Fairmode Technical Meeting

24.-25.6.2015, DAO, Univ. Aveiro

Validation of Complex Data Assimilation Methods

Hendrik Elbern,

Elmar Friese, Nadine Goris, Lars Nieradzik

and many others

Rhenish Institute for nvironmental Research at the University of Cologne

and

Institute for Energy and Climate Research (Troposphere)

Forschungszentrum Jülich

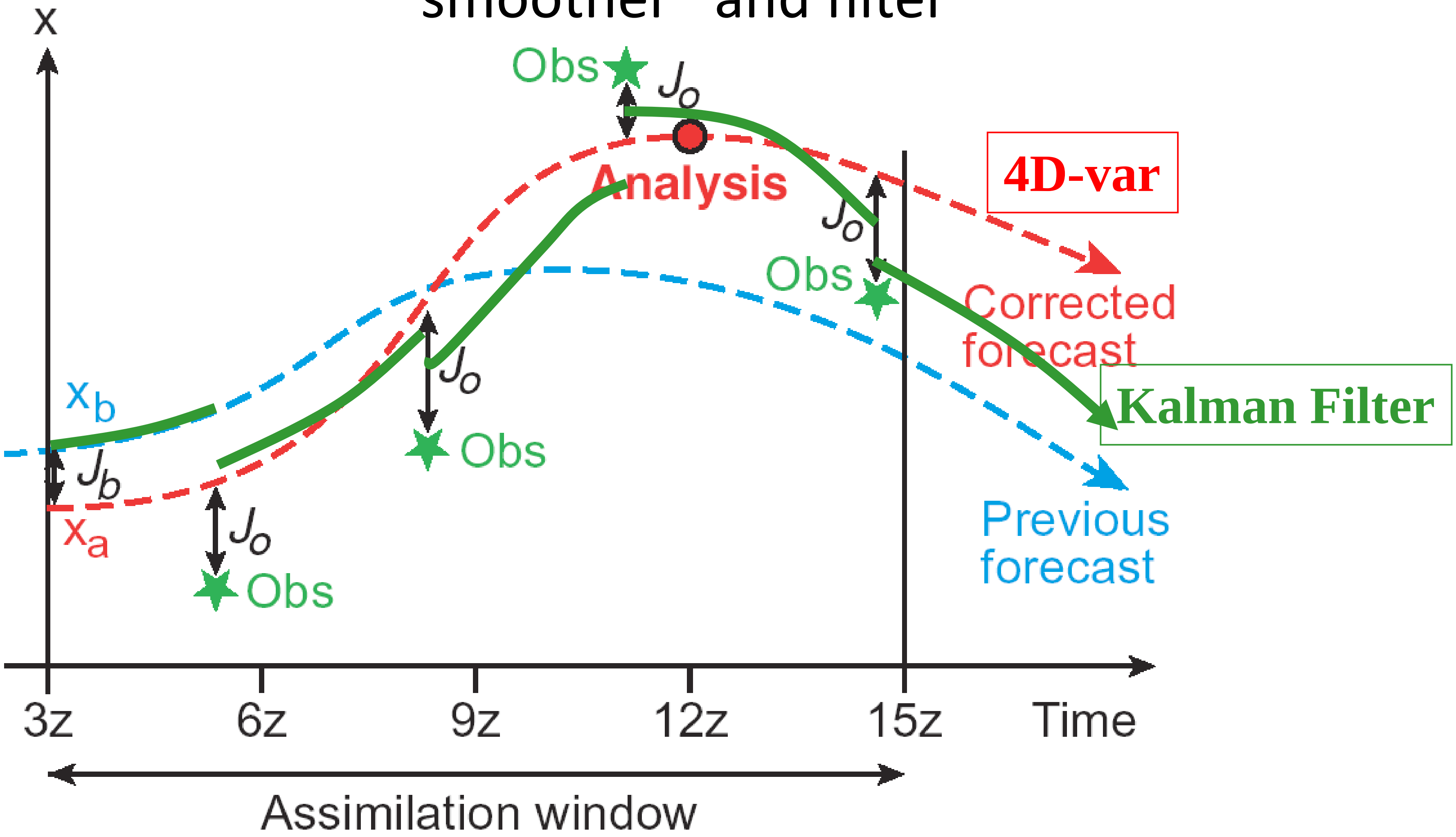
Contents

1. Intro.: What are *complex data assimilation methods*?
2. Observability: Do observations sustain assimilation results?
3. Practical verification: Validation by forecast skills
4. A posteriori Validation: Is the analysis consistent?

What are ***complex data assimilation methods***?

→ <u>spatio-temporal</u> techniques

2 types of assimilation algorithms: “smoother” and filter



The 4-dimensional variational technique: *Optimize over an assimilation window, then forecast*

Emission Rate Optimization

minimize cost function

$$
\begin{aligned}
J(\mathbf{x}(t_0), \mathbf{e}) =& \frac{1}{2}(\mathbf{x}^b(t_0) - \mathbf{x}(t_0))^T \mathbf{B}_0^{-1}(\mathbf{x}^b(t_0) - \mathbf{x}(t_0)) + \\
& \frac{1}{2}\int_{t_0}^{t_N} (\mathbf{e}_b(t) - \mathbf{e}(t))^T \mathbf{K}^{-1}(\mathbf{e}_b(t) - \mathbf{e}(t))dt + \\
& \frac{1}{2}\int_{t_0}^{t_N} \left(\mathbf{y}^0(t) - H[\mathbf{x}(t)]\right)^T \mathbf{R}^{-1}(\mathbf{y}^0(t) - H[\mathbf{x}(t)])dt
\end{aligned}
$$

deviations from background initial state

deviations from a priori emission rates

model deviations from observations

$\mathbf{x}^b(t_0)$	background state at $t = 0$
$\mathbf{x}(t)$	model state at time t
$\mathbf{e}_b(t_0)$	background emission rate at $t = 0$
$\mathbf{e}(t)$	emission rate field at time t
$\mathbf{K}$	emission rate error covariance matrix
$H[\quad]$	forward interpolator
$\mathbf{y}^0(t)$	observation at time t
$\mathbf{B}_0$	background error covariance matrix

Kalman filter: basic equations

Forecast steps:

a) the atmospheric state

$$\mathbf{x}^f(t_i) = \mathbf{M}(t_i, t_{i-1})\mathbf{x}^a(t_{i-1}) + \eta$$

b) the forecast error covariance matrix

$$\mathbf{P}_i^b = \mathbf{M}(t_i, t_{i-1})\mathbf{P}_{i-1}^a\mathbf{M}^T(t_i, t_{i-1}) + \mathbf{Q}$$

Analysis steps:

a) the atmospheric state

$$\mathbf{x}^a(t_i) = \mathbf{x}^b(t_i) + \mathbf{K}_i\mathbf{d}_i, \qquad (1)$$

$$\mathbf{K}_i := \mathbf{P}_i^b\mathbf{H}_i^T(\mathbf{H}_i\mathbf{P}_i^b\mathbf{H}_i^T + \mathbf{R}_i)^{-1} \quad \in \mathcal{R}^{n\times p_i} \qquad (2)$$

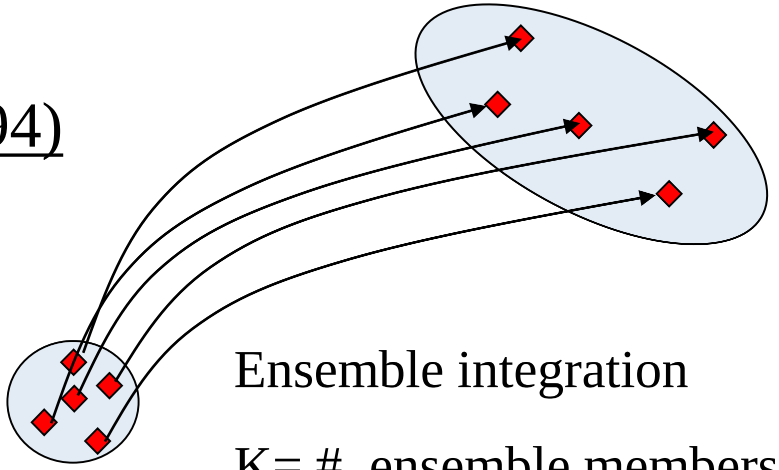
and b) the analysis error covariance matrix

$$\mathbf{P}_i^a = (\mathbf{I} - \mathbf{K}_i\mathbf{H}_i)\mathbf{P}_i^b. \qquad (3)$$

Computational challenge:
Background Error Covariance Matrix **P**$^{\mathbf{b}}$

1. Ensemble approach: (e.g. Evensen, 1994)

$$\boxed{B_{ij} = \frac{1}{K}\sum_{n=1}^{K}\left(x_i^n - \bar{x}_i\right)\left(x_j^n - \bar{x}_j\right)}$$



Ensemble integration

K= # ensemble members;

i,j grid cells

2. Observability: Do observations sustain assimilation results? Observation network design

Is the forecasted system sensitive to available observations?

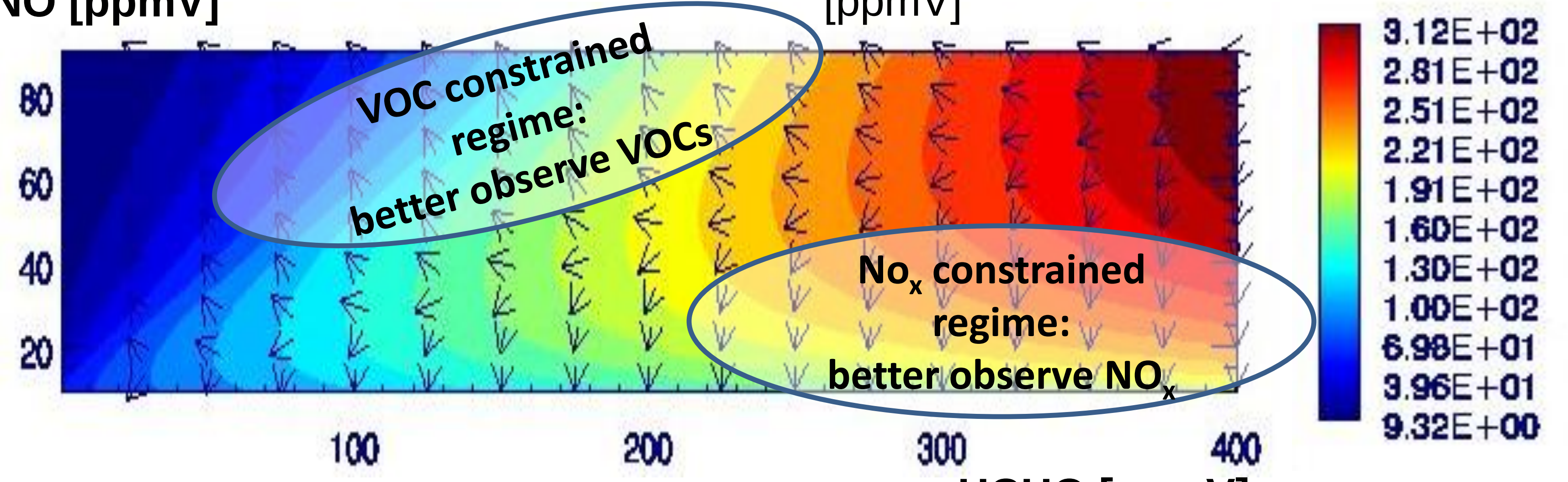
– Observation System Simulation Experiments (OSSEs)

– Targeted observations

Is NO$_x$ the key to ozone production?
And consequently, its observation^the key to better forecast?

