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

• Introduction
• An example of physical initialization
• A very simple variational assimilation problem
• 3D-Var assimilation
• The concept of adjoint
• 4D-Var assimilation
• 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 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,…).
Observations with errors

*a priori* information from model = background state with errors

Data assimilation system (e.g. 4D-Var)

Analysis

NWP model

Forecast

**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 (e.g. 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.
Flash flood over South of France (8-9 Sept 2002)

Ducrocq et al. (2004)

2.5-km resolution model Méso-NH

12h FC from modified analysis

12h FC from operational analysis

Rain gauges

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 $\sigma_b$.
- Simultaneous observation of 2m temperature: $y_o$ with error $\sigma_o$.

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

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

= quadratic distance to background and obs (weighted by their errors)

In other words:

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

And the analysis error, $\sigma_a$, 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

\[ J(x) = \frac{1}{2} (x - x_b)^T B^{-1} (x - x_b) + \frac{1}{2} (H(x) - y_o)^T R^{-1} (H(x) - y_o) \]

\( B \) is the background error covariance matrix, \( R \) is the observation error covariance matrix, \( H \) is the observation operator (used for converting model state vector \( x = (T, q_v, u, v) \) 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

\[ J = \frac{1}{2} (x - x_b)^T B^{-1} (x - x_b) + \frac{1}{2} (H(x) - y_o)^T R^{-1} (H(x) - y_o) \]
3D-Var assimilation

$$\mathcal{J}(x) = \frac{1}{2}(x - x_b)^T B^{-1}(x - x_b) + \frac{1}{2}(H(x) - y_o)^T R^{-1}(H(x) - y_o)$$

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

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

$$\nabla_x \mathcal{J}(x) = B^{-1}(x - x_b) + H^T R^{-1}(H(x) - y_o)$$

where $H^T$ 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 $B$ and $R$) but also on the Jacobian matrix ($H$) of the observation operator ($H$).
An example of observation operator

\( H \): input = model state \((T, q_v)\) \(\rightarrow\) 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)
• **Non-linear observation operator:**

\[ y = H(x) \]

• **Tangent linear operator:**

\[ \delta y = H(\delta x) \]

• **H is the Jacobian matrix derived from** \( H \):

\[
    H_{ij} = \frac{\partial y_i}{\partial x_j}
\]

\[
    \delta y_i = \sum_{j=1}^{N} \frac{\partial y_i}{\partial x_j} \delta x_j
\]
Adjoint technique

• Observation term of the cost-function:

\[ J_o = \frac{1}{2} (y - y_o)^T R^{-1} (y - y_o) \]

• Gradient with respect to y:

\[ \nabla_y J_o = R^{-1} (y - y_o) \]

• Gradient with respect to x:

\[ \frac{\partial J_o}{\partial x_i} = \sum_{j=1}^{M} \frac{\partial J_o}{\partial y_j} \frac{\partial y_j}{\partial x_i} \underbrace{H^T}_{H_{i,j}} \]

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

• Finally:

\[ \nabla_x J_o = H^T (\nabla_y J_o) = H^T R^{-1} (H(x) - y_o) \]
Solution of 3D-Var assimilation

- 3D-Var solution in the linear case:
  \[ x_a = x_b + BH^T(HBH^T + R)^{-1}(y_o - H(x_b)) \]

- with the analysis error covariance matrix \( A \) such as:
  \[ A^{-1} = B^{-1} + H^T R^{-1} H \]

- 3D-Var solution in the non-linear case:
  \[ x^{n+1} = x^n - \rho_n \nabla_x J(x^n) \]

which requires an iterative minimization algorithm (e.g. M1QN3, conjugate gradient)
0D-Var

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

\[ x_a = x_b + \frac{1}{\sigma_b^2 + \sigma_o^2} \left( y_o - x_b \right) \]

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

3D-Var

\[ J = \frac{1}{2} (x - x_b)^T B^{-1} (x - x_b) + \frac{1}{2} (H(x) - y_o)^T R^{-1} (H(x) - y_o) \]

\[ x_a = x_b + BH^T (HBH^T + R)^{-1} \left( y_o - H(x_b) \right) \]

\[ A^{-1} = B^{-1} + H^T R^{-1} H \]
The minimization of the cost function $J$ is usually performed using an iterative minimization procedure.

