ACC Frascati, 15.-17..6.2009

Chemical Data assimilation of satellite retrievals

H. Elbern[1,2], A. Strunk[1], E. Friese[1], L. Nieradzik[1]
[1]Rheinisches Institut für Umweltforschung
an der Universität zu Köln
and
[2]ICG-2, Research Centre Jülich

Contents

- Introduction: short reminder for the troposphere
- NO_2 trop. column assimilation
- PM assimilation

Objective of atmospheric data assimilation

The ambitious and elusive goal of data assimilation is to provide a ***dynamically consistent motion picture*** of the atmosphere and oceans, in three space dimensions, with known ***error bars***.

M. Ghil and P. Malanotte-Rizzoli (1991)

Terminology

Inverse Modelling

The inverse modelling problem consists of using the **actual** result of some **measurements** to **infer the values of the parameters** that characterize the system.

A. Tarantola (2005)

Objective of atmospheric data assimilation (2)

- "is to produce a regular, **physically consistent four dimensional** representation of the state of the atmosphere
- from a **heterogeneous** array of in situ and remote instruments
- which sample **imperfectly** and **irregularly** in space and time.

Data assimilation

- extracts the signal from noisy observations (**filtering**)
- interpolates in space and time (**interpolation**) and
- reconstructs state variables that are not sampled by the observation network (**completion**)." (Daley, 1997)

Characteristics in tropospheric chemistry data assimilation, mathematical viewpoints

- highly underdetermined system on 2 levels
  - (few observations (**y**) with respect to degrees of freedom of model: dim(**x**)>>dim(**y**))
  - scalar column value → profile vector

- regionally/locally highly nonlinear dynamics

- constraints by physical laws/models are insufficient

- assimilation or inversion problem to be solved?

Optimality criteria: Which property can be attributed to our analysis result?

(Need for quantification)

- maximum likelyhood:
  - maximum of probability density function
- minimal variance: l_2 norm
  - parameters optimal, for which analysis error spread is minimal (for Gaussian/normal and log-normal error distributions), Best Linear Unbiased Estimate (**BLUE**)
- minimax norm (discrete cases)
- maximum entropy

Optimal Interpolation

$$J(\mathbf{x}) = \frac{1}{2}[\mathbf{x}^b - \mathbf{x}]^T \mathbf{B}_0^{-1}[\mathbf{x}^b - \mathbf{x}] + \frac{1}{2}\left\{\mathbf{y}^0 - H[\mathbf{x}(t)]\right\}^T \mathbf{R}^{-1}\left\{\mathbf{y}^0 - H[\mathbf{x}]\right\}$$

The gradient then reads

$$\nabla J(\mathbf{x}) = \mathbf{B}_0^{-1}[\mathbf{x}^b - \mathbf{x}] + H^T \mathbf{R}^{-1}\left\{\mathbf{y}^0 - H[\mathbf{x} + (\mathbf{x}_b - \mathbf{x}_b)]\right\}$$

where a trivial expanion is introduced for later manipulation.

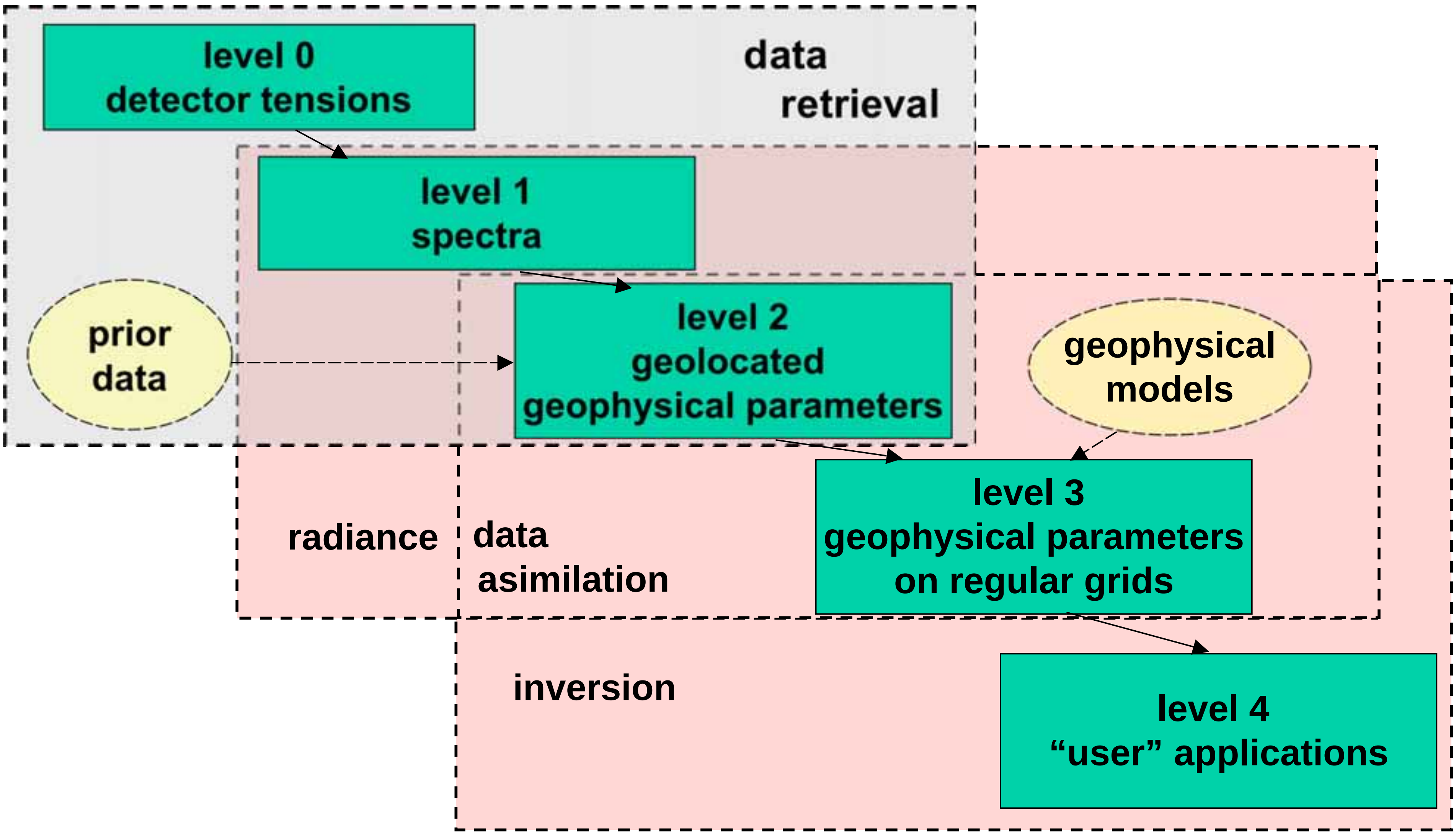
At the minimum $\mathbf{x} =: \mathbf{x}_a$

$$\begin{aligned}\mathbf{x}_a - \mathbf{x}_b &= (\mathbf{B}_0^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}\left\{\mathbf{y}^0 - H[\mathbf{x}_b]\right\} \\ &= \boxed{\mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}^T\mathbf{B}\mathbf{H})^{-1}\left\{\mathbf{y}^0 - H[\mathbf{x}_b]\right\}}\end{aligned}$$

with the latter result obtained after some manipulation.

Notation: Ide, K., P. Courtier, M. Ghil, and A. Lorenc,
Unified notation for data assimilation: operational sequential and variational,
J. Met. Soc. Jap., 75, 181--189, 1997.

DA in the satellite data application chain



Advanced spatio-temporal methods used in tropospheric chemistry data assimilation

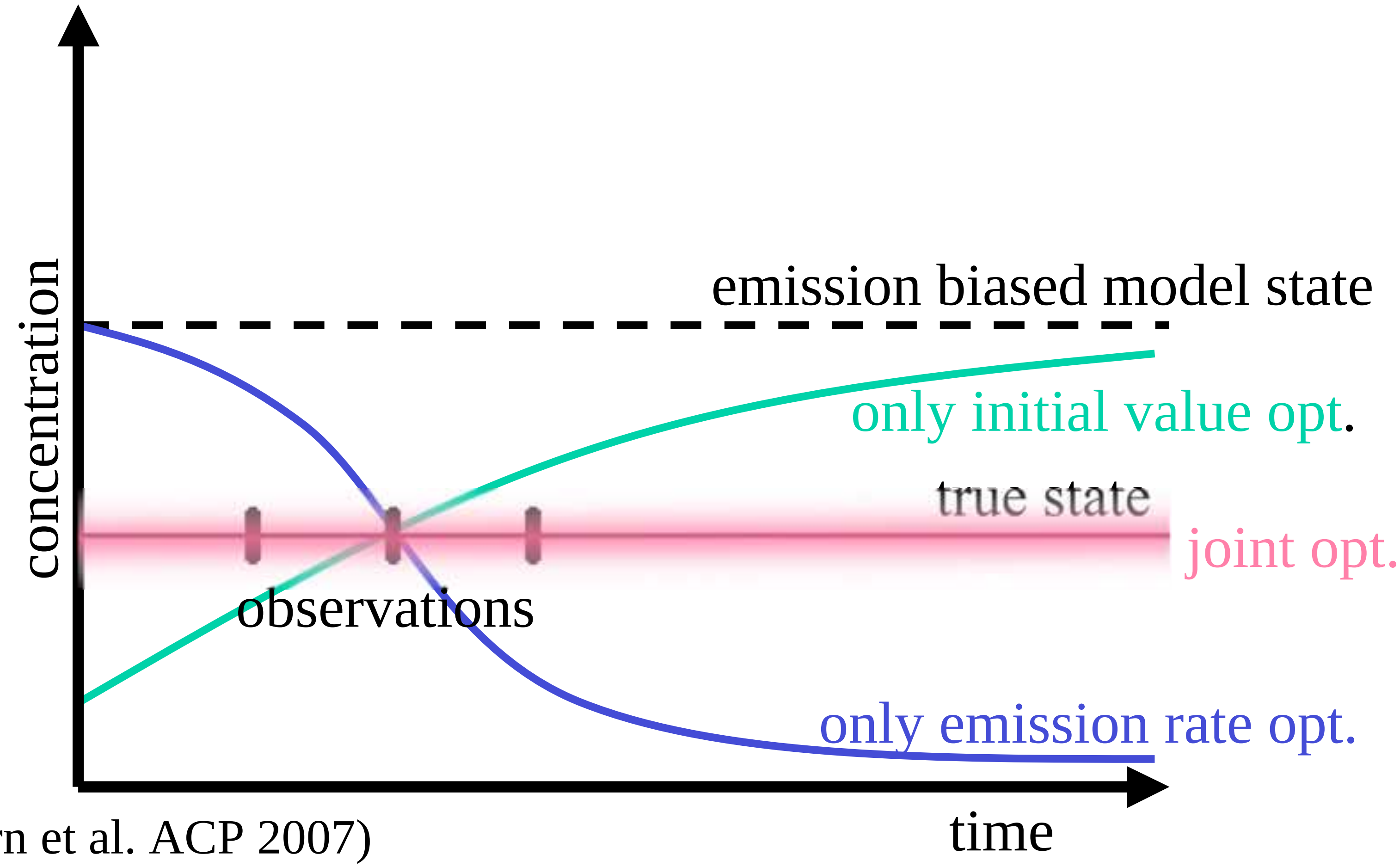
Spacio-temporal BLUEs applied in tropospheric chemistry data assimilation include:

- 4D var:
  - with EURAD (Elbern and Schmidt, 1999, 2001, Elbern et al., 2007),
  - with STEM 2K1 (Chai et al., 2006).
- Kalman Filter
  - with LOTOS model (van Loon et al, 2000), (RRSQR)
  - with EUROS model (Hanea et al. 2004) (En+RRSQRKF)

Question: Which parameter to be optimized?