Isopleths of ozone production [ppmV]

NO [ppmV]



VOC constrained regime: better observe VOCs

No$_x$ constrained regime: better observe NO$_x$

HCHO [ppmV]

Calculations

- ✓ within a fixed time span
- ✓ initial conventrations of NO / HCHO were varied
- ✓ change of final concentration is given by colour
- ✓gradients (SVs) of maximyl ozone production given by arrows

How can we optimize the observation configuration?

Given CTM (here RACM and EURAD-IM) acting as tan.-lin. model operator $\mathcal{L}$:

$$\delta\mathbf{c}(t_F) = \mathcal{L}_{t_I,t_F}\,\delta\mathbf{c}(t_I), \quad \mathcal{L}_{t_I,t_F} = \left.\frac{\partial\mathcal{M}_{t_I,t_F}}{\partial\mathbf{c}}\right|_{\mathbf{c}(t_I)}$$

1. Berliner et al., (1998) Statistical design:
"Minimize" the analysis error covariance matrix **A** (say, via trace):

$$\min_{\mathbf{H}} \mathbf{A} = \mathbf{B} - \underbrace{\mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{B}}_{\text{to be maximized by } \mathbf{H}}$$

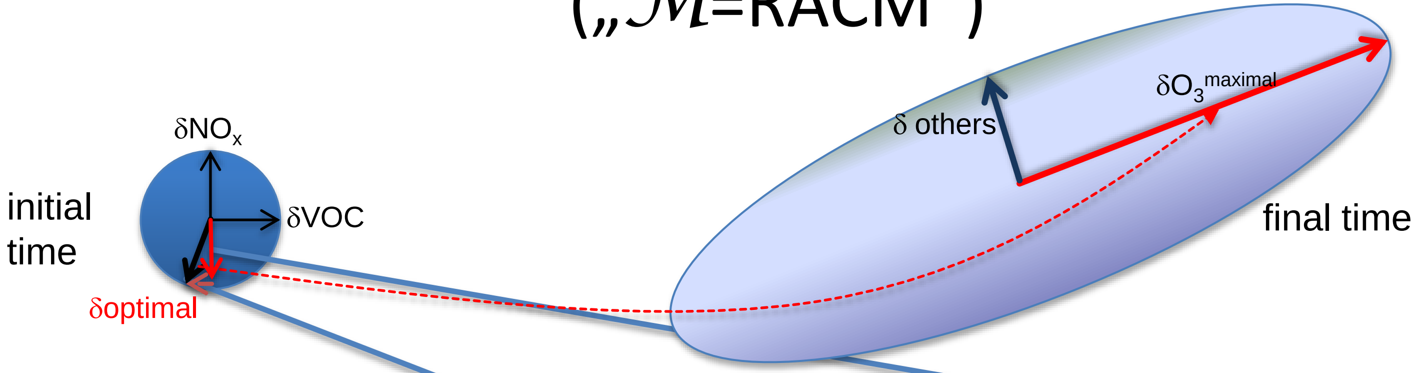
For this find maximal eigenvectors as observation operators **H,** which configure observations.

$$\mathcal{L}_{t_I,t_F}\mathbf{B}\mathcal{L}^T_{t_I,t_F}\mathbf{H}^T = \lambda\mathbf{H}^T$$

2. Palmer (1995) Singular vector analysis:
Observe maximal SV configuration:

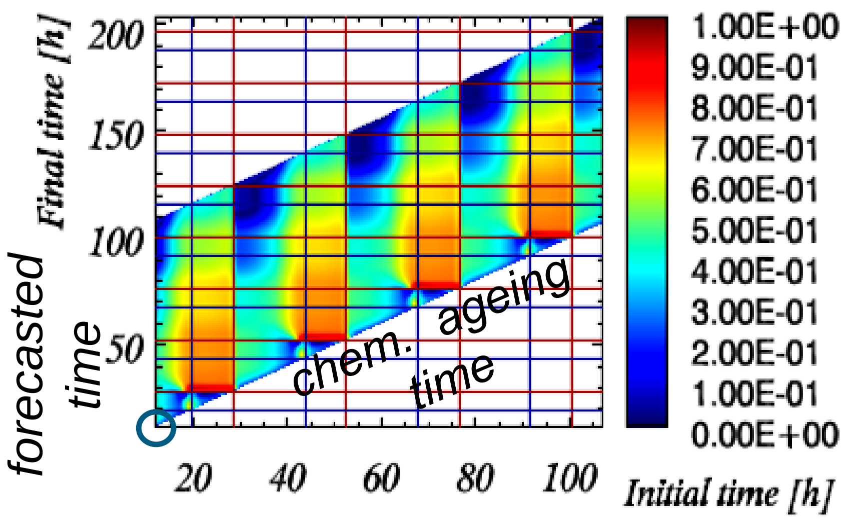
$$\max_{\delta\mathbf{c}(t_I)} \frac{\|\delta\mathbf{c}(t_F)\|^2_{\mathbf{B}}}{\|\delta\mathbf{c}(t_I)\|^2_{\mathbf{B}}} = \max_{\delta\mathbf{c}(t_I)} \frac{\delta\mathbf{c}(t_I)^T\mathcal{L}^T_{t_I,t_F}\mathbf{B}\,\mathcal{L}_{t_I,t_F}\delta\mathbf{c}(t_I)}{\delta\mathbf{c}(t_I)^T\mathbf{B}\,\delta\mathbf{c}(t_I)},$$

Basic 0-D Regional Atm. Chemistry Mechanism („$\mathcal{M}$=RACM“)

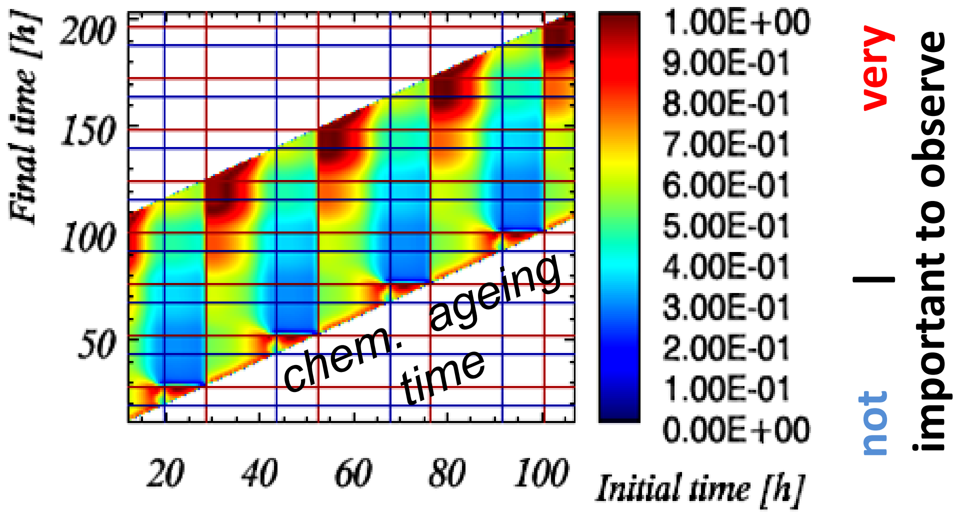


- **Optimal perturbations (Singular Vectors) for** scenario MARINE

1st Grouped Singular Vectors (δVOC)



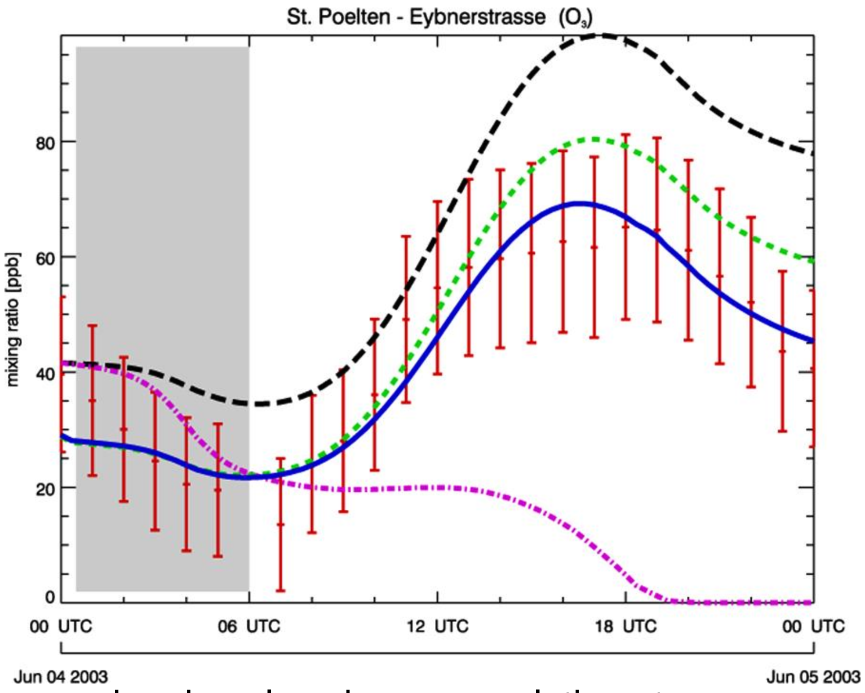
1st Grouped Singular Vectors (δNO_x)