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

Example with control vector $x = (x_1, x_2)$
4D-Var assimilation

\[ \mathcal{J}(x_0) = \frac{1}{2} (x_0 - x_b)^T B^{-1} (x_0 - x_b) + \frac{1}{2} \sum_{i=0}^{n} (H_i(x_i) - y_o_i)^T R_i^{-1} (H_i(x_i) - y_o_i) \]

where \( x_i \) is the model state at time step \( t_i \) such as:

\[ x_i = M(t_0, t_i) [x_0] \]

\( M \) 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 \( x_0 \) writes:

\[ \nabla_{x_0} \mathcal{J} = B^{-1} (x_0 - x_b) + \sum_{i=0}^{n} M^T(t_i, t_0) H_i^T R_i^{-1} (H_i(x_i) - y_o_i) \]

where \( M^T \) is the adjoint of the forecast model integrated between time \( t_i \) and time \( t_0 \).
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)

$x_a = \text{final analysis}$

$x_b = \text{model first-guess}$

$$
\begin{align*}
\min J &= \frac{1}{2} (x-x_b)^T B^{-1} (x-x_b) + \frac{1}{2} (H(x) - y_o)^T R^{-1} (H(x) - y_o) \\
\Leftrightarrow \nabla_x J &= B^{-1} (x-x_b) + H^T R^{-1} (H(x) - y_o) = 0
\end{align*}
$$
Adjoint of forecast model with simplified linearized physics

\[
\min J = \frac{1}{2} (x_0 - x_b)^T B^{-1} (x_0 - x_b) + \sum_{i=0}^{n} \left[ M^T [t_i, t_0] H_i R_i^{-1} \left( H_i (M[x_0]) - y_{oi} \right) \right]
\]

\[\Leftrightarrow \nabla_x J = B^{-1} (x_0 - x_b) + \sum_{i=0}^{n} [M^T [t_i, t_0] H_i^T R_i^{-1} (H_i (M[x_0]) - y_{oi})] = 0\]
Incremental 4D-Var

- **Model initial state**: \( x_0 = x_{0b} + \delta x_0 \)

- **Observation operator at time** \( t_i \): \( H_i(x_i) = H_i(x_{i\,b}) + H_i \delta x_i \)

  where \( \delta x_i = M_s(x_0, x_i)[\delta x_0] \)

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

  \[
  \mathcal{J}(x_0) = \frac{1}{2}(\delta x_0)^T B^{-1} \delta x_0 + \frac{1}{2} \sum_{i=0}^{n} (H_i(\delta x_i) - d_i)^T R_i^{-1} (H_i(\delta x_i) - d_i)
  \]

  where \( d_i = y_{o\,i} - H_i(x_{i\,b}) \) is the innovation vector.

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

  \[
  \nabla_{\delta x_0} \mathcal{J} = B^{-1} \delta x_0 + \sum_{i=0}^{n} M_s^T(t_i, t_0) H_i^T R_i^{-1} (H_i(\delta x_i) - d_i)
  \]
The analysis is obtained by adding the optimal $\delta x_a$ to the model background: $x_a = x_b + \delta x_a$

To account for non-linearities, the trajectory around which the model is linearized can be updated several times (using $x_a$ as a new $x_b$).

**In operational practice:**

- $d_i$ are computed with the non-linear model $M$ at high resolution (T1279 L137) with full physics.

- $\delta x_i$ are computed with the tangent linear model $M_s$ at low resolution (T255 L137) with simplified physics.

