Practical Session: data assimilation experiments

Patrick Laloyaux, Marcin Chrust, Massimo Bonavita, Xavier Abellan

EUROPEAN CENTRE FOR MEDIUM RANGE WEATHER FORECASTS

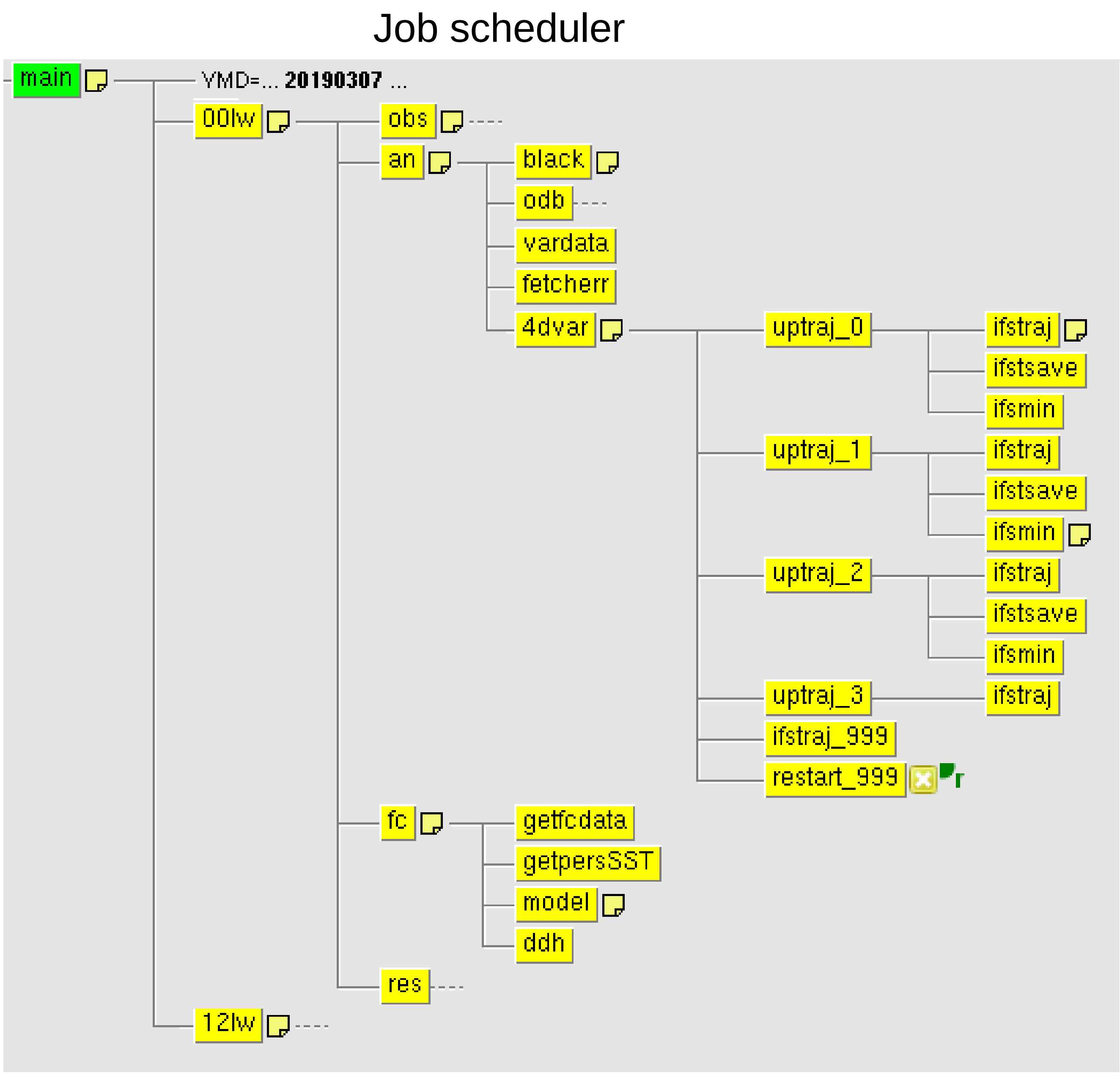
The operational IFS and 4D-Var system

The ECMWF model and assimilation software in numbers

- 18,000 source code files
- 1,000,000 lines of code
- 100,000 if statements
- 45 minutes to solve one 4D-Var problem… using 25,000 CPUs

```
!     ------------------------------------------------------------------

!     1.    Convert X into X-Xb : subtract background
!           ------------------------------------------

IF (LHOOK) CALL DR_HOOK('CVAR2',0,ZHOOK_HANDLE)
ASSOCIATE(YDDIM=>YDGEOMETRY%YRDIM, &
  & YDDIMV=>YDGEOMETRY%YRDIMV,YDGEM=>YDGEOMETRY%YRGEM, YDMP=>YDGEOMETRY%YRMP)
ASSOCIATE(NPROMA=>YDDIM%NPROMA, &
 & NFLEVG=>YDDIMV%NFLEVG, NFLEVL=>YDDIMV%NFLEVL, &
 & NGPTOT=>YDGEM%NGPTOT, &
 & MYLEVS=>YDMP%MYLEVS)
IF (YD_JB_STRUCT%JB_DATA%LSUBBG) CALL SBSBGS(YDGEOMETRY,YDGMV,YDGMV5,YDFIELDS)

!     3.    Apply B matrix
!           --------------