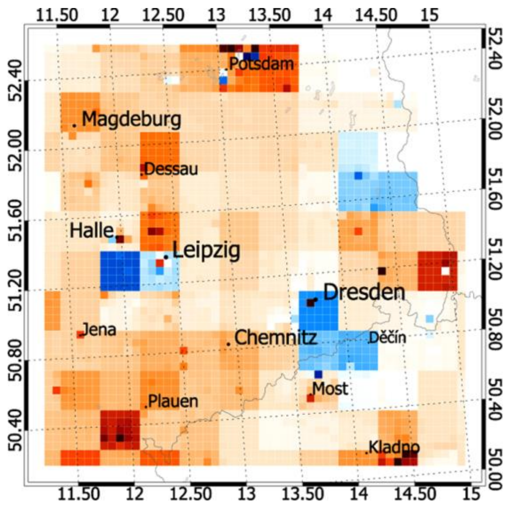
🟥 sunrise 🟦 sunset

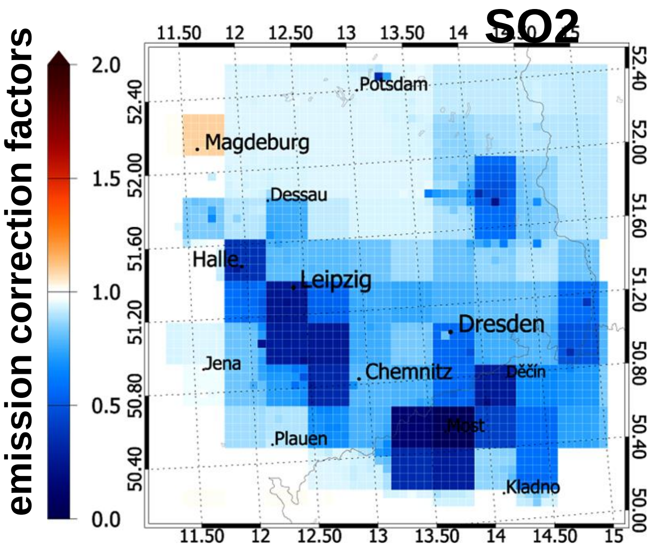
3. Practical verification: Validation by forecasts

Analysis of emissions by 4D-var (VERTIKO)



NO2



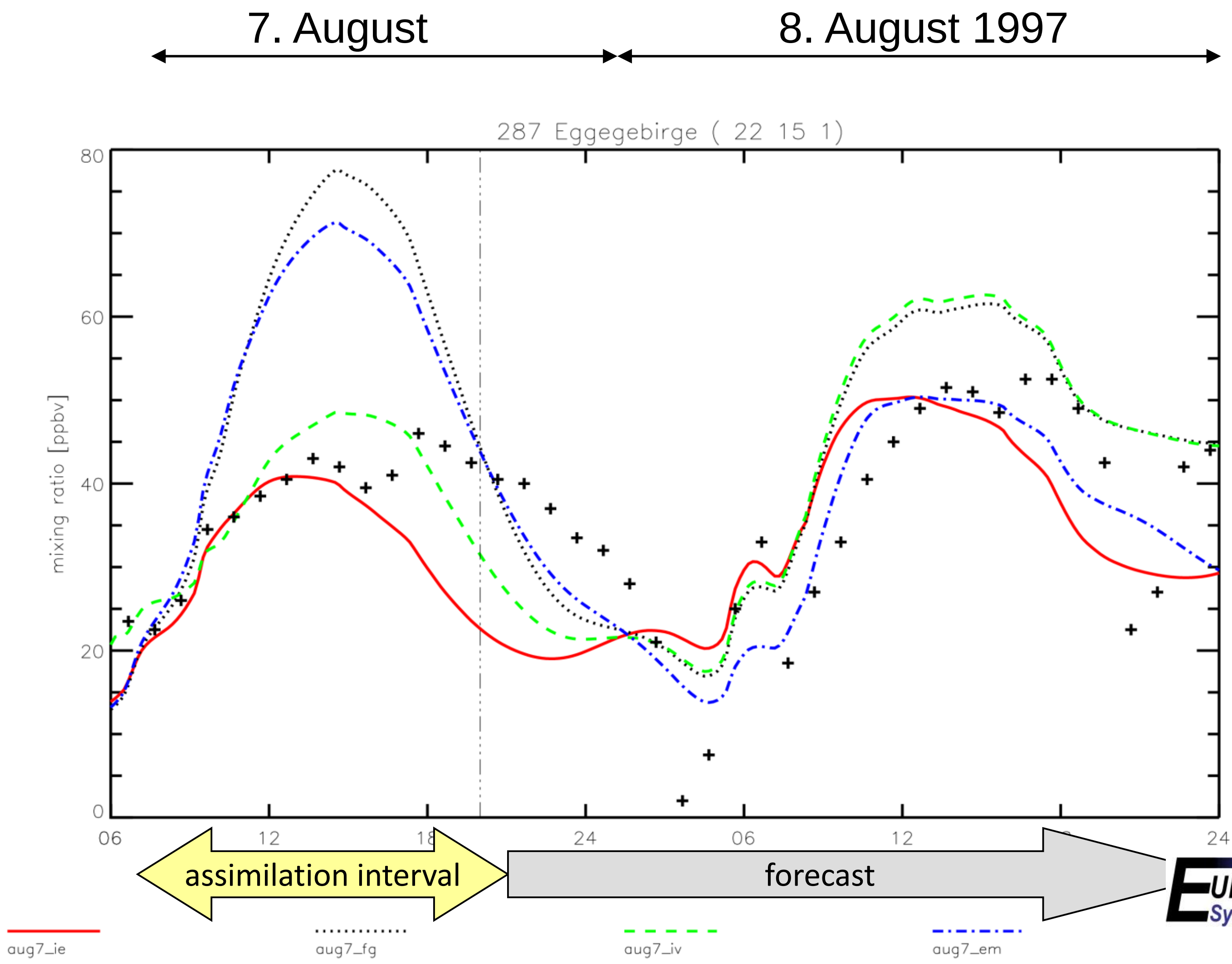


Observed and analysed ozone evolution at St. Poelten Vertical bars: ozone observations with error estimates.

- - - - -	Control run without data assimilation.
......	initial value optimisation.
-.-.-.-.-.	emission factor optimisation.
______	joint initial value and emission factor optimisation (Strunk et al., 2011)

4. Focus: joint emission rate initial value optimisation

Semi-rural measurement site **Eggegebirge**



+ observations

no optimisation

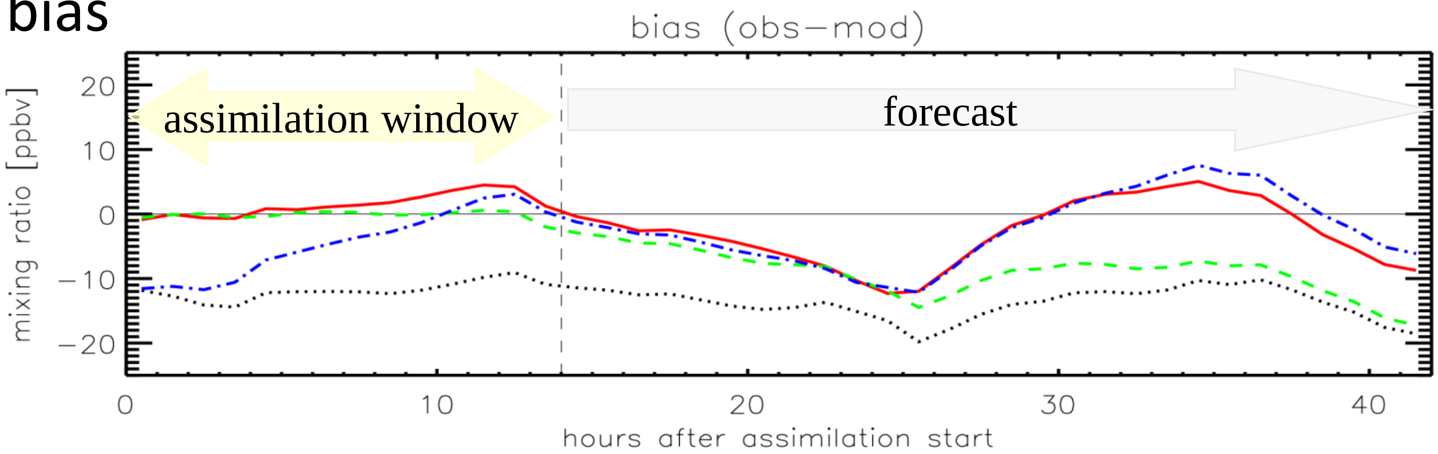
initial value opt.

emis. rate opt.

joint emis + ini val opt.

How long does data assimilation have an impact?
Answer gas phase
12-24 hours, dependent on optimisation

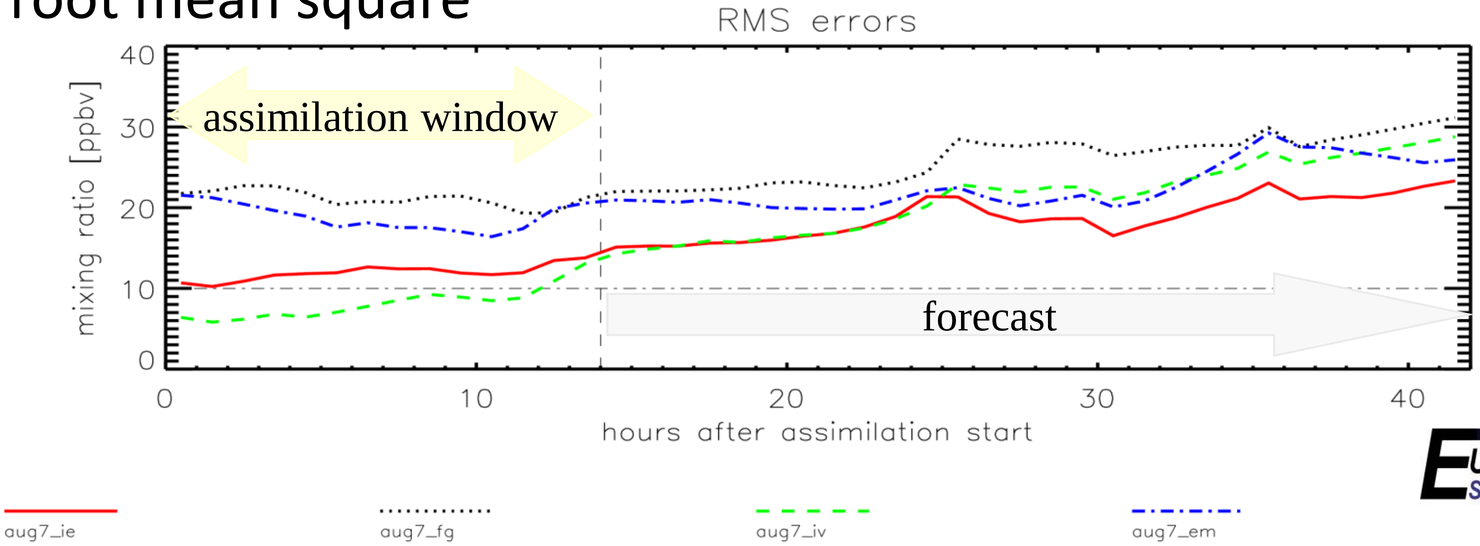
bias



+ observations
no optimisation

initial value opt.

root mean square



emis. rate opt.