Hypothesis:

initial state and **emission rates** are least known for a skillful forecast



(Elbern et al. ACP 2007)

In the troposphere, for **emission rates,** the product (*paucity of knowledge*importance)* is high

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
$H[\ \ ]$	forward interpolator
$\mathbf{y}^0(t)$	observation at time t
$\mathbf{B}_0$	background error covariance matrix

Adjoint integration "backward in time"

How to make the parameters of resolvents *i* $\mathbf{M}(t_{i-1},t_i)$ available in ***reverse*** order??

direct model

$$\frac{d\mathbf{x}}{dt} = \mathcal{M}(\mathbf{x})$$

tangent linear model

$$\delta\mathbf{x}(t_n) = \mathbf{M}'(t_n,t_0)\delta\mathbf{x}(t_0) = \prod_{i=n}^{1} \mathbf{M}'(t_i,t_{i-1})\delta\mathbf{x}(t_0)$$

adjoint model

$$-\frac{d\delta\mathbf{x}^*(t)}{dt} - \mathbf{M}'^T\delta\mathbf{x}^*(t) = \mathbf{R}^{-1}(\mathbf{y}^0(t) - H[\mathbf{x}(t)]).$$

gradient of the cost function

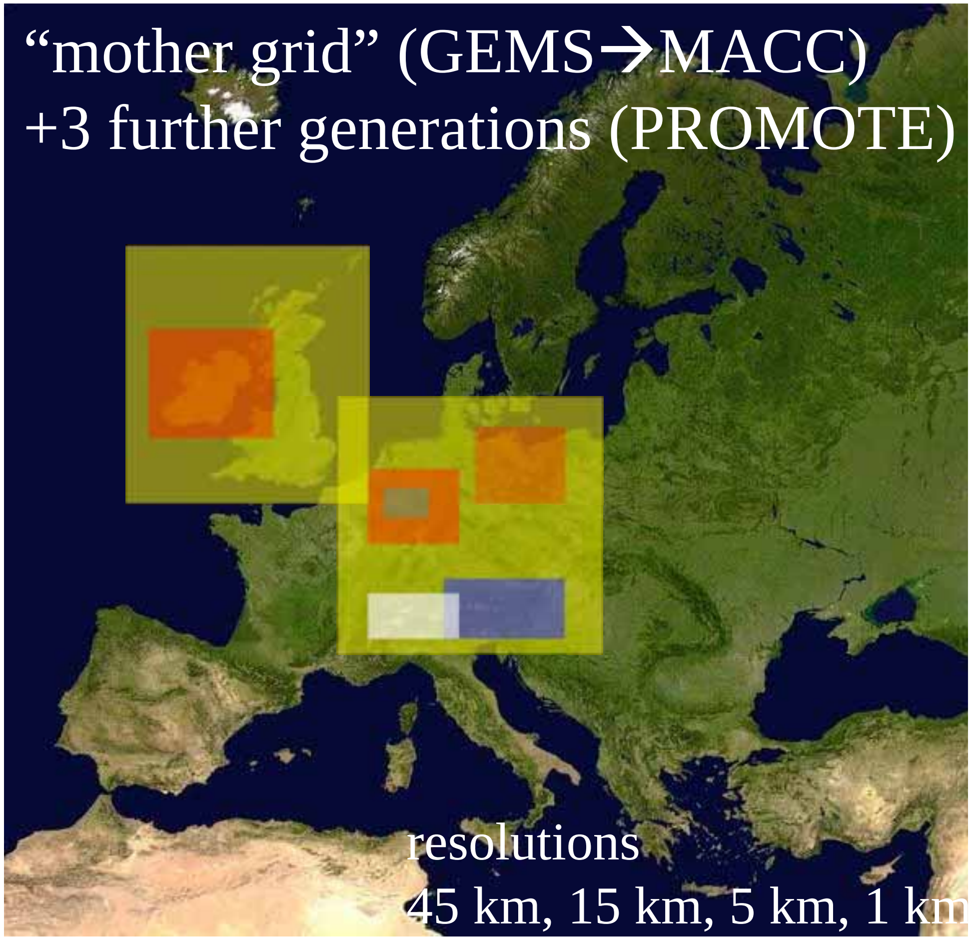
$$\nabla_{[\mathbf{x}(t_0),\mathbf{e}]}J = -\mathbf{B}_0^{-1}(\mathbf{x}^b(t_0) - \mathbf{x}(t_0)) - \mathbf{K}^{-1}(\mathbf{e}^b(t) - \mathbf{e}(t)) - \Sigma_{m=0}^{N} \Pi_{i=1}^{m} \mathbf{M}^T(t_{i-1},t_i)\mathbf{R}^{-1}(\mathbf{y}^0(t_m) - H[\mathbf{x}(t_m)])$$

Find minimum of $J(\mathbf{x}(t_0),\mathbf{e})$ with $\nabla_{[\mathbf{x}(t_0),\mathbf{e}]}J$ by use of a minimization routine

EURAD-IM
4D-var system (1)

EURAD-IM adjoints

- RACM
- implicit vertical diffusion
- explicit horizontal diffusion
- Bott 4th order advection
- emissions: EMEP, TNO



- MADE, SORGAM adjoint version under way

EURAD-IM
4D-var system (2)

- horizontal and vertical covariances: full anisotropy and inhomogeneity available by diffusion approach (Weaver and Courtier, 2001)
- preconditioning: options logarithmic, square root diffusion operator
- minimisation quasi-Newton by L-BFGS

Formulation of the background error covariance matrix:

Diffusion paradigm (Weaver and Courtier, 2001)

4D var needs the square root of the background error covariance matrix **B** (**O**=10^{12}):

Basic idea:

1. formulate covariances by Gaussians
2. approximate Gaussians by integration of the diffusion operator over time T
3. calculate $B^{1/2}$ by integration over time T/2 (comp. cheap), and
4. intermittent normalisation (comp. more challenging)

$$\mathcal{C} : \eta(z,0) \rightarrow (4\pi\kappa T)^{1/2}\eta(z,T)$$

with

$$\eta(z,T) = (4\pi\kappa T)^{-1/2} \int_{z'} \exp\left(-\frac{(z-z')^2}{4\kappa T}\right) \eta(z',0)dz'$$

and radius of influence

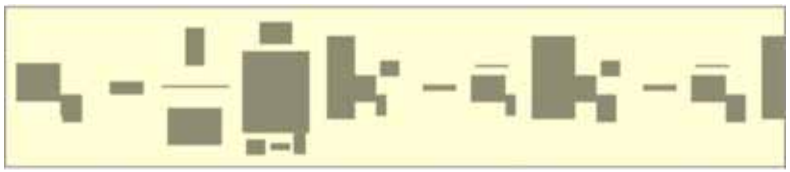
$$L^2 = 4\kappa T$$

Background Error Covariance Matrix B
(short design outline)

1. How to obtain the covariances?

K=# Ensembles; i,j neighboring cells

Ensemble/NMC approach:



2. How to process this information?

Translate into Diffusion coefficients→difusion paradigma

Correlation length L to neighboring gridcell:

$$B(r) = B(0)\exp\left(-\frac{r^2}{2L^2}\right), r = 1, B(1) = B_{ij}, B(0) = 1/2(B_{ii} + B_{jj})$$

$$\Rightarrow L = \left(2\ln\left(\frac{B(0)}{B(1)}\right)\right)^{-1/2}$$

diffusion coefficients κ: $L = \sqrt{2\kappa T}$

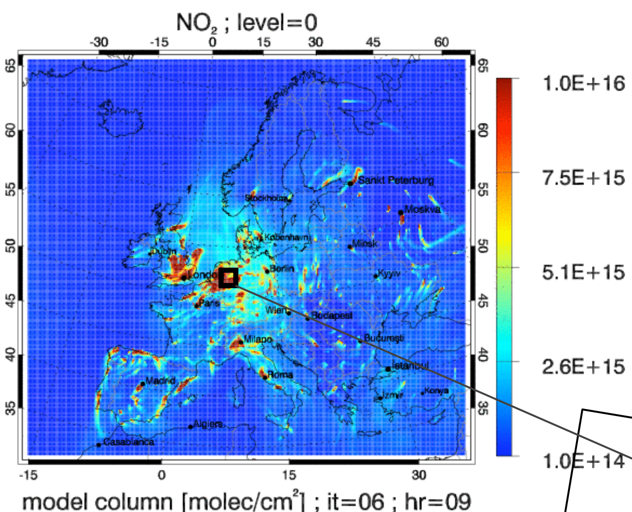
GLOBMODEL case study
NO2 column focussed

