Parametrizations in Data Assimilation

ECMWF Training Course
24 November 2023

Philippe Lopez

Physical Processes Group, Research Department, ECMWF
philippe.lopez@ecmwf.int

Parametrizations in Data Assimilation

Lecture 1

- **Introduction**
- **An example of physical initialization**
- **A very simple variational assimilation problem**
- **3D-Var assimilation**
- **The concept of adjoint**
- **4D-Var assimilation**

Lecture 2

- **Tangent-linear and adjoint coding**
- **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 three-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, microphysics,…)

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

Observations

Operationally assimilated for several decades:

* **Surface measurements** (SYNOP, SHIPS, buoys,…),
* **Vertical soundings** (TEMP, PILOT, aircraft reports, 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,…).

Over the past decade:

* **Satellite radiances/retrievals in cloudy and rainy regions** (SSM/I, TMI,…),
* Precipitation measurements from **ground-based radars** and **rain gauges.**
* Satellite measurements of aerosols and trace gases (e.g., CAMS analysis).

Still experimental:

* Satellite cloud/precipitation radar reflectivities or products (TRMM, CloudSat),
* Lidar backscattering/products (wind vectors, water vapour) (CALIPSO),
* Lightning optical signal (TRMM-LIS; more recently GOES-R series; soon MTG-LI).

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 (~ 60 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 more "exotic" observed quantities** (e.g. precipitation rates, radiances, radar 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.

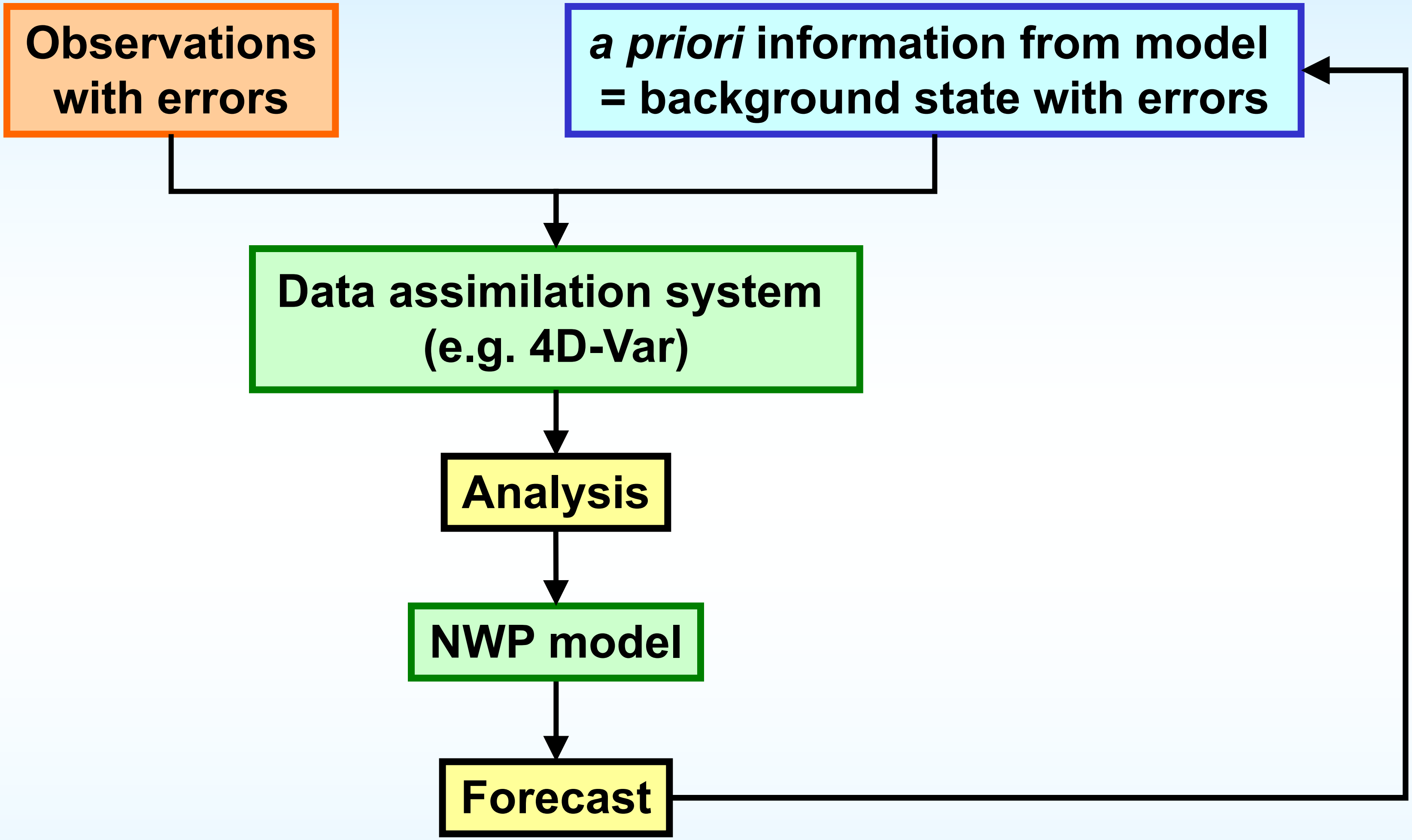
For instance, problems in the assimilation can arise from the discontinuous or non-linear nature of moist processes.

Improvements are still needed…

- More observations are needed to improve the analysis and forecast 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, snow on the ground, sea ice),
  - 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: 9 km globally, 137 levels),

To summarize…



Physical parametrizations are needed in data assimilation:

- to link the model variables to the observed quantities,

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

Empirical initialization

Example from Ducrocq *et al.* (2000), Météo-France:

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

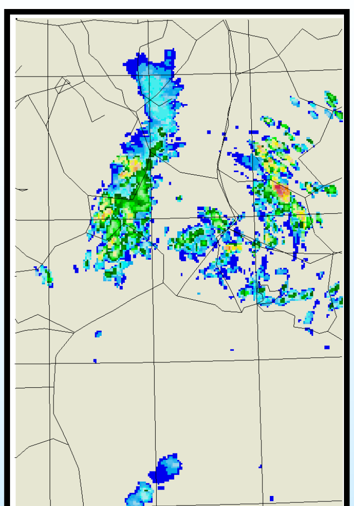
- Particular focus on strong convective events.

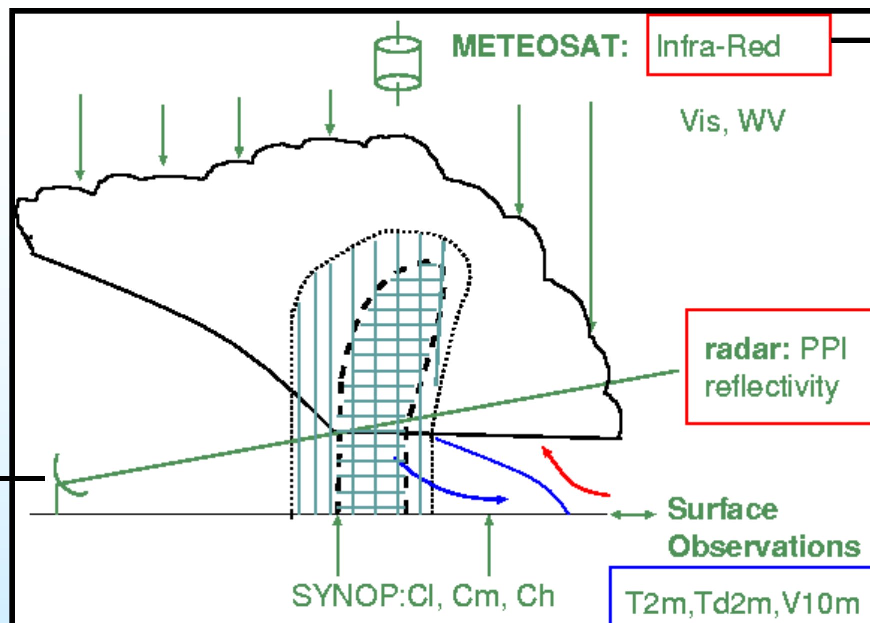
- **Method:** Before running the forecast:

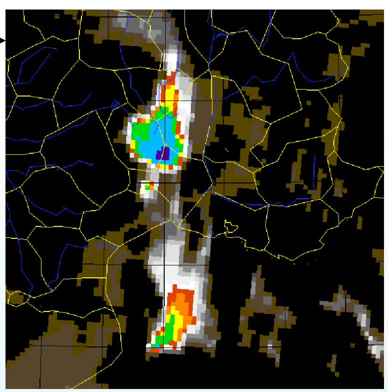
  1) A mesoscale surface analysis is performed (esp. to identify convective cold pools)

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

Radar

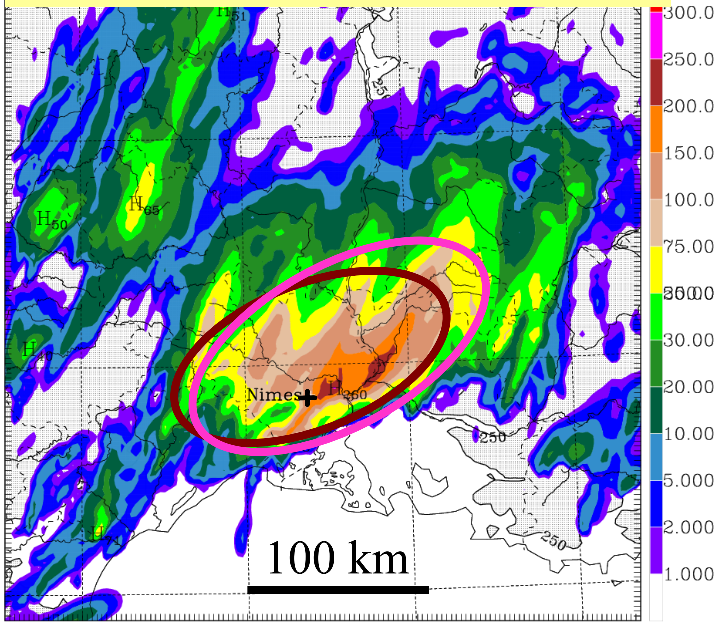




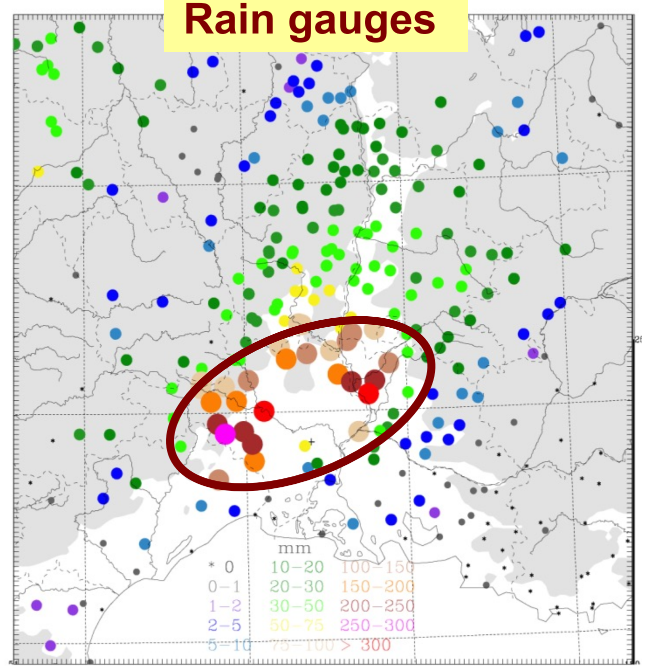


METEOSAT

12h FC from **modified analysis**



Rain gauges

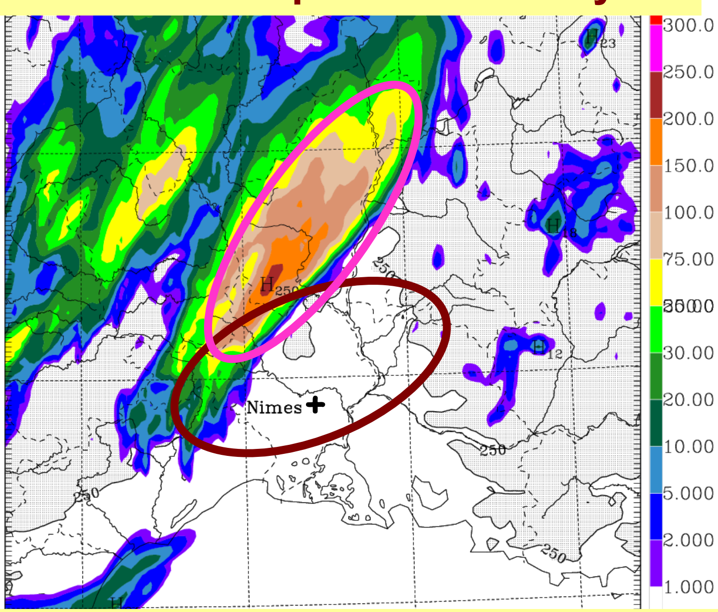


Ducrocq et al. (2004)

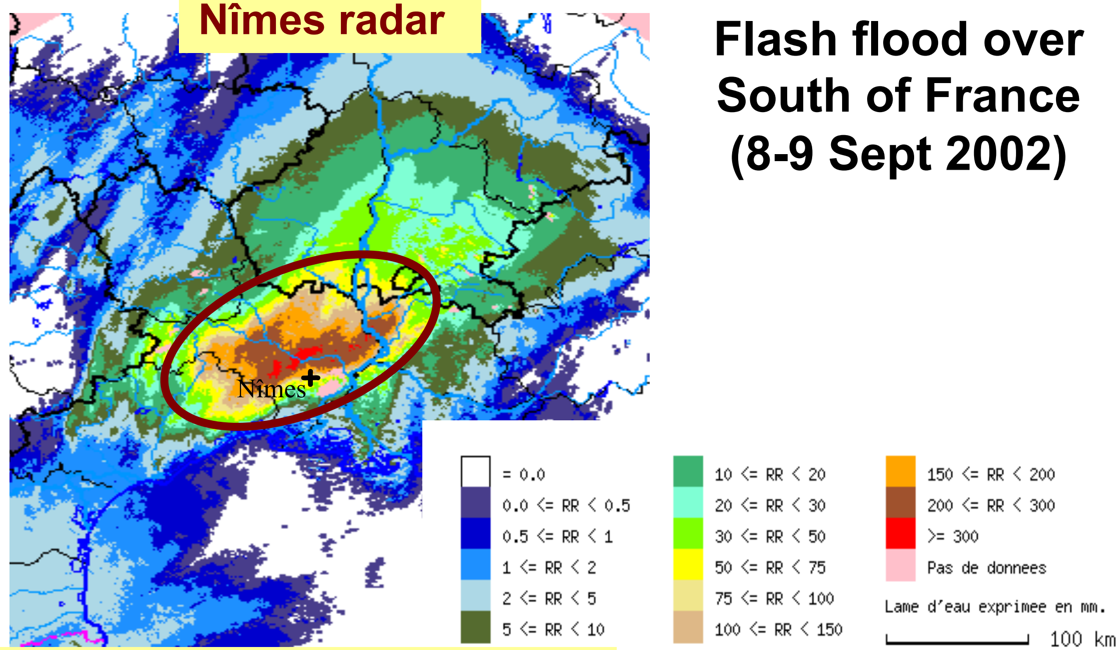
2.5-km resolution model Méso-NH

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

12h FC from **operational analysis**



Nîmes radar



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:

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

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.