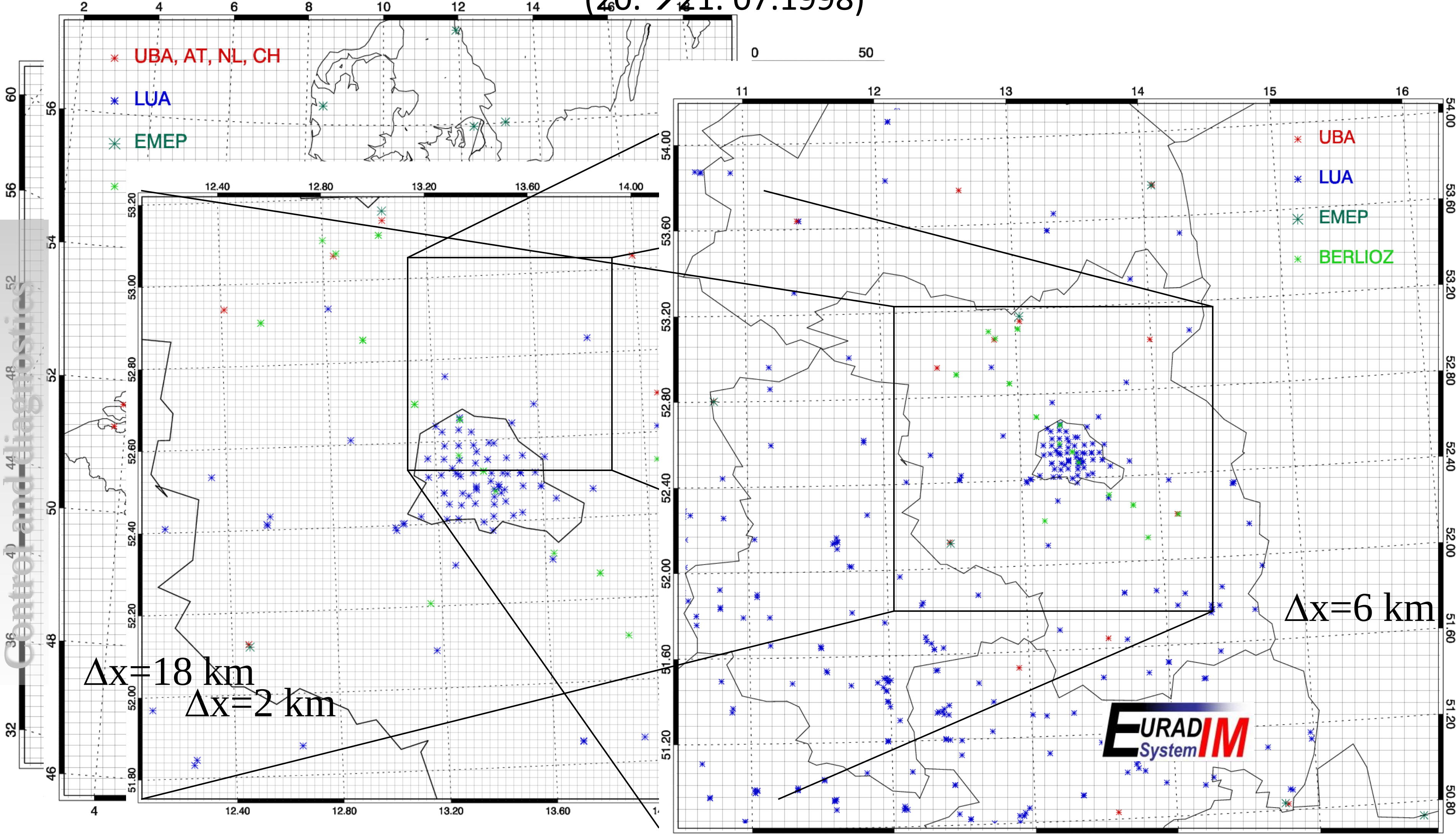
joint emis +
ini val opt.

EURAD IM System

Which is the requested resolution?

BERLIOZ grid designs and observational sites

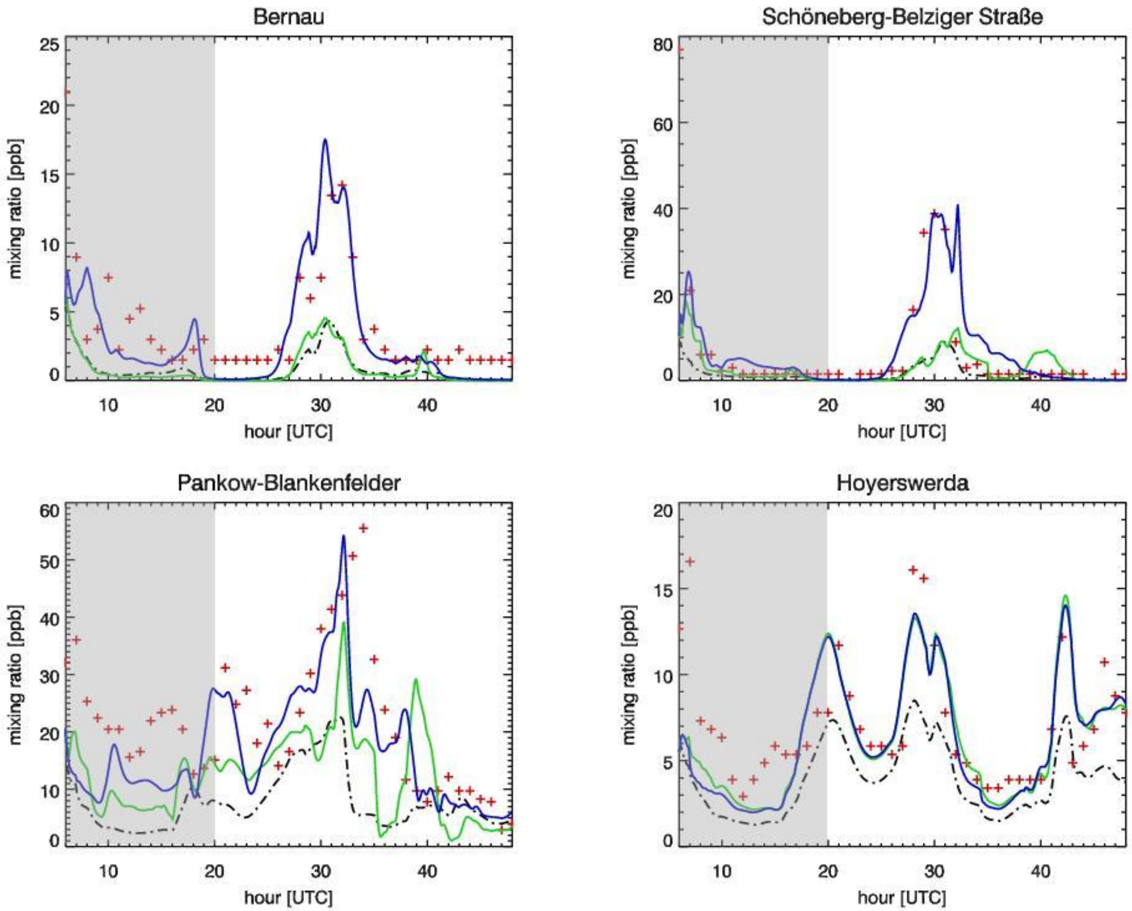
(20.→21. 07.1998)



Some BERLIOZ examples of NOx assimilation (20.→21. 07.1998)

NO

NO_2



Bernau

Schöneberg-Belziger Straße

Pankow-Blankenfelder

Hoyerswerda

Time series for selected NOx stations on nest 2.

+ observations,

-- - no assimilation,

-_____ N1 assimilation (**18 km**),

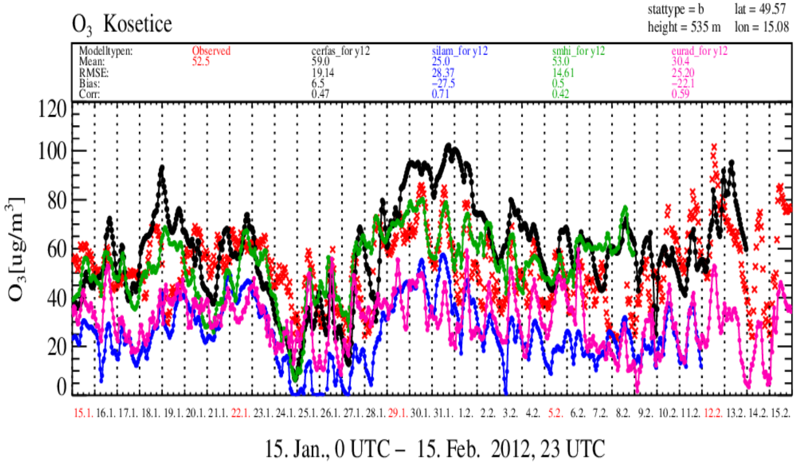
-_____N2 assimilation.(**6 km**),

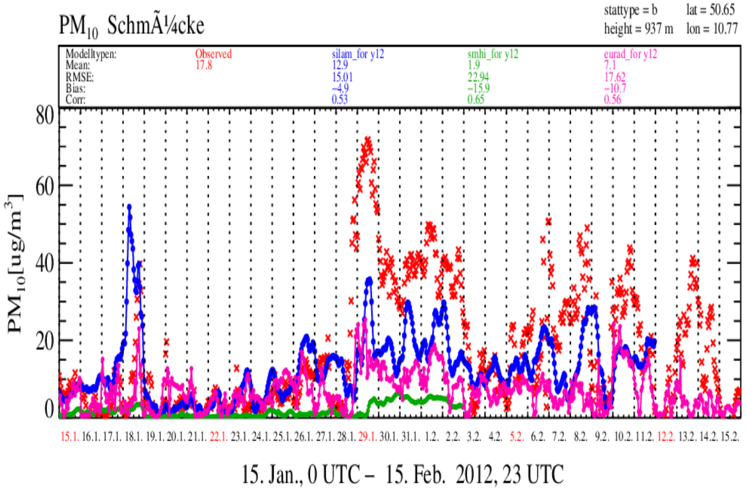
-grey shading: assimilated observations, others forecasted.