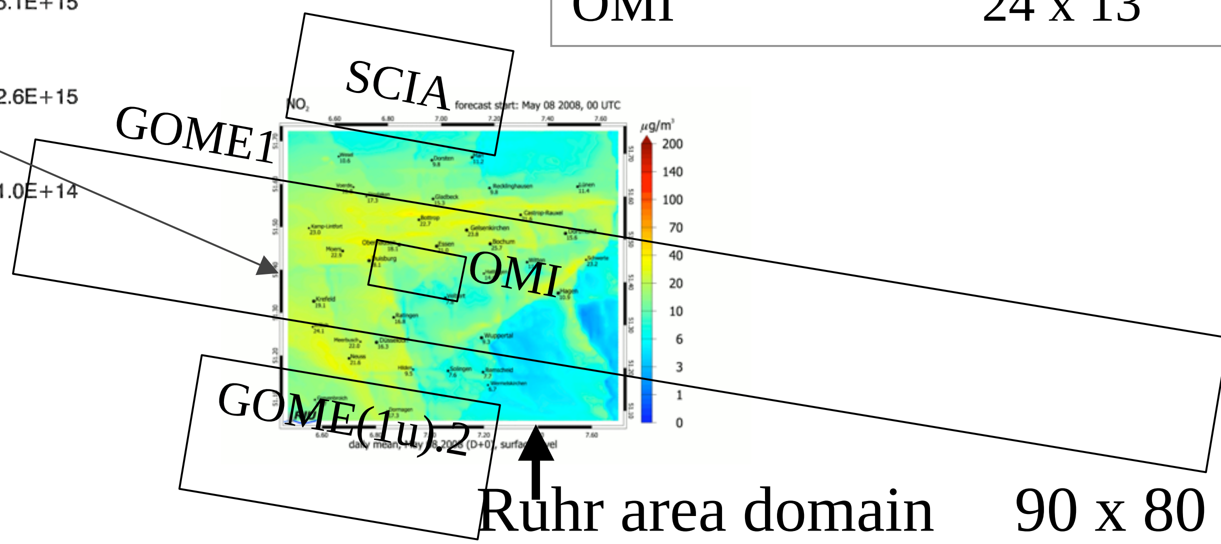
- resolution to meet OMI:
  → **15 km** horizontal resolution selected
- attention to forecast error covariance design:
  →spatial correlation exploitation via **inhomogeneous and anisotropic** radii of influence,
- DA method: chemical 4D-var as BLUE, incl emissions, with **externally** provided a prioris:
  → NO2 columns errors from data provider, model error from other case studies, i.e. **no “tuning”** introduced

Satellite information:
ESA UV-VIS satellite footprints Ruhr area comparison

minimal areas:

GOME 1	320 x 40 km^2
(special mode)	80 x 40 “
SCIAMACHY	60 x 30 “
GOME 2	80 x 40 “
OMI	24 x 13 “





Ruhr area domain   90 x 80 km^2

1 km resolution

(~12 000 000 inhabitants)

Error variances applied for period 1.-10.7.2006 over model domain

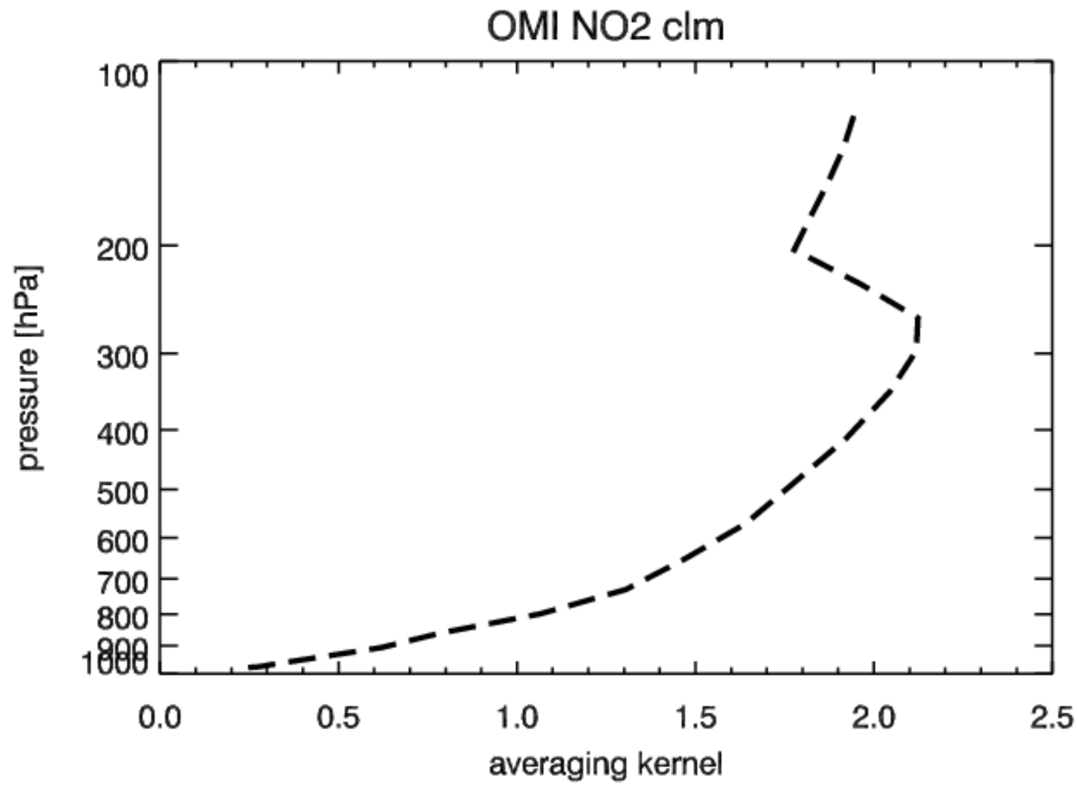
NO_2 columns from KNMI data files: **R** (diagonal)

molecules/cm^2	E(y)	E(Δy)
OMI	$1.4*10^{15}$	$0.8*10^{15}$
SCIAMACHY	$1.2*10^{15}$	$0.9*10^{15}$

Forecast error covariances **B** schematic formula

$\mathbf{B}_{ii}$ (spec,lev)= max{1 ppb,0.8*var(spec,lev), 0.5 max(spec,lev)}

Average OMI averaging kernel profile over model domain for July 9th, 2006



model domain mean averaging kernel.

Exploitation of NO2 column averaging kernel information

- shape largely dependent on optical properties of the atmosphere (cioud cover), rather than NO2
- typical maximal sensitivity above the boundary layer
- does not allow a clear distinction between PBL or lower free troposphere pollution burden

How to proceed to obtain benefit from trop. column integral information?

(A typical problem of Inverse Modelling by Integral Equations)

Two more specific questions:

- When is it justified to project averaging kernel information to the surface?
- Can this be done without destroying the BLUE property of the assimilation algorithm?

Partial observation operator $\mathbf{H}$

Formally an integral equation to be solved for vertical MO_2 molecule density function x

$$y = \int_1^0 w(\sigma)x(\sigma)d\sigma$$

Discretisation

$$y = \sum_{k=1}^{K} h_k x_k$$

At the minimum $\mathbf{x} =: \mathbf{x}_a$

$$d\mathbf{x}_a := \mathbf{x}_a - \mathbf{x}_b = (\mathbf{B}_0^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}\left\{\mathbf{y}^0 - H[\mathbf{x}_b]\right\}$$

$$= \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}\left\{\mathbf{y}^0 - H[\mathbf{x}_b]\right\}$$

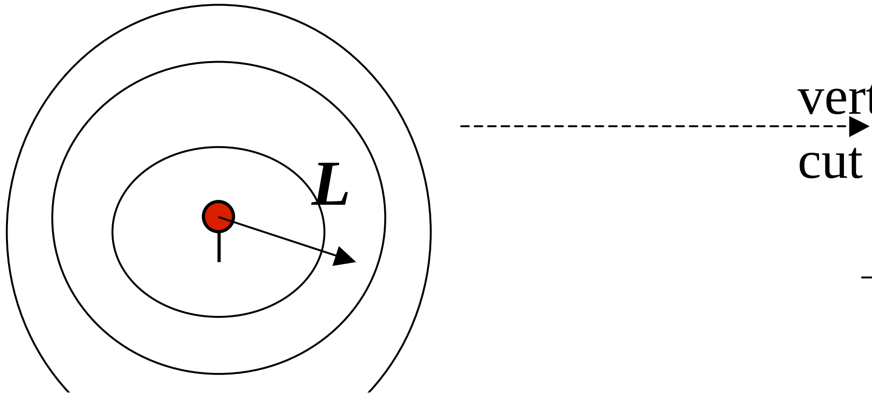
For scalar column retrieval:

$$\boxed{dx_a = \underbrace{\mathbf{B}\mathbf{h}^T(r+b)^{-1}}\left\{y^0 - H[\mathbf{x}_b]\right\}}$$

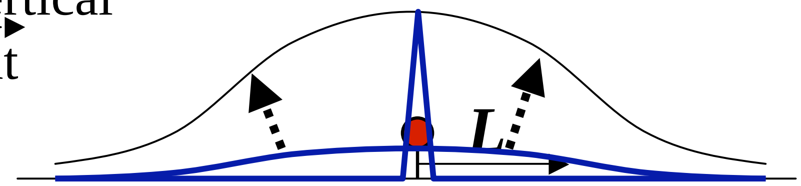
adjoint "representer" (oceanographic DA parlance)

Radius of Influence ((de-)correlation length): Extending the information from an observation location

Textbook:
horizontal influence radius ***L***
around a measurement site,
to be based on a priori
statistical assessments

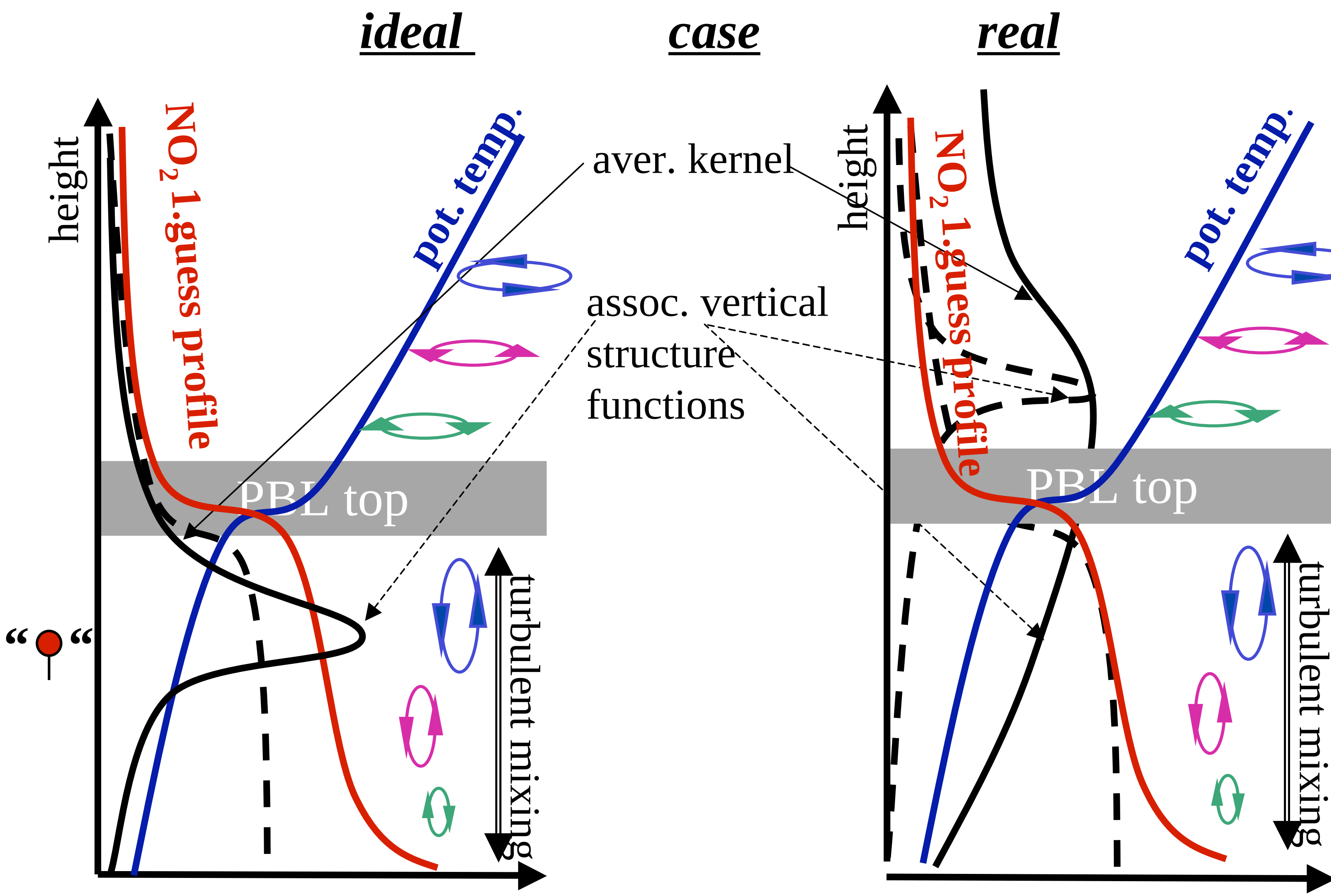


1D horizontal structure function,
to be stored as a column of the
forecast error covariance matrix