! Parameters to be estimated

IF (LVARBC) CALL YDVARBC%PARAM_GET(YD_JB_STRUCT%JB_DATA%LSUBBG,YDVAZX%PARAMS)

IF (YDVAZX%LAM1D) THEN
  II=0
  DO JF=1,CVA_DATA%NVA1D
    DO JS=1,NFLEVL
      IL=MYLEVS(JS)
      II=II+1
      IF (YD_JB_STRUCT%JB_DATA%TMEANUVER(IL,JF)/=0.) THEN
        YDVAZX%LAMCV(II) = YD_JB_STRUCT%JB_DATA%SPJB%SP1D(JS,JF) /&
                      & YD_JB_STRUCT%JB_DATA%TMEANUVER(IL,JF)
      ELSE
        YDVAZX%LAMCV(II) = 0.0_JPRB
      ENDIF
    ENDDO
  ENDDO
ENDIF
```

The operational IFS and 4D-Var system



Retrieve observations

Preprocess observations

Read B matrix

Run nonlinear model

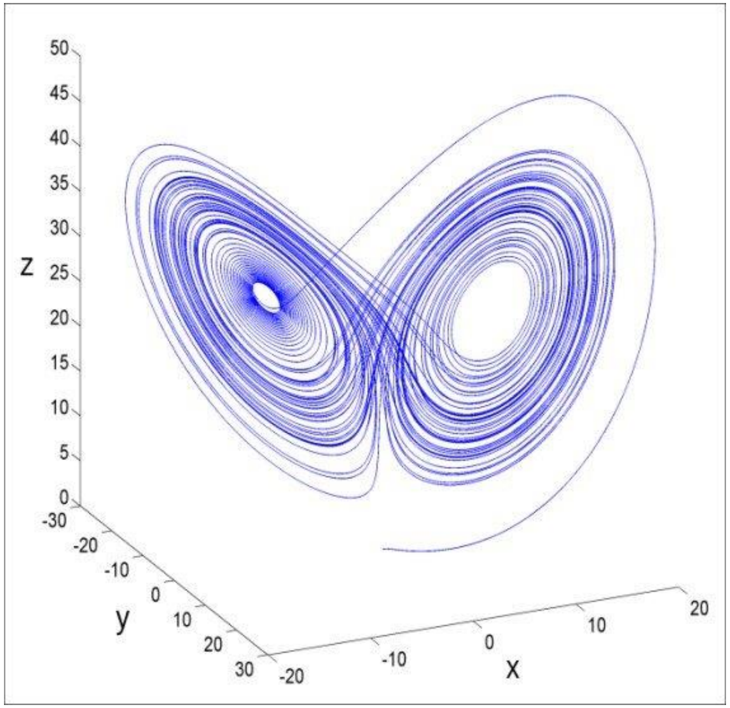
Minimise quadratic cost function

Produce forecast

The Lorenz Model

The Lorenz system is a system of ordinary differential equations.

It is famous for having chaotic solutions for certain parameter values and initial conditions



In the practical sessions, we will use a more complex model

- Lorentz-95 model (40 variables)
- Twin experiments (the true state is known)
- 3D-Var and 4D-Var systems