Validation by measurements withheld

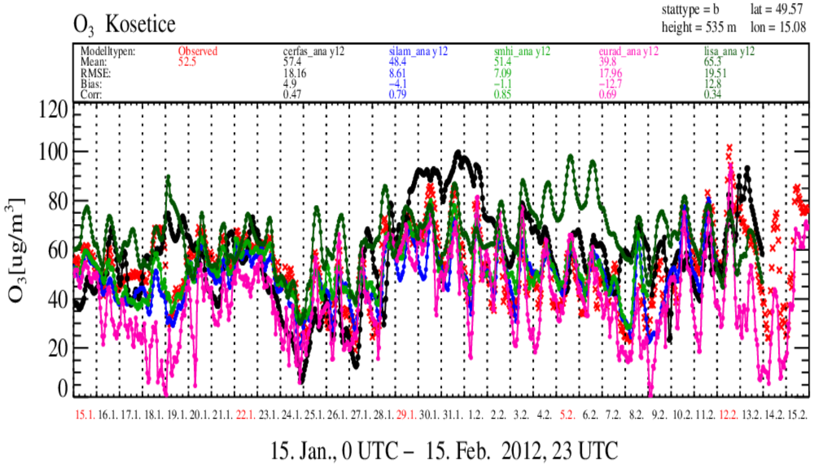
(extract from MACC III EDA report draft)

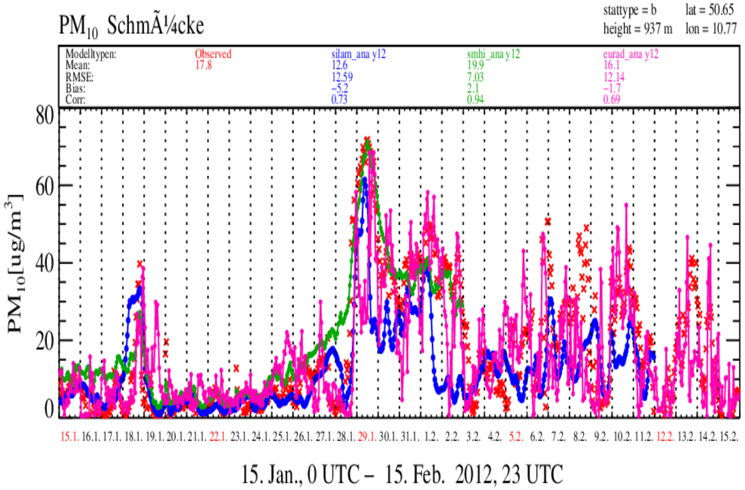
Forecast



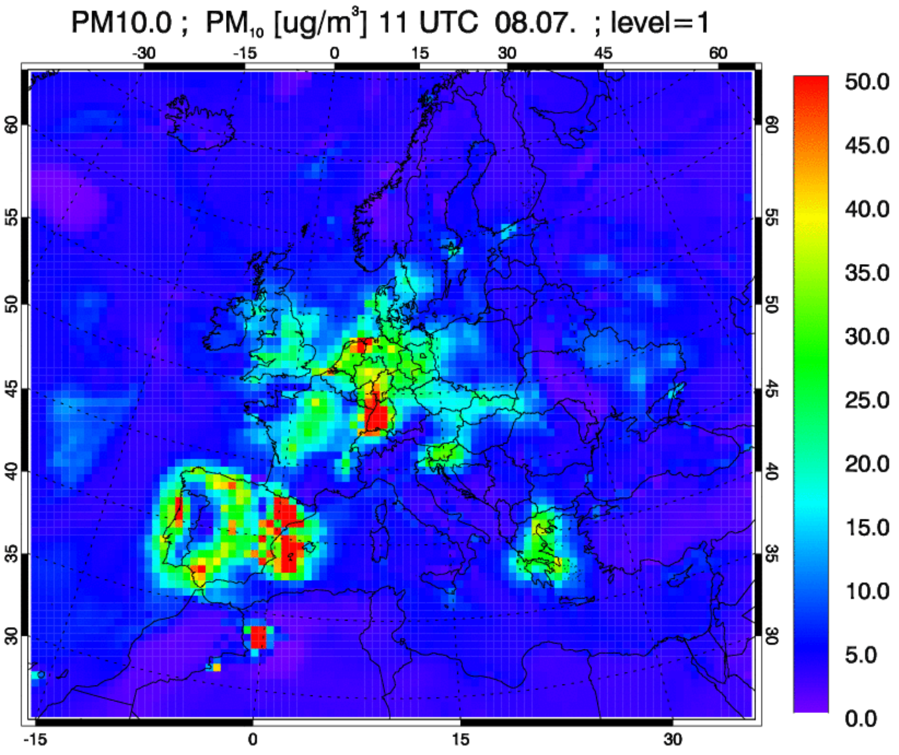


Analyses



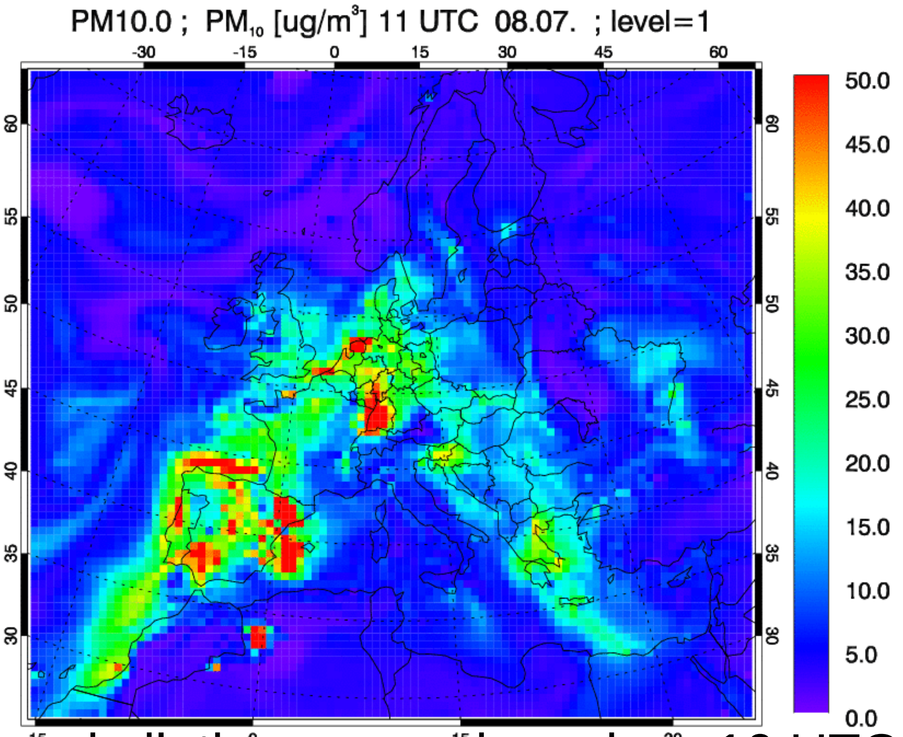


How long does data assimilation have an impact? Answer aerosol phase aerosol data assimilation effects accumulate



No previous assimilation

only 1 day: 14. July 2003



assimilation on previous days 10 UTC

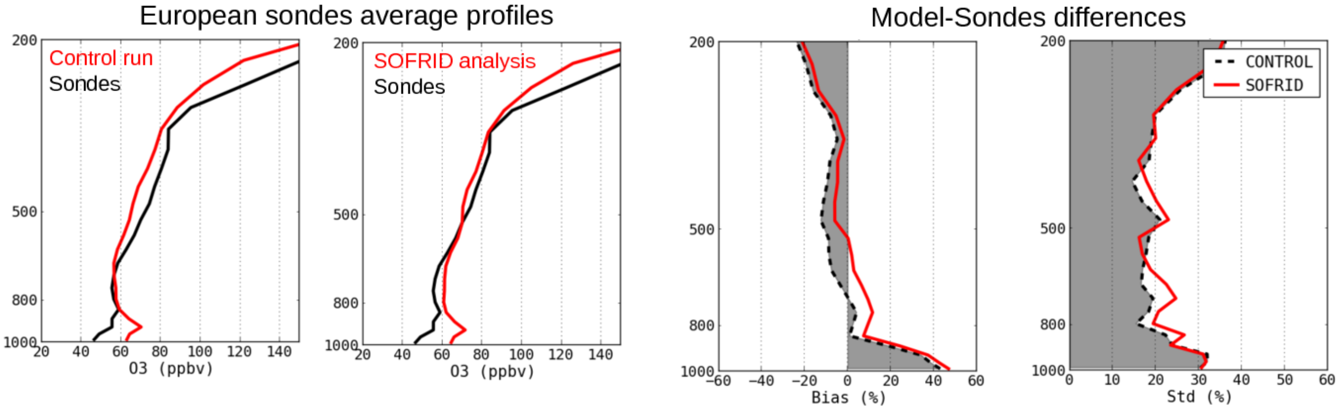
Accumulation of retrieval information over **14 days**

MOCAGE satellite data assimilation: IASI SOFRID O_3 re-analysis (CERFACS)