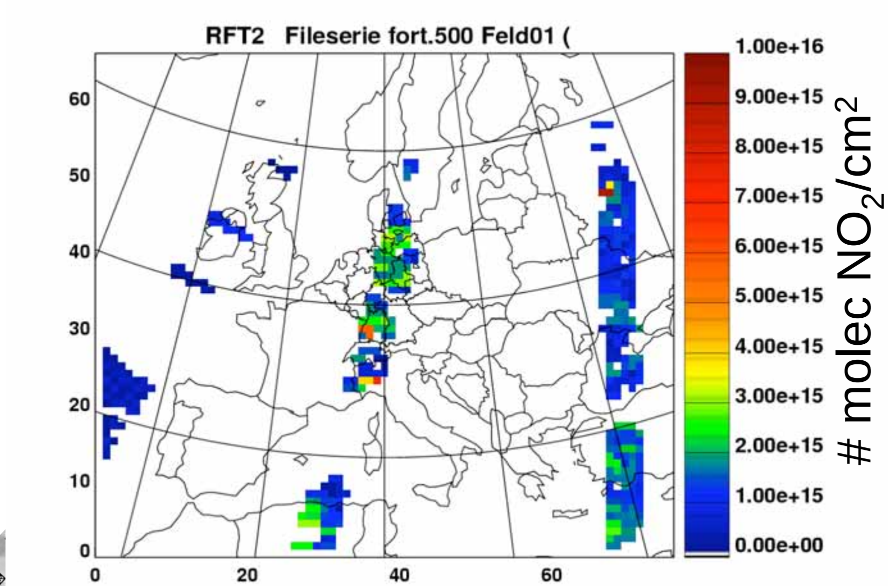
vertical Radius of Influence:

Extending the information from observation location



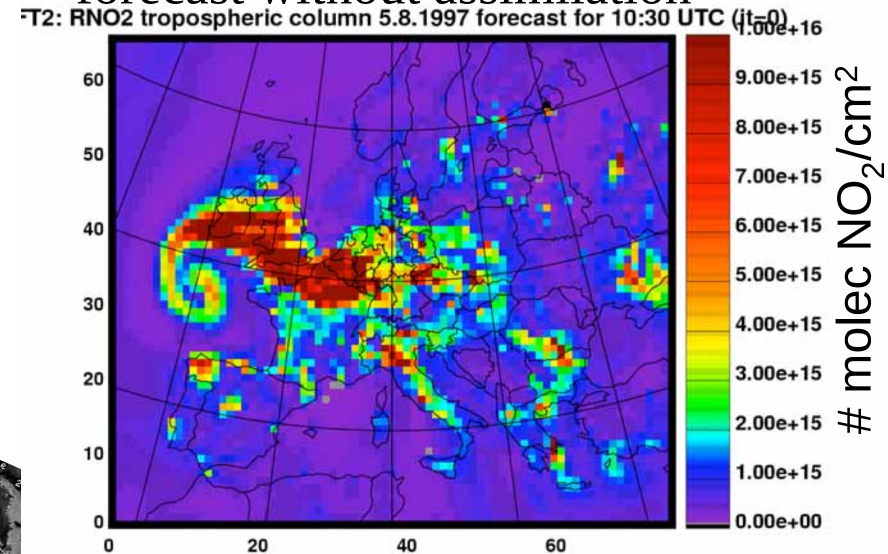
Assimilation of GOME NO_2 tropospheric columns, 5.8.1997

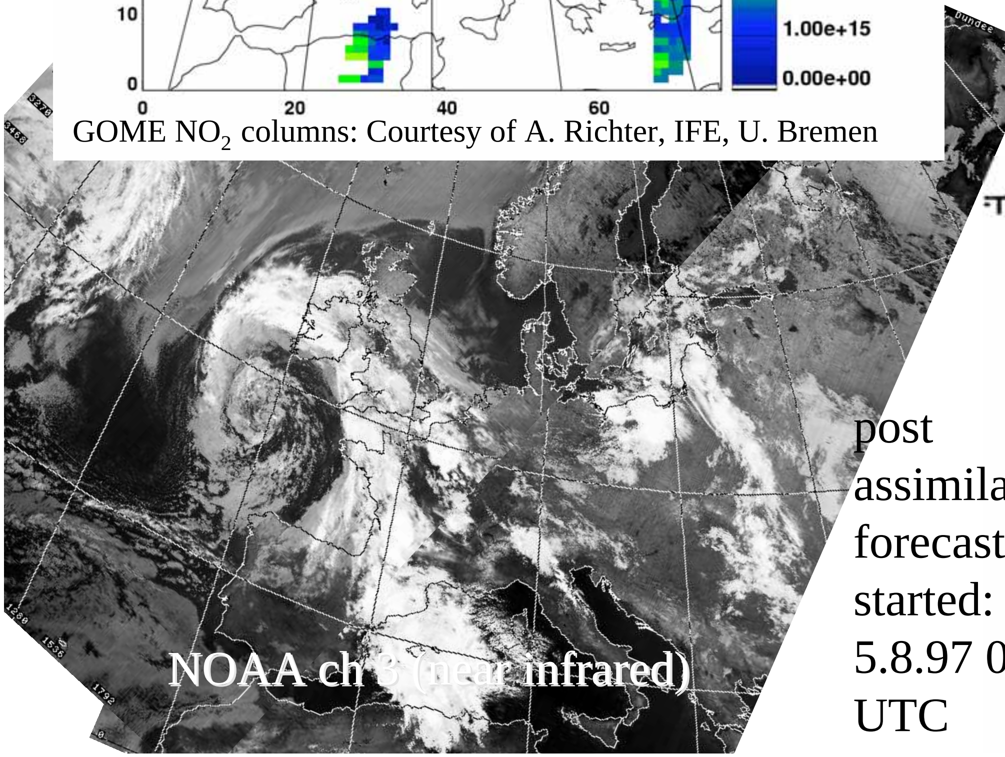
A case of lifted NO2 maxima



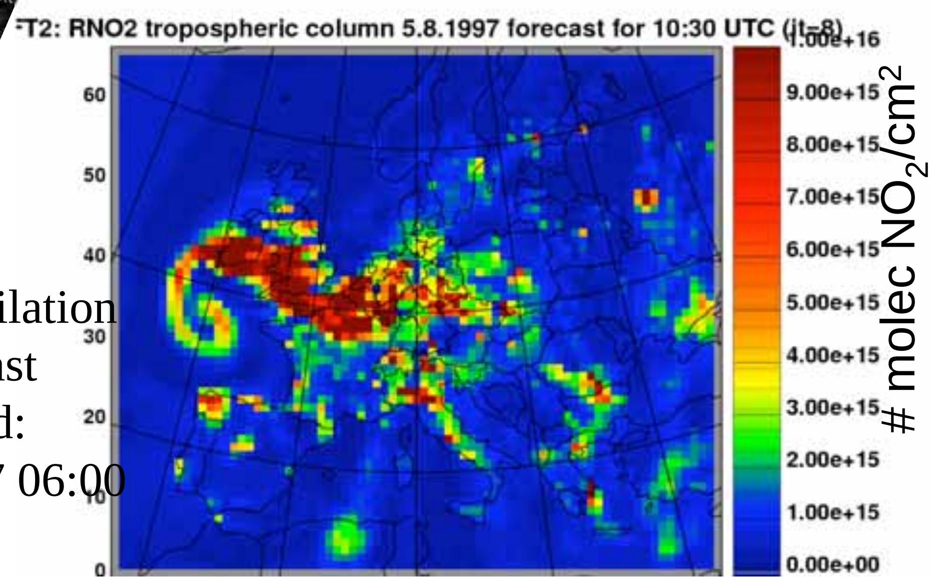
GOME NO_2 columns: Courtesy of A. Richter, IFE, U. Bremen