Matrix $\mathbf{H}^T$ = Adjoint of operator H

Observation operators in data assimilation

Example of non-linear observation operator:

$$\text{model state} \quad \mathbf{x} = \begin{bmatrix} T \\ q_v \\ u \\ v \\ P_s \\ q_{liq} \\ q_{ice} \end{bmatrix} \xrightarrow{H} \tilde{\mathbf{y}}_i = \begin{bmatrix} Rad_{ch1} \\ Rad_{ch2} \\ Rad_{ch3} \end{bmatrix}$$

observation equivalent
= satellite cloudy radiances

H = radiative transfer model

Tangent-linear and adjoint operators

The tangent-linear operator is applied to perturbations:

$$\boldsymbol{\delta}\mathbf{x} = \begin{bmatrix} \delta T \\ \delta q_v \\ \delta u \\ \delta v \\ \delta P_s \\ \delta q_{liq} \\ \delta q_{ice} \end{bmatrix} \xrightarrow{\mathbf{H}} \boldsymbol{\delta}\tilde{\mathbf{y}}_i = \begin{bmatrix} \delta Rad_{ch1} \\ \delta Rad_{ch2} \\ \delta Rad_{ch3} \end{bmatrix}$$

The adjoint operator is applied to the cost function gradient:

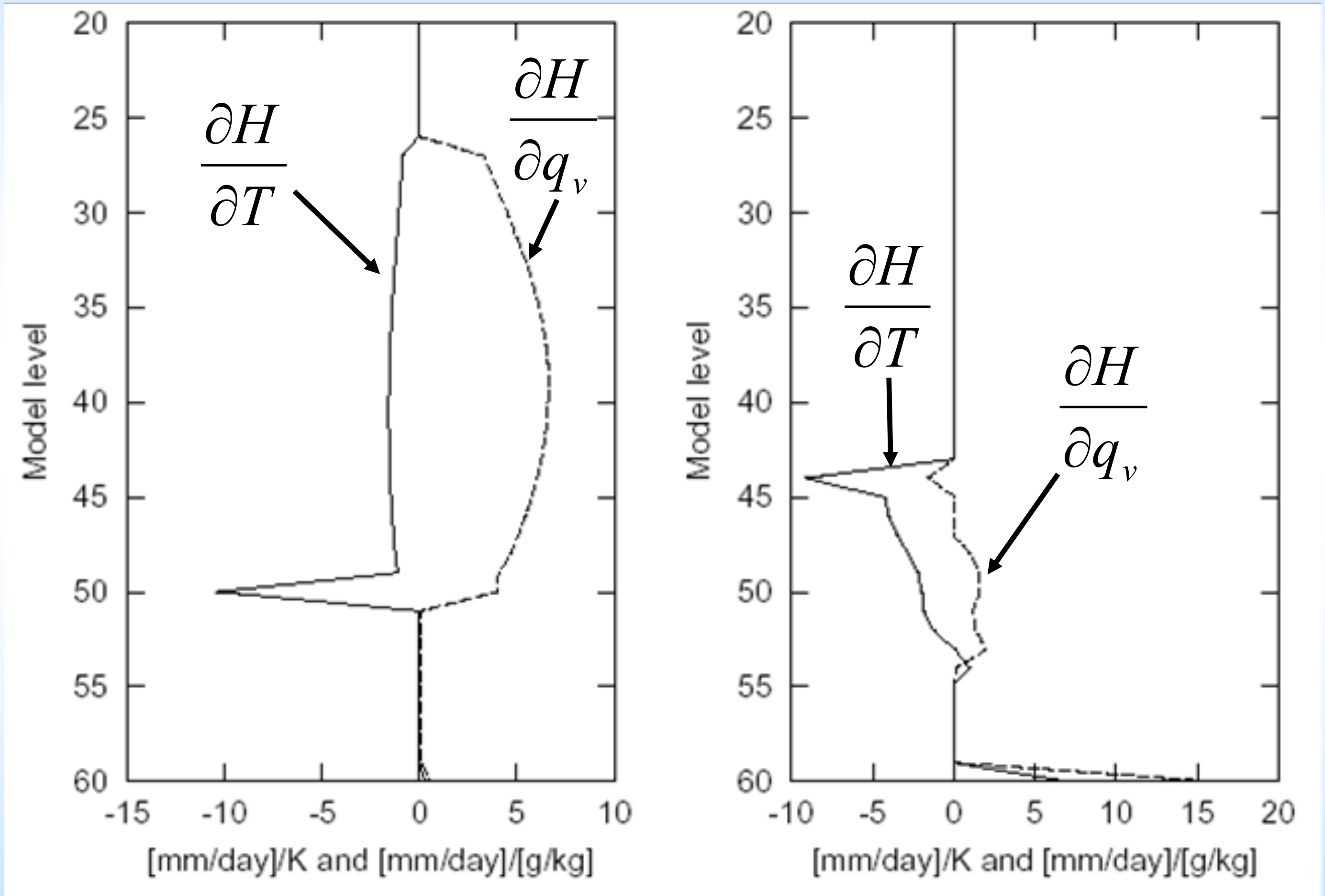
$$\nabla_{\tilde{\mathbf{y}}_i} J_o = \begin{bmatrix} \partial J_o / \partial Rad_{ch1} \\ \partial J_o / \partial Rad_{ch2} \\ \partial J_o / \partial Rad_{ch3} \end{bmatrix} \xrightarrow{\mathbf{H}^T} \nabla_{\mathbf{x}} J_o = \begin{bmatrix} \partial J_o / \partial T \\ \partial J_o / \partial q_v \\ \partial J_o / \partial u \\ \partial J_o / \partial v \\ \partial J_o / \partial P_s \\ \partial J_o / \partial q_{liq} \\ \partial J_o / \partial q_{ice} \end{bmatrix}$$

ECMWF

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

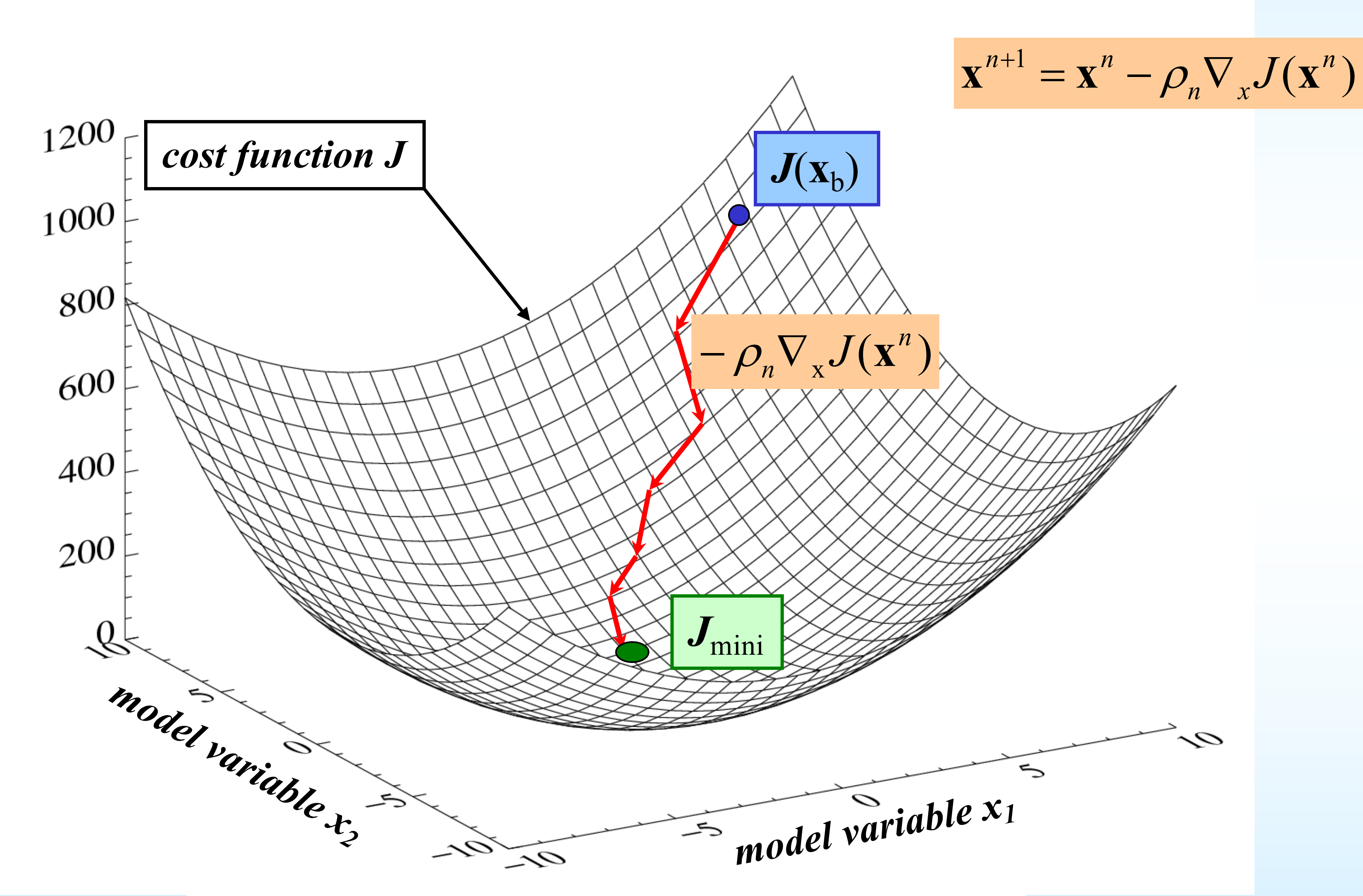


Marécal and Mahfouf (2002)

Betts-Miller (adjustment scheme)

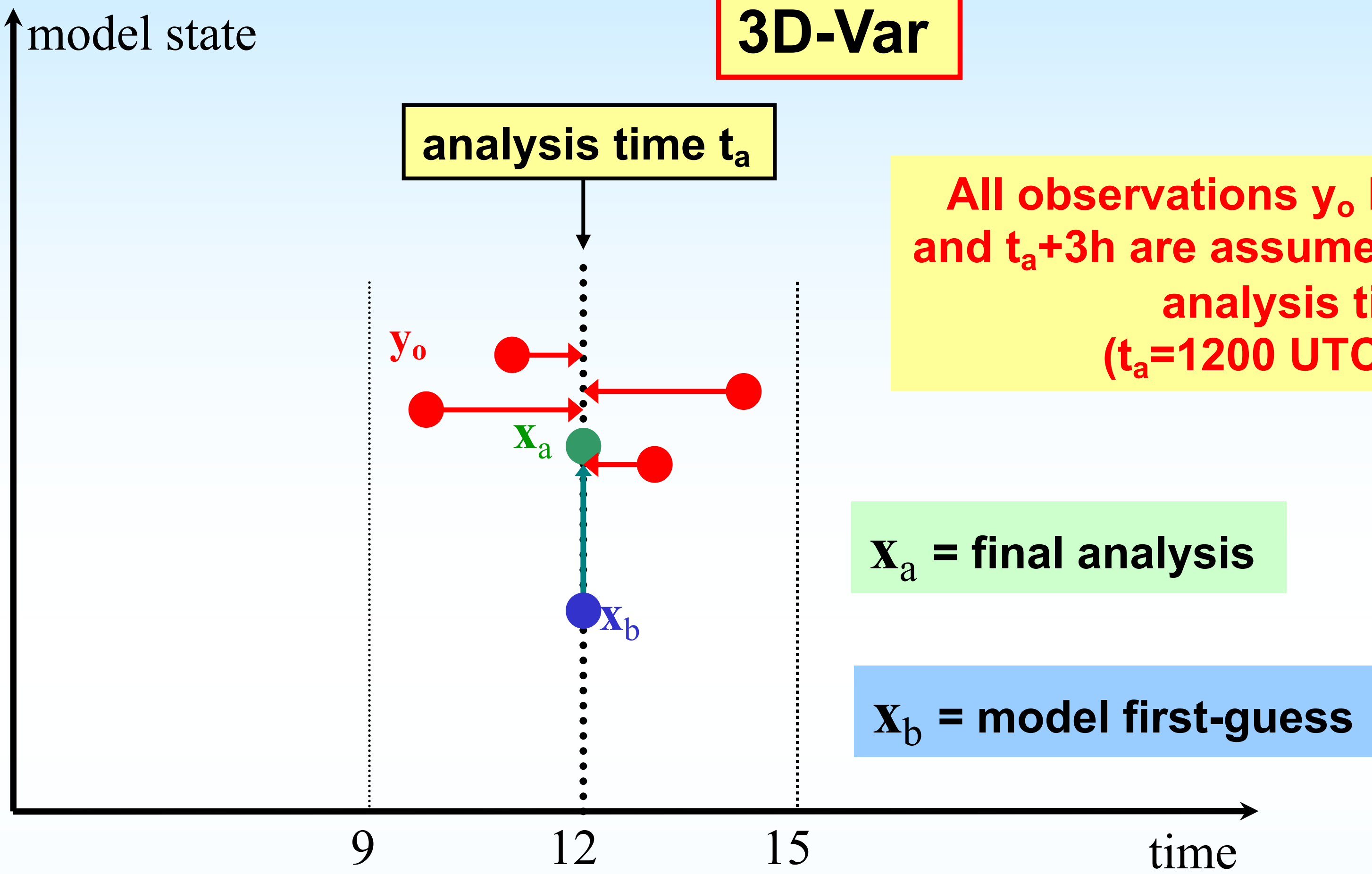
Tiedtke (ECMWF's oper mass-flux scheme)

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



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

3D-Var

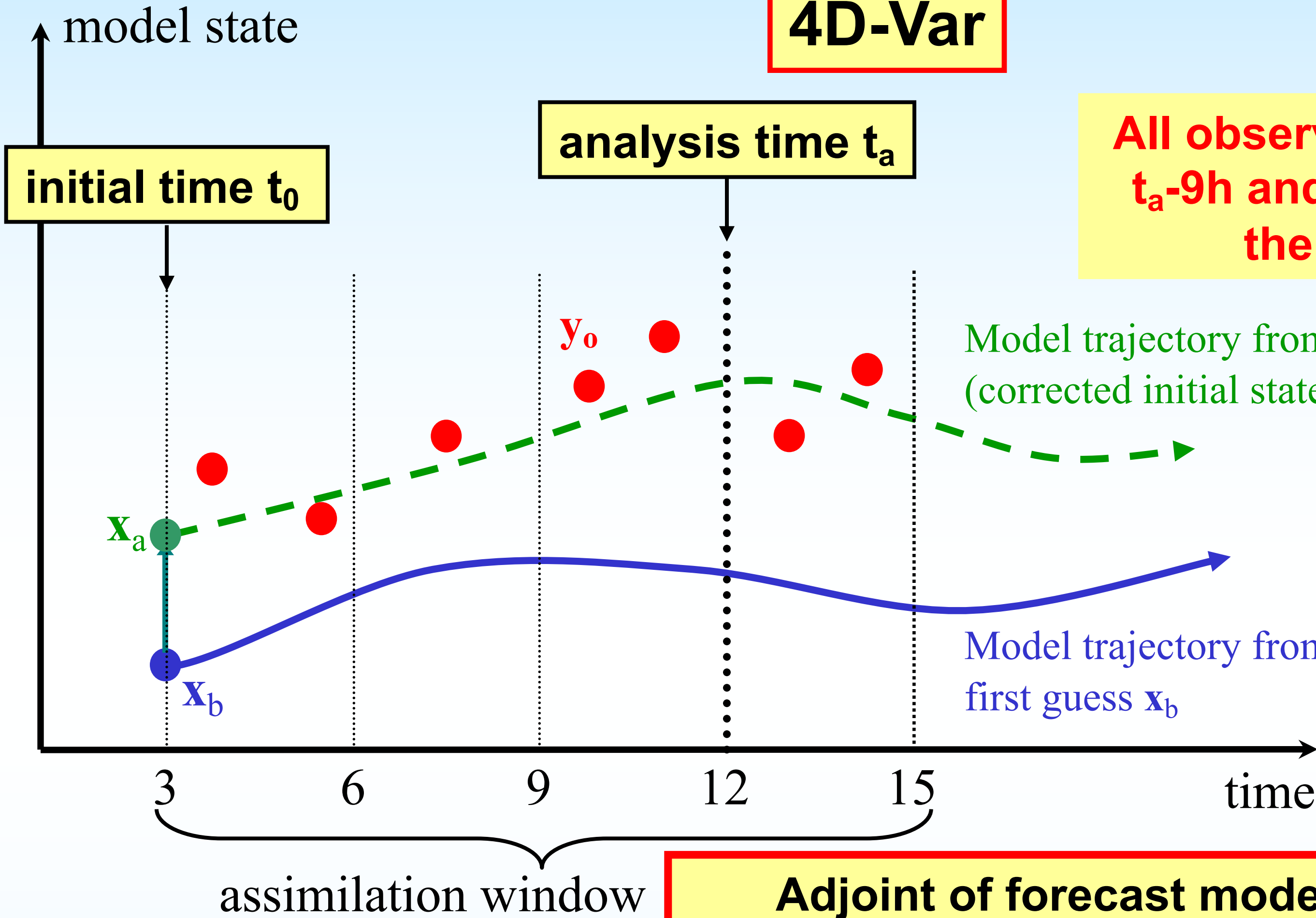


All observations y_o between t_a-3h and t_a+3h are assumed to be valid at analysis time (t_a=1200 UTC here)

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

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

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

Adjoint of forecast model with simplified linearized physics

$$\min J = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_b) + \frac{1}{2}\sum_{i=0}\left(H_i(M[\mathbf{x}_0]) - \mathbf{y}_{oi}\right)^T \mathbf{R}_i^{-1}\left(H_i(M[\mathbf{x}_0]) - \mathbf{y}_{oi}\right)$$

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

EXAMPLE OF ADJOINT CODING

- **non-linear statement**

$$x = y + z^2$$

EXAMPLE OF ADJOINT CODING

- **non-linear statement**

$$x = y + z^2$$

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

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.