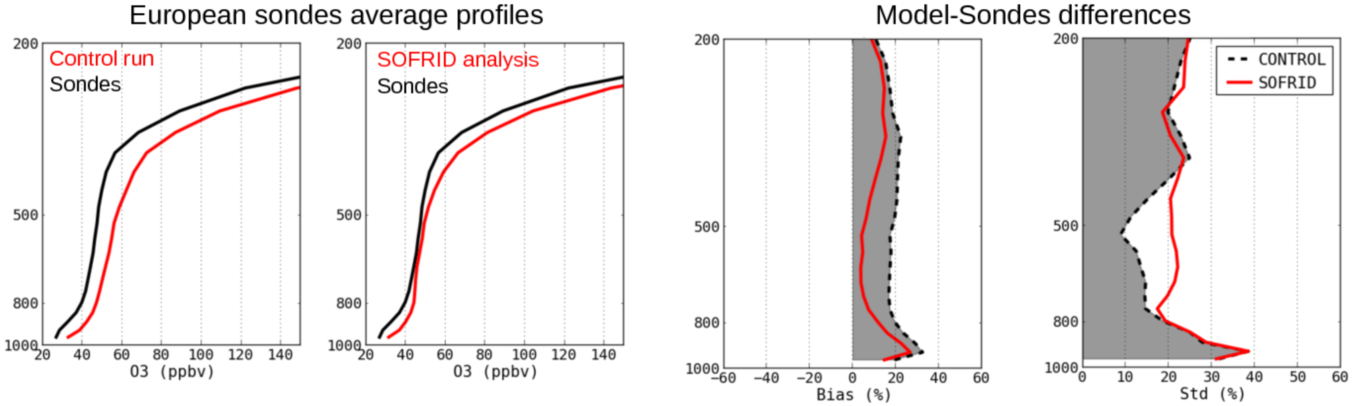
Validation of IASI analysis with ozonesonde data:

BIAS = model minus observations

O_3 profiles in July 2010:



O_3 profiles in Jan 2012:



Courtesy E. Emili, CERFACS

- Bias reduced in the free troposphere
- Surface ozone impact is minor
- MOZAIC-IAGOS as additional validation? (only 2012 available)

4. A posteriori Validation: Is the analysis consistent?

a posteriori validation of data assimilation results

Assumptions:

- Gaussian error distribution assumption sufficiently valid
- First guess not too far from “solution” (tangent-linear approximation must hold)
- A priori defined error covariances (background, observations)

Necessary condition for a posteriori validation:
adjust B and R such that:

at the minimum:

$$J_{min} = 1/2 d^T (\mathbf{HBH^T} + \mathbf{R})^{-1} d$$

$d := y - Hx^a$

p number of observations

Expectation

$$\mathcal{E}[J_{min}] = p/2$$

Variance

$$\mathcal{V}[J_{min}] = p/2$$

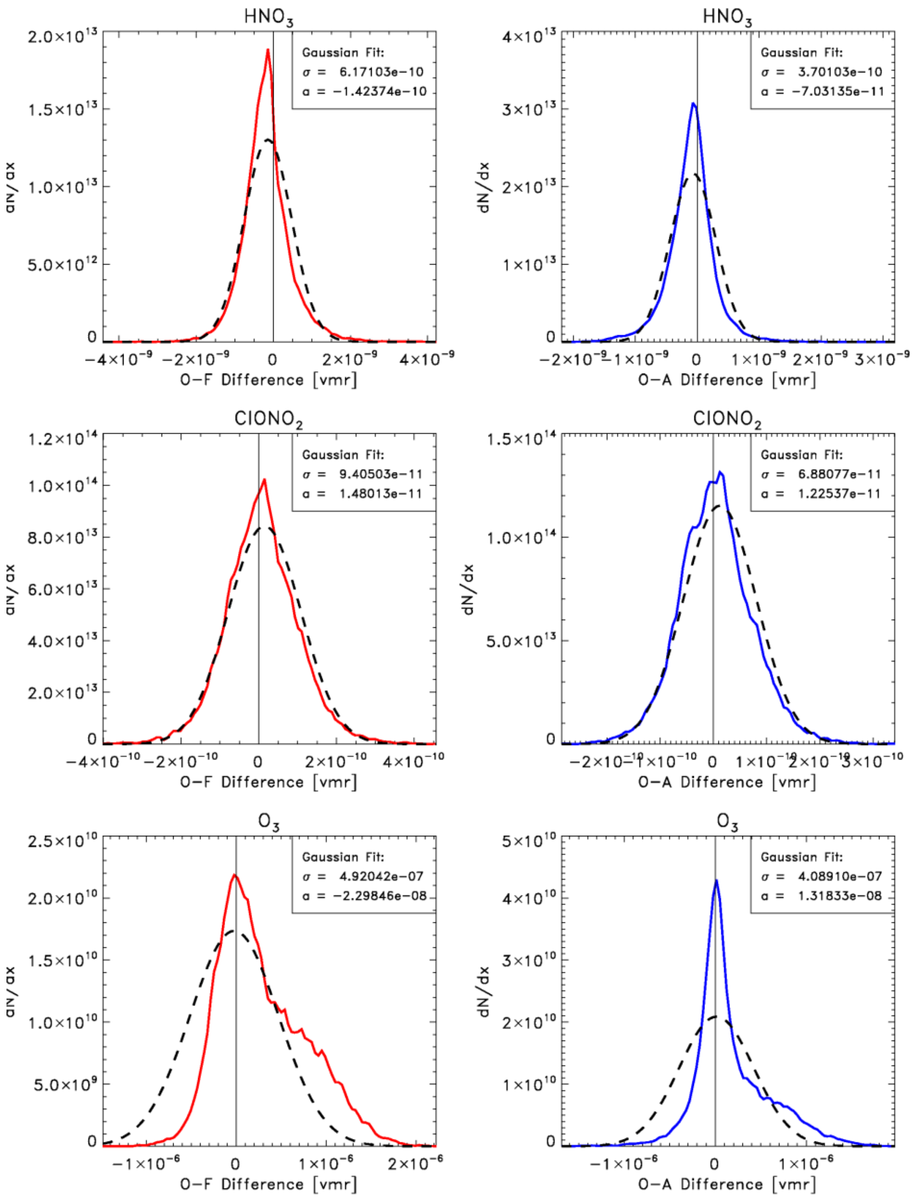
Evaluating the Gaussian error distribution assumption

SACADA

O-F differences (left column) and

O-A differences (right column)

Dotted line represents a Gaussian with same variance as the data

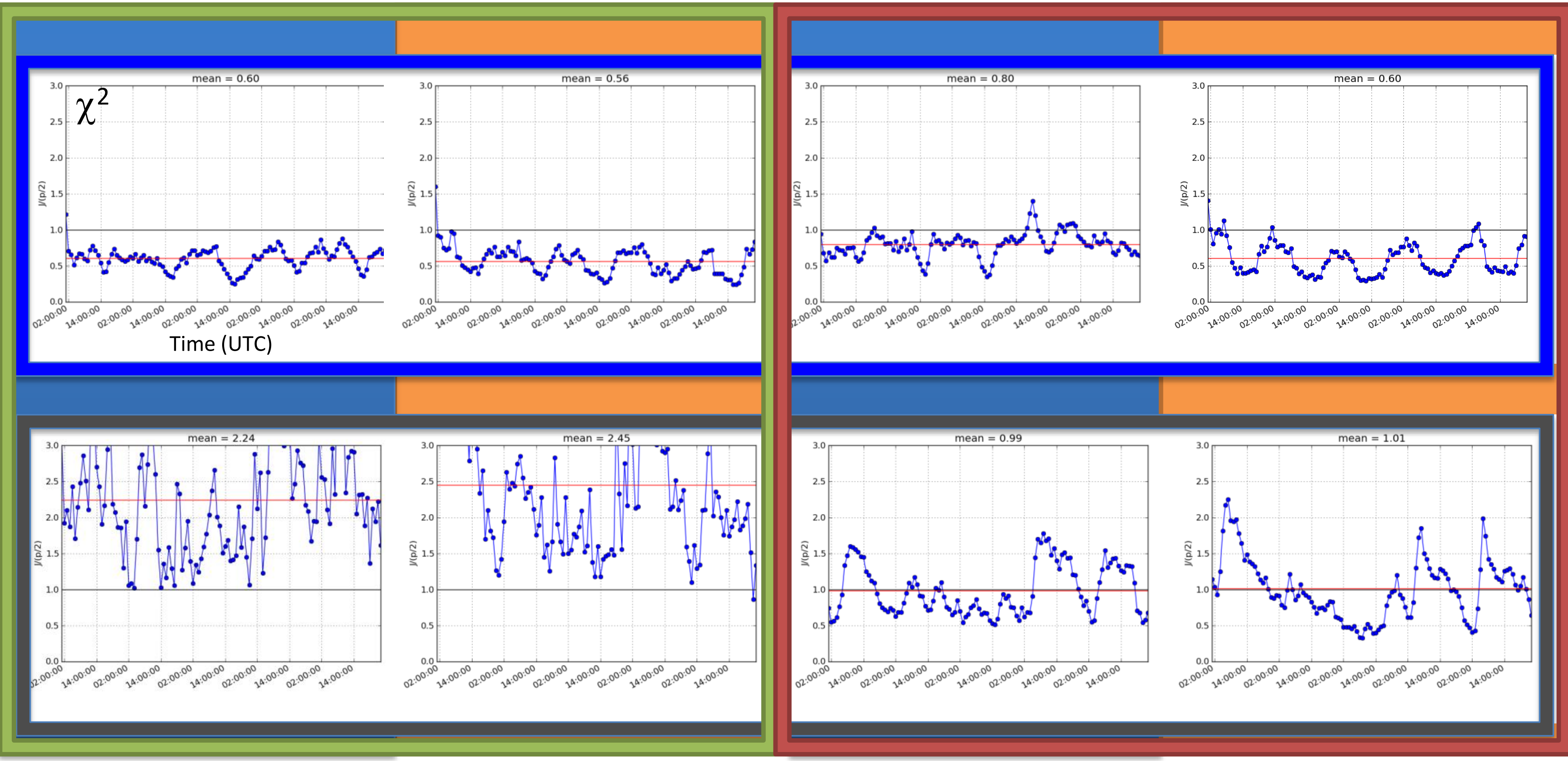


HNO_3

$ClONO_2$

O_3

χ^2 validation MOCAGE



🟦	Surface O_3 assimilated
⬛	Surface NO_2 assimilated
🟦	Winter period (1-2-2008, 6-8,2008)
🟧	Summer period (1-8-2008, 6-8-2008)
🟩	Only rural background sites assimilated
🟥	Only urban background sites assimilated