forecast without assimilation



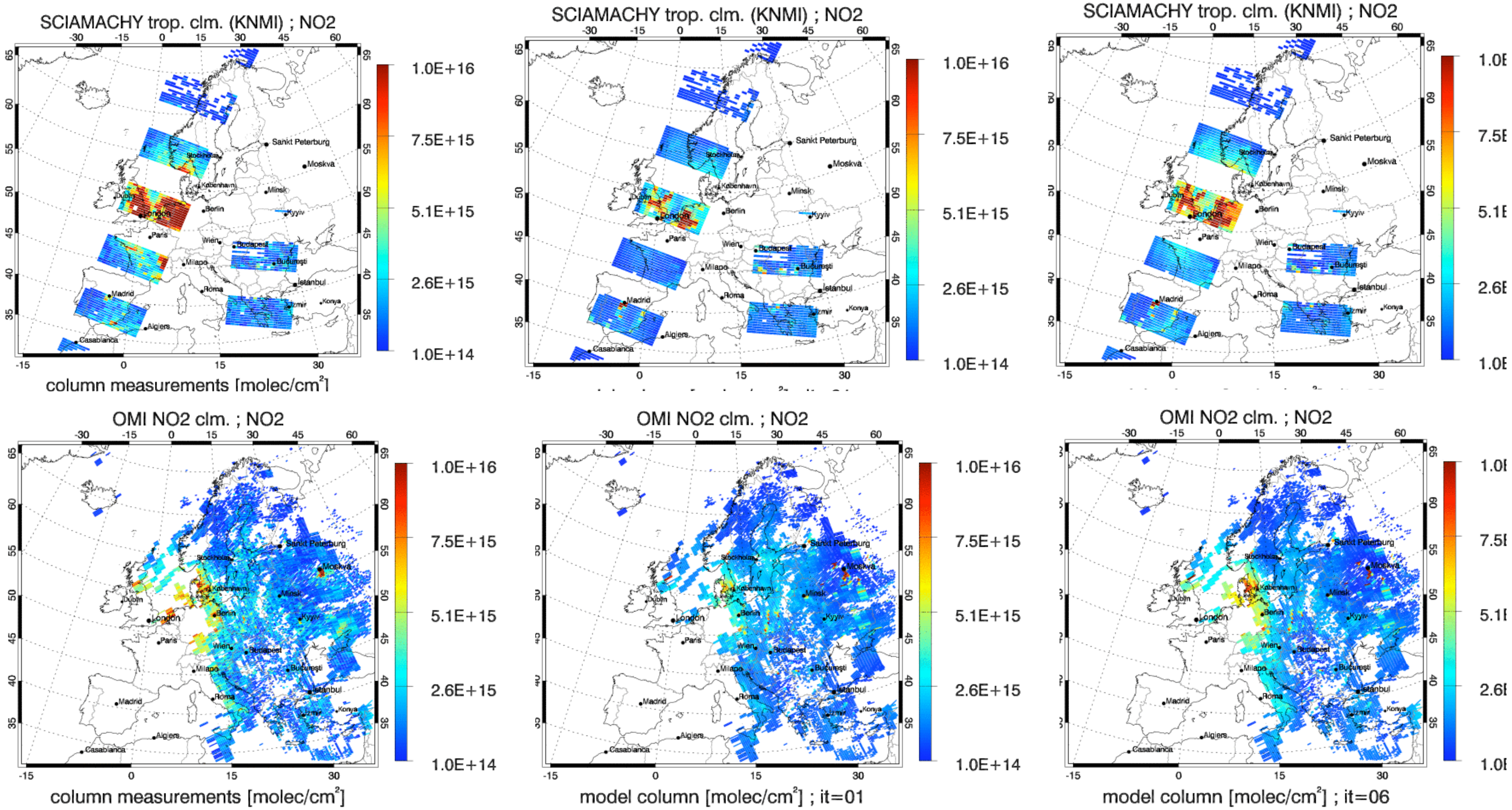


post assimilation forecast started: 5.8.97 06:00 UTC



Comparison of NO2 tropospheric columns

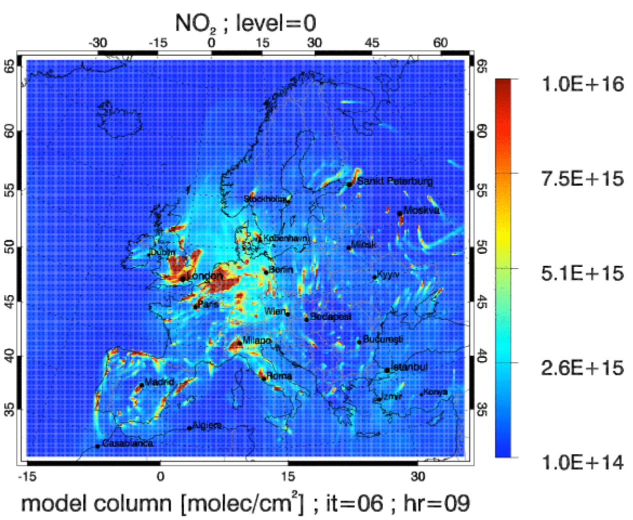
in molecules/cm^2 for July 6th, 2006, 09-12 UTC.

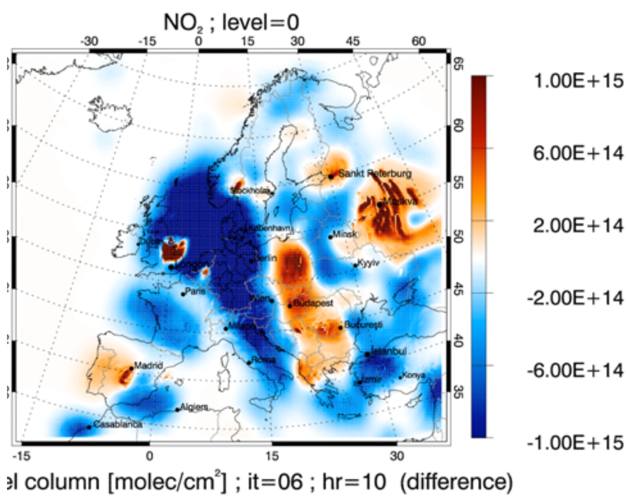


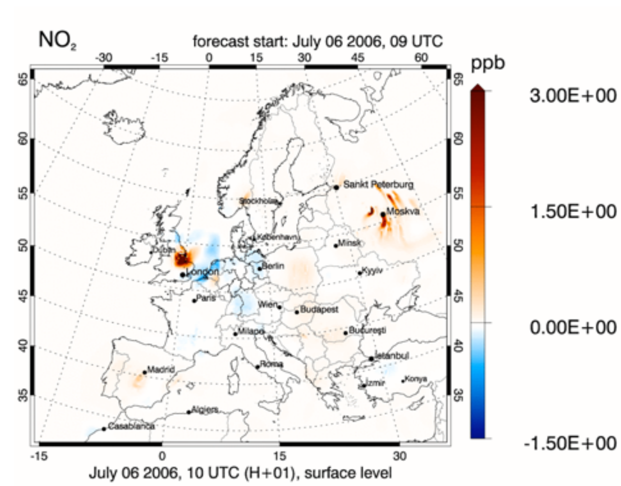
assimilated values (**y**) | EURAD forecasted ($\mathbf{Hx}_b$) | column analyses ($\mathbf{Hx}_a$)

Data assimilation result from tropospheric columns for **July 6th, 2006**.

NO2 model columns by OMI and SCIAMACHY assimilation interval 09-12 UTC.





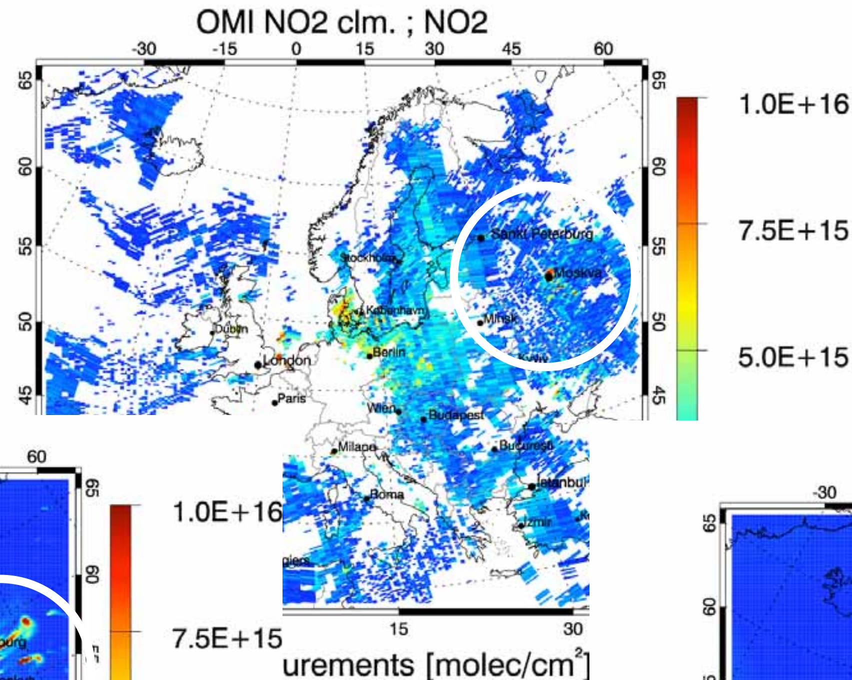


NO2 ppb

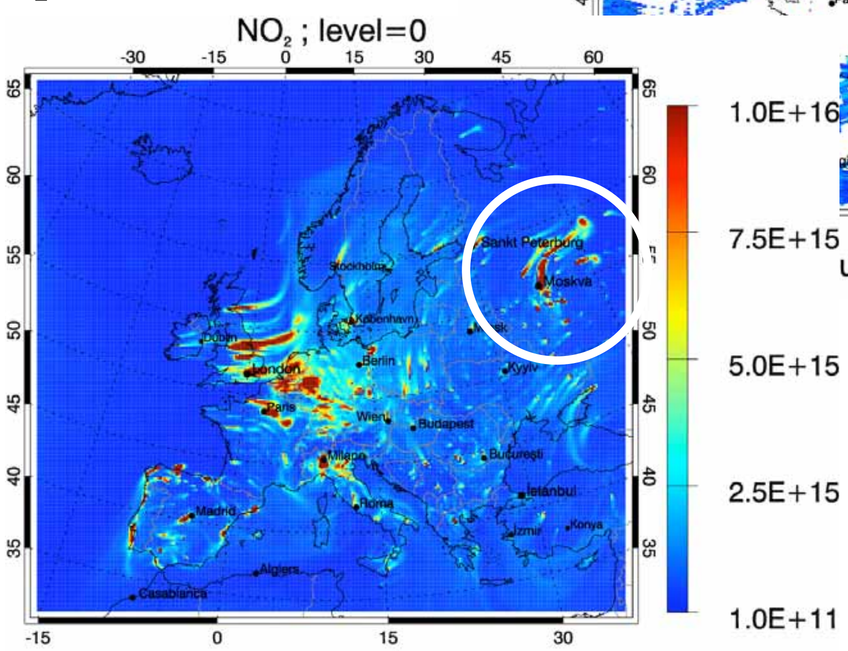
molecules/cm^2

Analysed NO2 colum | Difference field | surface concentration changes

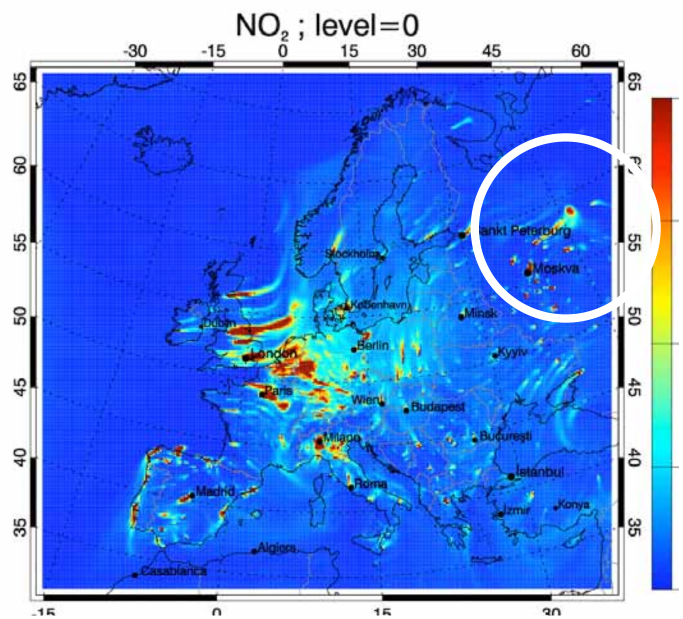
Data assimilation result in terms of tropospheric columns for **July 7th**, 2006. NO2 model columns based on OMI and SCIAMACHY assimilation within the assimilation interval, 09-12 UTC.