Testing the tangent-linear code

The correctness of the tangent-linear model, $\mathbf{M}$, must be assessed by checking that the first-order Taylor approximation is valid:

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

Example of output from a successful tangent-linear test:

Tiny perturbations (top of table)

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

Larger perturbations (bottom of table)

Machine precision reached (first four rows)

Improvement when perturbation size decreases

Testing the adjoint code

The correctness of the adjoint model needs to be assessed by checking that it satisfies the mathematical relationship:

$$\forall \delta \mathbf{x}, \delta \mathbf{y} \;\; \langle \mathbf{M}\delta \mathbf{x}, \delta \mathbf{y} \rangle = \langle \delta \mathbf{x}, \mathbf{M}^T \delta \mathbf{y} \rangle$$

where $\mathbf{M}$ is the tangent-linear model and $\mathbf{M}^T$ is the adjoint model.

Example of output from a successful adjoint test:

$$< \mathbf{M}\, \delta \mathbf{x}, \quad \delta \mathbf{y} > = -0.13765102625164E\text{-}01$$

$$< \quad \delta \mathbf{x}, \mathbf{M}^T \delta \mathbf{y} > = -0.13765102625168E\text{-}01$$

The difference is 11.351 times the zero of the machine

The adjoint test should be correct at the level of machine precision (typ. 13 to 15 common digits for the entire model).
If not, there must be a bug in the code!

Testing the adjoint code

At ECMWF, the adjoint is tested for multiple dates and configurations of the linearized physics → tabulated number of common digits:

AD test, exper ca**
TL399 L137 +12, cy47r1, 20 members

Number of matching digits

Members \ Physics configurations	1	2	3	4	5	6	7	8	9	10
20	7	7	13	10	13	12	11	14	11	12
19	8	8	13	11	14	13	14	14	13	12
18	9	9	13	10	13	11	11	12	11	11
17	13	13	13	13	10	13	13	12	12	14
16	9	9	14	11	13	14	15	15	14	9
15	10	10	15	11	15	14	14	13	14	14
14	11	11	14	13	14	14	14	15	15	12
13	10	10	14	12	14	15	14	15	14	13
12	9	9	14	10	14	13	13	14	12	12
11	9	9	14	11	13	12	11	14	12	13
10	9	9	11	10	10	12	14	12	13	12
9	8	8	14	10	14	12	11	14	12	11
8	8	8	14	9	14	13	9	14	13	13
7	9	9	15	11	15	13	12	14	13	13
6	10	10	12	10	12	10	12	12	10	12
5	9	9	13	10	13	11	12	13	11	12
4	10	10	14	12	13	13	13	15	13	12
3	10	10	15	11	14	12	13	14	13	12
2	9	9	14	11	14	12	13	15	12	12
1	10	10	13	11	13	11	11	13	11	14

AD test, exper au**
TL399 L137 +12, cy47r1, 20 members

Number of matching digits

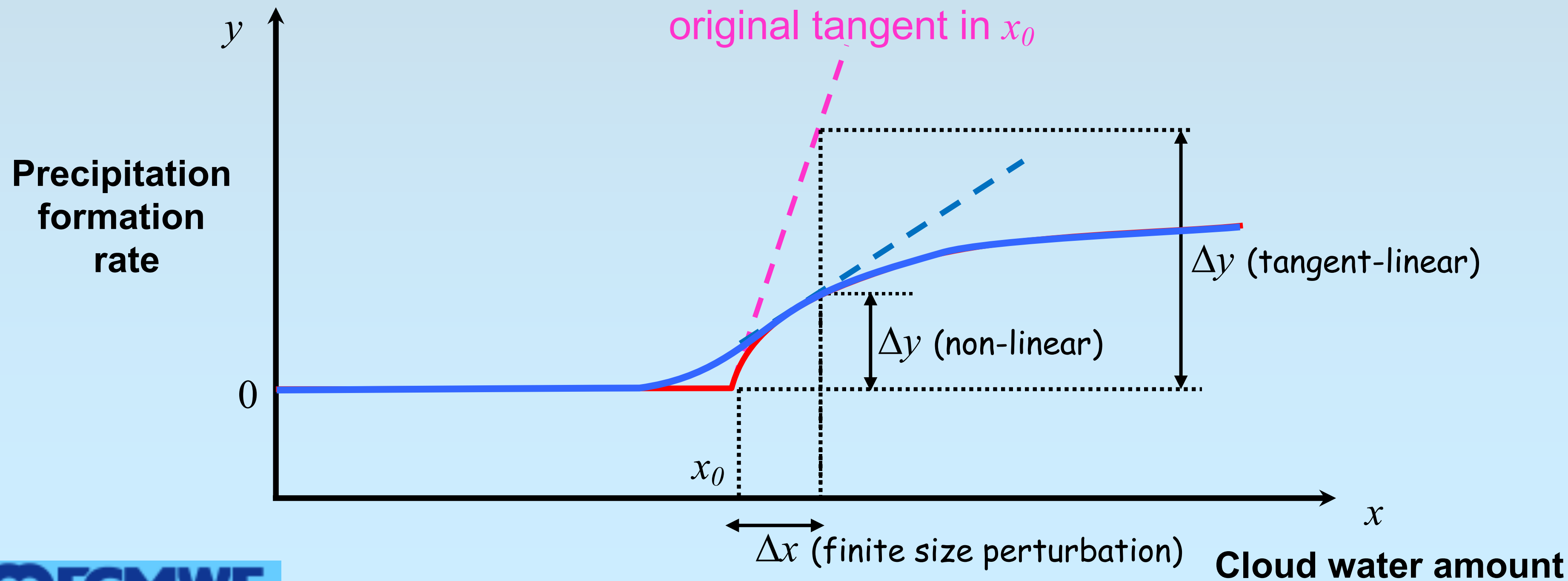
Members \ Physics configurations	1	2	3	4	5	6	7	8	9	10
20	11	11	13	13	13	13	13	14	13	12
19	14	14	13	14	14	14	14	13	13	14
18	12	13	12	12	12	12	12	13	12	12
17	12	12	12	13	12	13	13	13	14	13
16	11	11	14	12	15	15	14	15	14	11
15	13	13	14	14	14	15	13	14	14	13
14	13	13	15	14	13	15	15	14	15	13
13	12	13	15	14	13	15	14	15	14	13
12	12	12	14	13	14	14	14	14	14	11
11	13	13	14	13	14	14	14	14	13	12
10	10	11	11	12	11	11	13	12	12	11
9	14	14	13	13	14	14	13	12	13	13
8	13	13	14	12	15	15	14	15	15	13
7	12	13	14	13	14	14	14	14	15	12
6	13	12	12	13	12	12	12	12	12	12
5	12	12	13	13	13	13	13	13	13	12
4	12	12	13	13	13	14	13	13	14	12
3	12	12	14	14	14	15	15	13	14	13
2	13	13	14	12	15	15	13	15	14	13
1	13	13	13	15	13	13	13	13	12	13

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.**



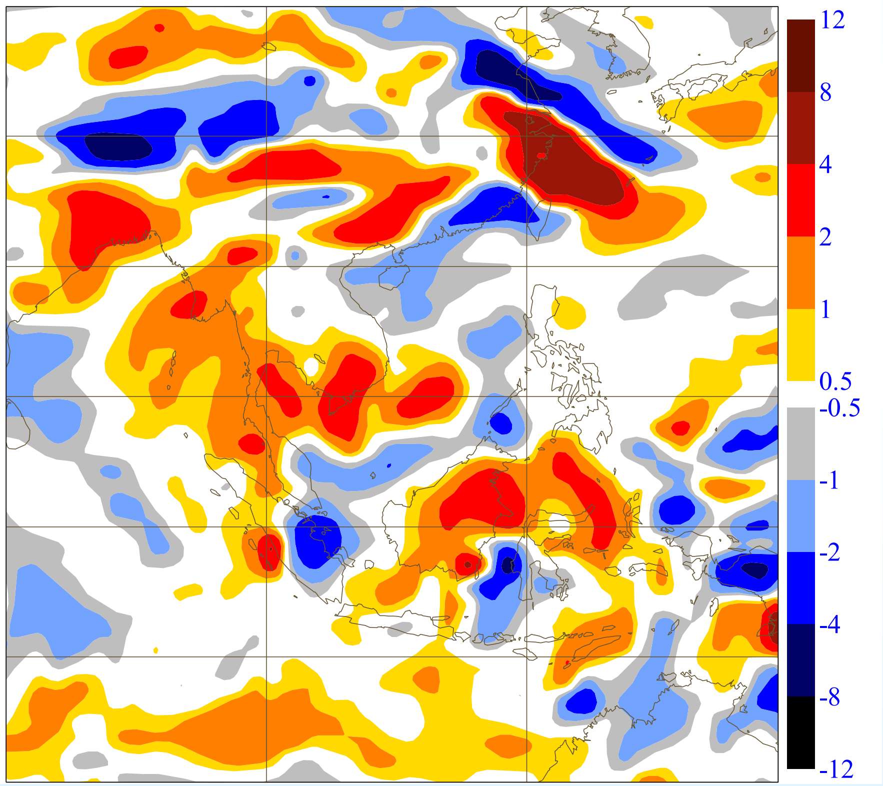
ECMWF

An example of spurious TL noise caused by a threshold in the autoconversion formulation of the large-scale cloud scheme.

~700 hPa zonal wind increments [m/s] from 12h model integration.

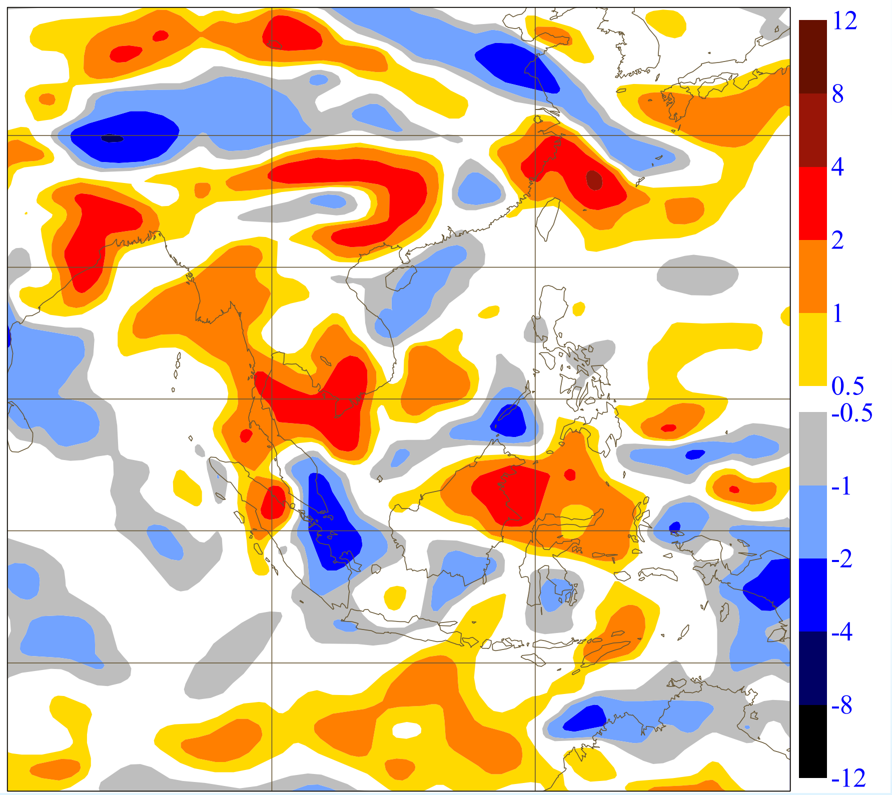
Nonlinear finite difference:

$M(\mathbf{x}+\delta\mathbf{x}) - M(\mathbf{x})$



from M. Janisková

Tangent-linear integration: $\mathbf{M}\delta\mathbf{x}$



with perturbation reduction in autoconversion

ECMWF

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)

Impact of linearized physics on TL approximation (1)

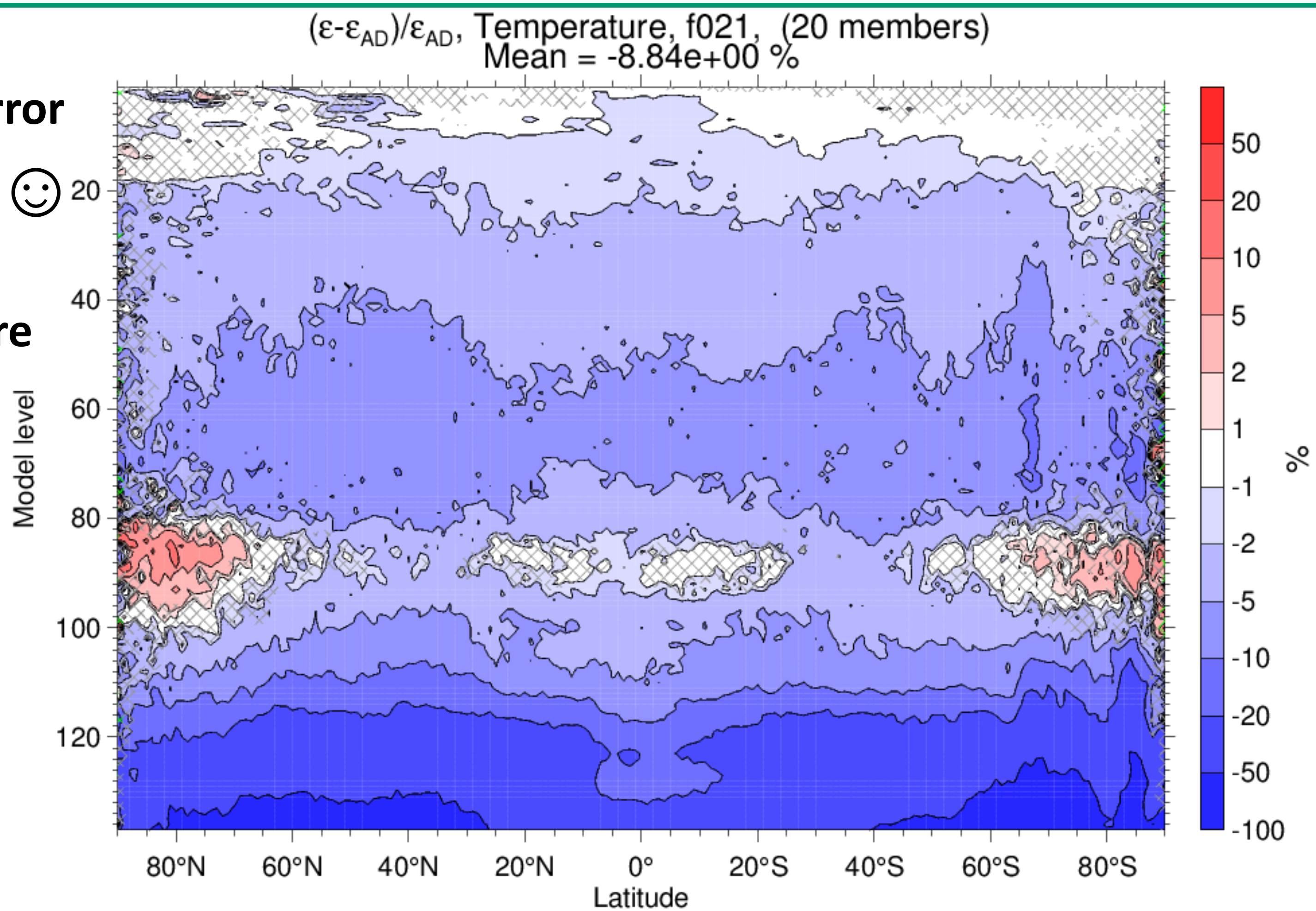
Zonal mean cross-section of change in TL error when TL includes:

VDIF + orog. GWD + SURF

Relative to adiabatic TL run (50-km resol.; 20 runs; after 12h integr.).

