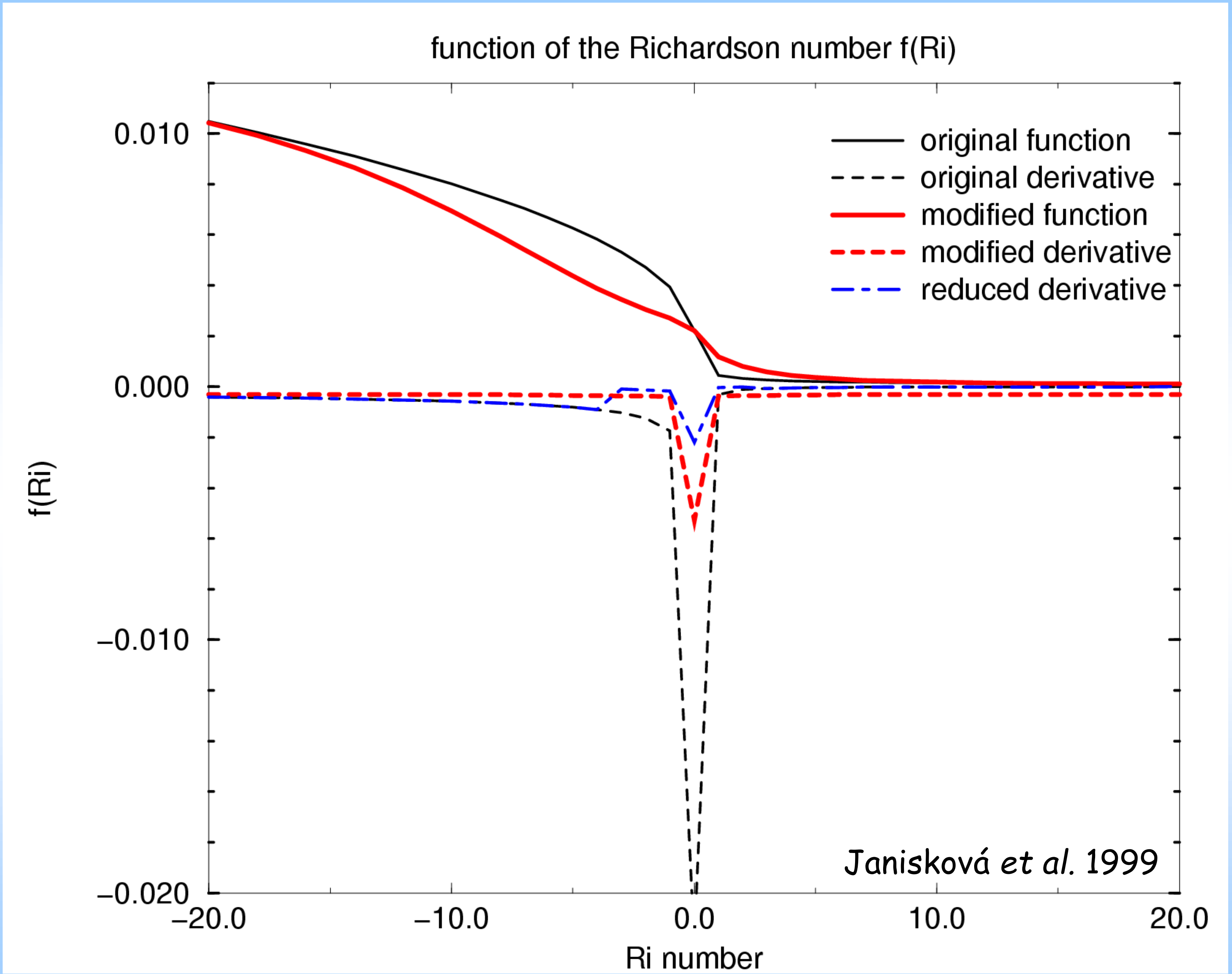
  $Ri = \frac{g}{c_p T} \frac{\frac{\partial s}{\partial z}}{\left\| \frac{\partial \vec{v}}{\partial z} \right\|^2}$ **modified as**

  $$Ri' = \frac{g}{c_p T} \frac{\frac{\partial s}{\partial z}}{\| \frac{\partial \vec{v}}{\partial z} \|^2 + c}$$ with a time constant $c = \frac{a}{(\Delta t_{phys})^2}$

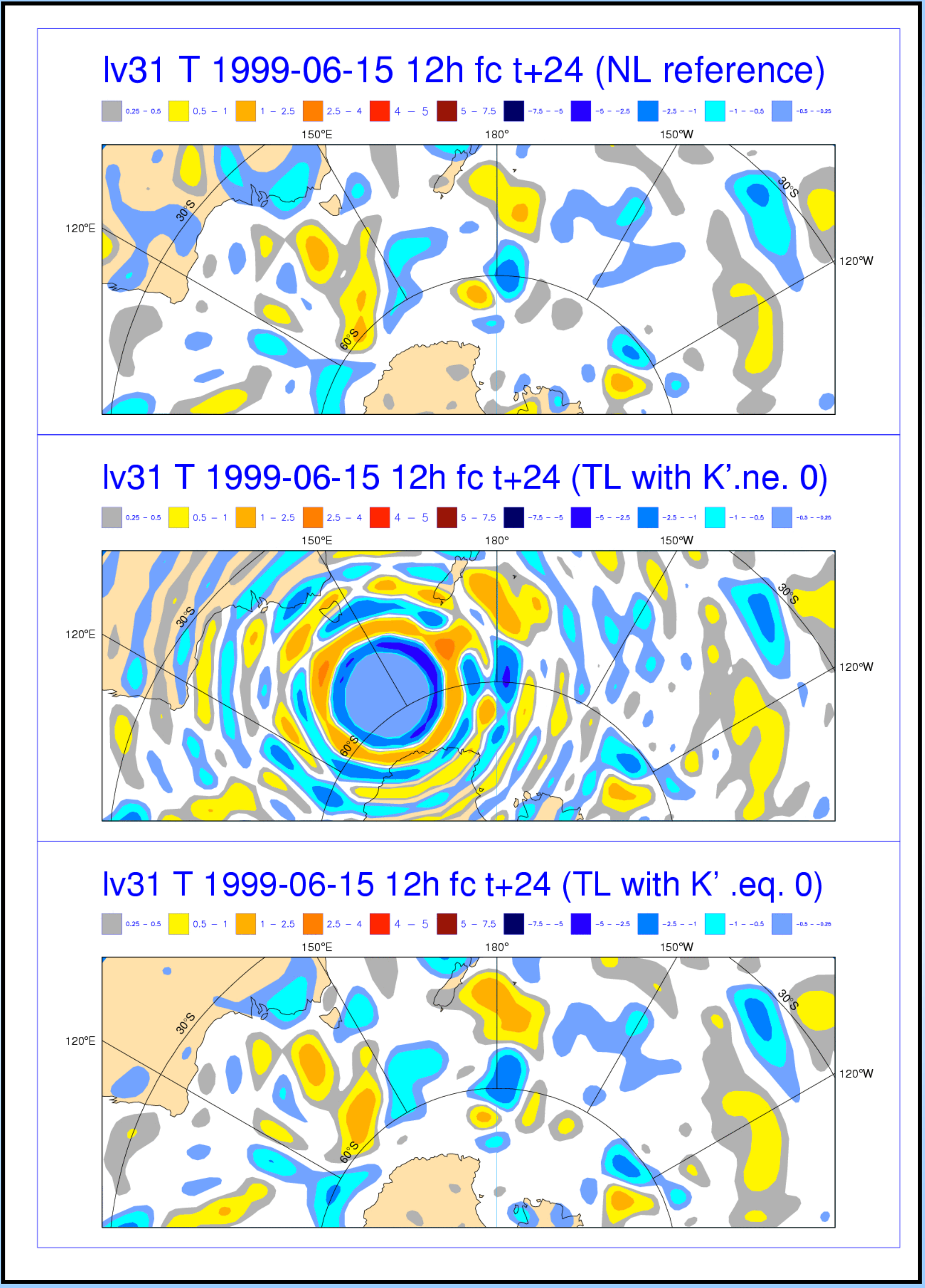
where Δt_{phys} is the physical time step

a is a tuning parameter of the regularization step

- **reducing a derivative** $f(Ri)$ **by a factor 10 in the central part (around the point of singularity) -** ***(Janisková et al., 1999)***


function of the Richardson number f(Ri)
0.010
0.000
−0.010
−0.020
f(Ri)
−20.0
−10.0
0.0
10.0
20.0
Ri number
original function
original derivative
modified function
modified derivative
reduced derivative
Janisková et al. 1999

Importance of regularization to prevent instabilities in tangent-linear model

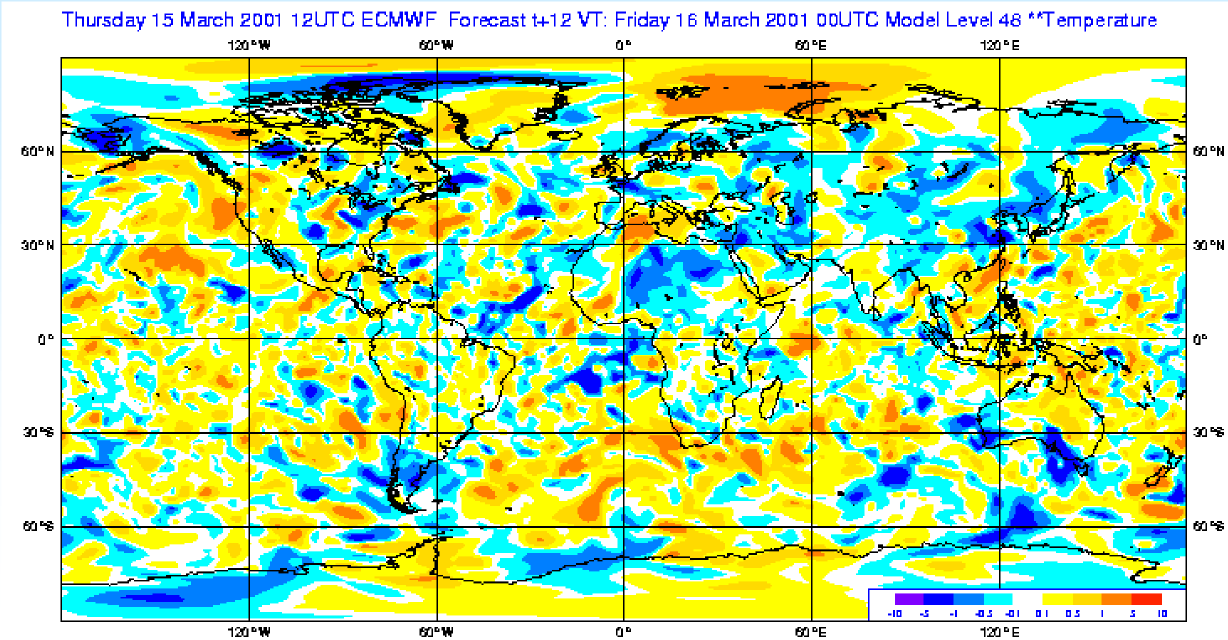


Evolution of temperature increments (24-hour forecast) with the tangent linear model using different approaches for the exchange coefficient K in the vertical diffusion scheme.