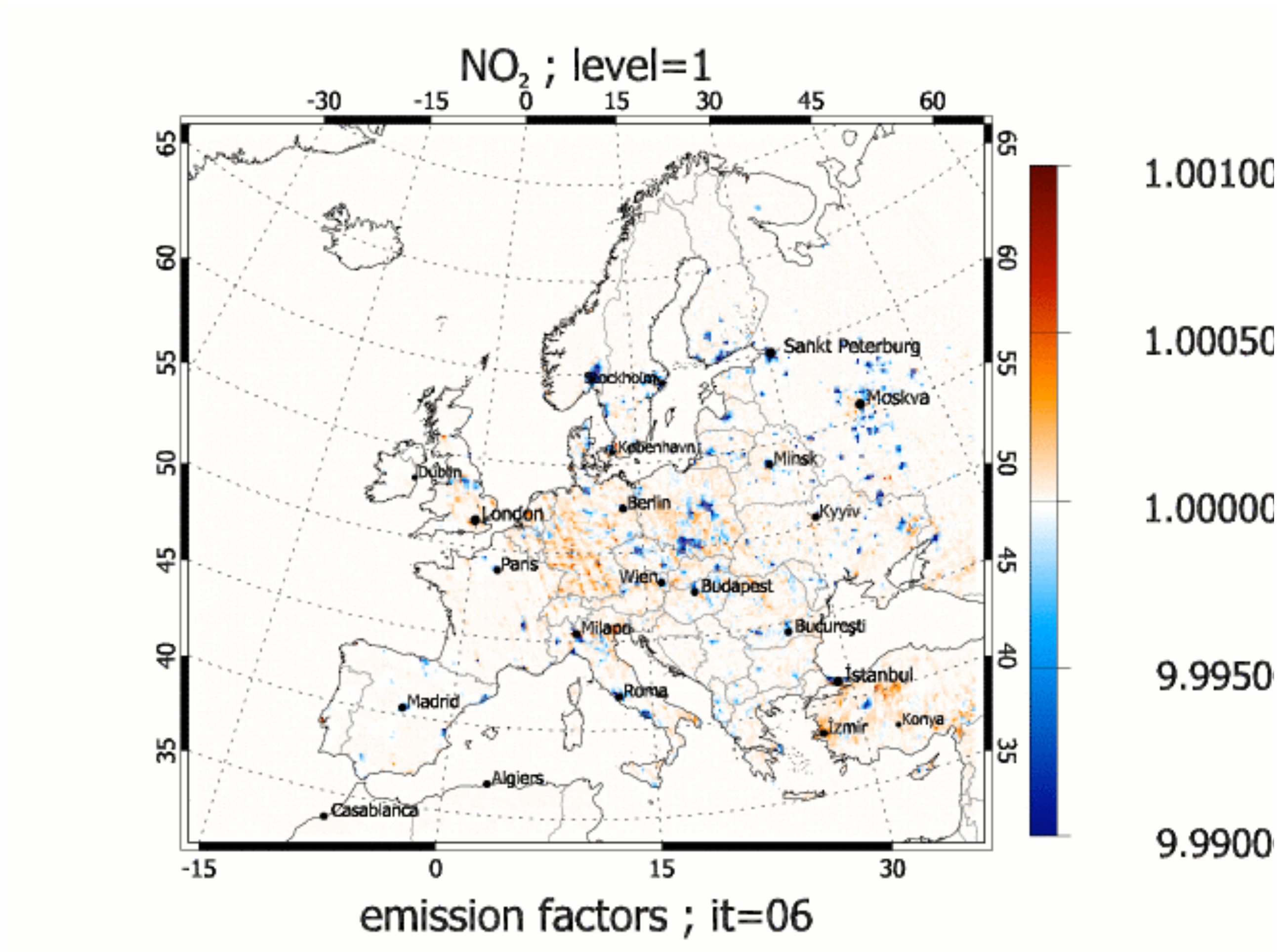
pure forecast



assimilation based forecast



Qualitative assessment of emission correction factors for **July 7th**, 2006



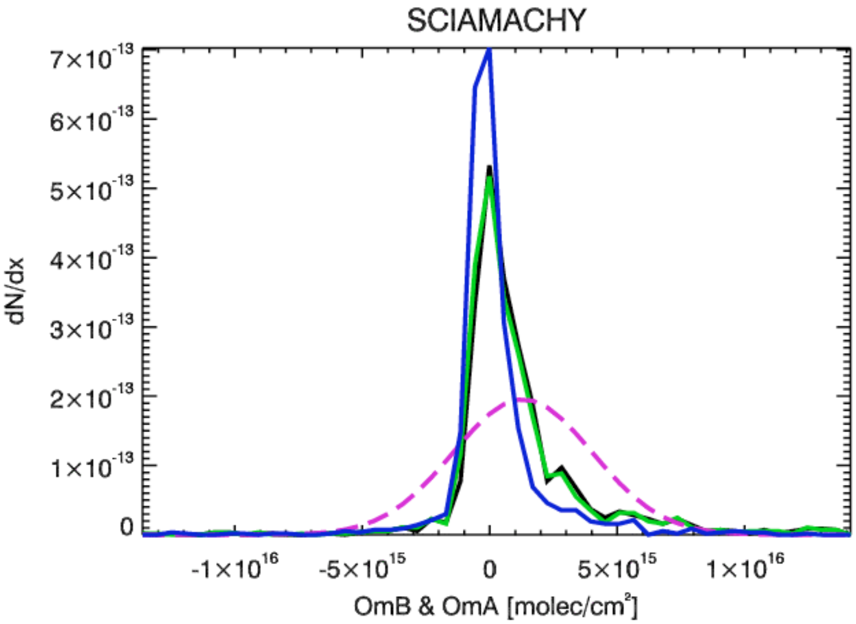
Emission inventory values

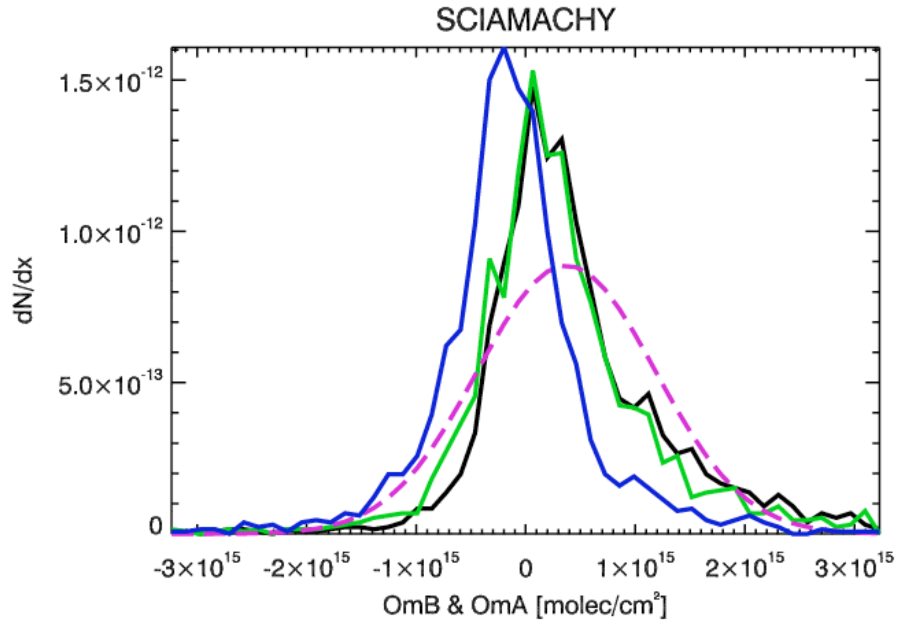
to be increased

to be reduced

SCIAMACHY
$O_{SCIA}mX_{OMI}$ probability density functions

for July 6th (left), and July 8th, (right).



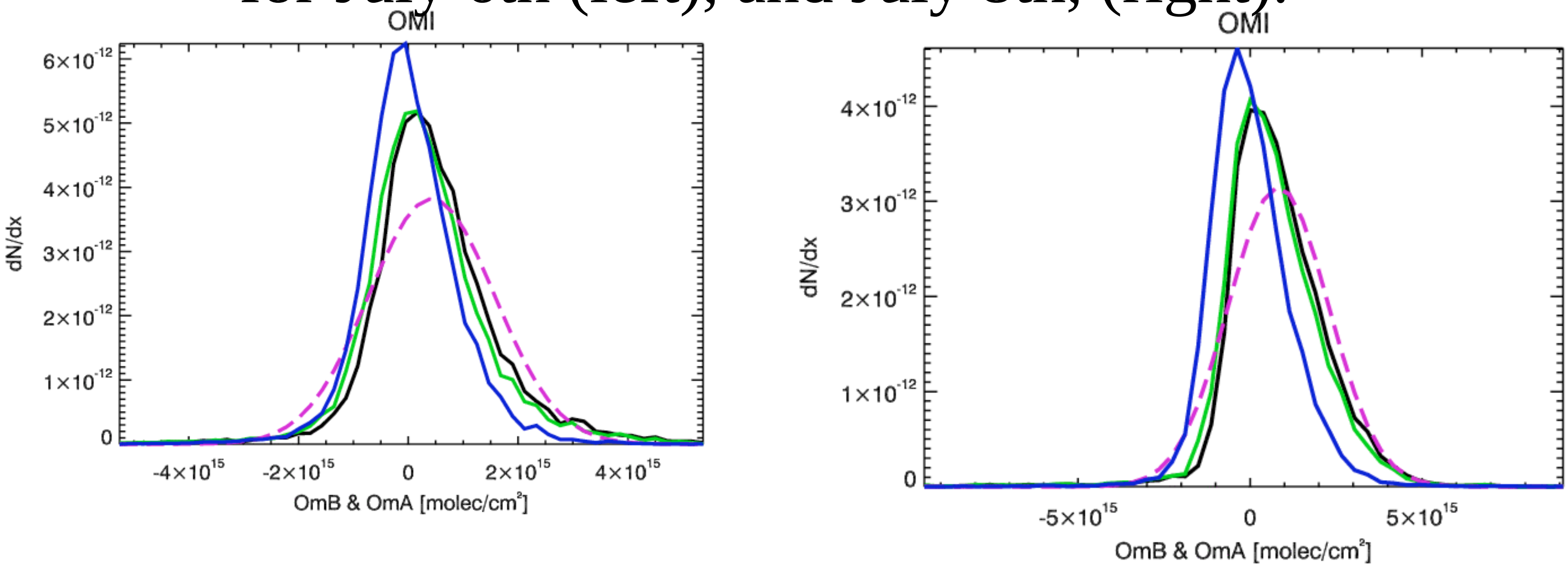


Control run (OmC) (no data assimilation at all,) black bold line,
assimilation based forecasted values (OmF) green bold line,
analyses (OmA) blue bold line.
For comparison: Gaussian fit to OmF pdf by
mean and standard deviation given by broken purple line.