Blue = TL error reduction = ☺

Temperature



Impact of linearized physics on TL approximation (2)

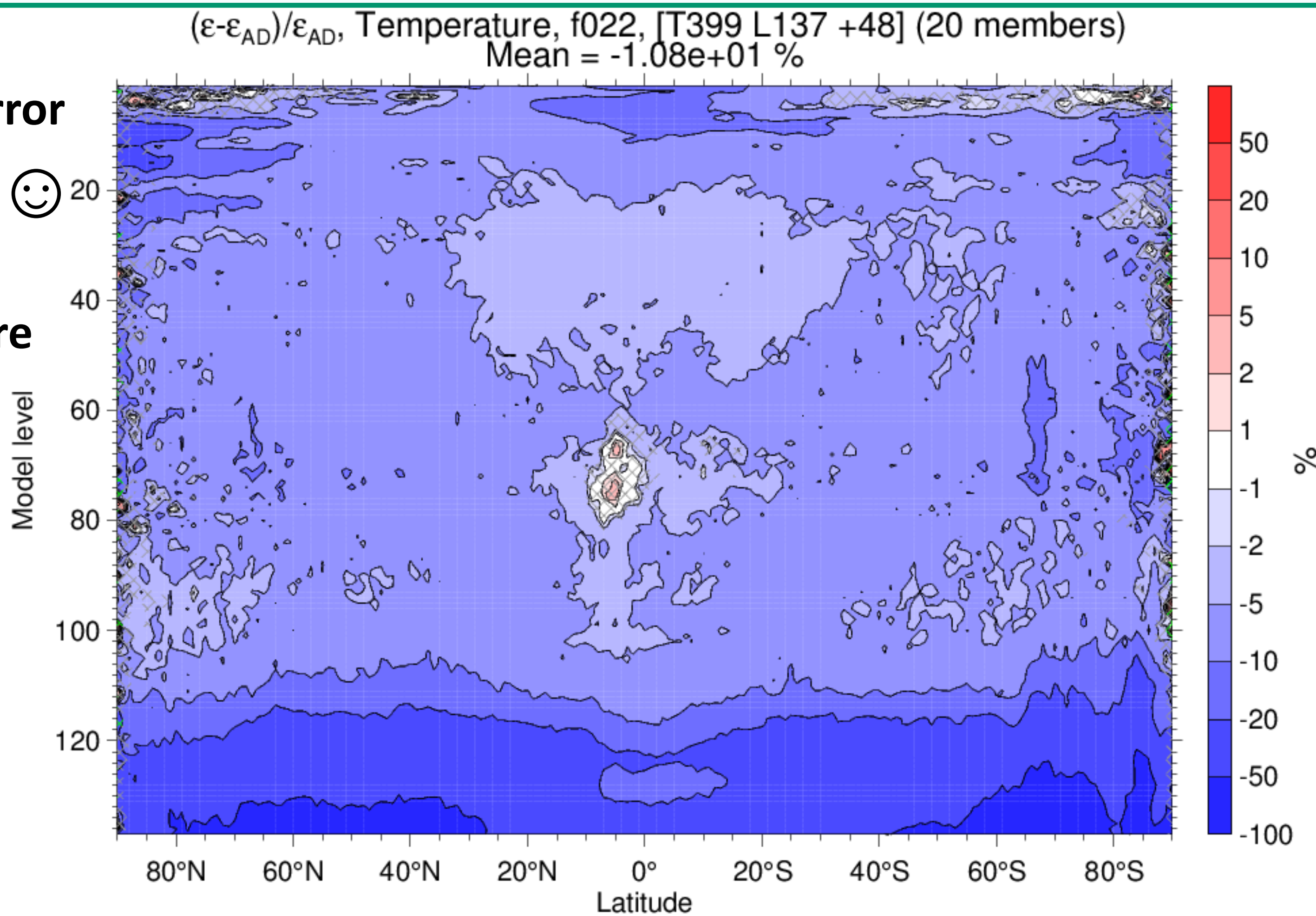
Zonal mean cross-section of change in TL error when TL includes:

VDIF + orog. GWD + SURF + RAD

Relative to adiabatic TL run (50-km resol.; 20 runs; after 12h integr.).

Blue = TL error reduction = ☺

Temperature



Impact of linearized physics on TL approximation (3)

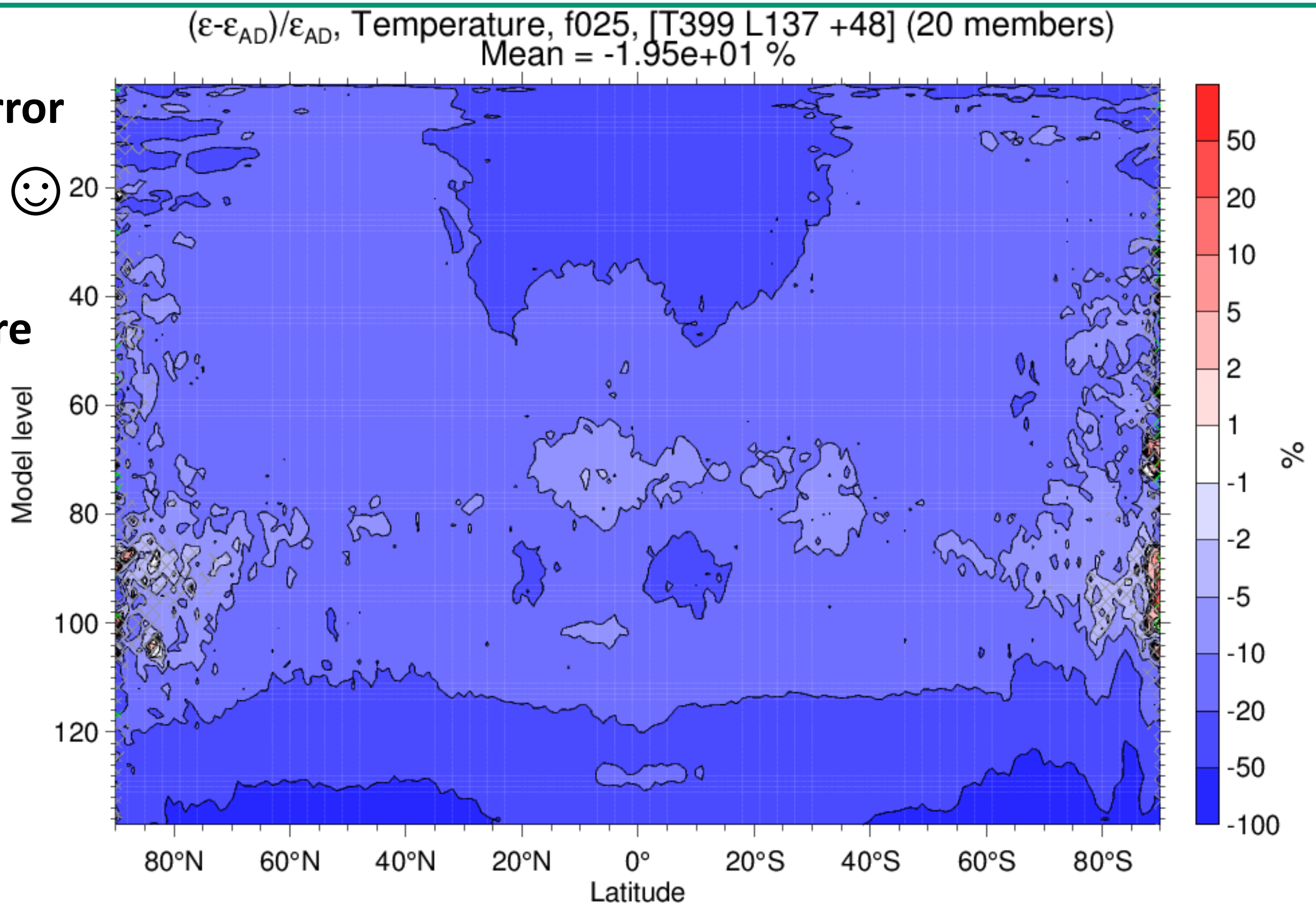
Zonal mean cross-section of change in TL error when TL includes:

VDIF + orog. GWD + SURF + RAD + non-orog GWD + moist physics

Relative to adiabatic TL run (50-km resol.; 20 runs; after 12h integr.).

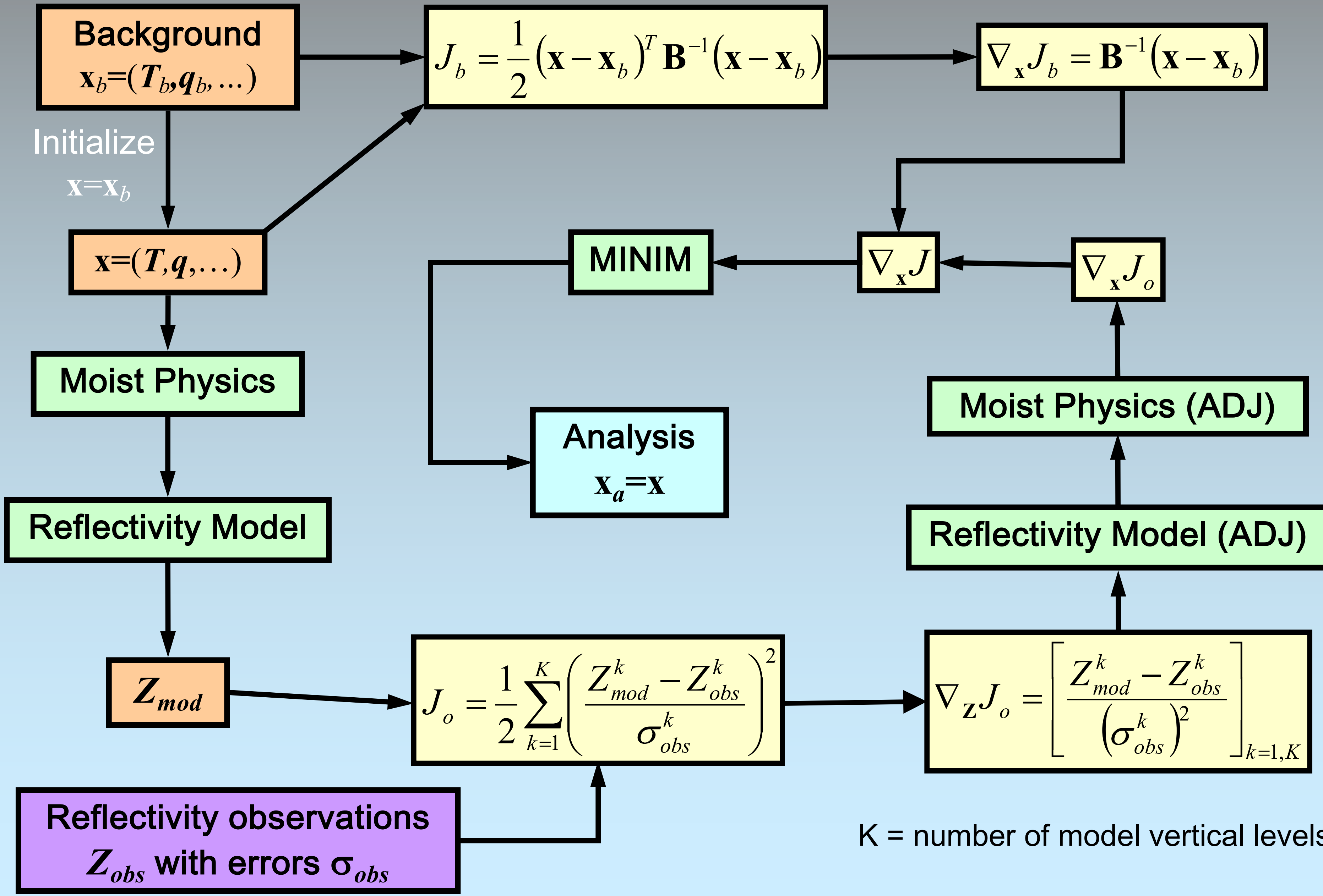
Blue = TL error reduction = ☺

Temperature



Applications

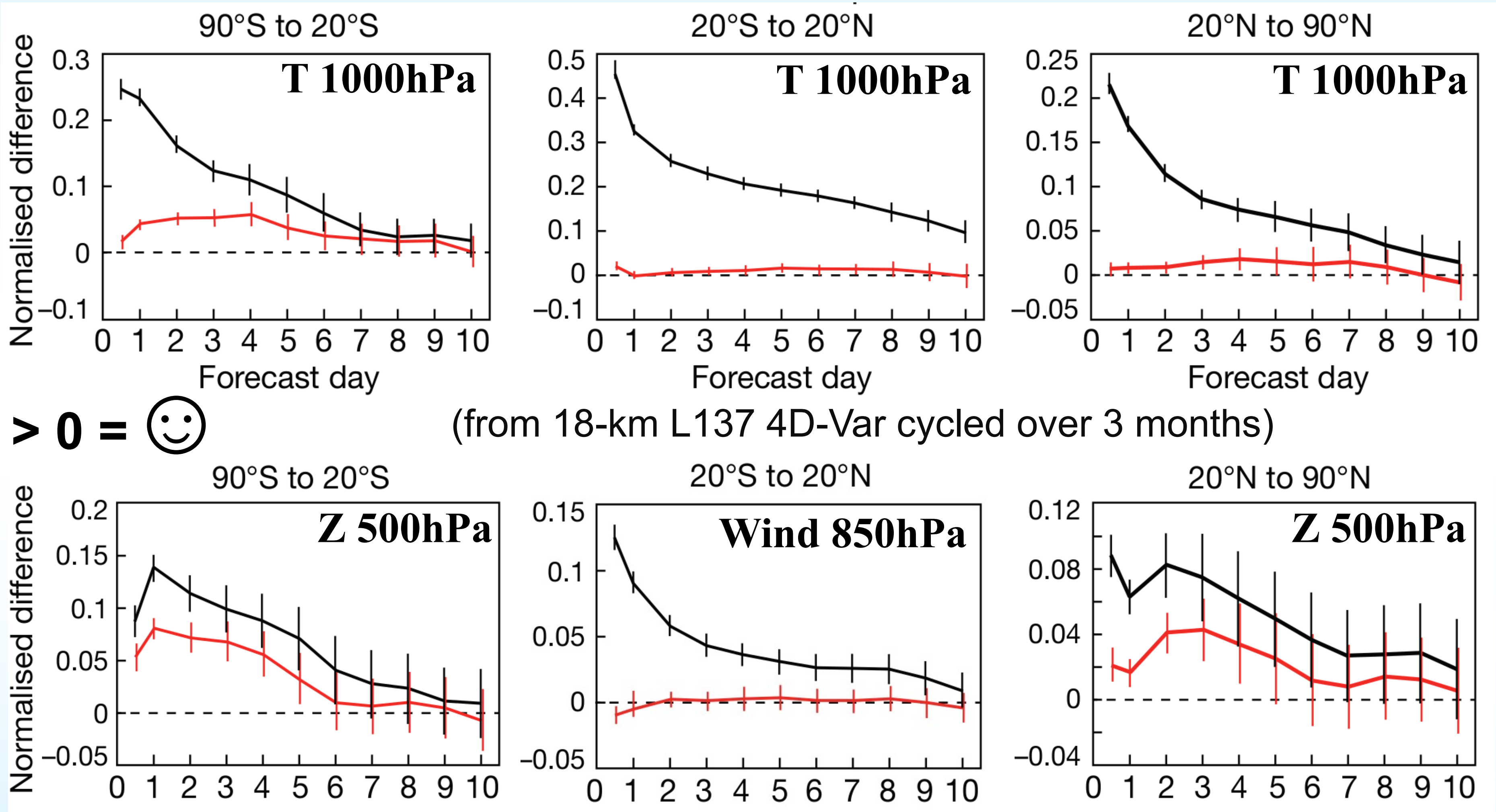
1D-Var with radar reflectivity profiles



K = number of model vertical levels

Impact of linearized physics on analyses and forecasts

Relative change in forecast RMS error due to the inclusion of linearized physics (as well as physics-related observations) in 4D-Var assimilation.



Black line: impact of physics-related observations + linearized physics.
Red line: impact of physics-related observations alone.

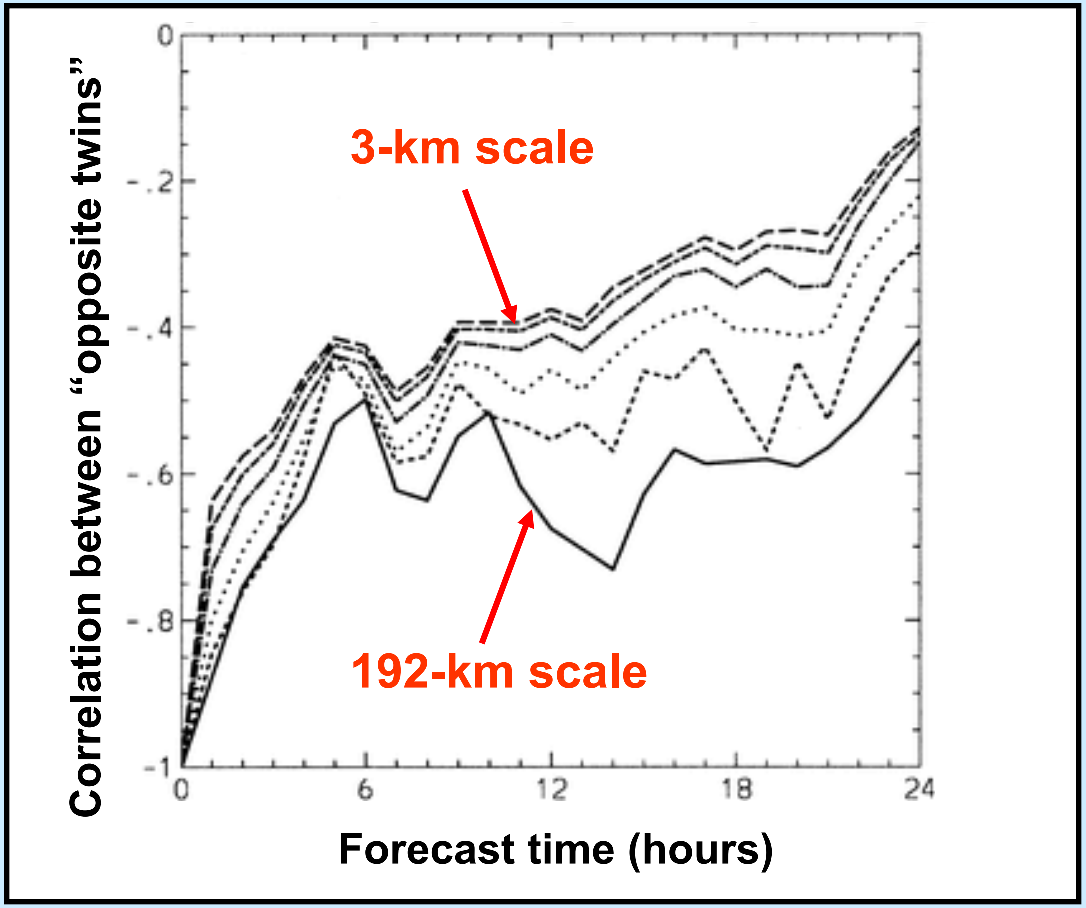
Janisková and Lopez (2023)

ECMWF

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.