Perturbations of K included in TL

Perturbations of K set to zero in TL

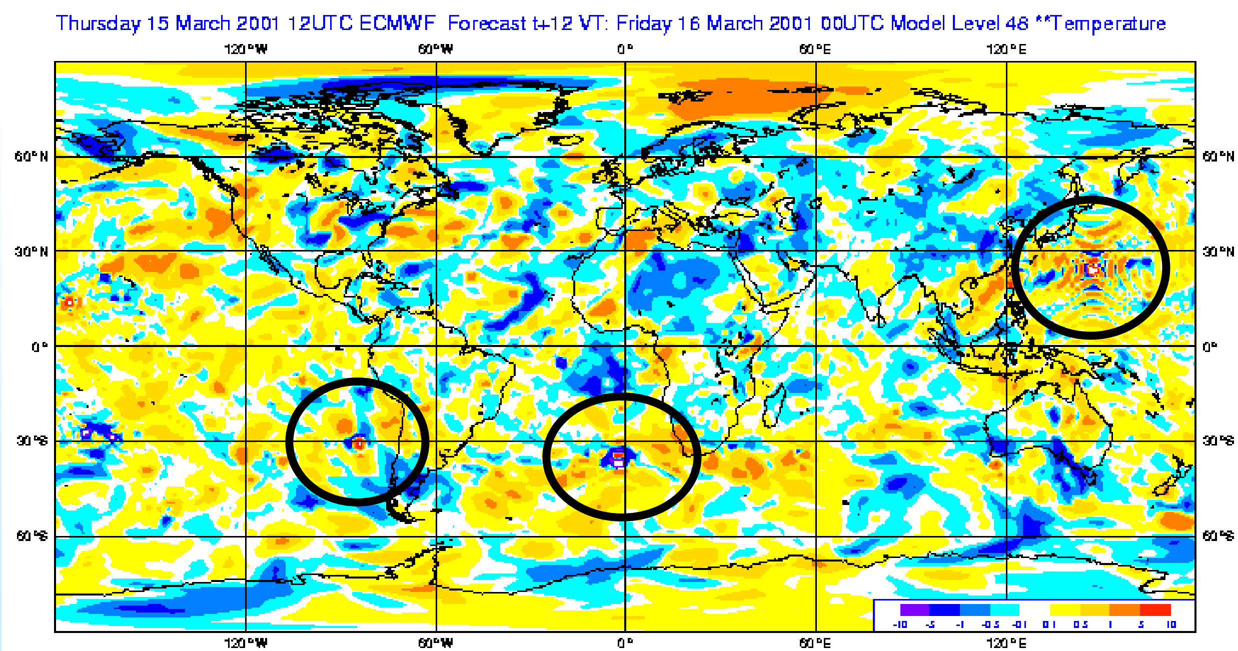
Importance of regularization to prevent instabilities in tangent-linear model



12-hour ECMWF model integration (T159 L60)

Temperature on level 48 (approx. 850 hPa)

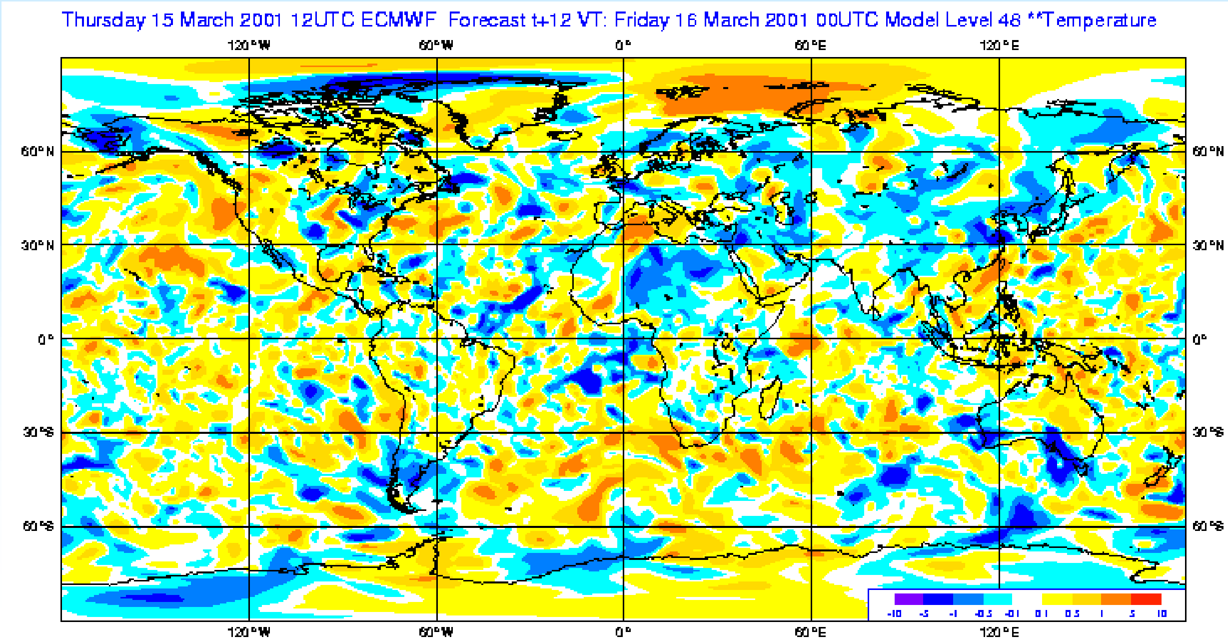
Finite difference between two non-linear model integrations



Corresponding perturbations evolved with tangent-linear model

No regularization in convection scheme

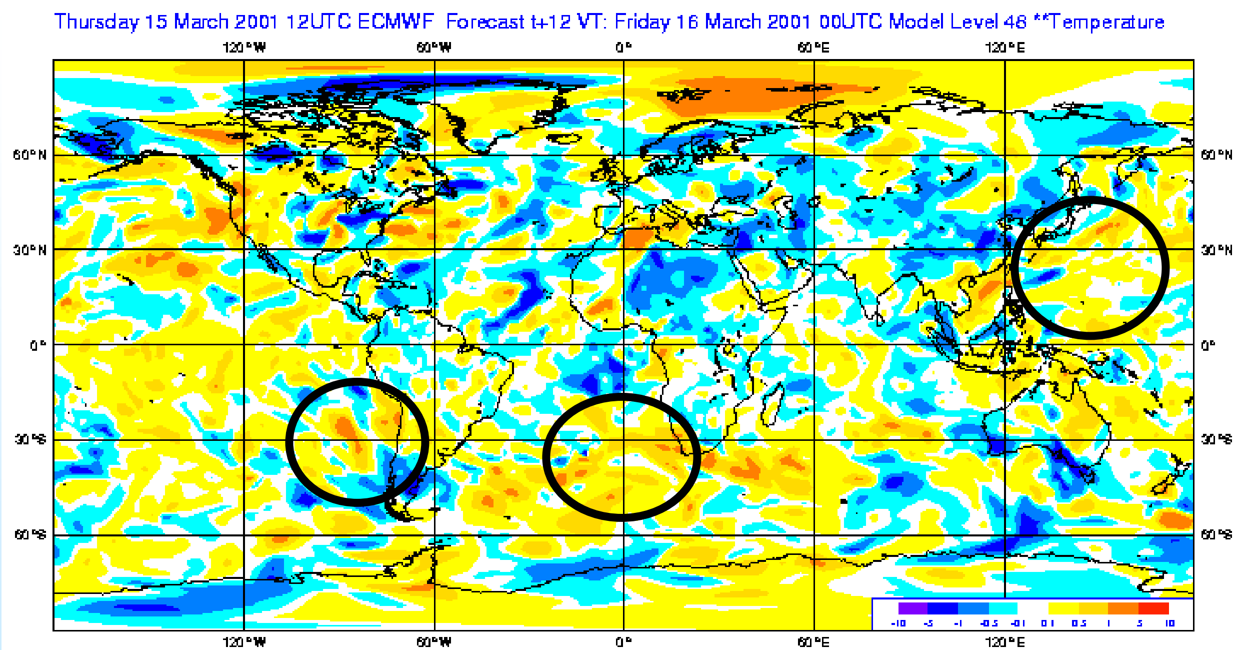
Importance of regularization to prevent instabilities in tangent-linear model



12-hour ECMWF model integration (T159 L60)