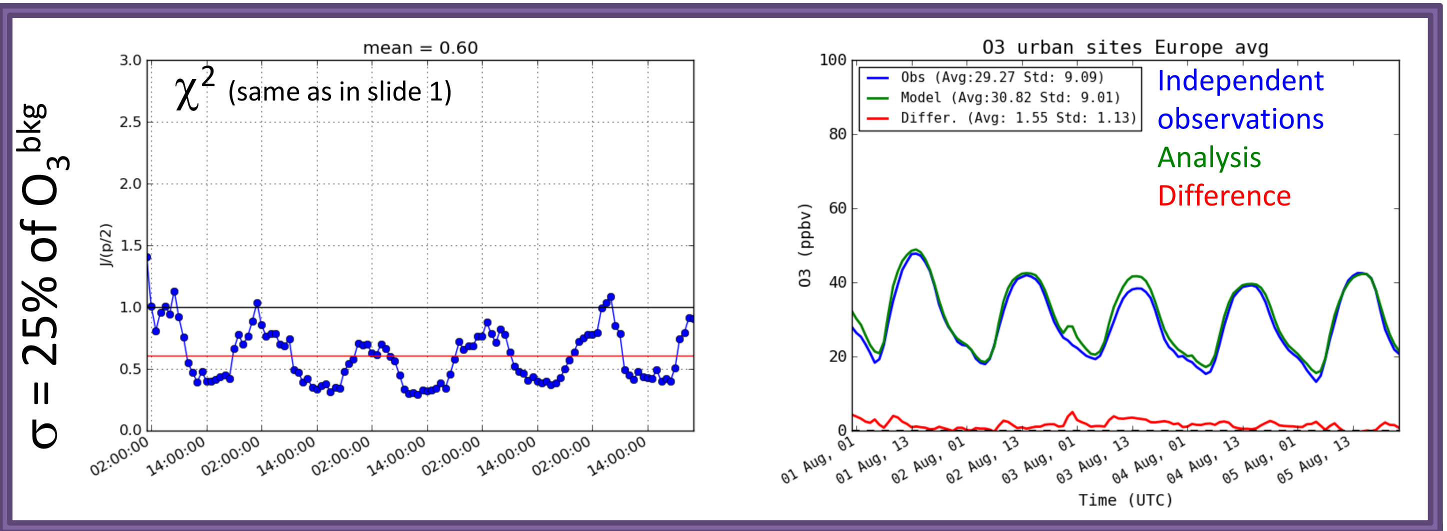
Courtesy E. Emili, CERFACS

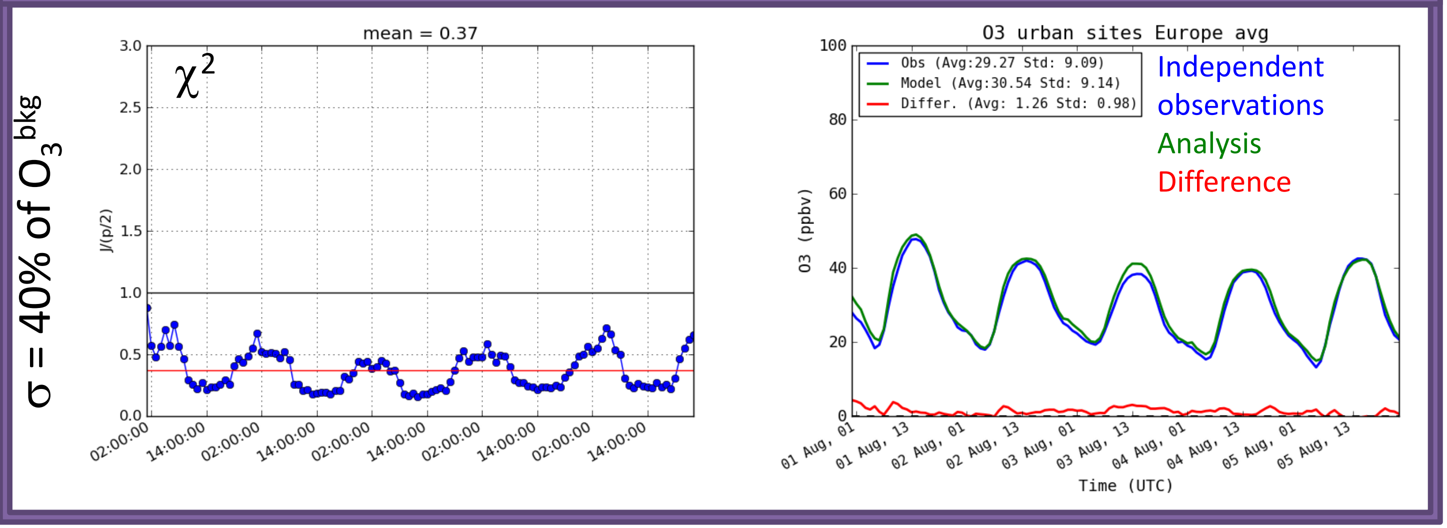
Comments:

- O_3
  - the urban case is the only case with a distinct winter-summer behavior (higher χ^2 in winter)
  - presence of diurnal variability in all cases
- NO_2
  - large differences between rural/urban cases
  - strong variations in the rural case
  - presence of diurnal variability in all cases
  - no evidence of significant seasonality

χ^2 validation MOCAGE

What is the impact of a low χ^2 in terms of validation with an independent dataset?
Example: O_3 background urban sites assimilated in summer, validation against sites kept out from the assimilation, two choices of the background error variance σ





Comments:
Case 2 (σ = 40%) has lower χ^2 but better analysis scores. A better χ^2 does not always imply a better analysis, because χ^2 stats do not consider model biases.

Conclusions

- Atmospheric chemistry is a highly coupled nonlinear dynamic system, which is best adressed by spatio-temporal data assimilation
- the system must be observed with respect to ist sensitivity (NOx-VOX interaction)
- Forecasts must be shown to improve
- the assimilation result must be consistent: proper baöance between a priori and a posteriori knowledge (χ^2-validation)

Additional illustrations

2. Focus: Can we identify flaws? A posteriori evaluation

1. χ^2 – validation

2. a posteriori validation in observation space

Theoretical baclground on a posteriori evaluation

$$J_{min} = \frac{1}{2} d^T E \breve{\ } \; d \; d^T \quad d$$

$$E(J_{min}) = \frac{p}{2}$$

Aposteriori validation in observation space

Extended Kalman filter equations

Forecast step: $\boldsymbol{x}^{\mathrm{b}}(t_i) = M_{i-1}\left[x^{\mathrm{a}}(t_{i-1})\right]$

$$\mathbf{B}(t_i) = \mathbf{L}_{i-1}\mathbf{A}(t_{i-1})\mathbf{L}_{i-1}^{T} + \mathbf{Q}(t_{i-1})$$

Analysis step: $\boldsymbol{x}^{\mathrm{a}}(t_i) = \boldsymbol{x}^{\mathrm{b}}(t_i) + \mathbf{K}(t_i)\left(\boldsymbol{y} - H\left[\boldsymbol{x}^{\mathrm{b}}(t_i)\right]\right)$

$$\mathbf{A}(t_i) = \left(\mathbf{I} - \mathbf{K}(t_i)\mathbf{H}\right)\mathbf{B}(t_i)$$

where

$M_i :=$ Model operator

$\mathbf{L}_i :=$ Tangent linear model operator

$\mathbf{K}(t_i) := \mathbf{B}(t_i)\mathbf{H}^{T}\left[\mathbf{R} + \mathbf{H}\mathbf{B}(t_i)\mathbf{H}^{T}\right]^{-1}$

$\mathbf{Q}(t_i) :=$ Model error covariance matrix

$\mathbf{B}(t_i) :=$ Background error covariance matrix

$\mathbf{A}(t_i) :=$ Analysis error covariance matrix

$\mathbf{R}(t_i) :=$ Observation error covariance matrix

optimize R and B directly,
and A indirectly

2. Focus: a posteriori validation

Diagnosis and Tuning of Error Covariances

(Desroziers et al. 2005)

$$E\left\{\boldsymbol{d}^{\mathrm{a}}_{\mathrm{b}}\boldsymbol{d}^{\mathrm{o}T}_{\mathrm{b}}\right\} = \mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^{T}$$

makes the difference

$$E\left\{\boldsymbol{d}^{\mathrm{o}}_{\mathrm{a}}\boldsymbol{d}^{\mathrm{o}T}_{\mathrm{b}}\right\} = \tilde{\mathbf{R}}$$

$$\begin{aligned} \boldsymbol{d}^{\mathrm{a}}_{\mathrm{b}} &:= H(\boldsymbol{x}^{\mathrm{a}}) - H(\boldsymbol{x}^{\mathrm{b}}) \\ \boldsymbol{d}^{\mathrm{o}}_{\mathrm{a}} &:= \boldsymbol{y} - H(\boldsymbol{x}^{\mathrm{a}}) \\ \boldsymbol{d}^{\mathrm{o}}_{\mathrm{b}} &:= \boldsymbol{y} - H(\boldsymbol{x}^{\mathrm{b}}) \end{aligned}$$

If $\mathbf{B}$ and $\mathbf{R}$ are **consistently** specified, then $\mathbf{B} = \tilde{\mathbf{B}}$ and $\mathbf{R} = \tilde{\mathbf{R}}$ and