OMI

$O_{OMI}mX_{OMI}$ probability density functions

for July 6th (left), and July 8th, (right).



Control run (OmC) (no data assimilation at all,) black bold line,

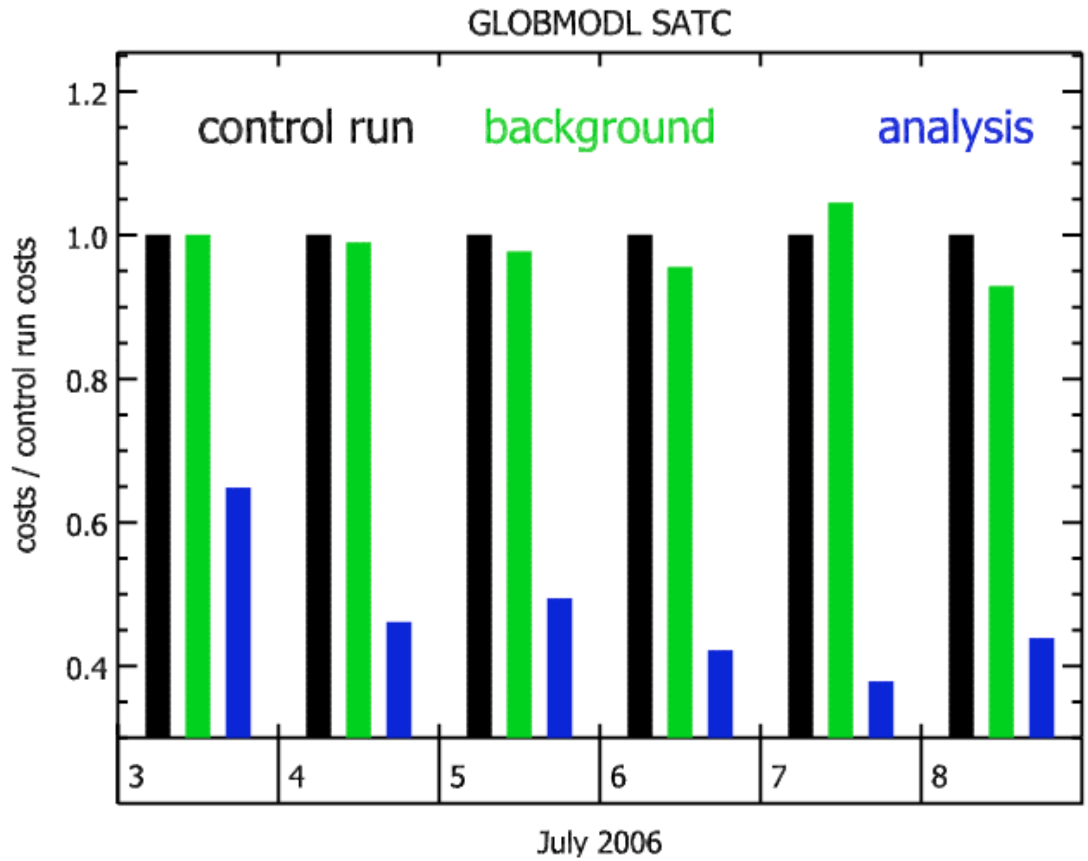
forecasted values (OmF) green bold line,

analyses (OmA) blue bold line.

For comparison: Gaussian fit to OmF pdf by

mean and standard deviation given by broken purple line.

Objective function based normalized costs of combined OMI and SCIAMACHY assimilation runs between July 3rd -8th, 2006.



Black bars: control run without any data assimilation,
for reference and normalisation to value 1 only.

Green bars: **one day forecast costs**. Blue bars: **analyses costs**.

Lessons

- The unfavourable kernel profiles can and must be balanced by more sophisticated vertical covariance design: “forcing information down to the surface”
- additional statistics for tuning this needed
- longer assimilation intervals needed: also to improve emission correction factors on a sound basis, with respect to working days, saturdays, and sundays
- ensemble generation should not be confined to emission perturbations, but also perturbed meteorological fields (from ECMWF), initial values, and j-values

Assimilation of Aerosol observations

- In situ:

  EEA Airbase: Database of groundstations of EU member countries & states:

  - 450 stations for PM_{10} (2003)
  - No $PM_{2.5}$. (4 stations in UK only)

- Satellite measurements:

  SYNAER (SYNergetic AErosol Retrieval, DLR-DFD, [Holzer-Popp, 2001])*

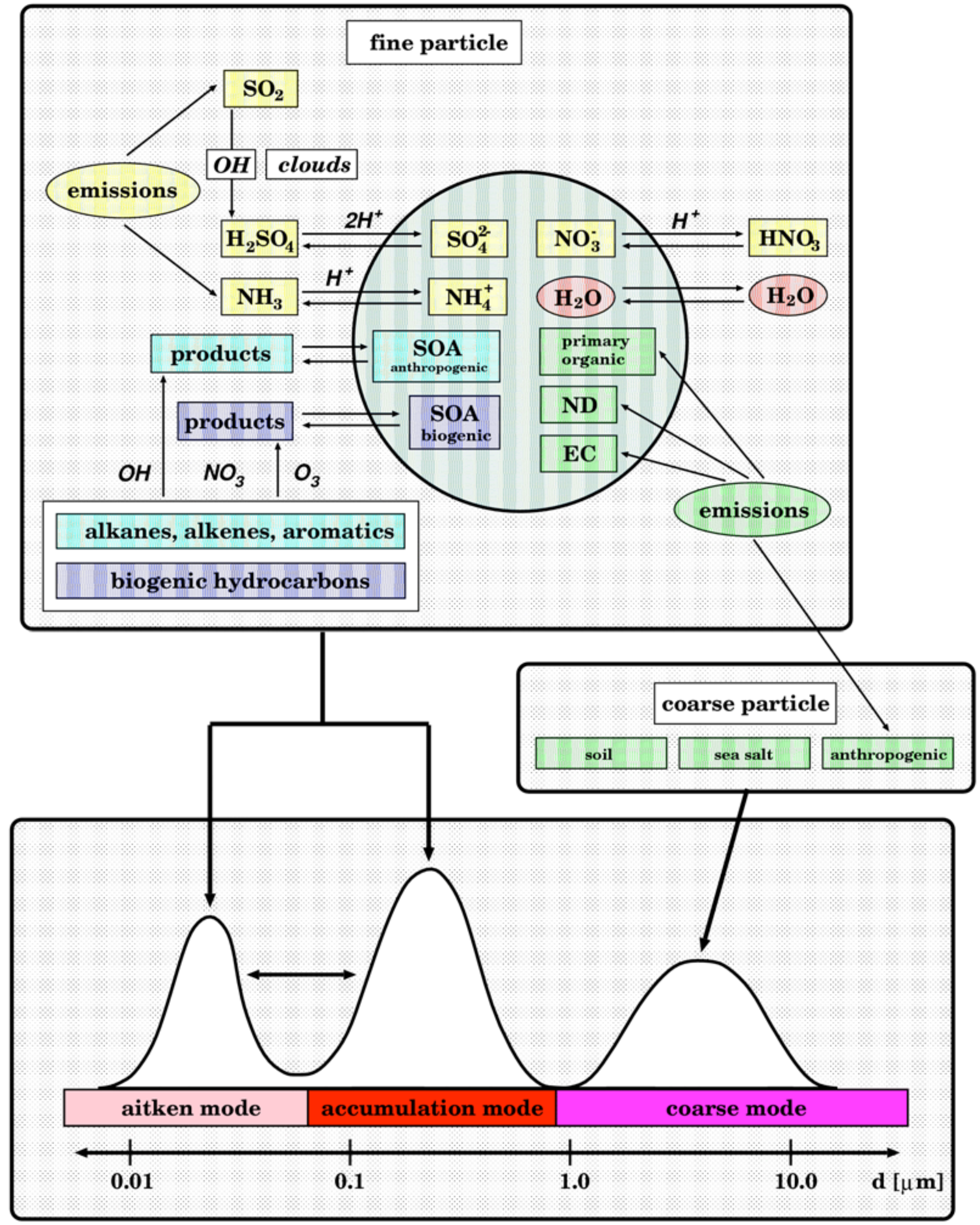
  - combines GOME&ATSR-2, SCIAMACHY&AATSR measurements aboard ERS-2/ENVISAT
  - ATSR-2/AATSR:

    dark field detection, BLAOT (Boundary Layer Aerosol Optical Thickness) and albedo are calculated
  - GOME/SCIAMACHY:

    Provides $PM_{0.5}$, $PM_{2.5}$ and PM_{10} columns and its composition (6 intrinsic species)