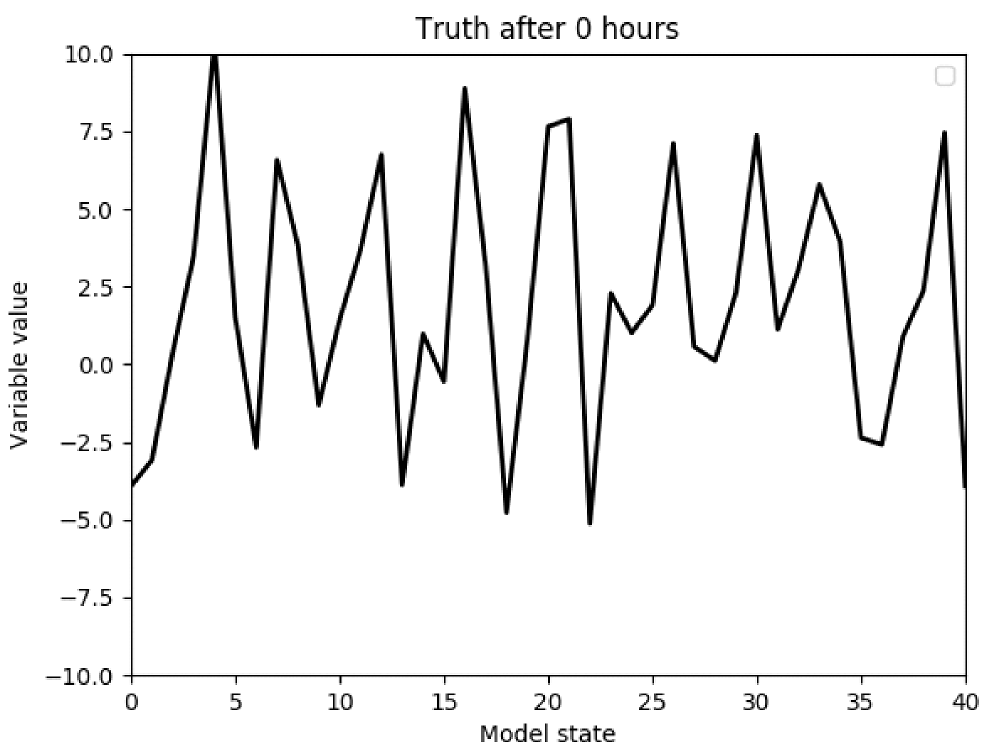
It is good enough to illustrate to most important concepts and issues

The Lorenz Model

- The Lorenz-95 model is a widely-used low-dimensional dynamical system for data assimilation studies
- The system is defined by a set of coupled ordinary differential equations

$$dx_i/dt = -x_{i-2}\, x_{i-1} + x_{i-1}\, x_{i+1} - x_i + F \quad \text{for } i = 1, 2 \ldots N$$

- For a range of values of F ≈ 8, the system is chaotic, and has similar characteristics as an operational NWP system



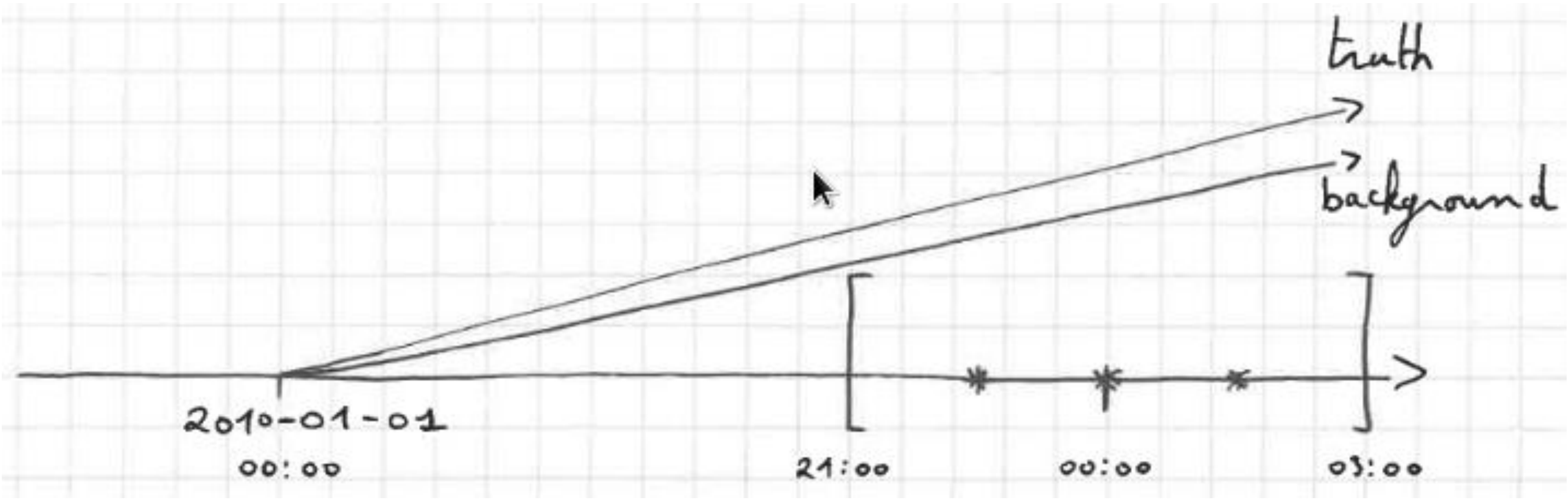
The model state **x** includes 40 variables