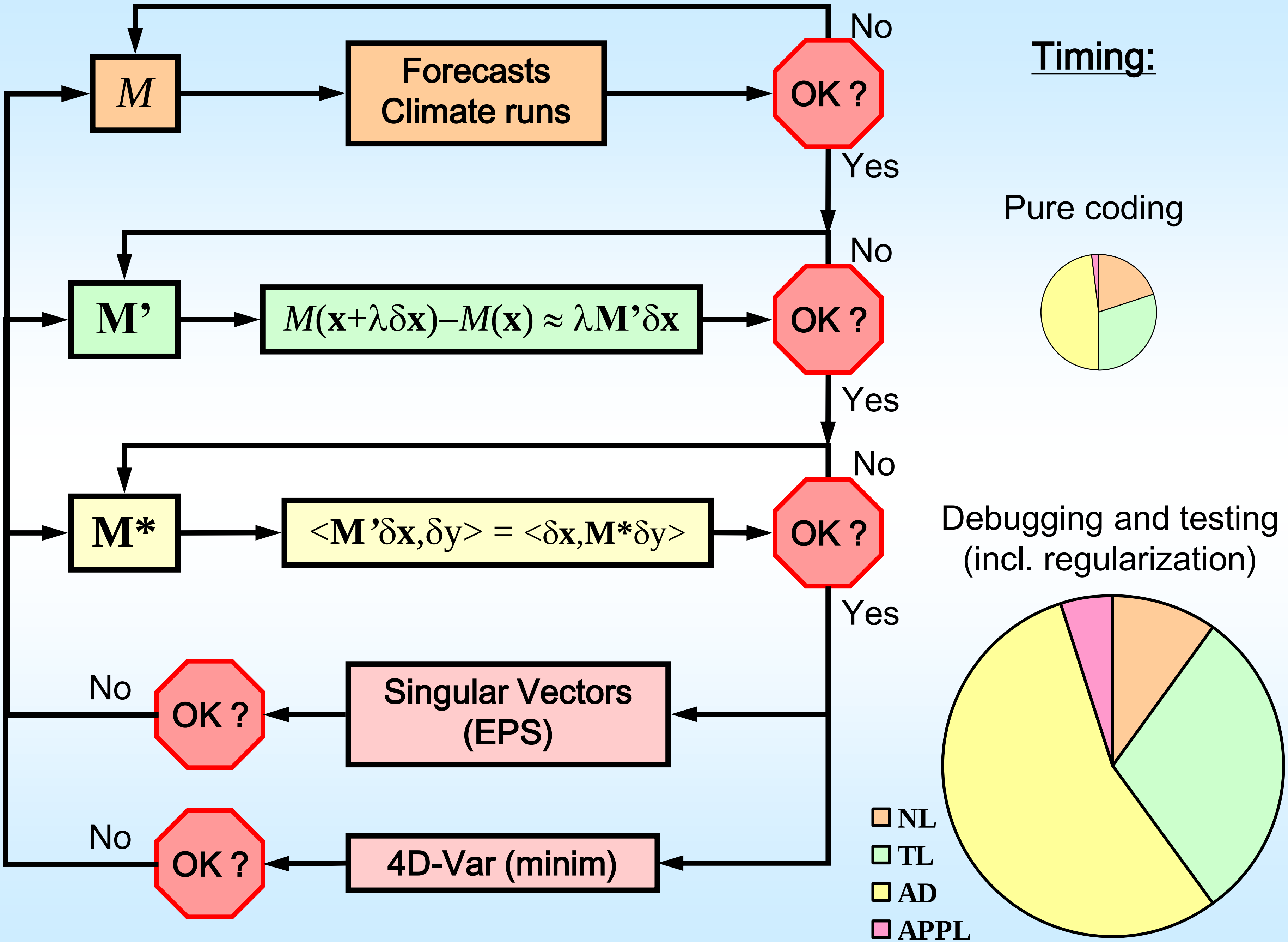
Temperature on level 48 (approx. 850 hPa)

Finite difference between two non-linear model integrations



Corresponding perturbations evolved with tangent-linear model

Regularization in convection scheme (buoyancy & updraught velocity reduced perturb.)


M
Forecasts
Climate runs
OK ?
No
Yes
M'
M(x+λδx)−M(x) ≈ λM'δx
OK ?
No
Yes
M*
<M'δx,δy> = <δx,M*δy>
OK ?
No
Yes
Singular Vectors
(EPS)
OK ?
No
4D-Var (minim)
OK ?
No
Timing:
Pure coding
Debugging and testing
(incl. regularization)
NL
TL
AD
APPL

ECMWF operational LP package (operational 4D-Var)

Currently used in ECMWF operational 4D-Var minimizations (main simplifications with respect to the full non-linear versions are highlighted in red):

- **Large-scale condensation scheme:** *[Tompkins and Janisková 2004]*
  - based on a uniform PDF to describe subgrid-scale fluctuations of total water,
  - melting of snow included,
  - precipitation evaporation included,
  - **reduction of cloud fraction perturbation and in autoconversion of cloud into rain.**

- **Convection scheme:** *[Lopez and Moreau 2005]*
  - mass-flux approach *[Tiedtke 1989]*,
  - deep convection (CAPE closure) and shallow convection (q-convergence) are treated,
  - perturbations of all convective quantities are included,
  - coupling with cloud scheme through detrainment of liquid water from updraught,
  - **some perturbations (buoyancy, initial updraught vertical velocity) are reduced.**

- **Radiation:** TL and AD of longwave and shortwave radiation available *[Janisková et al. 2002]*
  - **shortwave:** based on *Morcrette (1991)*, **only 2 spectral intervals** (instead of 6 in non-linear version),
  - **longwave:** based on *Morcrette (1989)*, **called every 2 hours only.**

ECMWF operational LP package (operational 4D-Var)

- **Vertical diffusion:**
  - mixing in the surface and planetary boundary layers,
  - based on K-theory and Blackadar mixing length,
  - exchange coefficients based on *Louis et al. [1982]*, near surface,
  - Monin-Obukhov higher up,
  - mixed layer parametrization and PBL top entrainment recently added.
  - **Perturbations of exchange coefficients are smoothed (esp. near the surface).**

- **Orographic gravity wave drag:** *[Mahfouf 1999]*
  - subgrid-scale orographic effects *[Lott and Miller 1997]*,
  - **only low-level blocking part is used.**

- **Non-orographic gravity wave drag:** *[Oor et al. 2010]*
  - isotropic spectrum of non-orographic gravity waves *[Scinocca 2003]*,
  - **Perturbations of output wind tendencies below 200 hPa reset to zero.**

- RTTOV is employed to simulate radiances at individual frequencies (infrared, longwave and microwave, with cloud and precipitation effects included) to compute model–satellite departures in observation space.

Impact of linearized physics on tangent-linear approximation

Comparison:

Finite difference of two NL integrations $\leftrightarrow$ TL evolution of initial perturbations

→ Examination of the accuracy of the linearization for typical analysis increments:

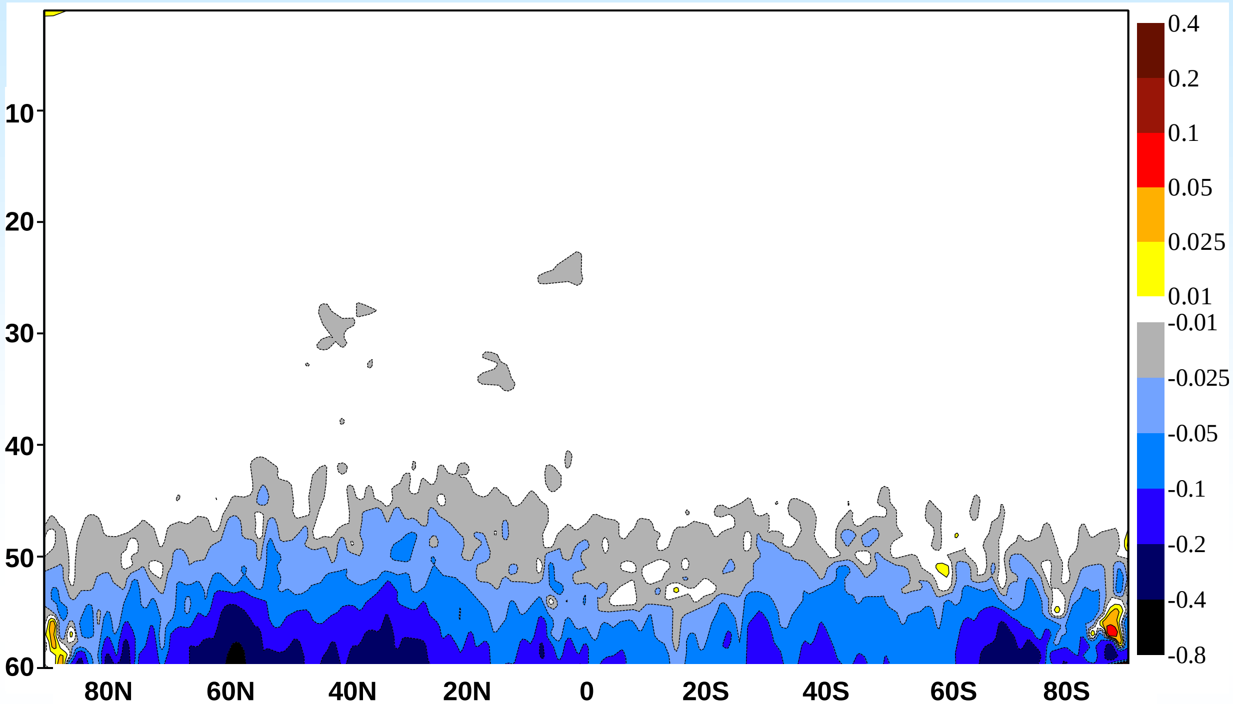
$$M(\mathbf{x}_{an}) - M(\mathbf{x}_{bg}) \;\leftrightarrow\; \mathbf{M}'\underbrace{(\mathbf{x}_{an} - \mathbf{x}_{bg})}_{\text{typical size of 4D-Var analysis increments}}$$

Diagnostics:

- mean absolute errors: $\varepsilon = \overline{\left| \left[M(\mathbf{x}_{an}) - M(\mathbf{x}_{bg}) \right] - \mathbf{M}'(\mathbf{x}_{an} - \mathbf{x}_{bg}) \right|}$
- relative error change: $\dfrac{\varepsilon_{\text{EXP}} - \varepsilon_{\text{REF}}}{\varepsilon_{\text{REF}}} \times 100\%$ (improvement if < 0)
- here: REF = adiabatic run (i.e. no physical parametrizations in tangent-linear)