- $\nabla J$ is computed with the adjoint model $M_s^T$ 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.
• TANGENT LINEAR MODEL
  If $M$ is a model such as:
  \[ x(t_{i+1}) = M[x(t_i)] \]
  then the tangent linear model of $M$, called $M'$, is:
  \[ \delta x(t_{i+1}) = M'[x(t_i)]\delta x(t_i) = \frac{\partial M[x(t_i)]}{\partial x} \delta x(t_i) \]

• ADJOINT MODEL
  The adjoint of a linear operator $M'$ is the linear operator $M^*$ such that, for the inner product $<,>$,
  \[ \forall x, \forall y \quad < M'x, y > = < x, M^*y > \]

Remarks:
– with the euclidian inner product, $M^* = M'^T$.
– in variational assimilation, $\nabla_x J = M^* \nabla_y J$, where $J$ is the cost function.
EXAMPLE OF ADJOINT CODING

- non-linear statement

\[
\begin{align*}
  x & = y + z^2 \\
  z & = z \\
  y & = y \\
  x & = y + z^2
\end{align*}
\]

- tangent linear statement

\[
\begin{align*}
  \delta z & = \delta z \\
  \delta y & = \delta y \\
  \delta x & = \delta y + 2z\delta z
\end{align*}
\]

or in a matrix form:

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

### 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:

\[
\delta z^* = \delta z^* + 2z\delta x^* \\
\delta y^* = \delta y^* + \delta x^* \\
\delta x^* = 0
\]
Basic rules for line-by-line adjoint coding (1)

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

<table>
<thead>
<tr>
<th>Tangent linear code</th>
<th>Adjoint code</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta x = 0$</td>
<td>$\delta x^* = 0$</td>
</tr>
</tbody>
</table>
| $\delta x = A \delta y + B \delta z$ | $\delta y^* = \delta y^* + A \delta x^*$  
|                     | $\delta z^* = \delta z^* + B \delta x^*$  
|                     | $\delta x^* = 0$ |
| $\delta x = A \delta x + B \delta z$ | $\delta z^* = \delta z^* + B \delta x^*$  
|                     | $\delta x^* = A \delta x^*$ |
| do $k = 1, N$       | do $k = N, 1, -1$ (Reverse the loop!) |
| $\delta x(k) = A \delta x(k-1) + B \delta y(k)$ | $\delta x^*(k-1) = \delta x^*(k-1) + A \delta x^*(k)$  
| end do              | $\delta y^*(k) = \delta y^*(k) + B \delta x^*(k)$  
|                     | $\delta x^*(k) = 0$ |
| 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.

The most common sources of error in adjoint coding are:
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:

<table>
<thead>
<tr>
<th>Tangent linear code</th>
<th>Trajectory and adjoint code</th>
</tr>
</thead>
<tbody>
<tr>
<td>if $(x &gt; x_0)$ then</td>
<td></td>
</tr>
<tr>
<td>$\delta x = A \frac{\delta x}{x}$</td>
<td></td>
</tr>
<tr>
<td>$x = A \text{Log}(x)$</td>
<td></td>
</tr>
<tr>
<td>end if</td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>$x_{\text{store}} = x$ (storage for use in adjoint)</td>
<td></td>
</tr>
<tr>
<td>if $(x &gt; x_0)$ then</td>
<td></td>
</tr>
<tr>
<td>$x = A \text{Log}(x)$</td>
<td></td>
</tr>
<tr>
<td>end if</td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>$---------------------$ Trajectory $---------------------$</td>
<td></td>
</tr>
<tr>
<td>$---------------------$ Adjoint $---------------------$</td>
<td></td>
</tr>
<tr>
<td>if $(x_{\text{store}} &gt; x_0)$ then</td>
<td></td>
</tr>
<tr>
<td>$\delta x^* = A \frac{\delta x^*}{x_{\text{store}}}$</td>
<td></td>
</tr>
<tr>
<td>end if</td>
<td></td>
</tr>
</tbody>
</table>
TEST FOR TANGENT LINEAR MODEL