Numerical experiments

Twin experiments

- generate a true trajectory
- compute perturbed observations (adding a white Gaussian noise, mean=0, stdv=σ)
- compute a background trajectory (error introduced by changing slightly the forcing F)

Understanding strengths and limitations of 3D-Var and 4D-Var



Task 1: Compilation and make the executables

Run the magic script!

source DAcourse/DA_TC_2021/makeoops

→ the file TC_oops_2021.pdf is a copy of the slides

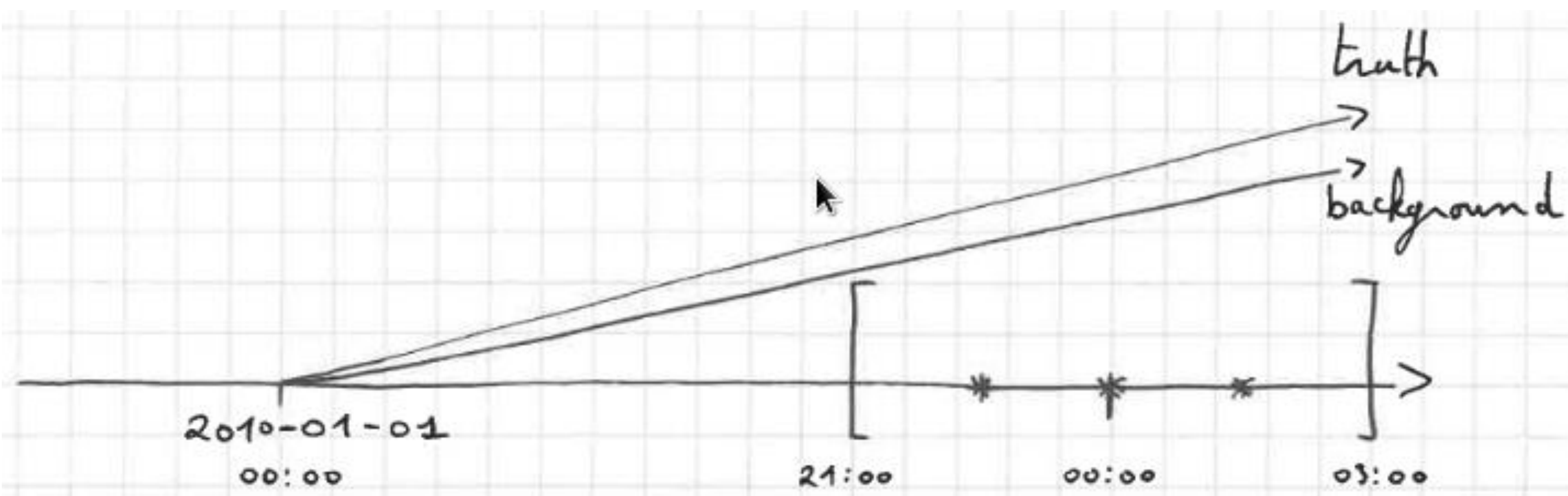
Task 2: Generate input data for the assimilation

Generate the ”truth”:
l95_forecast.x l95_truth.xml

Generate observations from the truth
l95_makeobs.x l95_makeobs_6h.xml

Generate the background trajectory:
l95_forecast.x l95_forecast.xml

Plot the true model trajectory over the first three days
python l95_plotTrajMod.py



Task 3: Run a cycle of 6-hour 3D-Var with many observations

Run a 3D-Var analysis

l95_4dvar.x l95_3dvar_6h.xml

Plot the truth, background, 3dvar analysis in the middle of the assimilation window (2010-01-02T00:00:00)

python l95_plotTraj.py 3dvar_6h &

Plot the background error, analysis error and analysis increment in the middle of the assimilation window (2010-01-02T00:00:00)

python l95_plotDiffs.py 3dvar_6h &

Questions:

- Where are the largest differences between the background and the analysis?
- What is the relationship between background error, analysis error and analysis increment?

Task 4: Run a cycle of 6-hour 3D-Var with a single observation

Generate a single observation from the truth at the beginning of the window
l95_makeobs.x l95_makeobs_6h_single_begin.xml

Run a 3D-Var analysis with a single observation
l95_4dvar.x l95_3dvar_6h_single_begin.xml

Plot the truth, background, 3dvar analysis in the middle of the assimilation window
python l95_plotTraj.py 3dvar_6h_single &

Plot the background error, analysis error and analysis increment
python l95_plotDiffs.py 3dvar_6h_single &

Task 4: Run a cycle of 6-hour 3D-Var with a single observation

Edit the file **l95_3dvar_6h_single_begin.xml** and change the parameters of the covariance matrix:

- standard deviation
- length scale

Questions:

- How does the increment size evolve when the background standard deviation is increased/decreased?
- How does the increment spread when the background correlation lengthscale is increased/decreased?
- Would you have the same results if the same observation was available at the end of the window?
- What is the size of the increment at the observation location when $\sigma_b = 0.6$ ($\sigma_o = 0.4$ and $d^o =$ -0.155 in your experiment)? Derive a mathematical formula and confirm your result running the experiment!