Temperature

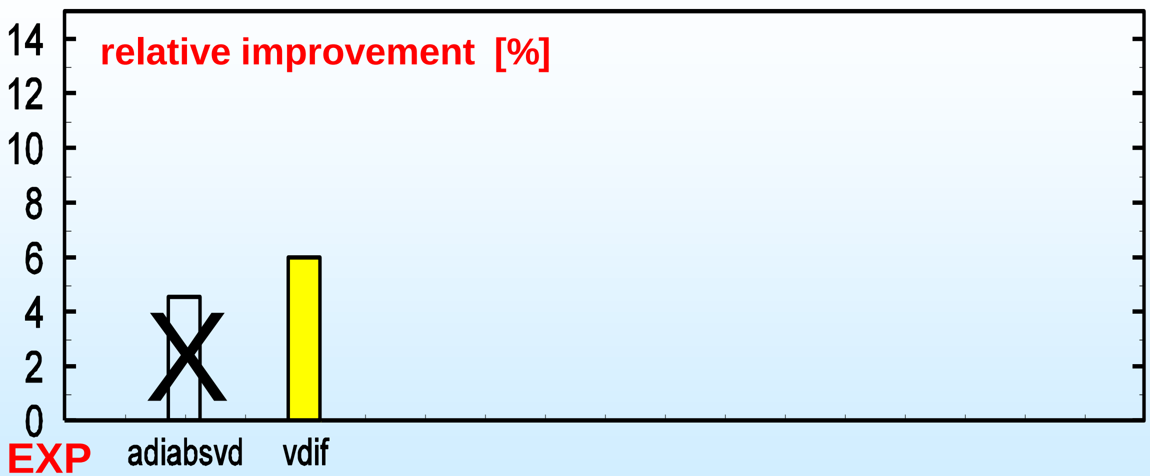
Impact of operational vertical diffusion scheme



$\varepsilon_{EXP} - \varepsilon_{REF}$

REF = ADIAB

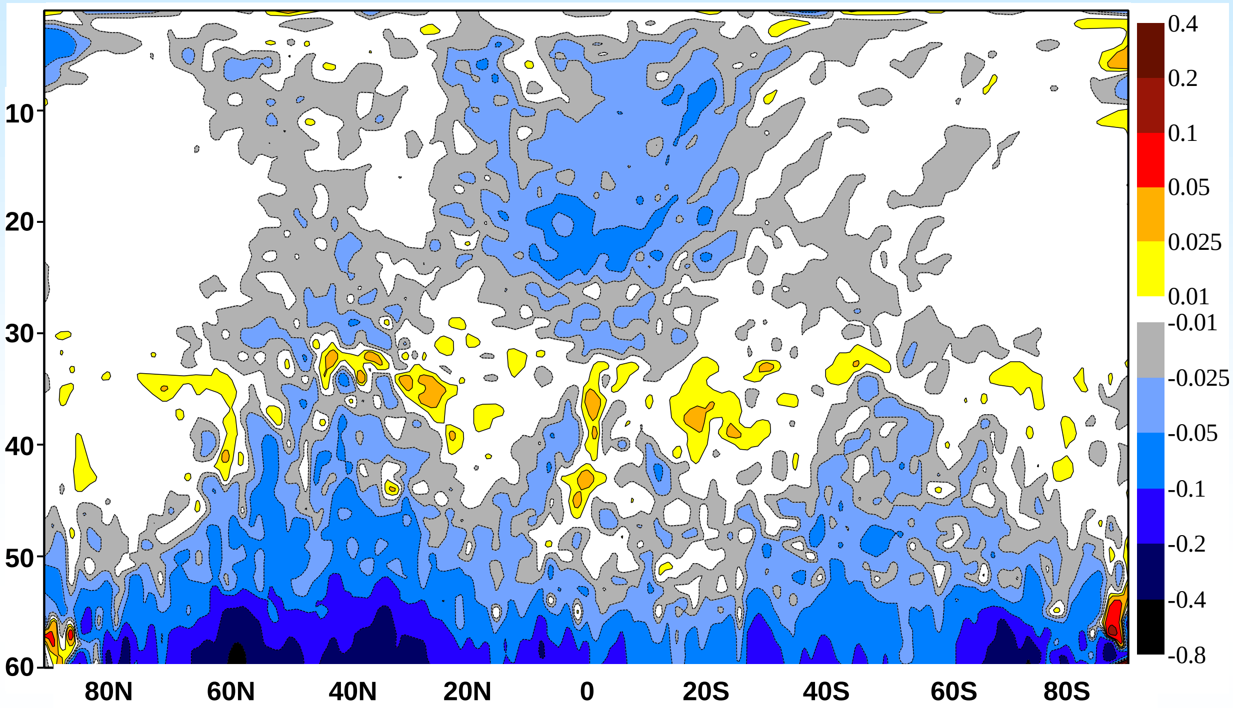
12-hour T159 L60 integration



Adiab simp vdif | vdif

Temperature

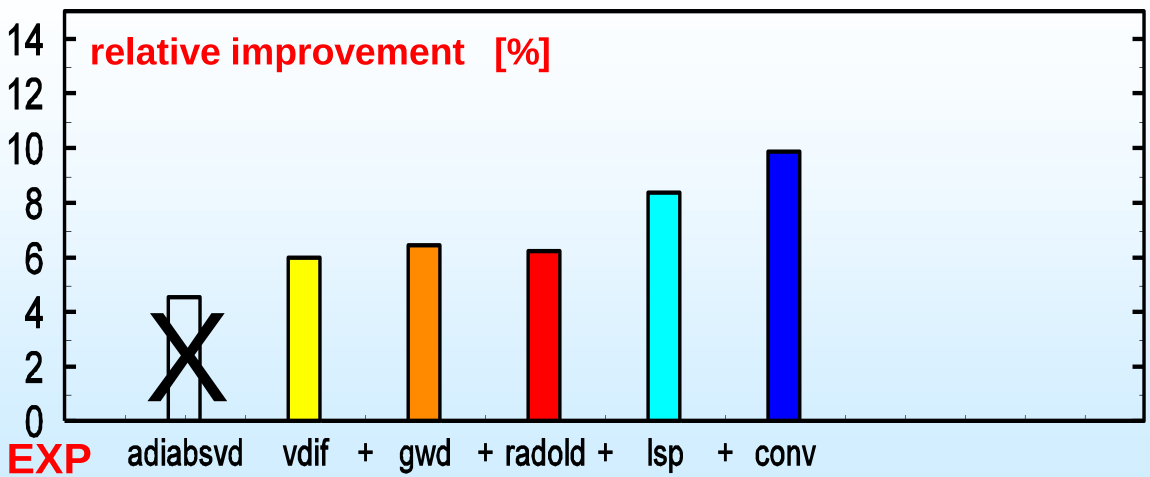
Impact of dry + moist physical processes



$\varepsilon_{EXP} - \varepsilon_{REF}$

REF = ADIAB

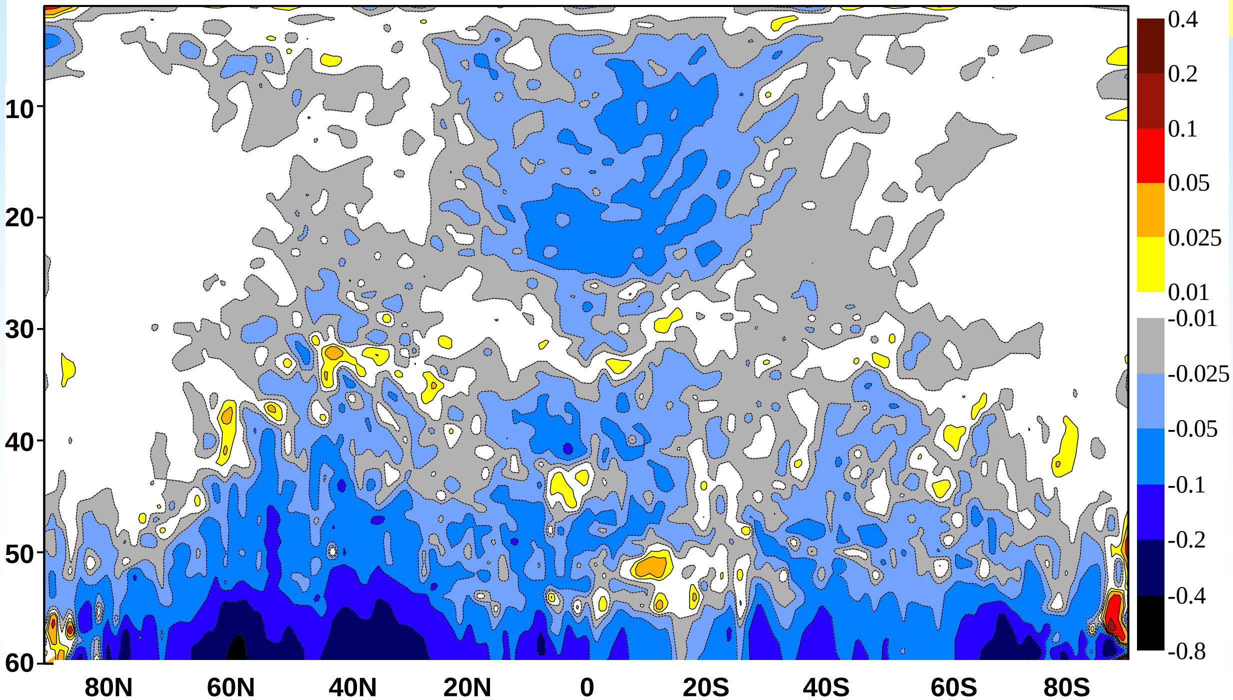
12-hour T159 L60 integration



Adiab simp vdif | vdif + gwd + radold + lsp + conv

Temperature

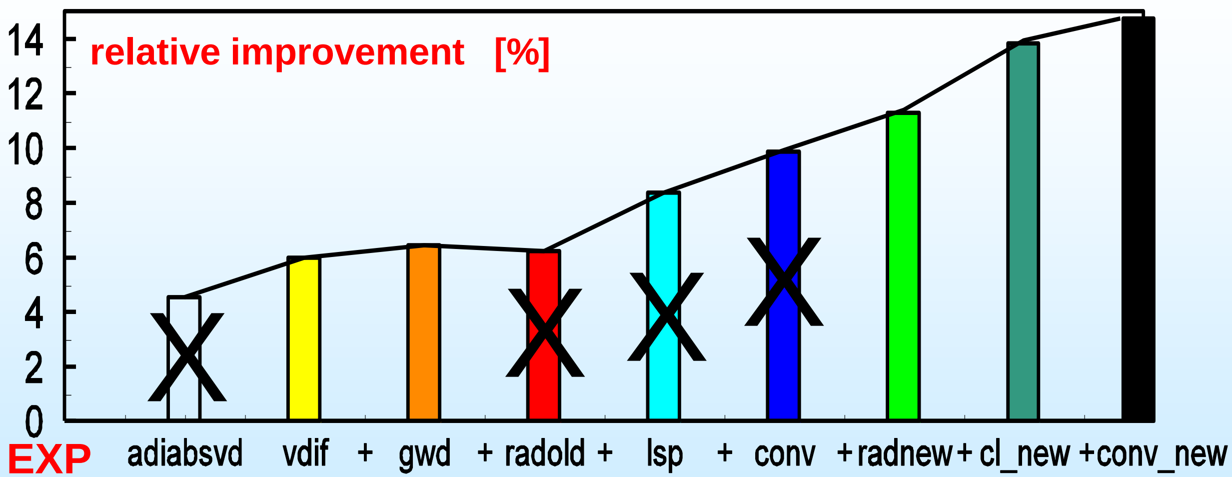
Impact of all physical processes (including new moist physics & radiation)