Summary and prospects (1)

- **Linearized physical parameterizations have become essential components of variational data assimilation systems (4D-Var):**

→ **Better representation of the evolution of the atmospheric state during the minimization of the cost function (via the adjoint model integration).**

→ **Extraction of information from observations that are strongly affected by physical processes (e.g. by clouds or precipitation).**

- **However, there are some limitations to the LP approach:**

1) Theoretical:

The domain of validity of the linear hypothesis shrinks with increasing resolution and integration length.

2) Technical:

Linearized models require sustained & time-consuming attention:

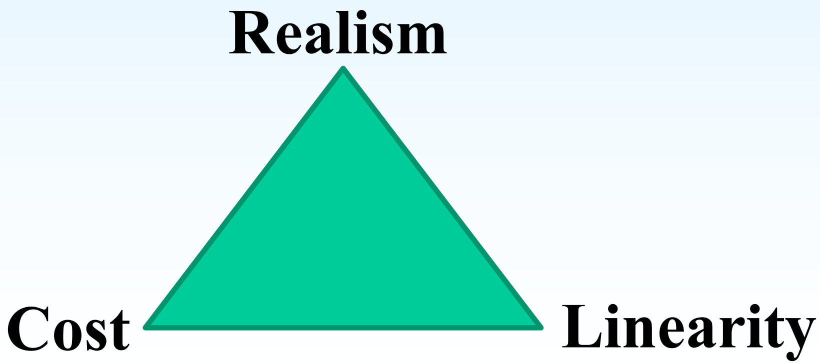
→ **Testing tangent-linear approximation and adjoint code.**

→ **Regularizations / simplifications to eliminate any source of instability.**

→ **Revisions to ensure good match with reference non-linear forecast model.**

Summary and prospects (2)

- In practice, it all comes down to achieving the best compromise between:



- Alternative data assimilation methods exist that do not require the development of linearized code, but so far none of them has been able to outperform 4D-Var, especially in global models:
  → Ensemble Kalman Filter (EnKF; still relies on the linearity assumption),
  → Particle filters (difficult to implement for high-dimensional problems).

- So what is the future of LP?


From a small challenge…


… to a much bigger challenge…
na Wildlife Park / Barcroft I

Summary and prospects (3)

- **Eventually, it might become impractical or even impossible to make LP work efficiently at resolutions of a few kilometres, even if the linearity constraint can be relaxed (e.g. by using shorter 4D-Var window or weak-constraint 4D-Var).**

- **If the current 4D-Var becomes too expensive at very-high resolution, Artificial Intelligence might offer a solution by replacing some of the physical parametrizations with much cheaper equivalents (e.g., based on neural networks). But this is still ongoing research…**

Thank you!

Extra material

A few references…

REFERENCES

- **Variational data assimilation:**

Lorenc, A., 1986, *Quarterly Journal of the Royal Meteorological Society*, **112**, 1177-1194.

Courtier, P. *et al.*, 1994, *Quarterly Journal of the Royal Meteorological Society*, **120**, 1367-1387.

Bouttier, F. and P. Courtier, 1999: Data assimilation concepts and methods, ECMWF lecture notes, available at ***http://www.ecmwf.int/newsevents/training/lecture_notes/LN_DA.html***.

Rabier, F. *et al.*, 2000, *Quarterly Journal of the Royal Meteorological Society*, **126**, 1143-1170.

- **The adjoint technique:**

Errico, R.M., 1997, *Bulletin of the American Meteorological Society*, **78**, 2577-2591.

- **Tangent-linear approximation:**

Errico, R.M. *et al.*, 1993, *Tellus*, **45A**, 462-477.

Errico, R.M., and K. Reader, 1999, *Quarterly Journal of the Royal Meteorological Society*, **125**, 169-195.

Janisková, M. *et al.*, 1999, *Monthly Weather Review*, **127**, 26-45..

- **Physical parameterizations for data assimilation:**

Mahfouf, J.-F., 1999, Tellus, **51A**, 147-166.

Janisková, M. *et al.*, 2002, *Quarterly Journal of the Royal Meteorological Society*, **128**, 1505-1527.

Tompkins, A. M., and M. Janisková, 2004, *Quarterly Journal of the Royal Meteorological Society*, **130**, 2495-2518.

Lopez, P., and E. Moreau, 2005, *Quarterly Journal of the Royal Meteorological Society*, **131**, 409-436.

Janisková, M., and P. Lopez, 2012: Linearized physics for data assimilation at ECMWF, *Technical Memorandum 666*, available from ECMWF, Reading, UK.

REFERENCES

Janisková, M., and P. Lopez, 2023: Linearised physics: the heart of ECMWF's 4D-Var, *ECMWF Newsletter 175*, https://www.ecmwf.int/en/elibrary/81350-newsletter-no-175-spring-2023.

- **Microwave Radiative Transfer Model for data assimilation:**

Bauer, P., May 2002, *Satellite Application Facility for Numerical Weather Prediction NWPSAF*-EC-TR-005, 27 pages.

- **Assimilation of observations affected by precipitation and/or clouds:**

Lopez, P. 2007: *Journal Atmospheric Sciences, Special Issue on the Assimilation of Satellite Cloud and Precipitation Observations in NWP Models*, **64**, 3770-3788 (Review paper on the use of moist physical parameterizations in data assimilation).

Bauer, P., P. Lopez, A. Benedetti, D. Salmond, and E. Moreau, 2006a, *Quarterly Journal of the Royal Meteorological Society*, **132**, 2277-2306.

Bauer, P., P. Lopez, A. Benedetti, D. Salmond, S. Saarinen, and M. Bonazzola, 2006b, *Quarterly Journal of the Royal Meteorological Society*, **132**, 2307-2332.

Fillion, L., and R.M. Errico, 1997, *Monthly Weather Review*, **125**, 2917-2942.

Ducrocq, V., *et al.*, 2000, *Quarterly Journal of the Royal Meteorological Society*, **126**, 3041-3065.

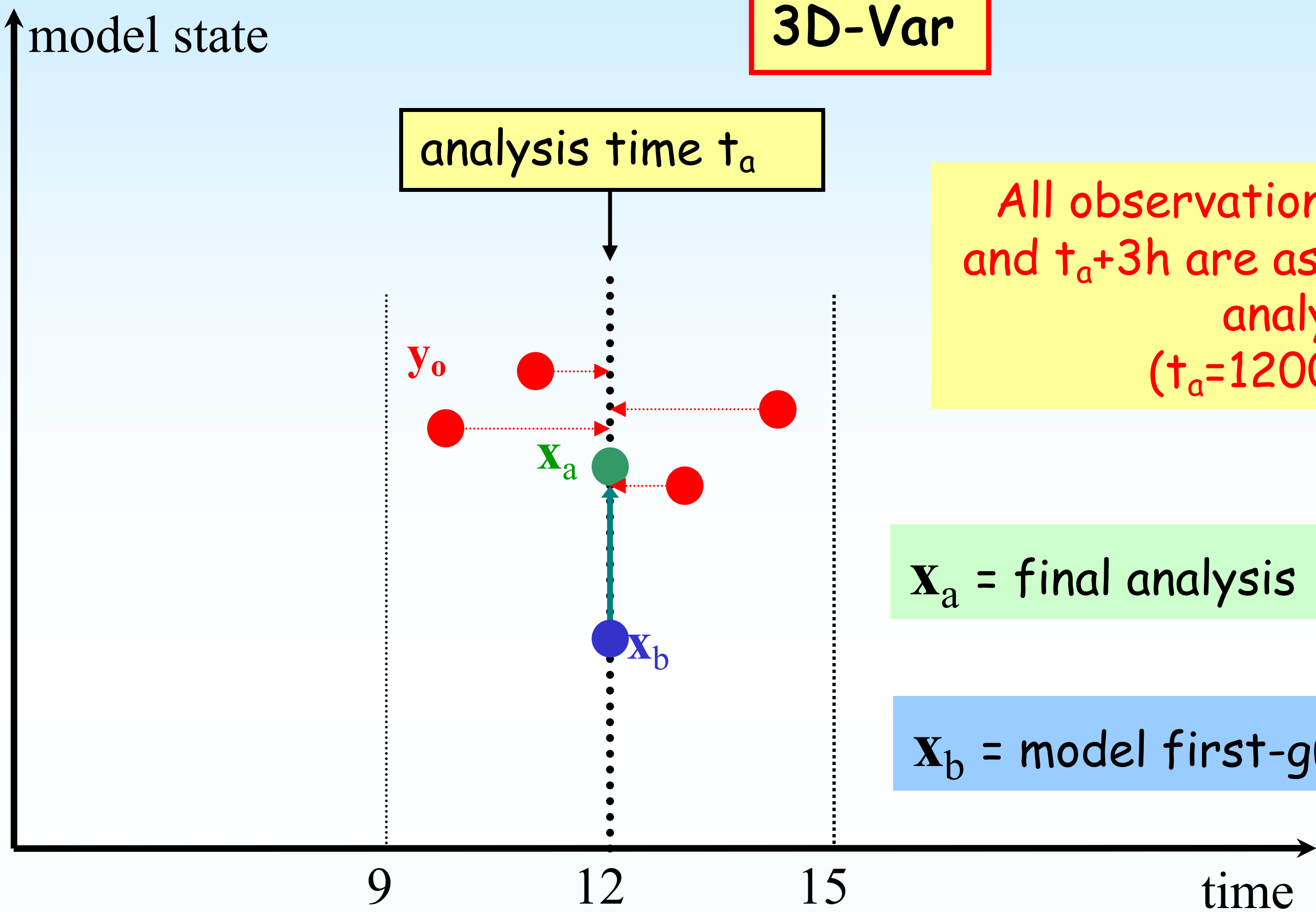
Hou, A. *et al.*, 2000, *Monthly Weather Review*, **128**, 509-537.

Marécal, V., and. J.-F. Mahfouf, 2002, *Monthly Weather Review*, **130**, 43-58.

Marécal, V., and. J.-F. Mahfouf, 2003, *Quarterly Journal of the Royal Meteorological Society*, **129**, 3137-3160.

Some slides of summary…

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)

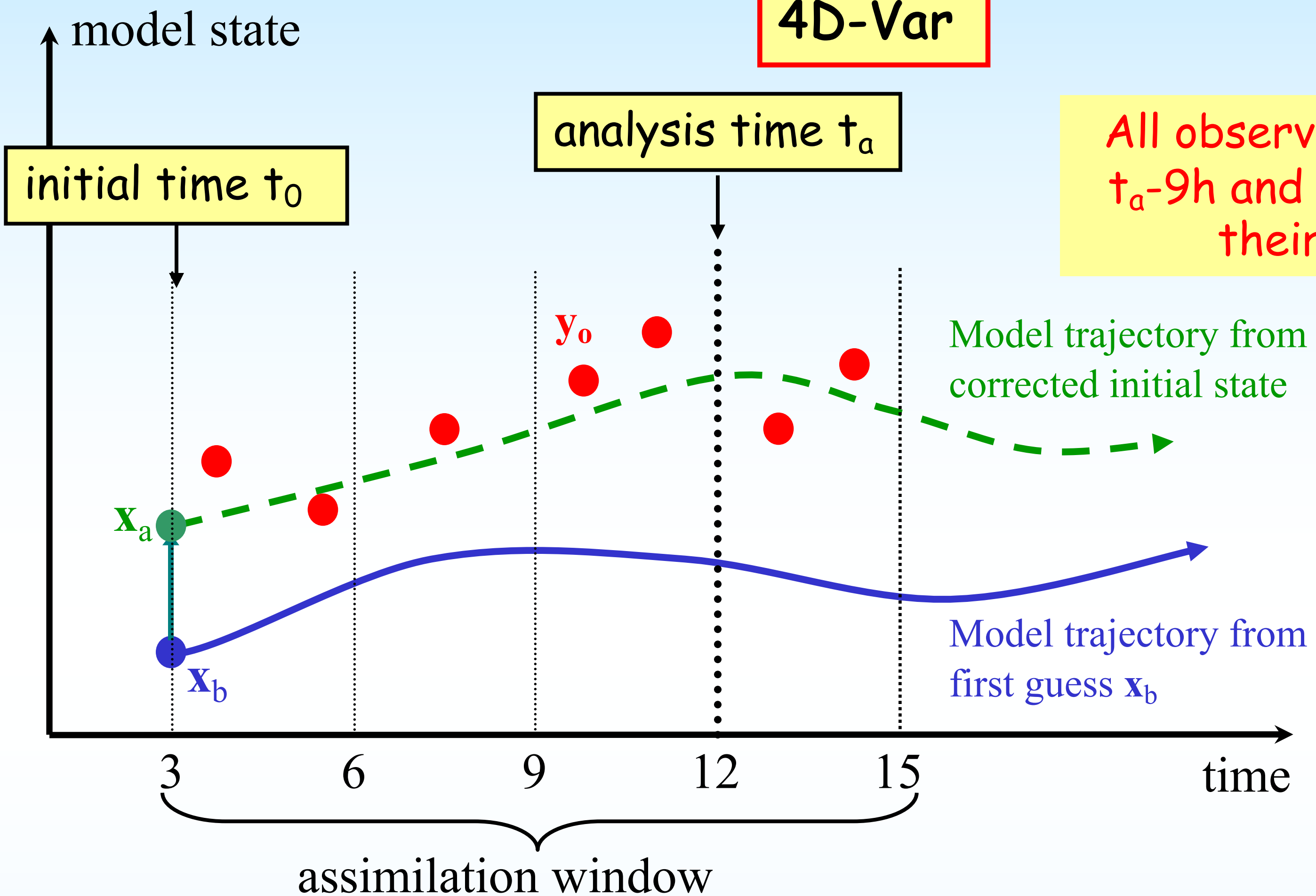
$\mathbf{x}_a$ = final analysis

$\mathbf{x}_b$ = model first-guess

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

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

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

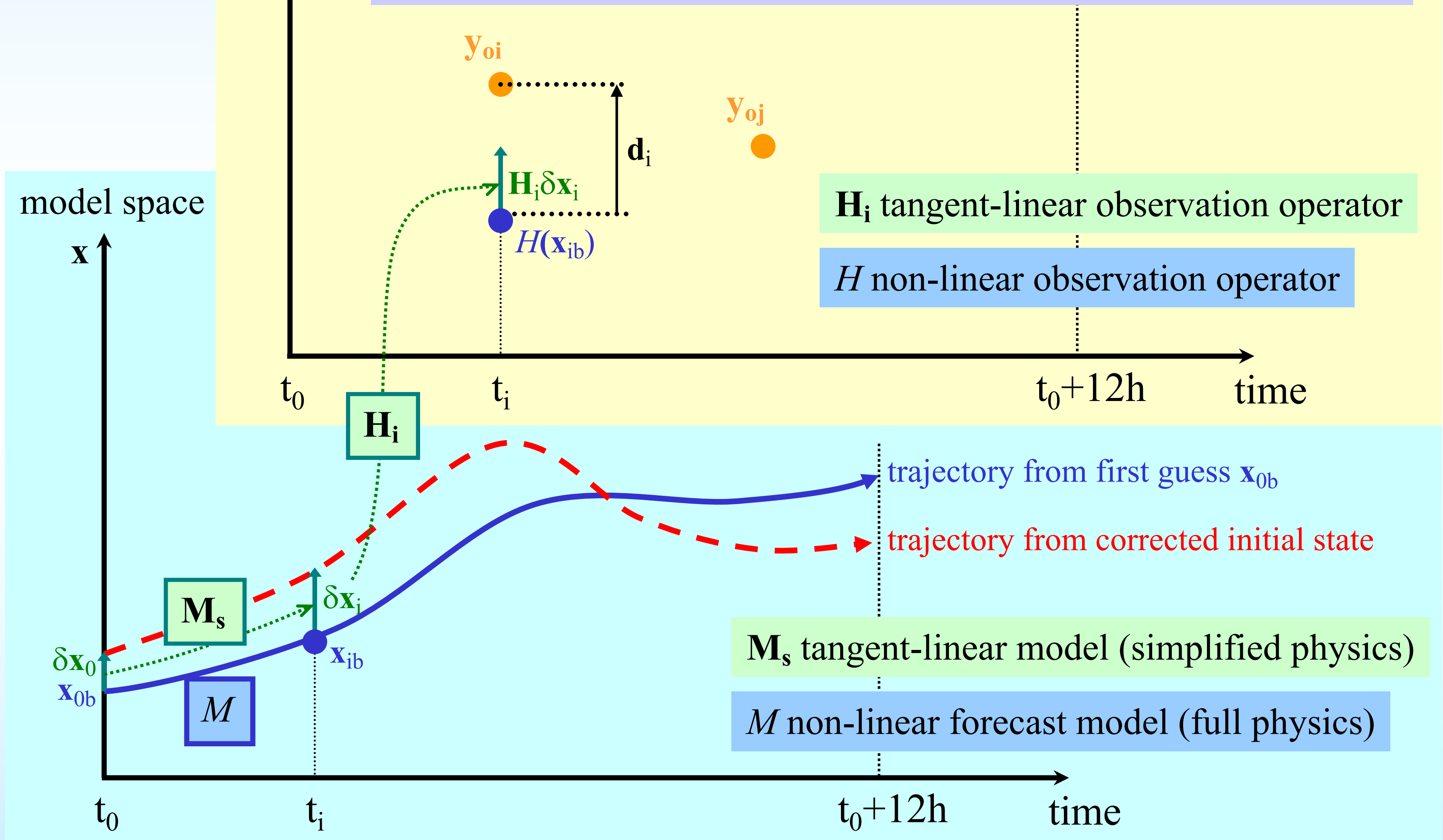
$$\min J = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_b)^T \ldots {}_{oi})^T \mathbf{R}_i^{-1}\left(H_i(M[\mathbf{x}_0]) - \mathbf{y}_{oi}\right)$$

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

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



Summary

- Variational data assimilation relies on some essential assumptions:
  - Gaussian and unbiased model background and observation errors,
  - Quasi-linearity of all operators involved (H, M).