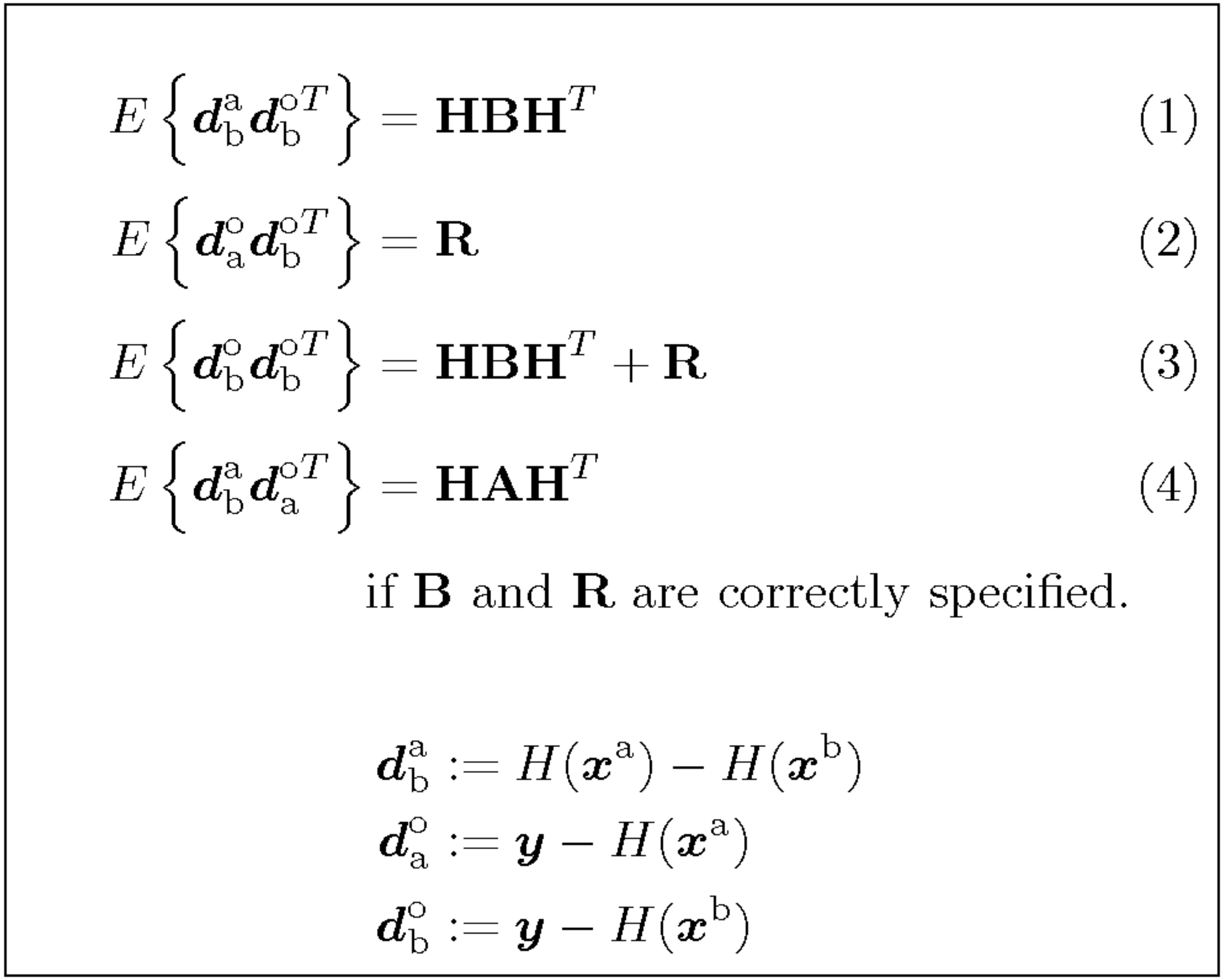
$$E\left\{\boldsymbol{d}^{\mathrm{a}}_{\mathrm{b}}\boldsymbol{d}^{\mathrm{o}T}_{\mathrm{a}}\right\} = \mathbf{H}\mathbf{A}\mathbf{H}^{T}$$

Only a necessary, but not a sufficient condition is fulfilled: no unique solution

2. Focus: a posteriori validation

Tuning of Error Covariances in observation space

(Desroziers et al. 2005)



$$E\left\{\boldsymbol{d}_{\mathrm{b}}^{\mathrm{a}}\boldsymbol{d}_{\mathrm{b}}^{\mathrm{o}T}\right\} = \mathbf{HBH}^T \tag{1}$$

$$E\left\{\boldsymbol{d}_{\mathrm{a}}^{\mathrm{o}}\boldsymbol{d}_{\mathrm{b}}^{\mathrm{o}T}\right\} = \mathbf{R} \tag{2}$$

$$E\left\{\boldsymbol{d}_{\mathrm{b}}^{\mathrm{o}}\boldsymbol{d}_{\mathrm{b}}^{\mathrm{o}T}\right\} = \mathbf{HBH}^T + \mathbf{R} \tag{3}$$

$$E\left\{\boldsymbol{d}_{\mathrm{b}}^{\mathrm{a}}\boldsymbol{d}_{\mathrm{a}}^{\mathrm{o}T}\right\} = \mathbf{HAH}^T \tag{4}$$

if $\mathbf{B}$ and $\mathbf{R}$ are correctly specified.

$$\begin{aligned}
\boldsymbol{d}_{\mathrm{b}}^{\mathrm{a}} &:= H(\boldsymbol{x}^{\mathrm{a}}) - H(\boldsymbol{x}^{\mathrm{b}}) \\
\boldsymbol{d}_{\mathrm{a}}^{\mathrm{o}} &:= \boldsymbol{y} - H(\boldsymbol{x}^{\mathrm{a}}) \\
\boldsymbol{d}_{\mathrm{b}}^{\mathrm{o}} &:= \boldsymbol{y} - H(\boldsymbol{x}^{\mathrm{b}})
\end{aligned}$$

in practice: Iterative approach

2. Focus: a posteriori validation

Practical estimate of diagonal elements of R and B

$$
\begin{aligned}
(\tilde{\sigma}_i^{\mathrm{b}})^2 &= (\mathbf{d}_{\mathrm{b}}^{\mathrm{a}})_i^{\mathrm{T}}(\mathbf{d}_{\mathrm{b}}^{\mathrm{o}})_i = \textstyle\sum_{j=1}^{p_i}(\mathbf{y}_j^{\mathrm{a}} - \mathbf{y}_j^{\mathrm{b}})(\mathbf{y}_j^{\mathrm{o}} - \mathbf{y}_j^{\mathrm{b}})/p_i \\
(\tilde{\sigma}_i^{\mathrm{o}})^2 &= (\mathbf{d}_{\mathrm{a}}^{\mathrm{o}})_i^{\mathrm{T}}(\mathbf{d}_{\mathrm{b}}^{\mathrm{o}})_i = \textstyle\sum_{j=1}^{p_i}(\mathbf{y}_j^{\mathrm{o}} - \mathbf{y}_j^{\mathrm{a}})(\mathbf{y}_j^{\mathrm{o}} - \mathbf{y}_j^{\mathrm{b}})/p_i
\end{aligned}
$$

Estimate of off-diagonal elements of B

$$
(\tilde{\sigma}_{ij}^{\mathrm{b}})^2 = \sum_{\substack{i,j=1 \\ i\neq j}}^{p_{ij}} (\mathbf{y}_i^{\mathrm{a}} - \mathbf{y}_i^{\mathrm{b}})(\mathbf{y}_j^{\mathrm{o}} - \mathbf{y}_j^{\mathrm{b}})/p_{ij} \,,
$$

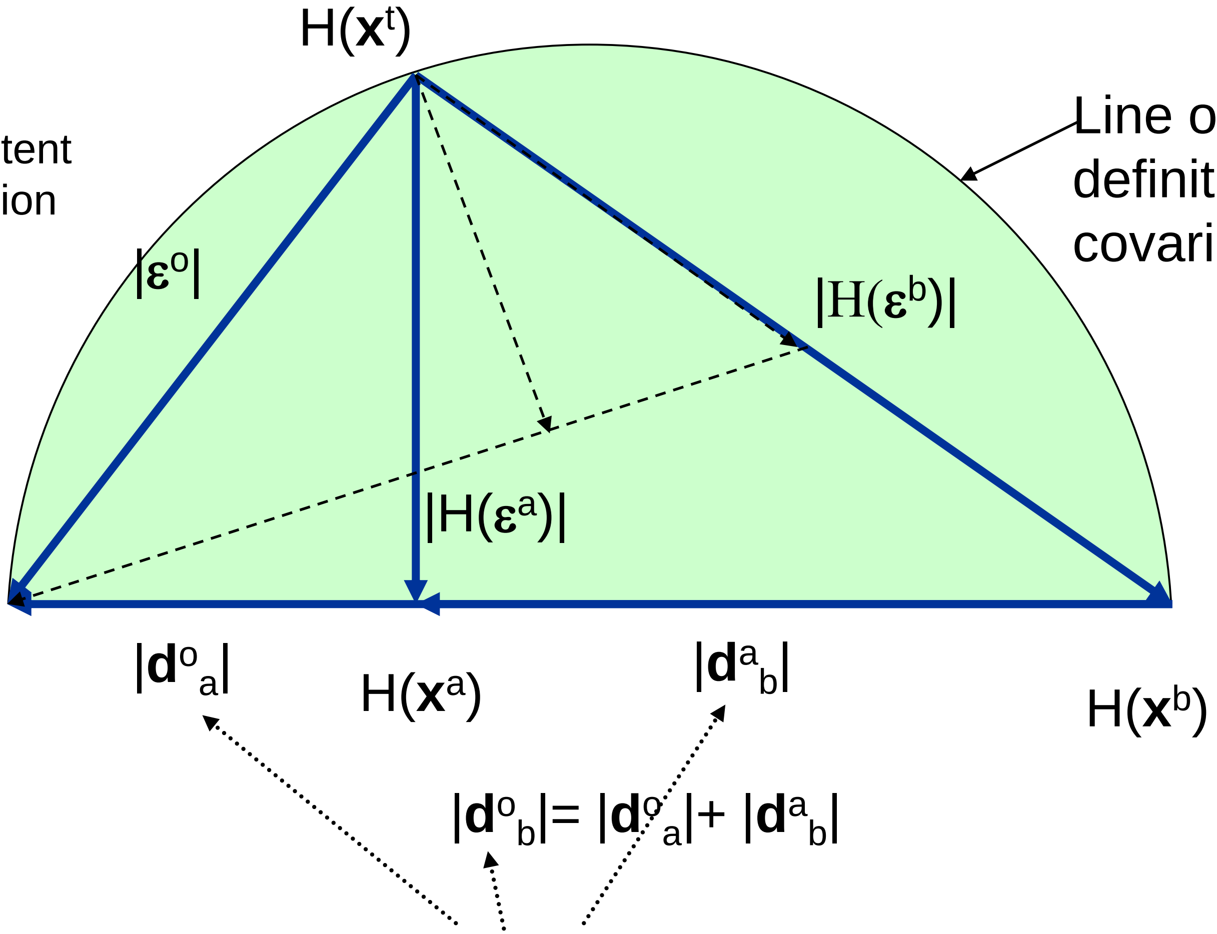
Applied only along orbits in observation space

Δt < 10 min

2. Focus: a posteriori validation

Geometrical representation of error components

inconsistent formulation


$H(\mathbf{x}^t)$
$|\boldsymbol{\varepsilon}^o|$
$|H(\boldsymbol{\varepsilon}^b)|$
$|H(\boldsymbol{\varepsilon}^a)|$
$|\mathbf{d}^o{}_a|$
$H(\mathbf{x}^a)$
$|\mathbf{d}^a{}_b|$
$H(\mathbf{x}^b)$

Line of consistent definition of error covariance matrices

$$|\mathbf{d}^o{}_b| = |\mathbf{d}^o{}_a| + |\mathbf{d}^a{}_b|$$

amenable for a posteriori check