$\varepsilon_{EXP} - \varepsilon_{REF}$

REF = ADIAB

12-hour T159 L60 integration

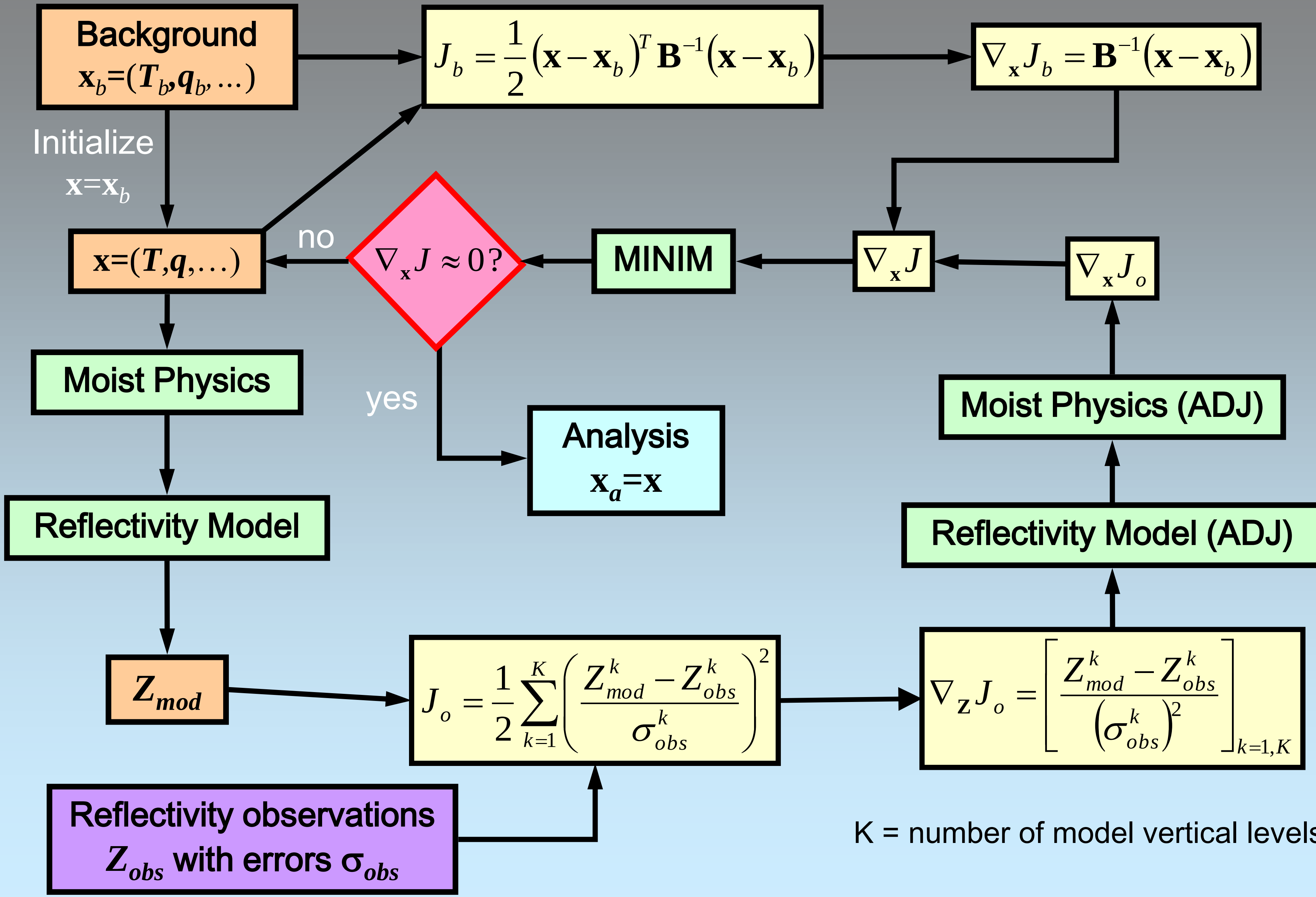


Adiab simp vdif | vdif + gwd + radnew + cl_new + conv_new

ECMWF

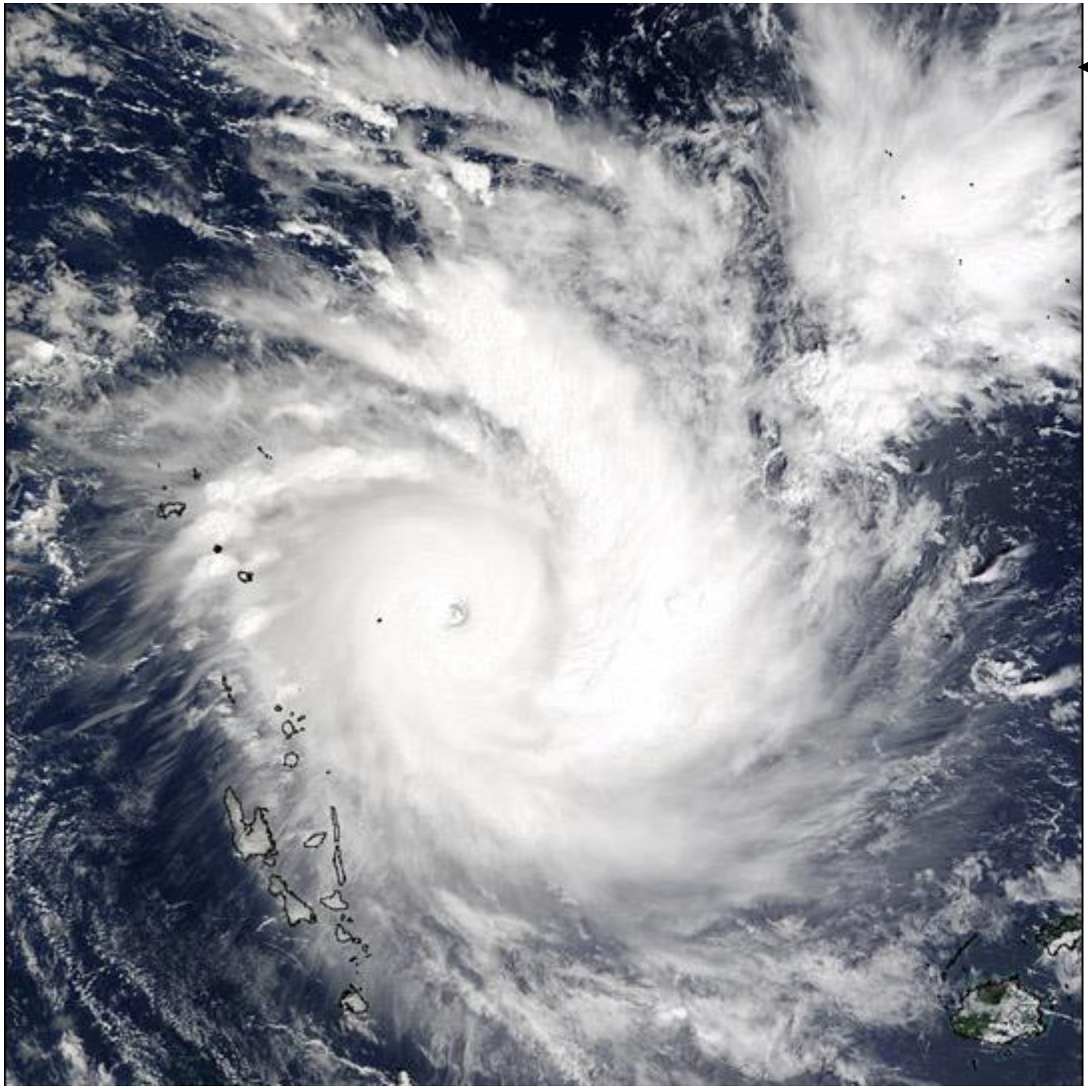
Applications

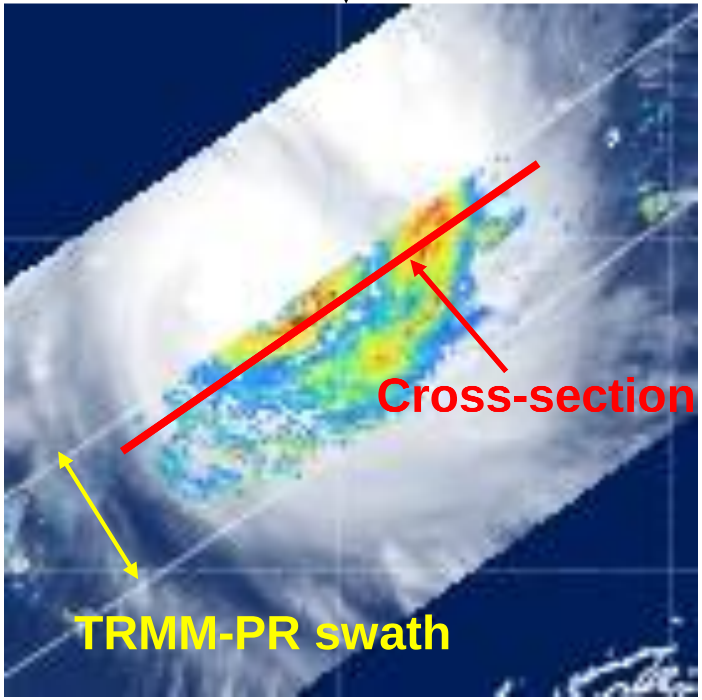
1D-Var with radar reflectivity profiles



1D-Var with TRMM/Precipitation Radar data

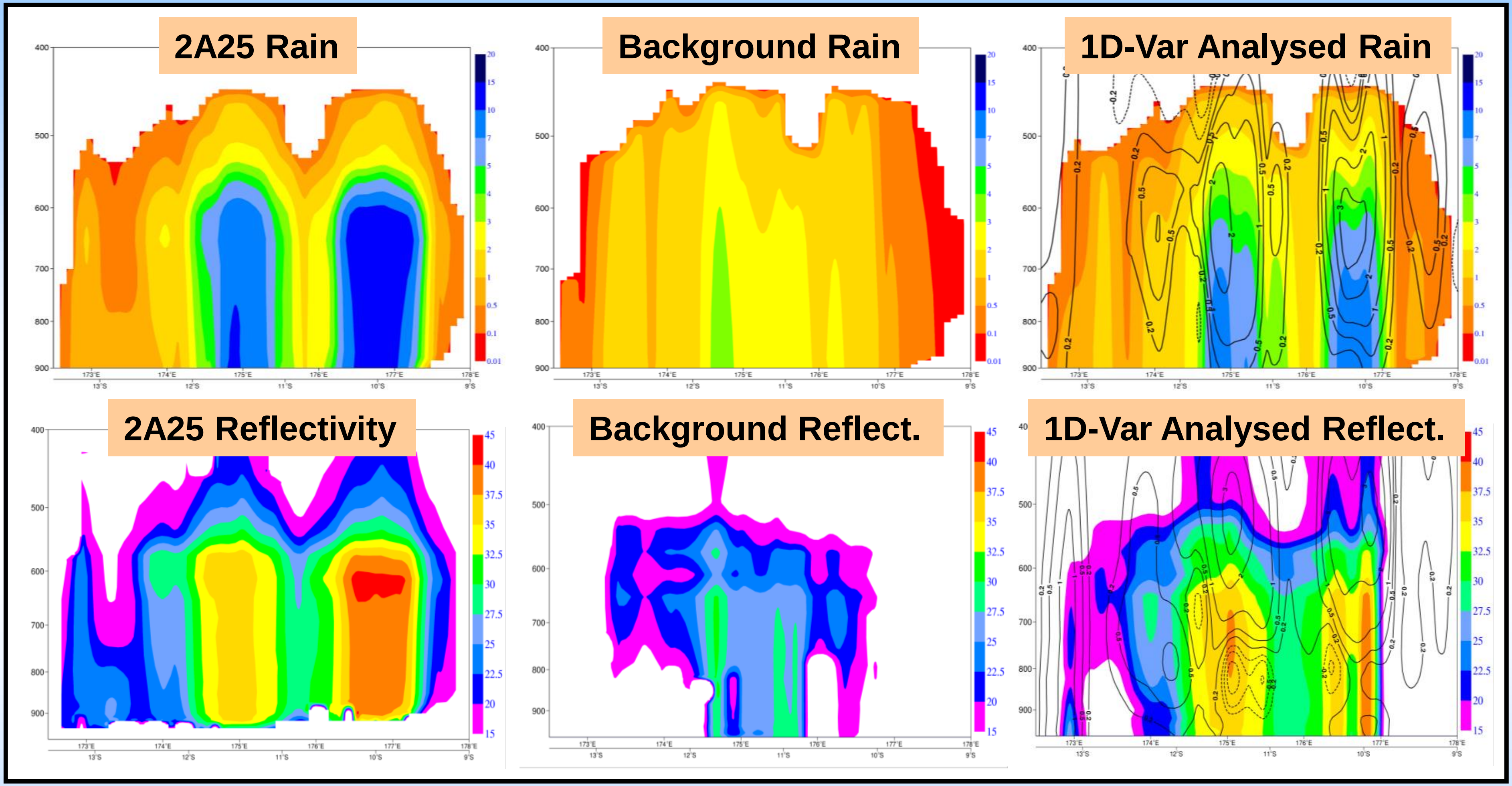
Tropical Cyclone Zoe (26 December 2002 @1200 UTC; Southwest Pacific)





ECMWF

1D-Var with TRMM/Precipitation Radar data



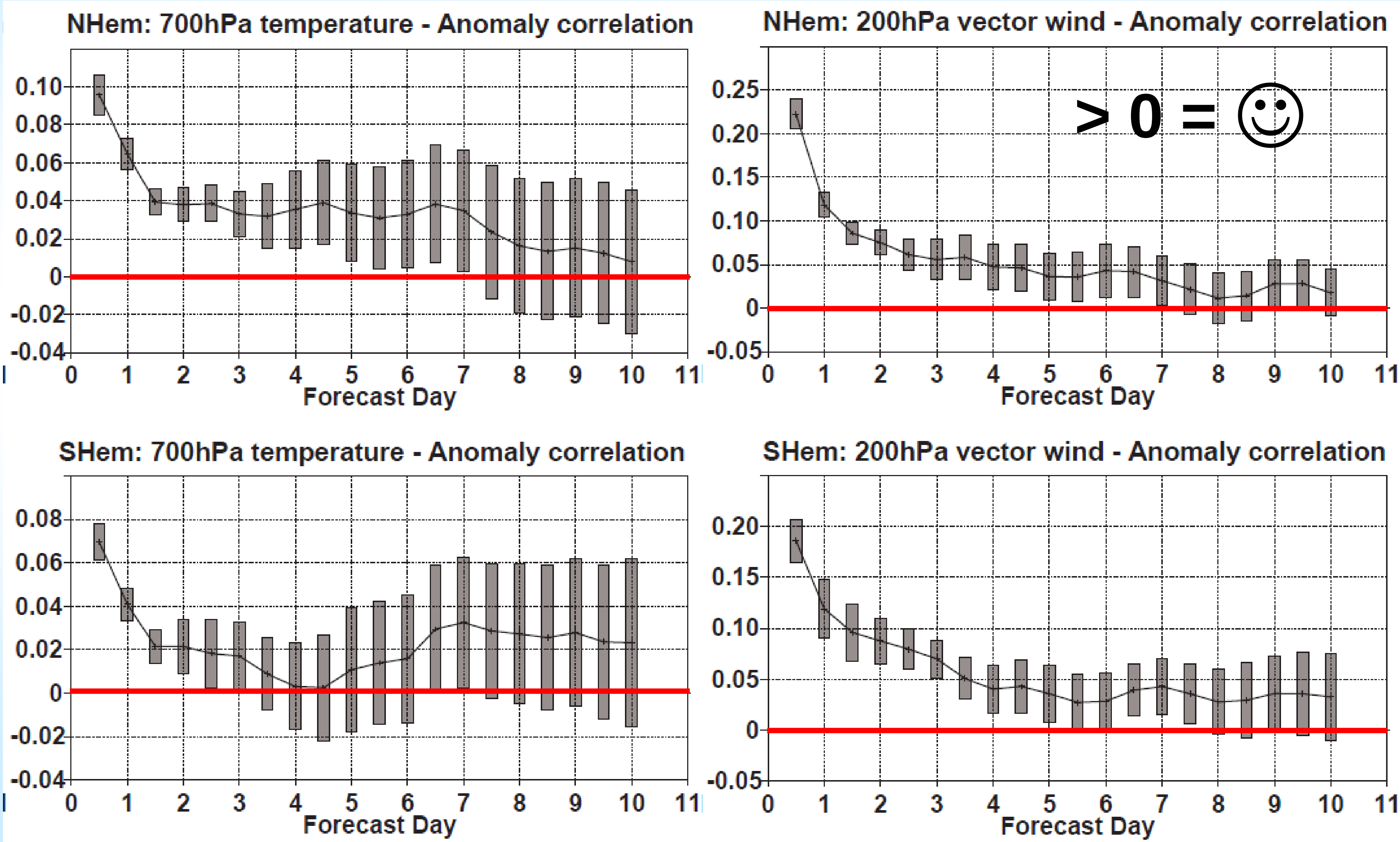
Tropical Cyclone Zoe (26 December 2002 @1200 UTC)

Vertical cross-section of rain rates (top, mm h^{-1}) and reflectivities (bottom, dBZ): observed (left), background (middle), and analyzed (right).