- Taylor formula:

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

<table>
<thead>
<tr>
<th>Perturbation scaling factor</th>
<th>RATIO</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1E-09</td>
<td>0.9994875881543574E+00</td>
</tr>
<tr>
<td>0.1E-08</td>
<td>0.9999477148855701E+00</td>
</tr>
<tr>
<td>0.1E-07</td>
<td>0.9999949234236705E+00</td>
</tr>
<tr>
<td>0.1E-06</td>
<td>0.9999993501022509E+00</td>
</tr>
<tr>
<td>0.1E-05</td>
<td>0.9999999949619013E+00</td>
</tr>
<tr>
<td>0.1E-04</td>
<td>0.99999999611338369E+00</td>
</tr>
<tr>
<td>0.1E-03</td>
<td>0.999999996317919371E+00</td>
</tr>
<tr>
<td>0.1E-02</td>
<td>0.999999724488345042E+00</td>
</tr>
<tr>
<td>0.1E-01</td>
<td>0.99999727842790062E+00</td>
</tr>
<tr>
<td>0.1E+00</td>
<td>0.9978007454264978E+00</td>
</tr>
<tr>
<td>0.1E+01</td>
<td>0.9583066504549524E+00</td>
</tr>
</tbody>
</table>

\{ machine precision reached \}
TEST FOR ADJOINT MODEL

• adjoint identity:

\[ \forall x, \forall y \quad < M'.x, y > = < x, M*.y > \]

\[ < F(X), Y > = -1.3765102625251640000E-01 \]
\[ < X, F*(Y) > = -1.3765102625251680000E-01 \]

ratio of norms = 1.0000000000000005

THE DIFFERENCE IS 11.351 TIMES THE ZERO OF THE MACHINE
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

\[ \Delta y (tangent-linear) = 0 \]

\[ \Delta y (non-linear) \]

Threshold

\[ \Delta x (finite size perturbation) \]

Precipitation formation rate

Cloud water amount
Illustration of discontinuity effect on cost function shape:

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

Simple parametrization of rain rate:

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

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

- **Saturated background**: Single minimum of cost function
- **Dry background**: Several local minima of cost function
- **No convergence!**
REGULARIZATION OF VERTICAL DIFFUSION SCHEME

- perturbation of the exchange coefficients is neglected, $K' = 0$ \textit{(Mahfouf, 1999)}

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

- original computation of the Richardson number $Ri$

  \[ Ri = \frac{g}{cpT} \left\| \frac{\partial \bar{w}}{\partial z} \right\| \]

  \[ Ri' = \frac{g}{cpT} \frac{\partial s}{\partial z} \left\| \frac{\partial \bar{w}}{\partial z} \right\|^2 + c \]

  with a time constant $c = \frac{a}{(\Delta t_{phys})^2}$

  where $\Delta t_{phys}$ 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) - \textit{(Janísková et al., 1999)}
function of the Richardson number $f(Ri)$

- **original function**
- **original derivative**
- **modified function**
- **modified derivative**
- **reduced derivative**

Janisková et al. 1999
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.

Importance of regularization to prevent instabilities in tangent-linear model.
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.)
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 ↔ TL evolution of initial perturbations

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

\[ M(x_{an}) - M(x_{bg}) \leftrightarrow M'(x_{an} - x_{bg}) \]

typical size of 4D-Var analysis increments

Diagnostics:
• mean absolute errors:

\[ \varepsilon = \left| \left[ M(x_{an}) - M(x_{bg}) \right] - M'(x_{an} - x_{bg}) \right| \]

• relative error change:

\[ \frac{\varepsilon_{EXP} - \varepsilon_{REF}}{\varepsilon_{REF}} \times 100\% \] (improvement if < 0)

• here: REF = adiabatic run (i.e. no physical parametrizations in tangent-linear)
Impact of operational vertical diffusion scheme

Temperature

Relative improvement [%]