Task 4: Run a cycle of 3D-Var with a single observation

Remember the Linear Analysis Equation:

$$x_a = x_b + K(y - Hx_b)$$
$$\text{where } K = BH^T \left(HBH^T + R\right)^{-1}$$

For a single observation, located at a gridpoint:
$H = (0, \ldots, 0, 1, 0, \ldots, 0)$
Hence

$$\begin{aligned} x_a - x_b &= K(y - Hx_b) \\ &= B \begin{pmatrix} 0 \\ \vdots \\ z \\ \vdots \\ 0 \end{pmatrix} \quad \text{where } z = \left(HBH^T + R\right)^{-1} (y - Hx_b) \end{aligned}$$

That is, $x_a - x_b \propto$ a column of B

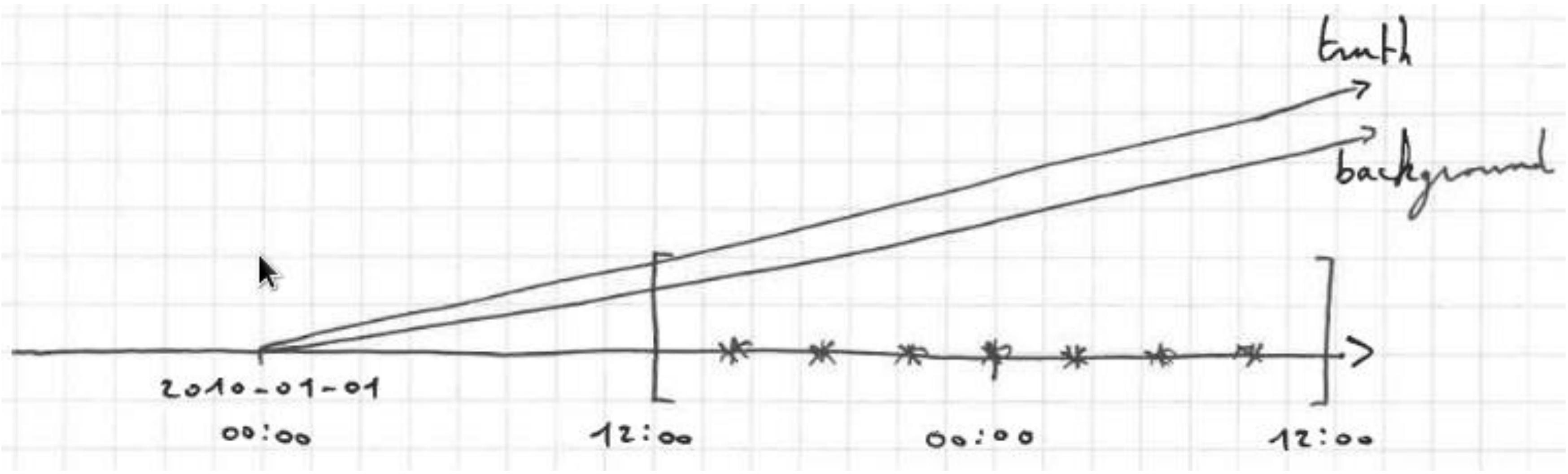
Task 5: Run a cycle of 24-hour 3D-Var with many observations

Generate observations from the truth:
l95_makeobs.x l95_makeobs_24h.xml

Run a 3D-Var analysis
l95_4dvar.x l95_3dvar_24h.xml

Plot the truth, background, 3dvar analysis in the middle of the assimilation window
python l95_plotTraj.py 3dvar_24h &

Plot the background error, analysis error and analysis increment
python l95_plotDiffs.py 3dvar_24h &



Task 5: Run a cycle of 24-hour 3D-Var with many observations

Compare the analysis error from the 6-h and the 24-h 3dvar

python l95_plotErr.py 3dvar_6h 3dvar_24h &

Plot the histogram of departures for the 6h and the 24h windows

python l95_plotHist.py 3dvar_6h 3dvar_24h &

Plot the departures with respect to the time in the assimilation window

python l95_plotTime.py 3dvar_6h 3dvar_24h &

Questions:

- What is the impact of having more observations over a longer window on the analysis error?
- What is the problem to run a 3D-Var with a long assimilation window?
- How can we assess the performance of an assimilation system when the true state is unknown?