Black isolines on right panels = 1D-Var specific humidity increments.

Impact of ECMWF linearized physics on forecast scores

Comparison of two T511 L91 4D-Var 3-month experiments with & without full linearized physics: Relative change in forecast anomaly correlation.



Own impact of NCEP Stage IV hourly precipitation data over the U.S.A. (combined ground-based radar & rain gauge observations)

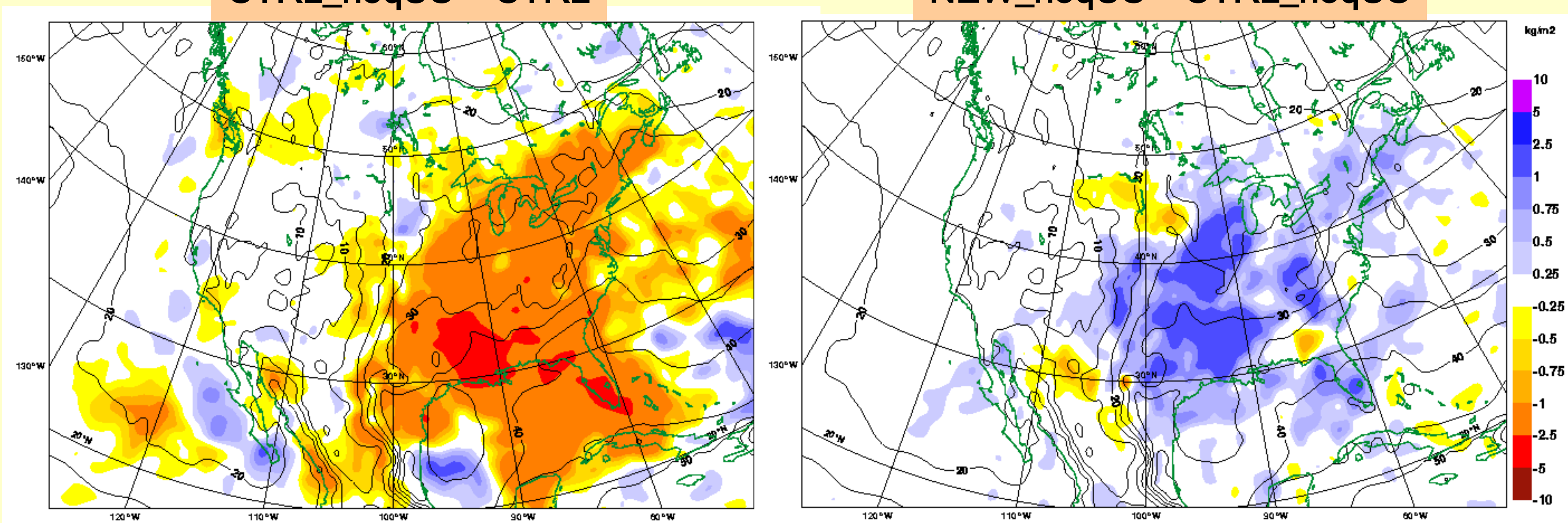
Three 4D-Var assimilation experiments (20 May - 15 June 2005):

CTRL = all standard observations.
CTRL_noqUS = all obs except no moisture obs over US (surface & satellite).
NEW_noqUS = CTRL_noqUS + NEXRAD hourly rain rates over US ("1D+4D-Var").