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

REF = ADIAB

12-hour T159 L60 integration

Adiab simp vdif | vdiff
Impact of dry + moist physical processes

Temperature

Relative improvement [%]

Error difference: WSPHYSold_oper - ADIAB: -9.72 %

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

12-hour T159 L60 integration

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

\[ \varepsilon_{\text{EXP}} - \varepsilon_{\text{REF}} \]

Relative improvement [%]

12-hour T159 L60 integration

Adiab simp vdif | vdif + gwd + radnew + cl_new + conv_new
Applications
1D-Var with radar reflectivity profiles

**Background**
\[ x_b = (T_b, q_b, \ldots) \]

Initialize
\[ x = x_b \]

**Moist Physics**

**Reflectivity Model**

**Z_{mod}**

**Reflectivity observations**
\[ Z_{obs} \text{ with errors } \sigma_{obs} \]

**J_b = \frac{1}{2} (x - x_b)^T B^{-1} (x - x_b)\]

\[ \nabla_x J_b = B^{-1} (x - x_b) \]

**MINIM**

**Analysis**
\[ x_a = x \]

**Reflectivity Model (ADJ)**

**Moist Physics (ADJ)**

\[ J_o = \frac{1}{2} \sum_{k=1}^{K} \left( \frac{Z_{mod}^k - Z_{obs}^k}{\sigma_{obs}^k} \right)^2 \]

\[ \nabla_z J_o = \left[ \frac{Z_{mod}^k - Z_{obs}^k}{(\sigma_{obs}^k)^2} \right]_{k=1,K} \]

K = number of model vertical levels
Tropical Cyclone Zoe (26 December 2002 @1200 UTC; Southwest Pacific)
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.
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

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:

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

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

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

Adjoint model incl. physics

$$\frac{\partial J}{\partial x_j} = \text{sensitivity of rain criterion (}J\text{) to model input variables (}x_j\text{) several hours earlier}$$
Adjoint sensitivities for a European winter storm:

\[ J = \text{mean 3h precipitation accumulation inside black box.} \]

RR3h and MSLP, Exper: ka12, 2009021000 T159 L91

Mean precipitation inside target box = 16.39 mm/day

![Map of European winter storm with precipitation data]
\( \frac{\partial J}{\partial x} \) 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)

06/03/2003 12 UTC
Tropical cyclone Kalunde

Tropical singular vectors
VDIF only in TL/AD

Tropical singular vectors
Full physics in TL/AD

ECMWF
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):

$$
\mathbf{x} = \begin{bmatrix}
T_1 \\
\vdots \\
T_n \\
q_1 \\
\vdots \\
q_n
\end{bmatrix}
\quad
\xrightarrow{H}
\quad
\mathbf{y} = \begin{bmatrix}
\text{Rad}_{ch1} \\
\text{Rad}_{ch2} \\
\text{Rad}_{ch3}
\end{bmatrix}
$$

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

$$
\mathbf{H} = \begin{bmatrix}
\frac{\partial \text{Rad}_{ch1}}{\partial T_1} & \ldots & \frac{\partial \text{Rad}_{ch1}}{\partial T_n} & \frac{\partial \text{Rad}_{ch1}}{\partial q_1} & \ldots & \frac{\partial \text{Rad}_{ch1}}{\partial q_n} \\
\frac{\partial \text{Rad}_{ch2}}{\partial T_1} & \ldots & \frac{\partial \text{Rad}_{ch2}}{\partial T_n} & \frac{\partial \text{Rad}_{ch2}}{\partial q_1} & \ldots & \frac{\partial \text{Rad}_{ch2}}{\partial q_n} \\
\frac{\partial \text{Rad}_{ch3}}{\partial T_1} & \ldots & \frac{\partial \text{Rad}_{ch3}}{\partial T_n} & \frac{\partial \text{Rad}_{ch3}}{\partial q_1} & \ldots & \frac{\partial \text{Rad}_{ch3}}{\partial q_n}
\end{bmatrix}
$$