Task 6: Run a cycle of 6-hour 4D-Var with many observations

Run 4D-Var

l95_4dvar.x l95_4dvar_6h.xml

Plot the truth, background, 4dvar analysis in the middle of the assimilation window

python l95_plotTraj.py 4dvar_6h &

Compare the analysis error for 3dvar and 4dvar

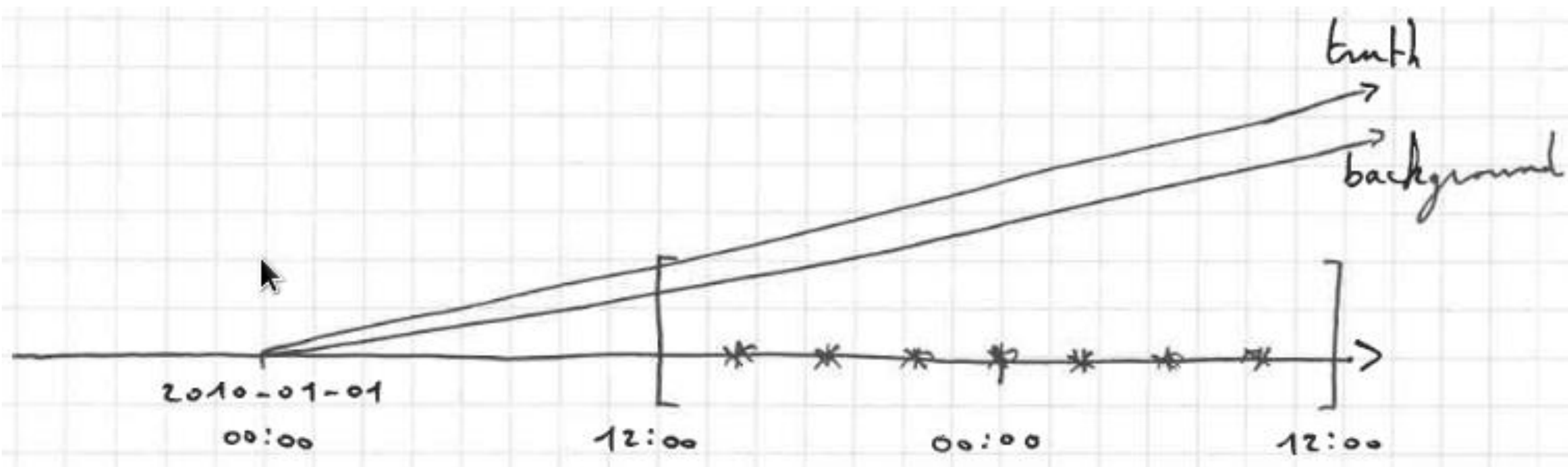
python l95_plotErr.py 3dvar_6h 4dvar_6h &

Plot the histogram of departures for 3dvar and 4dvar

python l95_plotHist.py 3dvar_6h 4dvar_6h &

Task 6: Run a cycle of 6-hour 4D-Var with many observations

4D-Var produces an increment and an analysis at the beginning of the window. To compare with the 3D-Var analysis (valid at 00:00) the 4D-Var analysis is integrated in time from 12:00 to 00:00



Questions:

- Does 4D-Var produce a better analysis compared to 3D-Var?

Task 7: Run a cycle of 6-hour 4D-Var with a single observation

Generate the observations at the beginning of the window

l95_makeobs.x l95_makeobs_6h_single_begin.xml

Run 4D-Var

l95_4dvar.x l95_4dvar_6h_single_begin.xml

Plot the background error, analysis error and analysis increment

python l95_plotDiffs.py 4dvar_6h_single_begin &

Generate the observations at the end of the window

l95_makeobs.x l95_makeobs_6h_single_end.xml

Run 4D-Var

l95_4dvar.x l95_4dvar_6h_single_end.xml

Plot the background error, analysis error and analysis increment

python l95_plotDiffs.py 4dvar_6h_single_end &

Questions:

- Does the observation time matter in 4D-Var?
- Which mathematical operators control the information spread in 4D-VAR?

Task 7: Run a cycle of 24-hour 4D-Var with a single observation

Confirm your last answer by running a 24-hour 4D-Var with a single observation at the end of the assimilation window

Generate the observations at the beginning of the window

l95_makeobs.x l95_makeobs_24h_single_end.xml

Run 4D-Var

l95_4dvar.x l95_4dvar_24h_single_end.xml

Plot the background error, analysis error and analysis increment

python l95_plotDiffs.py 4dvar_24h_single_end &

Task 7: Run a cycle of 24-hour 4D-Var with a single observation

The increment computed by the 4D-Var is valid at the begining of the window and can be expressed as:

$$\delta x \quad = \quad BM^T H^T \left(HMBM^T H^T + R \right)^{-1} (y - \mathcal{G}(x_b))$$