Mean differences of TCWV analyses at 00UTC

CTRL_noqUS – CTRL

NEW_noqUS – CTRL_noqUS



ECMWF

Lopez and Bauer (*Monthly Weather Review*, 2007)

Adjoint sensitivities

Idea: The time integration of the adjoint model allows the computation of adjoint sensitivities of any physical aspect (J) inside a target geographical domain to the model control variables several hours earlier.

Here:

J = 3h total surface precipitation averaged over a selected domain (N_{points}).

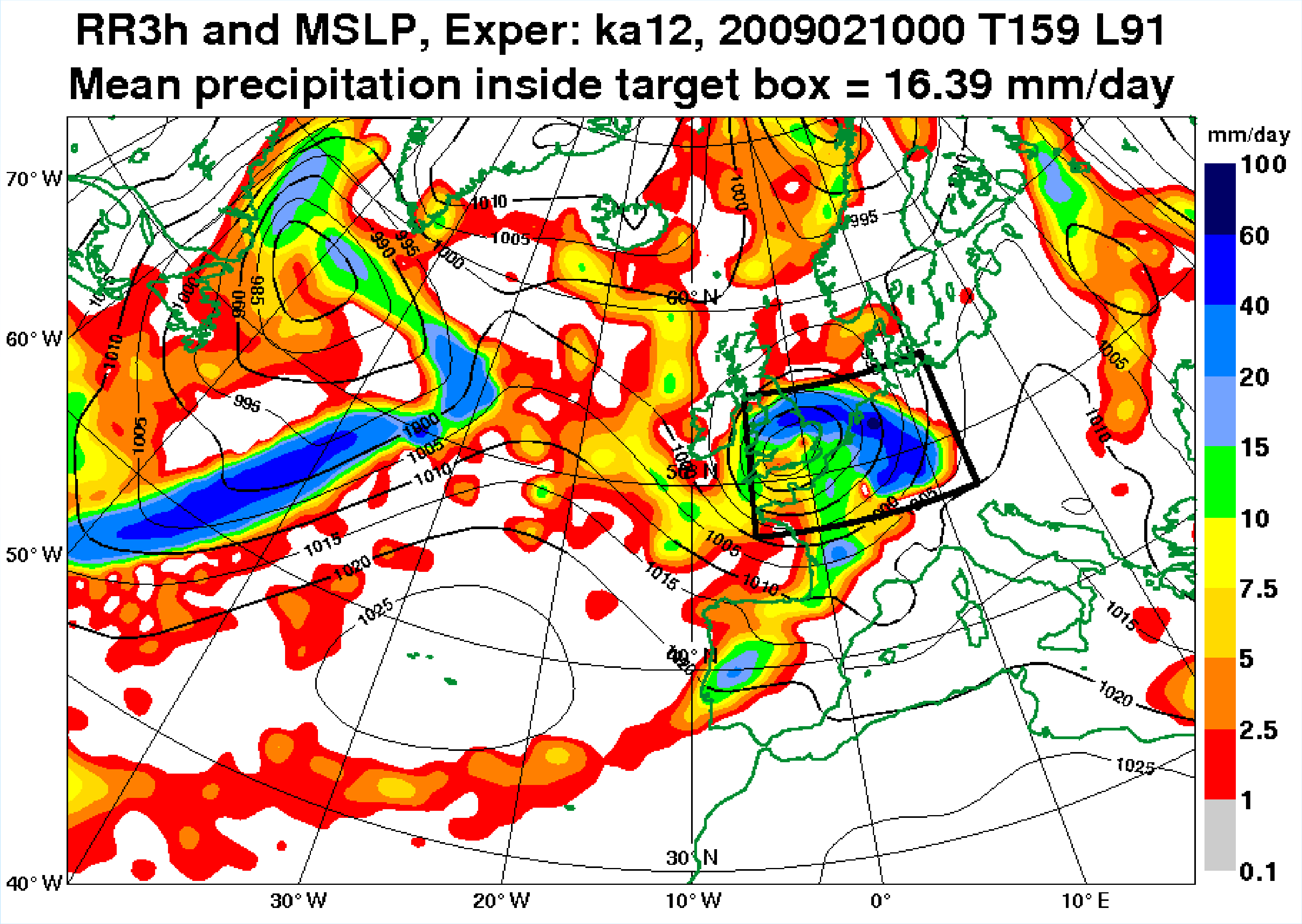
$$J = \frac{1}{N_{steps} \sum_{i=1}^{N_{points}} S_i} \sum_{t=1}^{N_{steps}} \sum_{i=1}^{N_{points}} S_i \, R_{i,t}$$

$$\frac{\partial J}{\partial R_{i,t}} = \frac{S_i}{N_{steps} \sum_{i=1}^{N_{points}} S_i} \xrightarrow{\text{Adjoint model incl. physics}} \frac{\partial J}{\partial \mathbf{x}_j}$$

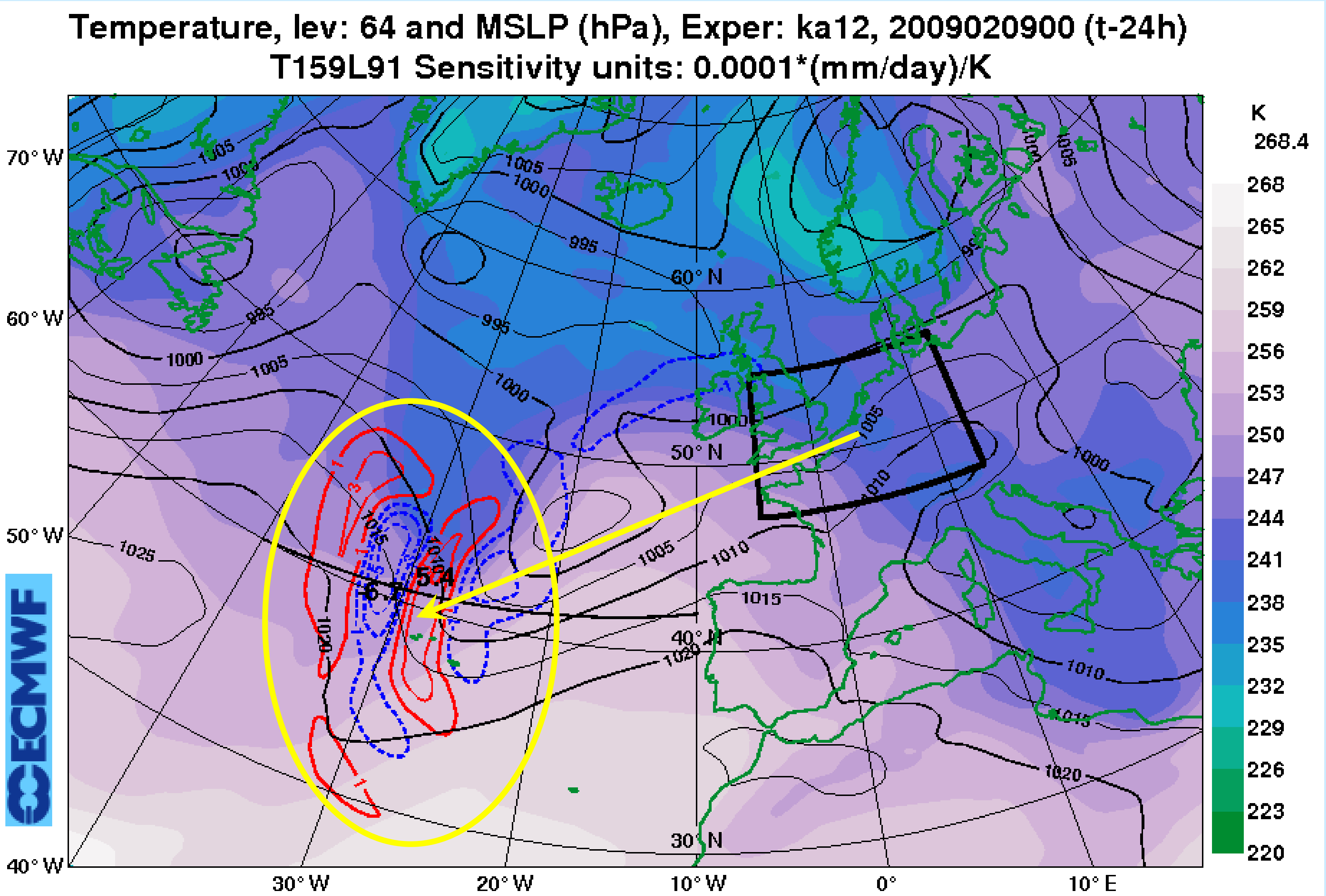
= sensitivity of rain criterion (J) to model input variables ($\mathbf{x}_j$) several hours earlier

ECMWF

Adjoint sensitivities for a European winter storm: *J* = mean 3h precipitation accumulation inside black box.