- Given some background fields and a very large set of asynchronous observations available within a certain time window (6 or 12h-long), 4D-Var searches the statistically optimal initial model state $\mathbf{x_0}$ that minimizes the cost function:

$$J(\mathbf{x_0}) = J_{\mathrm{b}}(\mathbf{x_0}) + J_{\mathrm{o}}(HM(\mathbf{x_0}))$$

- The calculation of $\nabla_{\mathbf{x0}} J$ requires the coding of tangent-linear and adjoint versions of the observation operator H and of the full nonlinear forecast model M (including physical parameterizations).

- The tangent-linear and adjoint forecast models, $\mathbf{M}$ and $\mathbf{M}^{\mathrm{T}}$, are usually based on a simplified version of the full nonlinear model, M, to reduce computational cost in the iterative minimization and to avoid nonlinearities.

Summary

- The aim of data assimilation is to produce a statistically optimal model state (the **analysis**) which can be used to initialize a forecast model.

- In variational DA, this is achieved by minimizing a **cost function**, J, that measures the distance to the **model background** and **observations**, weighted by their respective **error statistics.**

  In 3D-Var:

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

- **Parameterizations** are needed during the minimization to convert the model control variables (T,q,u,v,P_s) into observed equivalents (e.g. reflectivities, radiances,…) ("**observation operator**" H).

- <u>Fundamental assumptions:</u>

  - **Background and observation errors** are **Gaussian** and **unbiased**.

  - Observation operator H is **not too non-linear**.

Summary

- The aim of a data assimilation system is to produce a statistically optimal model state (the analysis) that can be used to initialize a forecast model.
- In variational DA this is achieved by minimizing iteratively a cost function (J) that measures the distance to the model background and observations, weighted by their respective error statistics (Gaussian and unbiased).
- Parameterizations are needed during the minimization to:

  - convert the model variables (T,q,u,v,P_s) into observed equivalents (e.g. reflectivities, radiances,...) (observation operator H),
  - evolve the model state from analysis time to observation time (4D-Var).

- The tangent-linear and adjoint versions of these usually simplified parameterizations must be coded, tested, and some regularization is usually needed to eliminate discontinuities/non-linearities.
- The adjoint version of the parameterizations is needed to compute the gradient of the cost function with respect to the initial model state, $\mathbf{x}$:

$$\nabla_{\mathbf{x}_0} J = \mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + \mathbf{M}[t_i,t_0]^T \mathbf{H}^T \nabla_{\mathbf{y}} J_o \quad \text{with } \nabla_{\mathbf{y}} J_o = \mathbf{R}^{-1}(H(\mathbf{x}) - \mathbf{y}_o)$$

A few examples/exercises…

A simple analysis problem

Exercise

- 6-hour forecast of 2m temperature produced by the model:
  $\mathbf{x}_b$ with a standard deviation of forecast error σ_b

- observation of 2m temperature:
  $\mathbf{y}_o$ with a standard deviation of observation error σ_o

- The best estimate of the 2m temperature (analysis) minimizes the departure from the model first-guess and from the observation according to their relative accuracies:

$$\mathcal{J}(\mathbf{x}) = \frac{1}{2}\left(\frac{\mathbf{x}-\mathbf{x}_b}{\sigma_b}\right)^2 + \frac{1}{2}\left(\frac{\mathbf{x}-\mathbf{y}_o}{\sigma_o}\right)^2$$

Since the analysis $\mathbf{x}_a$ minimizes the cost function, then

$$\frac{\partial \mathcal{J}}{\partial \mathbf{x}}(\mathbf{x}_a) = 0$$

Analysis state can be written as:

$$\mathbf{x}_a = \mathbf{x}_b + \alpha(\mathbf{y}_o - \mathbf{x}_b)$$

A simple analysis problem

Exercise

- **Problem:**
  - Find the coefficient α.
  - Show that the variance of the analysis error is:

$$\frac{1}{{\sigma_a}^2} = \frac{1}{{\sigma_b}^2} + \frac{1}{{\sigma_o}^2}$$

(Note: $\sigma^2 = \overline{(\mathbf{x} - \mathbf{x}_t)^2}$, where $\mathbf{x}_t$ is the unknown true state).

A simple analysis problem
Solution

- Since the analysis $\mathbf{x}_a$ minimizes the cost function, then

$$\frac{\partial \mathcal{J}}{\partial \mathbf{x}}(\mathbf{x}_a) = 0$$

$$\frac{\partial \mathcal{J}}{\partial \mathbf{x}} = \frac{\mathbf{x} - \mathbf{x}_b}{{\sigma_b}^2} + \frac{\mathbf{x} - \mathbf{y}_o}{{\sigma_o}^2} = 0$$

$$\frac{\mathbf{x}{\sigma_o}^2 - \mathbf{x}_b{\sigma_o}^2 + \mathbf{x}{\sigma_b}^2 - \mathbf{y}_o{\sigma_b}^2}{{\sigma_b}^2{\sigma_o}^2} = 0$$

$$(*) \qquad \mathbf{x}({\sigma_o}^2 + {\sigma_b}^2) = \mathbf{x}_b{\sigma_o}^2 + \mathbf{y}_o{\sigma_b}^2$$

$$\mathbf{x} = \frac{\mathbf{x}_b{\sigma_o}^2 + \mathbf{y}_o{\sigma_b}^2 - \mathbf{x}_b{\sigma_b}^2 + \mathbf{x}_b{\sigma_b}^2}{{\sigma_o}^2 + {\sigma_b}^2}$$

$$\mathbf{x} = \frac{\mathbf{x}_b({\sigma_o}^2 + {\sigma_b}^2) + {\sigma_b}^2(\mathbf{y}_o - \mathbf{x}_b)}{{\sigma_o}^2 + {\sigma_b}^2}$$

$$\mathbf{x} = \mathbf{x}_b + \boxed{\frac{{\sigma_b}^2}{{\sigma_o}^2 + {\sigma_b}^2}}(\mathbf{y}_o - \mathbf{x}_b)$$

α

A simple analysis problem

Solution

- Analysis error:
  starting from equation (*) one gets

$$
\begin{aligned}
\mathbf{x}_a &= \mathbf{x}_b \frac{{\sigma_o}^2}{{\sigma_o}^2 + {\sigma_b}^2} + \mathbf{y}_o \frac{{\sigma_b}^2}{{\sigma_o}^2 + {\sigma_b}^2} \\
\mathbf{x}_a - \mathbf{x}_t &= (\mathbf{x}_b - \mathbf{x}_t) \frac{{\sigma_o}^2}{{\sigma_o}^2 + {\sigma_b}^2} + (\mathbf{y}_o - \mathbf{x}_t) \frac{{\sigma_b}^2}{{\sigma_o}^2 + {\sigma_b}^2} \\
\overline{(\mathbf{x}_a - \mathbf{x}_t)^2} &= \overline{(\mathbf{x}_b - \mathbf{x}_t)^2} \frac{{\sigma_o}^4}{({\sigma_o}^2 + {\sigma_b}^2)^2} + \overline{(\mathbf{y}_o - \mathbf{x}_t)^2} \frac{{\sigma_b}^4}{({\sigma_o}^2 + {\sigma_b}^2)^2} \\
&+ \overline{(\mathbf{x}_b - \mathbf{x}_t)(\mathbf{y}_o - \mathbf{x}_t)} \frac{{\sigma_o}^2 {\sigma_b}^2}{({\sigma_o}^2 + {\sigma_b}^2)^2}
\end{aligned}
$$

Since background and observation errors are assumed to be uncorrelated:

$$
Cov(\mathbf{x}_b, \mathbf{y}_o) = \overline{(\mathbf{x}_b - \mathbf{x}_t)(\mathbf{y}_o - \mathbf{x}_t)} = 0
$$

which gives

$$
\begin{aligned}
{\sigma_a}^2 &= \frac{{\sigma_b}^2 {\sigma_o}^4}{({\sigma_o}^2 + {\sigma_b}^2)^2} + \frac{{\sigma_b}^4 {\sigma_o}^2}{({\sigma_o}^2 + {\sigma_b}^2)^2} \\
{\sigma_a}^2 &= \frac{{\sigma_b}^2 {\sigma_o}^2}{{\sigma_o}^2 + {\sigma_b}^2} \quad \Longleftrightarrow \quad \boxed{\frac{1}{{\sigma_a}^2} = \frac{1}{{\sigma_b}^2} + \frac{1}{{\sigma_o}^2}}
\end{aligned}
$$

1D-Var assimilation of physical fluxes

Example

- observation operator = physical parametrization
  - example: thermal radiation at the surface *(Brunt, 1934)*

$$R_L = \sigma T^4 (a + b\sqrt{e})$$

    where T is the screen level temperature and e is the water vapour pressure

- model temperature and humidity (T_b, e_b) can be modified to better match an observation of thermal radiation $R_{L\,o}$

- the optimal values of T and e minimize the following cost function:

$$\mathcal{J}(T,e) = \frac{1}{2}\left(\frac{T - T_b}{\sigma_{T\,b}}\right)^2 + \frac{1}{2}\left(\frac{e - e_b}{\sigma_{e\,b}}\right)^2 + \frac{1}{2}\left(\frac{R_L - R_{L\,o}}{\sigma_o}\right)^2$$

gradient of the cost function:

$$\frac{\partial \mathcal{J}}{\partial T} = \frac{T - T_b}{\sigma_{T\,b}^2} + \frac{\partial R_L}{\partial T}\left(\frac{R_L - R_{L\,o}}{\sigma_o^2}\right)$$

$$\frac{\partial \mathcal{J}}{\partial e} = \frac{e - e_b}{\sigma_{e\,b}^2} + \frac{\partial R_L}{\partial e}\left(\frac{R_L - R_{L\,o}}{\sigma_o^2}\right)$$

1D-Var assimilation of physical fluxes

Example

- tangent-linear operator:

$$\delta R_L = \left(\frac{\partial R_L}{\partial T} \quad \frac{\partial R_L}{\partial e} \right) \cdot \begin{pmatrix} \delta T \\ \delta e \end{pmatrix}$$

- adjoint of the tangent-linear operator:

$$\left(\frac{\partial J_o}{\partial T} \quad \frac{\partial J_o}{\partial e} \right) = \begin{pmatrix} \frac{\partial R_L}{\partial T} \\ \frac{\partial R_L}{\partial e} \end{pmatrix} \cdot \left(\frac{\partial J_o}{\partial R_L} \right)$$

with $\frac{\partial R_L}{\partial T} = 4\sigma T^3(a + b\sqrt{e})$ and $\frac{\partial R_L}{\partial e} = \frac{b\sigma T^4}{2\sqrt{e}}$

EXERCISE 2

- write tangent linear (TL) and adjoint (AD) code of the following non-linear (NL) code (FORTRAN 90)

```
SUBROUTINE Longwave_Radiation (EA, TA, RL)

! Longwave radiation at the surface (RL in Watts/m2)
! Empirical expression from Brunt (1934) depending upon
! TA = air temperature (K)
! EA = water vapour pressure (hPa)
!
! Non-linear routine
! ------------------

IMPLICIT NONE
REAL , INTENT(IN)  :: EA, TA
REAL , INTENT(OUT) :: RL
REAL , PARAMETER   :: A=0.75, B=0.003
REAL , PARAMETER   :: SIGMA=5.67E-8
REAL               :: ZEMIS

ZEMIS = A+B*SQRT(EA)
RL    = ZEMIS*SIGMA*TA**4

END SUBROUTINE Longwave_Radiation
```

EXERCISE 2 - solution

- tangent linear code

```
SUBROUTINE Longwave_Radiation_TL (EA5, TA5, RL5, EA, TA, RL)

! Longwave radiation at the surface (RL in Watts/m2)
! Empirical expression from Brunt (1934) depending upon
! TA = air temperature (K)
! EA = water vapour pressure (hPa)
!
! Tangent-linear routine
! ----------------------

IMPLICIT NONE
REAL , INTENT(IN)  :: EA5, TA5         ! Trajectory
REAL , INTENT(OUT) :: RL5              ! Trajectory
REAL , INTENT(IN)  :: EA, TA           ! Perturbation
REAL , INTENT(OUT) :: RL               ! Perturbation
REAL , PARAMETER   :: A=0.75, B=0.003
REAL , PARAMETER   :: SIGMA=5.67E-8
REAL               :: ZEMIS5, ZEMIS

ZEMIS5 = A+B*SQRT(EA5)
ZEMIS  = B/(2.*SQRT(EA5))*EA
RL5    = ZEMIS5*SIGMA*TA5**4
RL     = ZEMIS *SIGMA*TA5**4 + 4.*ZEMIS5*SIGMA*TA5**3*TA

END SUBROUTINE Longwave_Radiation_TL
```

EXERCISE 2 - solution

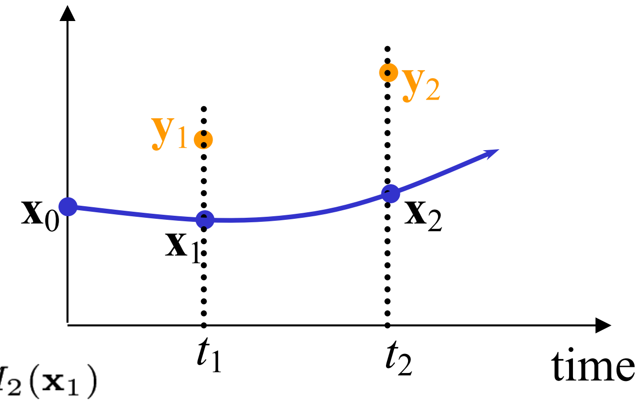
- adjoint code

```
 SUBROUTINE Longwave_Radiation_AD (EA5, TA5, RL5, EA, TA, RL)

! Longwave radiation at the surface (RL in Watts/m2)
! Empirical expression from Brunt (1934) depending upon
! TA = air temperature (K)
! EA = water vapour pressure (hPa)
!
! Adjoint routine
! ---------------

 IMPLICIT NONE
 REAL , INTENT(IN)  :: EA5, TA5         ! Trajectory
 REAL , INTENT(OUT) :: RL5              ! Trajectory
 REAL , INTENT(IN)  :: EA, TA           ! Perturbation
 REAL , INTENT(OUT) :: RL               ! Perturbation
 REAL , PARAMETER   :: A=0.75, B=0.003
 REAL , PARAMETER   :: SIGMA=5.67E-8
 REAL               :: ZEMIS5, ZEMIS

! Trajectory computations

 ZEMIS5 = A+B*SQRT(EA5)
 RL5    = ZEMIS5*SIGMA*TA5**4

! Initialization of local variables
```

```
ZEMIS = 0.

! Adjoint computation

TA = TA + 4.*ZEMIS5*SIGMA*TA5**3*RL
ZEMIS = ZEMIS + SIGMA*TA5**4*RL
RL = 0.
EA = EA + B/(2.*SQRT(EA5))*ZEMIS
ZEMIS = 0.

END SUBROUTINE Longwave_Radiation_AD
```

A simple 4D-Var analysis problem

Exercise



- Observations $\mathbf{y}_1$ and $\mathbf{y}_2$ at time t_1 and t_2
- Model first guess $\mathbf{x}_1$ and $\mathbf{x}_2$ at time t_1 and t_2
- Time evolution from the initial time t_0:

$$\mathbf{x}_1 = M_1(\mathbf{x}_0) \quad \text{and} \quad \mathbf{x}_2 = M_2(\mathbf{x}_1)$$

- Cost function:

$$\mathcal{J}(\mathbf{x}_0) = \frac{1}{2}\left(\frac{\mathbf{x}_1 - \mathbf{y}_1}{\sigma_1}\right)^2 + \frac{1}{2}\left(\frac{\mathbf{x}_2 - \mathbf{y}_2}{\sigma_2}\right)^2 = \mathcal{J}_1 + \mathcal{J}_2$$

- **Problem:**
  Estimate the gradient of $\mathcal{J}$ with respect to the initial state $\mathbf{x}_0$.