For a single observation at time k, located at a gridpoint:
$H = (0, \ldots, 0, 1, 0, \ldots, 0)$

Hence

$$\delta x \quad = \quad BM_k^T \begin{pmatrix} 0 \\ \vdots \\ z \\ \vdots \\ 0 \end{pmatrix} \text{where } z = \left(HM_k BM_k^T H^T + R \right)^{-1} (y_k - \mathcal{G}_k(x_b))$$

That is, $\delta \mathbf{x} \propto$ a column of BM_k^T

Practical Session: data assimilation experiments

Patrick Laloyaux, Marcin Chrust, Massimo Bonavita, Xavier Abellan

EUROPEAN CENTRE FOR MEDIUM RANGE WEATHER FORECASTS

Task 8: Run a cycle of 24-hour 4D-Var with many observations

Run a 4D-Var analysis
l95_4dvar.x l95_4dvar_24h.xml

Plot the truth, background, 3dvar analysis in the middle of the assimilation window
python l95_plotTraj.py 4dvar_24h &

Compare the analysis error from 3dvar and 4dvar
python l95_plotErr.py 3dvar_24h 4dvar_24h &

Plot the histogram of departures for 3dvar and 4dvar
python l95_plotHist.py 3dvar_24h 4dvar_24h &

Scatter diagram of model and observations
python l95_plotScatt.py 4dvar_24h &

Plot the departures with respect to the time in the assimilation window
python l95_plotTime.py 3dvar_24h 4dvar_24h &

Questions:
- Why does 4D-Var outperform 3D-Var with a long assimilation window?

Task 9: What's going on?

A new version of the model and the assimilation system has been developed and a new set of observations is available… but something is wrong…

Could you find the reason of the poor performance of 4D-Var?

Run a 4D-Var analysis
l95_4dvar.x l95_4dvar_24h_bad3.xml

Compare the analysis error with the previous 4dvar
python l95_plotErr.py 4dvar_24h 4dvar_24h_bad3 &

Plot the histogram of departures
python l95_plotHist.py 4dvar_24h 4dvar_24h_bad3 &

Plot the departures with respect to the time in the assimilation window
python l95_plotTime.py 4dvar_24h 4dvar_24h_bad3 &

Scatter diagram of model and observations
python l95_plotScatt.py 4dvar_24h_bad3 &

Plot the truth, background, 4dvar analysis in the middle of the assimilation window
python l95_plotTraj.py 4dvar_24h_bad3 &

Task 9: What's going on?

A new version of the model and the assimilation system has been developed and a new set of observations is available… but something is wrong…

Could you find the reason of the poor performance of 4D-Var?

Run a 4D-Var analysis
l95_4dvar.x l95_4dvar_24h_bad2.xml

Compare the analysis error with the previous 4dvar
python l95_plotErr.py 4dvar_24h 4dvar_24h_bad2 &

Plot the histogram of departures
python l95_plotHist.py 4dvar_24h 4dvar_24h_bad2 &

Plot the departures with respect to the time in the assimilation window
python l95_plotTime.py 4dvar_24h 4dvar_24h_bad2 &

Scatter diagram of model and observations
python l95_plotScatt.py 4dvar_24h_bad2 &

Plot the truth, background, 4dvar analysis in the middle of the assimilation window
python l95_plotTraj.py 4dvar_24h_bad2 &

Task 9: What's going on?

A new version of the model and the assimilation system has been developed and a new set of observations is available… but something is wrong…

Could you find the reason of the poor performance of 4D-Var?

Run a 4D-Var analysis

l95_4dvar.x l95_4dvar_24h_bad4.xml

Compare the analysis error with the previous 4dvar

python l95_plotErr.py 4dvar_24h 4dvar_24h_bad4 &

Plot the histogram of departures

python l95_plotHist.py 4dvar_24h 4dvar_24h_bad4 &

Plot the departures with respect to the time in the assimilation window

python l95_plotTime.py 4dvar_24h 4dvar_24h_bad4 &

Scatter diagram of model and observations

python l95_plotScatt.py 4dvar_24h_bad4 &

Plot the truth, background, 4dvar analysis in the middle of the assimilation window

python l95_plotTraj.py 4dvar_24h_bad4 &

Task 10: Number of outer and inner iterations in 4D-Var

The cost function can be written in terms of the increment $\delta\mathbf{x}^{(m)}$, and approximated by the quadratic function:

$$J\left(\delta\mathbf{x}^{(m)}\right) = \frac{1}{2}\left[\delta\mathbf{x}^{(m)} - \delta\mathbf{x}_b\right]^{\mathrm{T}} \mathbf{P}^{b^{-1}} \left[\delta\mathbf{x}^{(m)} - \delta\mathbf{x}_b\right] + \frac{1}{2}\left[\mathbf{d}^{(m)} - \mathbf{G}\delta\mathbf{x}^{(m)}\right]^{\mathrm{T}} \mathbf{R}^{-1} \left[\mathbf{d}^{(m)} - \mathbf{G}\delta\mathbf{x}^{(m)}\right]$$