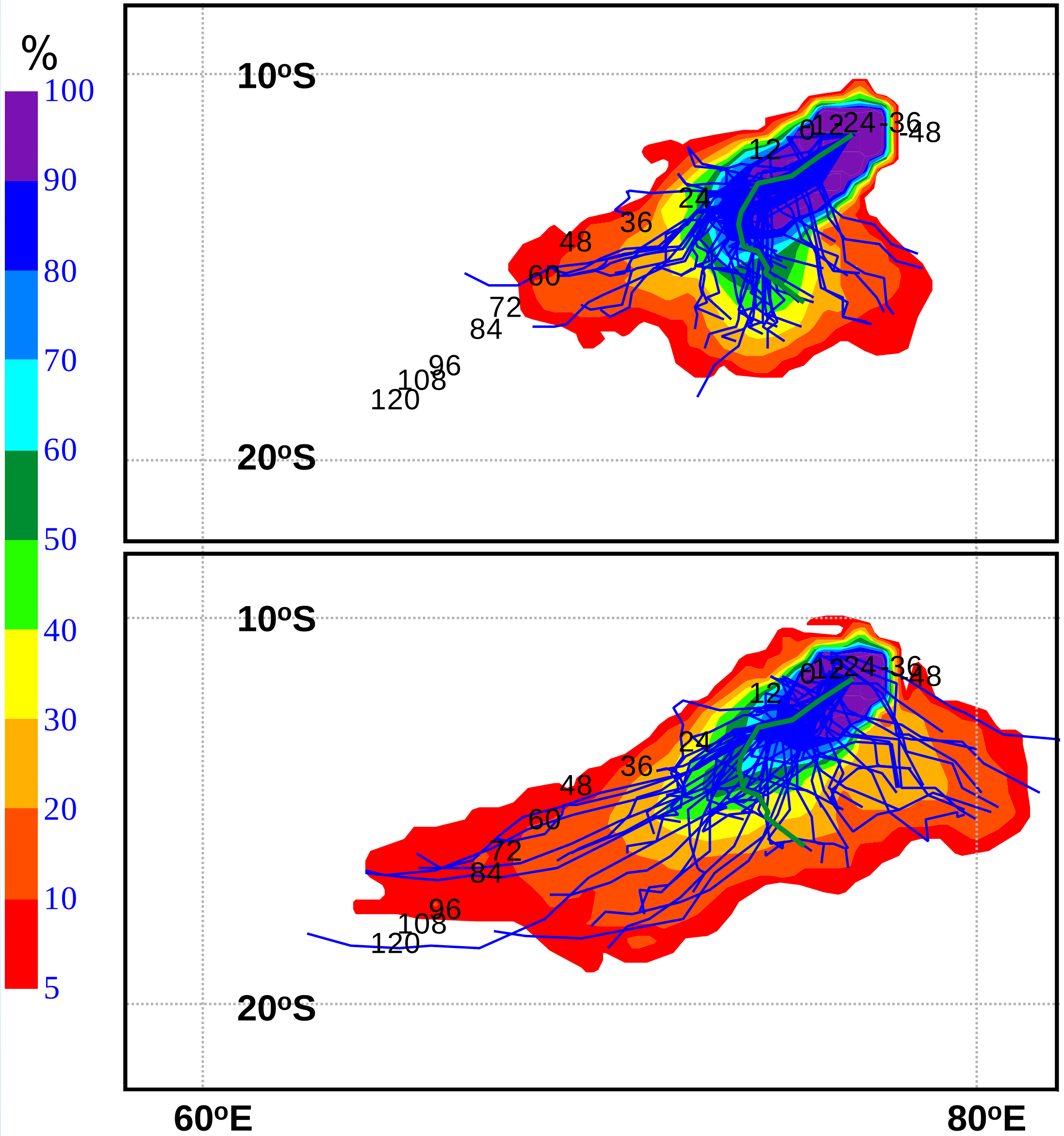


$\partial J/\partial x$ after 24 hours of "backward" adjoint integration



Tropical singular vectors in EPS *[Leutbecher and Van Der Grijn 2003]*

Probability of tropical cyclone passing within 120 km radius during next 120 hrs:



numbers – real position of the cyclone at the certain hour

green line – control T255 forecast (unperturbed member of ensemble)

Tropical singular vectors
VDIF only in TL/AD

06/03/2003 12 UTC

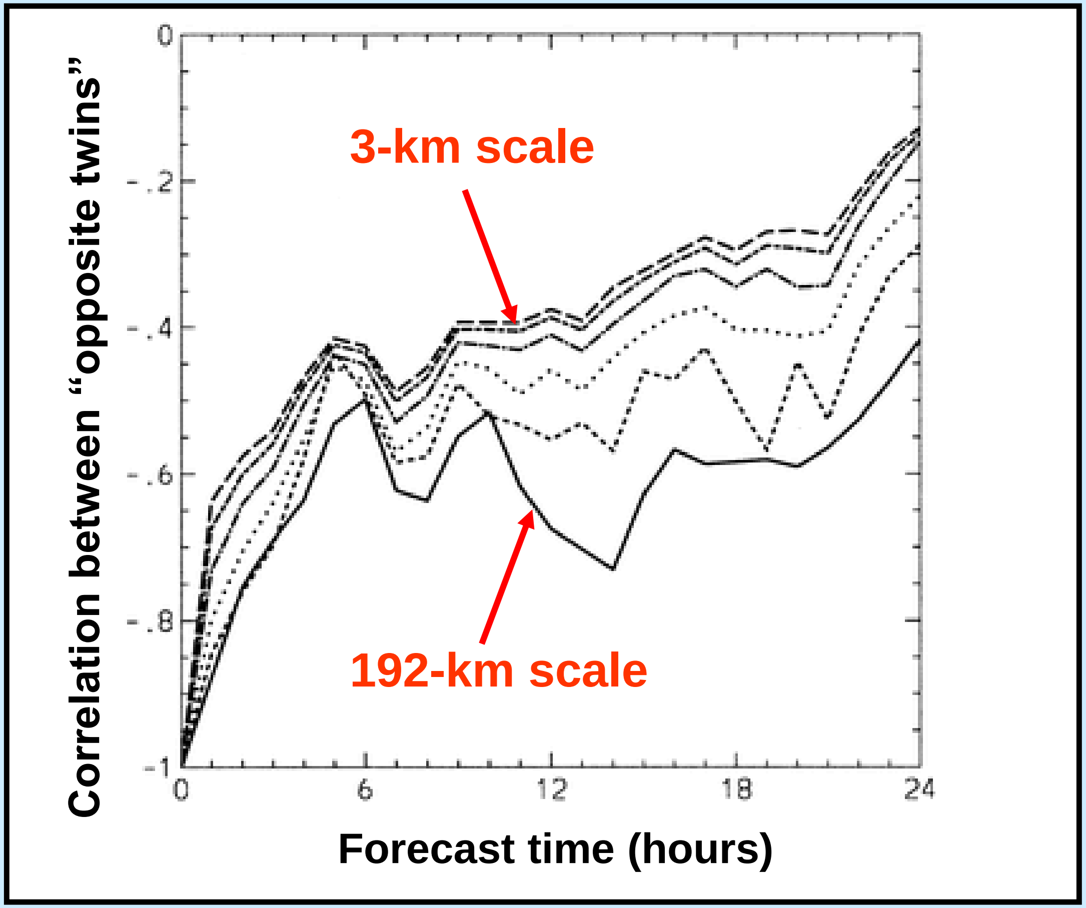
Tropical cyclone Kalunde

Tropical singular vectors
Full physics in TL/AD

Influence of time and resolution on linearity assumption in physics

Results from ensemble runs with the MC2 model (3 km resolution) over the Alps, from *Walser et al. (2004).*
Comparison of a pair of "opposite twin" experiments.

Linearity



→The validity of the linear assumption for precipitation quickly drops in the first hours of the forecast, especially for smaller scales.

General conclusions

- **Physical parametrizations have become important components in recent variational data assimilation systems.**

- **However, their linearized versions (tangent-linear and adjoint) require some special attention (regularizations/simplifications) in order to eliminate possible discontinuities and non-differentiability of the physical processes they represent.**

- **This is particularly true for the assimilation of observations related to precipitation, clouds and soil moisture, to which a lot of efforts are currently devoted.**

- **Developing new simplified parametrizations for data assimilation requires:**
  - **Compromise between realism, linearity and computational cost,**
  - **Evaluation in terms of Jacobians (not to noisy in space and time),**
  - **Validation of forward simplified code against observations,**
  - **Comparison to full non-linear code used in forecast mode (4D-Var trajectory),**
  - **Numerical tests of tangent-linear and adjoint codes for small perturbations,**
  - **Validity of the linear hypothesis for perturbations with larger size (typical of analysis increments).**
  - **Successful convergence of 4D-Var minimizations.**

Thank you!

Example of observation operator H (radiative transfer model):

$$\mathbf{x} = \begin{bmatrix} T_1 \\ \vdots \\ T_n \\ q_1 \\ \vdots \\ q_n \end{bmatrix} \xrightarrow{\;H\;} \mathbf{y} = \begin{bmatrix} Rad_{ch1} \\ Rad_{ch2} \\ Rad_{ch3} \end{bmatrix}$$

and its tangent-linear operator $\mathbf{H}$:

$$\mathbf{H} = \begin{bmatrix} \frac{\partial Rad_{ch1}}{\partial T_1} & \cdots & \frac{\partial Rad_{ch1}}{\partial T_n} & \frac{\partial Rad_{ch1}}{\partial q_1} & \cdots & \frac{\partial Rad_{ch1}}{\partial q_n} \\ \frac{\partial Rad_{ch2}}{\partial T_1} & \cdots & \frac{\partial Rad_{ch2}}{\partial T_n} & \frac{\partial Rad_{ch2}}{\partial q_1} & \cdots & \frac{\partial Rad_{ch2}}{\partial q_n} \\ \frac{\partial Rad_{ch3}}{\partial T_1} & \cdots & \frac{\partial Rad_{ch3}}{\partial T_n} & \frac{\partial Rad_{ch3}}{\partial q_1} & \cdots & \frac{\partial Rad_{ch3}}{\partial q_n} \end{bmatrix}$$