A simple 4D-Var analysis problem

Solution

- At time t_2:

$$\frac{\partial \mathcal{J}_2}{\partial \mathbf{x}_2} = \frac{\mathbf{x}_2 - \mathbf{y}_2}{\sigma_2^2}$$

- At time t_1:

$$\frac{\partial \mathcal{J}_2}{\partial \mathbf{x}_1} = \mathrm{M}_2{}^T \left(\frac{\partial \mathcal{J}_2}{\partial \mathbf{x}_2} \right) = \mathrm{M}_2{}^T \left(\frac{\mathbf{x}_2 - \mathbf{y}_2}{\sigma_2^2} \right)$$

$$\frac{\partial \mathcal{J}_1}{\partial \mathbf{x}_1} = \frac{\mathbf{x}_1 - \mathbf{y}_1}{\sigma_1^2}$$

- At time t_0:

$$\frac{\partial \mathcal{J}_2}{\partial \mathbf{x}_0} = \mathrm{M}_1{}^T \left(\frac{\partial \mathcal{J}_2}{\partial \mathbf{x}_1} \right) = \mathrm{M}_1{}^T \left[\mathrm{M}_2{}^T \left(\frac{\mathbf{x}_2 - \mathbf{y}_2}{\sigma_2^2} \right) \right]$$

$$\frac{\partial \mathcal{J}_1}{\partial \mathbf{x}_0} = \mathrm{M}_1{}^T \left(\frac{\partial \mathcal{J}_1}{\partial \mathbf{x}_1} \right) = \mathrm{M}_1{}^T \left(\frac{\mathbf{x}_1 - \mathbf{y}_1}{\sigma_1^2} \right)$$

- Finally:

$$\frac{\partial \mathcal{J}}{\partial \mathbf{x}_0} = \frac{\partial \mathcal{J}_1}{\partial \mathbf{x}_0} + \frac{\partial \mathcal{J}_2}{\partial \mathbf{x}_0} = \mathrm{M}_1{}^T \left(\frac{\mathbf{x}_1 - \mathbf{y}_1}{\sigma_1^2} \right) + \mathrm{M}_1{}^T \left[\mathrm{M}_2{}^T \left(\frac{\mathbf{x}_2 - \mathbf{y}_2}{\sigma_2^2} \right) \right]$$

Supporting slides

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:**

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

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 $<,>$,

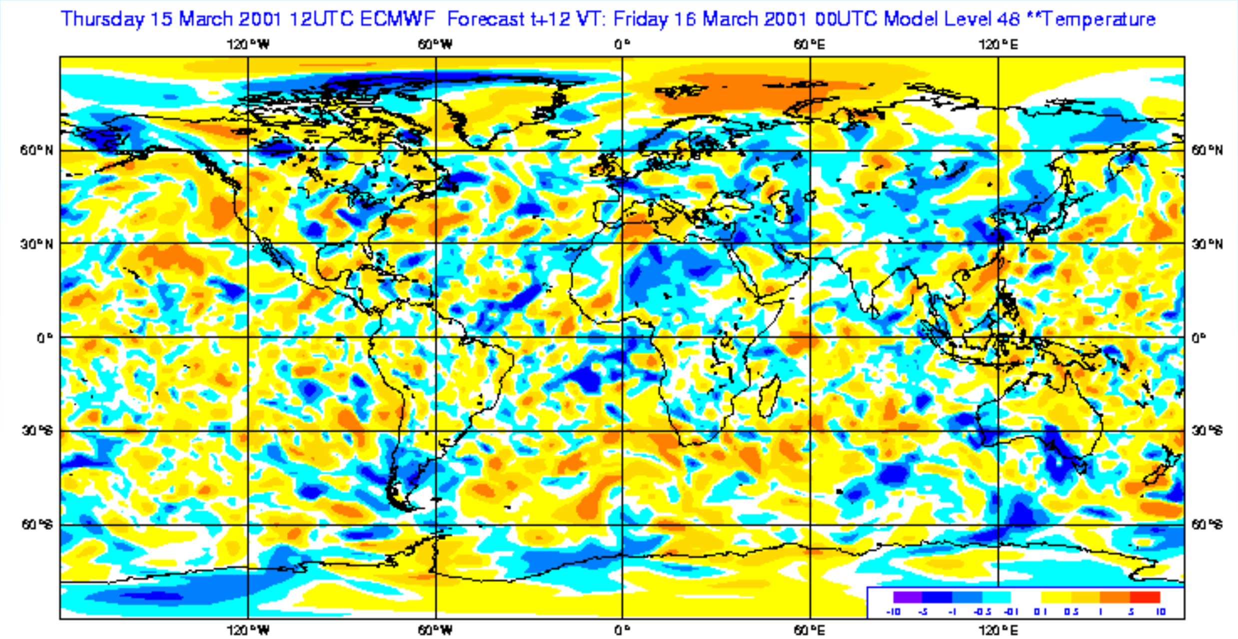
$$\forall \mathbf{x}, \forall \mathbf{y} \qquad < M'\mathbf{x}, \mathbf{y} > = < \mathbf{x}, \mathbf{M}^*\mathbf{y} >$$

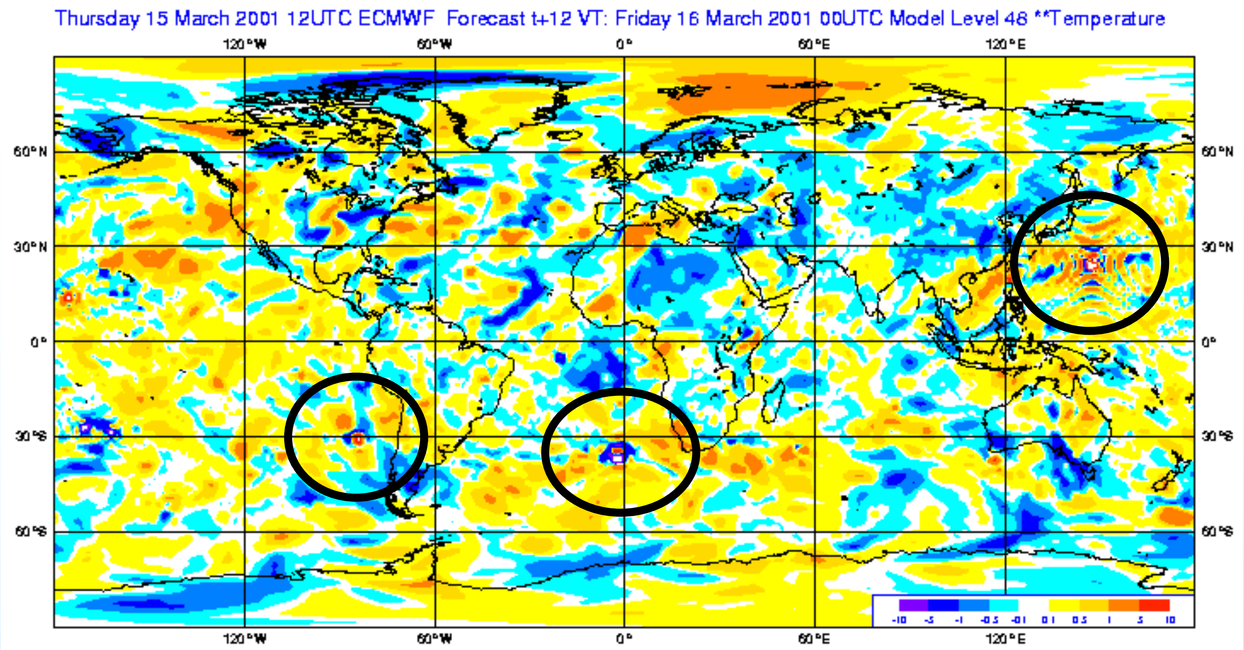
<u>Remarks:</u>

– 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.

Importance of regularization to prevent instabilities in tangent-linear model





12-hour ECMWF model integration (T159 L60)

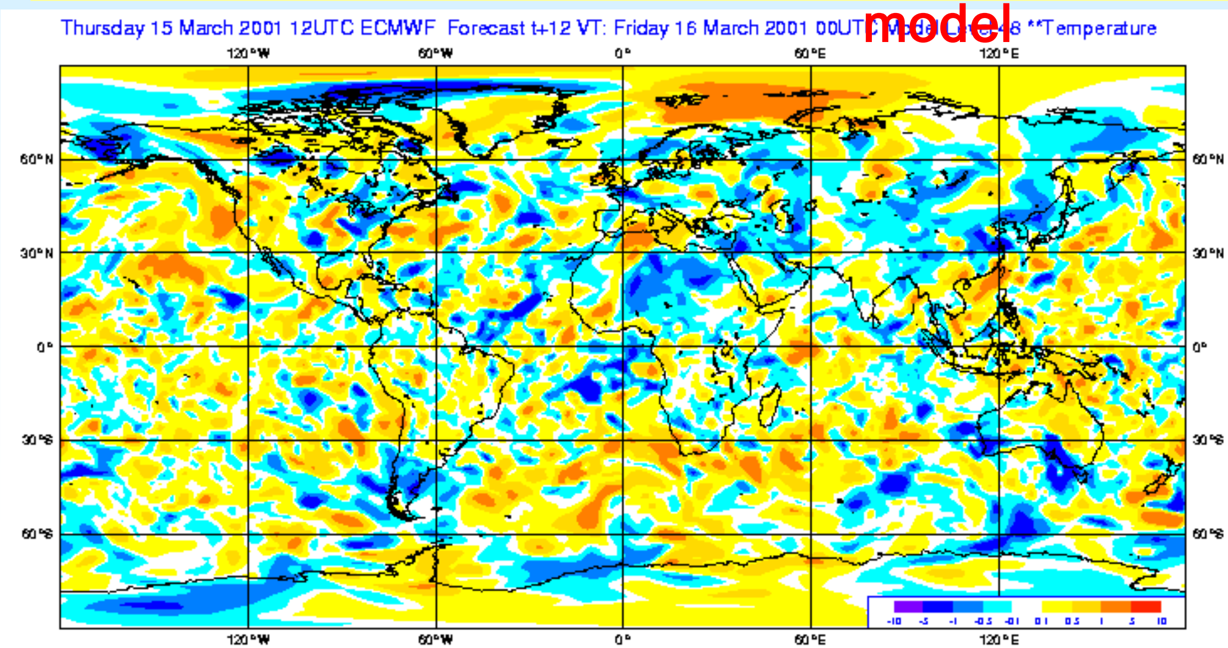
Temperature on level 48 (approx. 850 hPa)

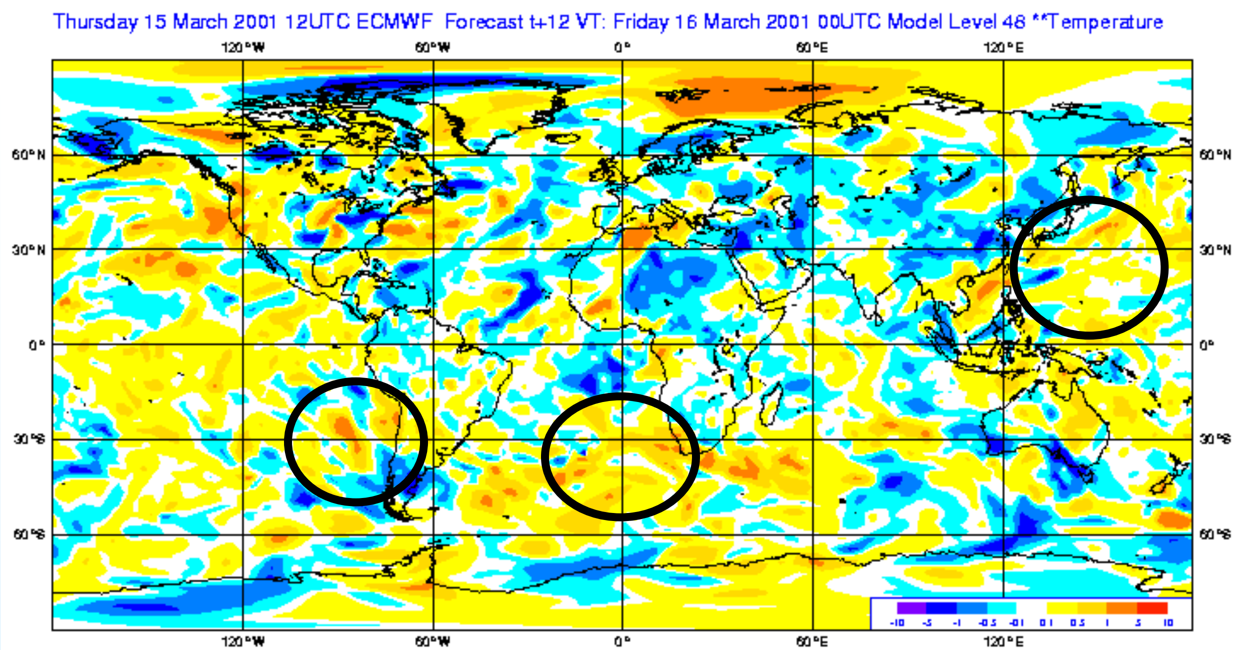
Finite difference between two non-linear model integrations

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)

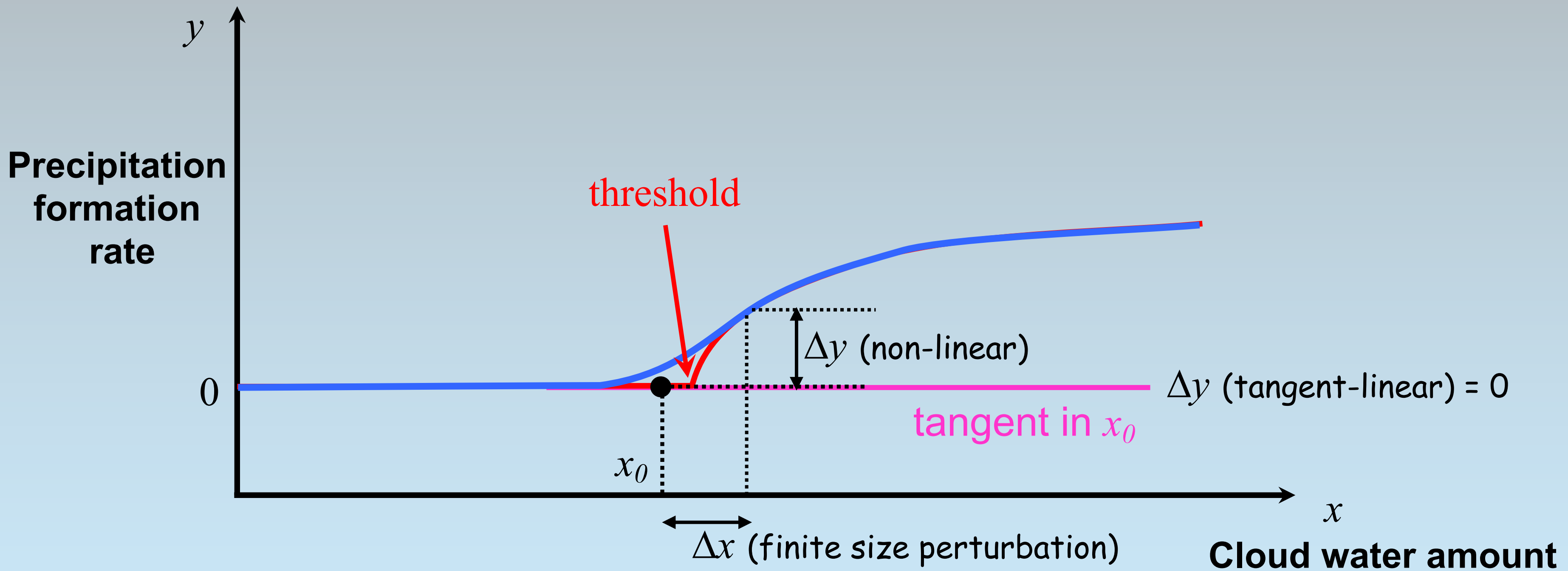
Temperature on level 48 (approx. 850 hPa)