The incremental method treats the minimisation of J as a sequence of quadratic problems:

for $m = 0, 1, \cdots, M$ **do**

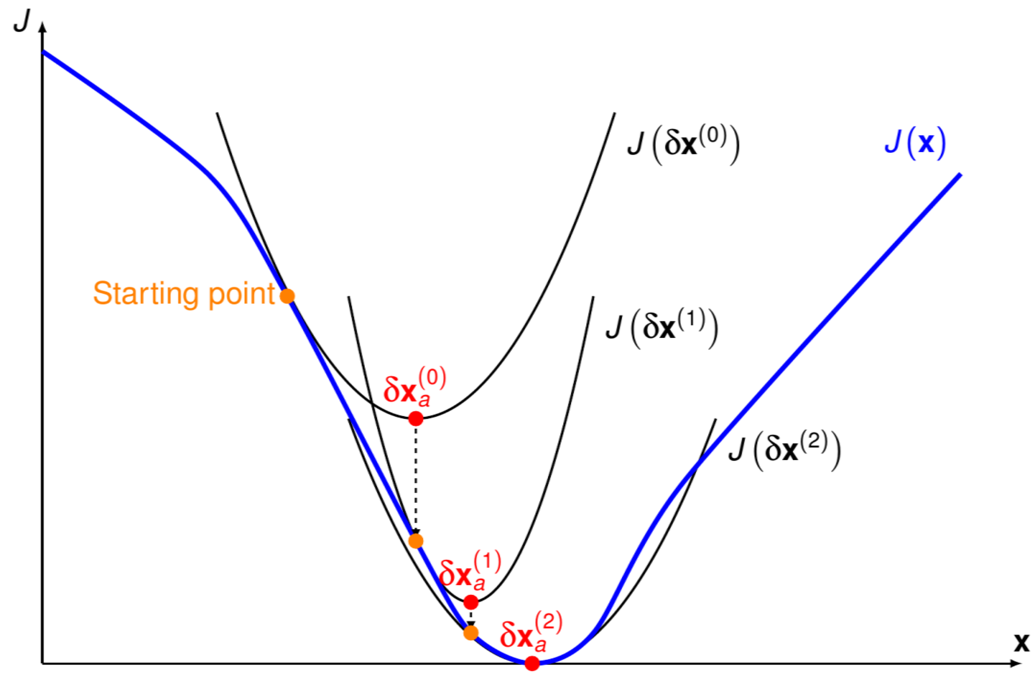
/* Minimise the quadratic cost function $J\left(\delta\mathbf{x}^{(m)}\right)$

$\delta\mathbf{x}_a^{(m)} = \arg\min_{\mathbf{x}} \left[J\left(\delta\mathbf{x}^{(m)}\right)\right]$;

/* Set the new linearisation state

$\mathbf{x}^{(m+1)} = \mathbf{x}^{(m)} + \delta\mathbf{x}_a^{(m)}$;

end



Task 10: Number of outer and inner iterations

Run a 24hour 4D-Var with one, two and three outer iterations (the number of inner iteration is set to 4)

l95_4dvar.x l95_4dvar_24h_outer1.xml > 4dvar_24h_outer1.out
l95_4dvar.x l95_4dvar_24h_outer2.xml > 4dvar_24h_outer2.out
l95_4dvar.x l95_4dvar_24h_outer3.xml > 4dvar_24h_outer3.out

Compare the analyses looking at the fit to the observations
python l95_plotTime3.py 4dvar_24h_outer1 4dvar_24h_outer2 4dvar_24h_outer3 &

Plot the evolution of the 4D-Var cost function during the minimization
python l95_plotConv.py 4dvar_24h_outer3 &

Change the number of inner iterations (from 4 to 10) by editing 3 times the variables ninner in l95_4dvar_24h_outer3.xml

Plot the evolution of the 4D-Var cost function during the minimization
python l95_plotConv.py 4dvar_24h_outer3 &

Questions:

- What's the effect of having more inner iterations?
- Is this always guaranteed?

Task 10: Number of outer and inner iterations

Run a 6hour 4D-Var with three outer iterations (the number of inner iteration is set to 4)

l95_4dvar.x l95_4dvar_6h_outer3.xml > 4dvar_6h_outer3.out

Plot the evolution of the 4D-Var cost function during the minimization

python l95_plotConv.py 4dvar_6h_outer3 &

Questions:

- How does the convergence speed evolve with respect to the assimilation window? Why?