Finite difference between two non-linear model integrations

Corresponding perturbations evolved with tangent-linear model

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

Linearity issue



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

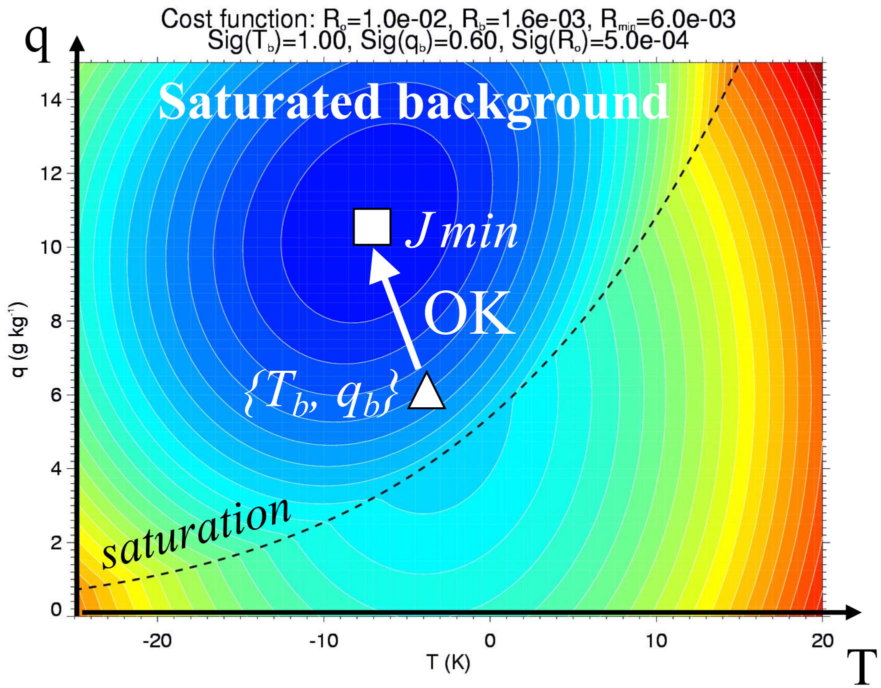
Illustration of discontinuity effect on cost function shape:

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

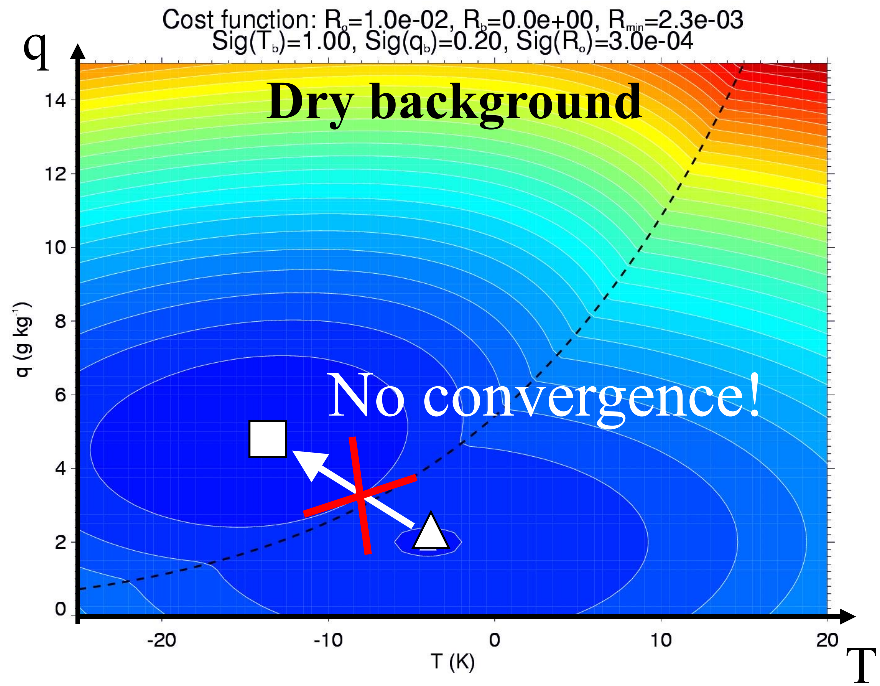
Simple parametrization of rain rate:

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

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



Single minimum of cost function



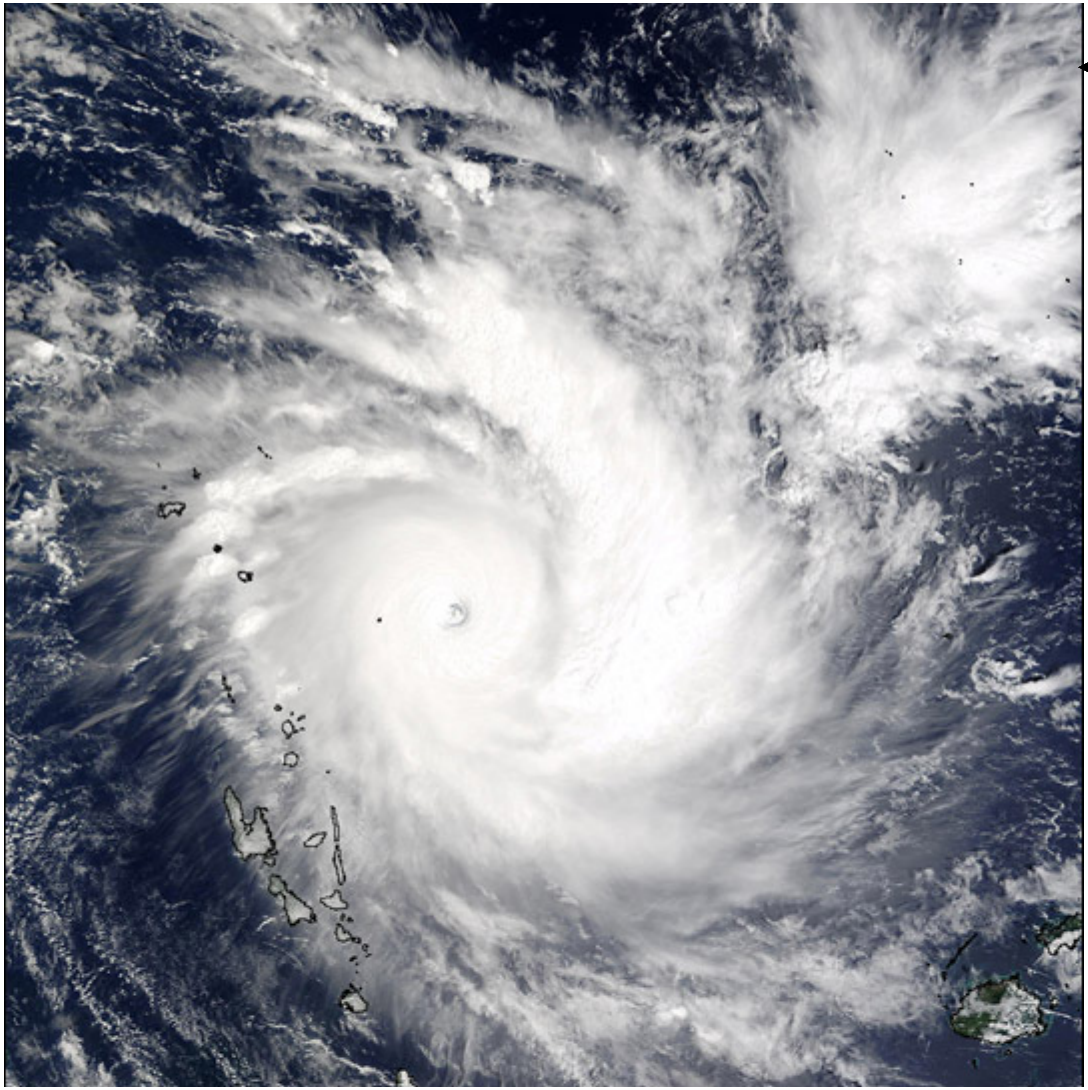
Several local minima of cost function

A short list of existing LP packages used in operational DA

- Tsuyuki (1996): Kuo-type convection and large-scale condensation schemes (FSU 4D-Var).
- Mahfouf (1999): full set of simplified physical parametrizations (gravity wave drag currently used in ECMWF operational 4D-Var and EPS).
- Janisková et al. (1999): full set of simplified physical parametrizations (Météo-France operational 4D-Var).
- Janisková et al. (2002): linearized radiation (ECMWF 4D-Var).
- Lopez (2002): simplified large-scale condensation and precipitation scheme (Météo-France).
- Tompkins and Janisková (2004): simplified large-scale condensation and precipitation scheme (ECMWF).
- Lopez and Moreau (2005): simplified mass-flux convection scheme (ECMWF).
- Mahfouf (2005): simplified Kuo-type convection scheme (Environment Canada).

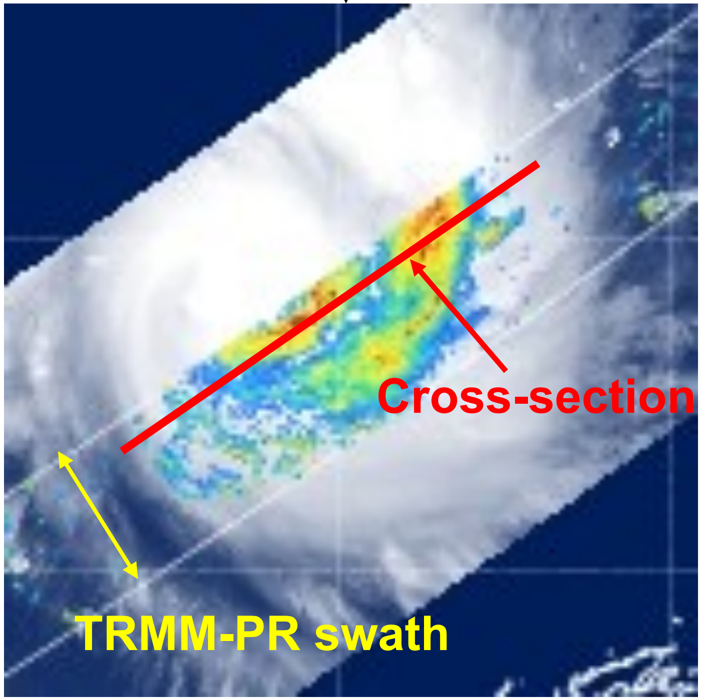
1D-Var with TRMM/Precipitation Radar data

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



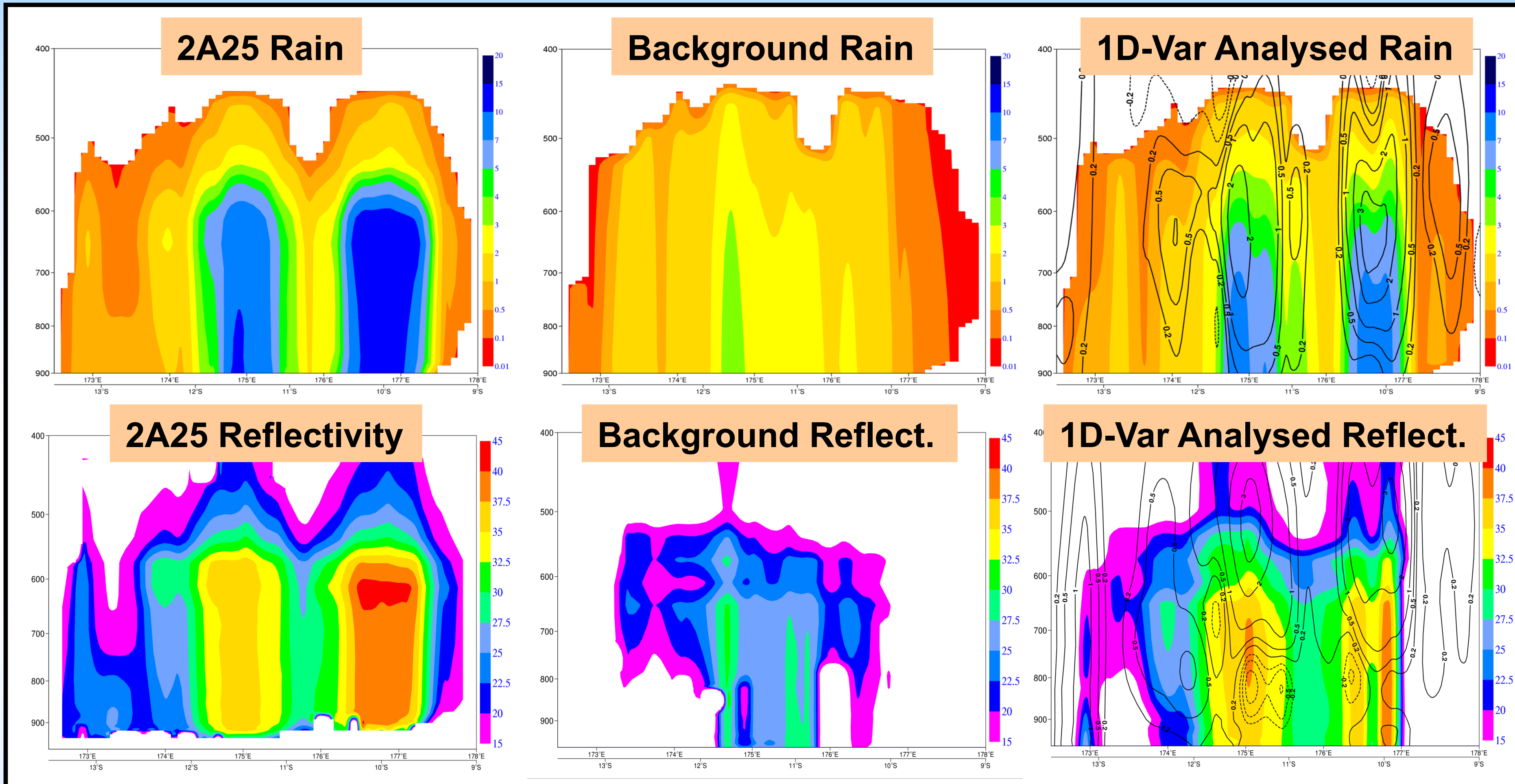
AQUA MODIS image

TRMM Precipitation Radar



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.

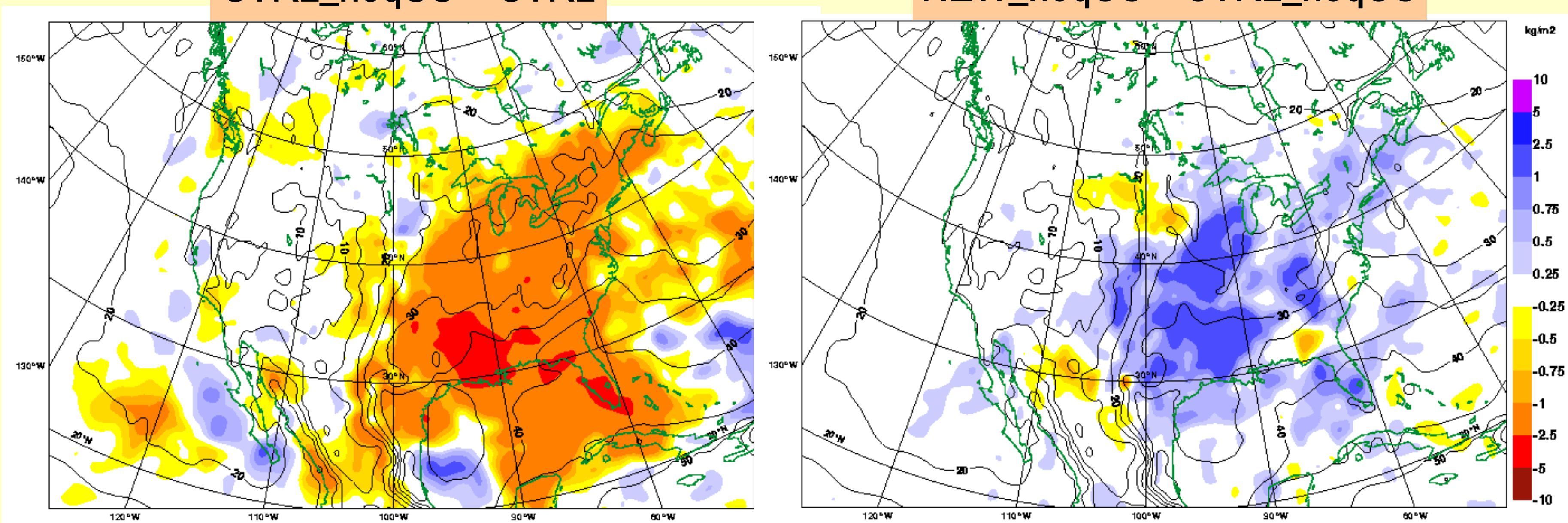
Own impact of NCEP Stage IV hourly precipitation data over the U.S.A. (combined ground-based radar & rain gauge observations)

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**



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

ECMWF