Aerosol Chemistry in **MADE**

Modal **A**erosol **D**ynamics for **E**URAD/**E**urope (Ackerman et al., 1998, Schell 2000)



$$dM_i^k/dt = nuk_i^k + coag_{ii}^k + coag_{ij}^k + cond_i^k + emi_i^k$$

M_i^k:=k^{th} Moment of i^{th} Mode

Bridge from optical to chemical properties

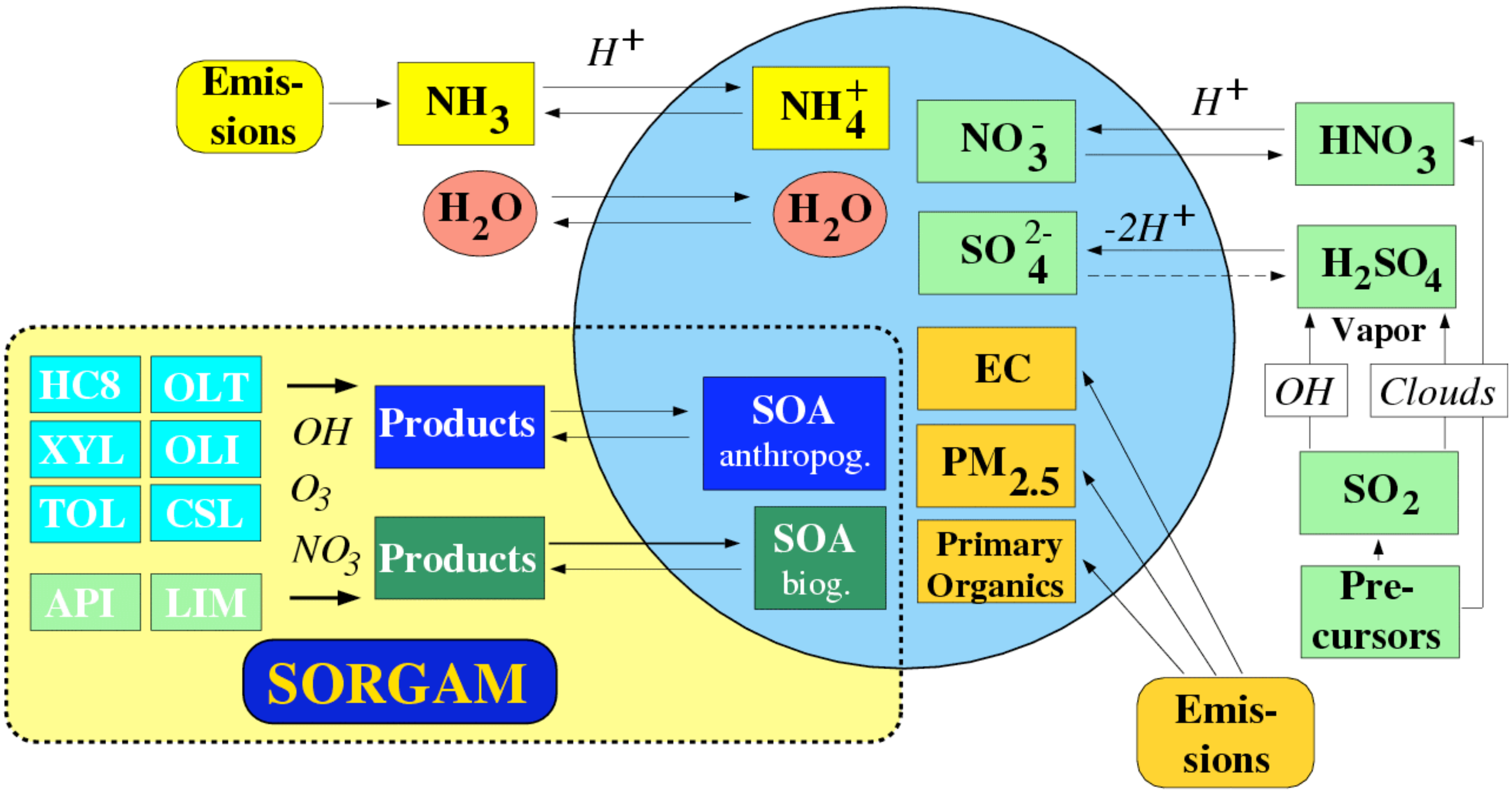
assimilation of aerosol

By sattelite retrievals: e.g.

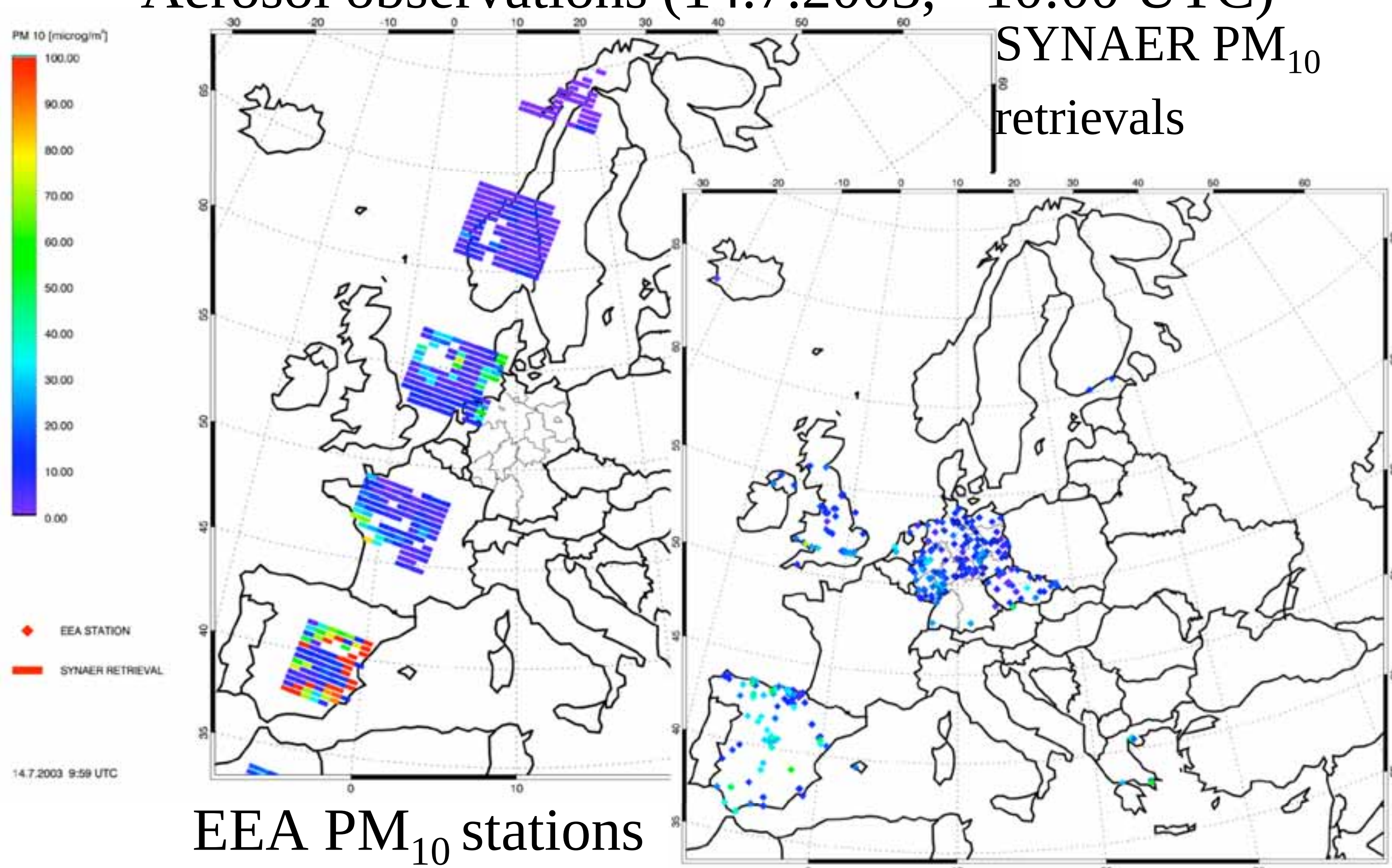
MERIS MODIS

AATSR+SCIAMACHY,...

Example: chemical complexity:
The EURAD Secondary ORGanic Aerosol Model (SORGAM)



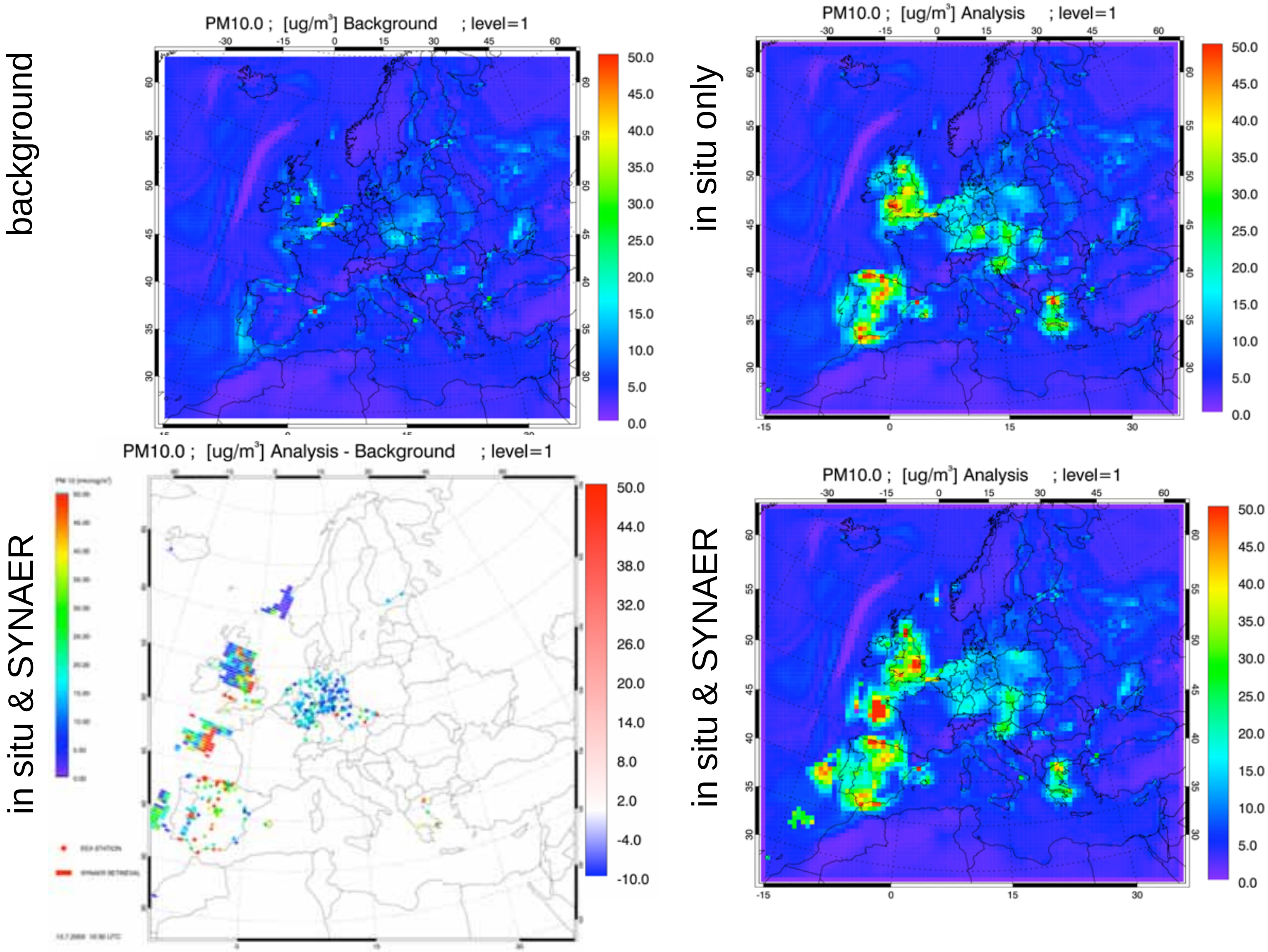
Aerosol observations (14.7.2003, ~10:00 UTC)



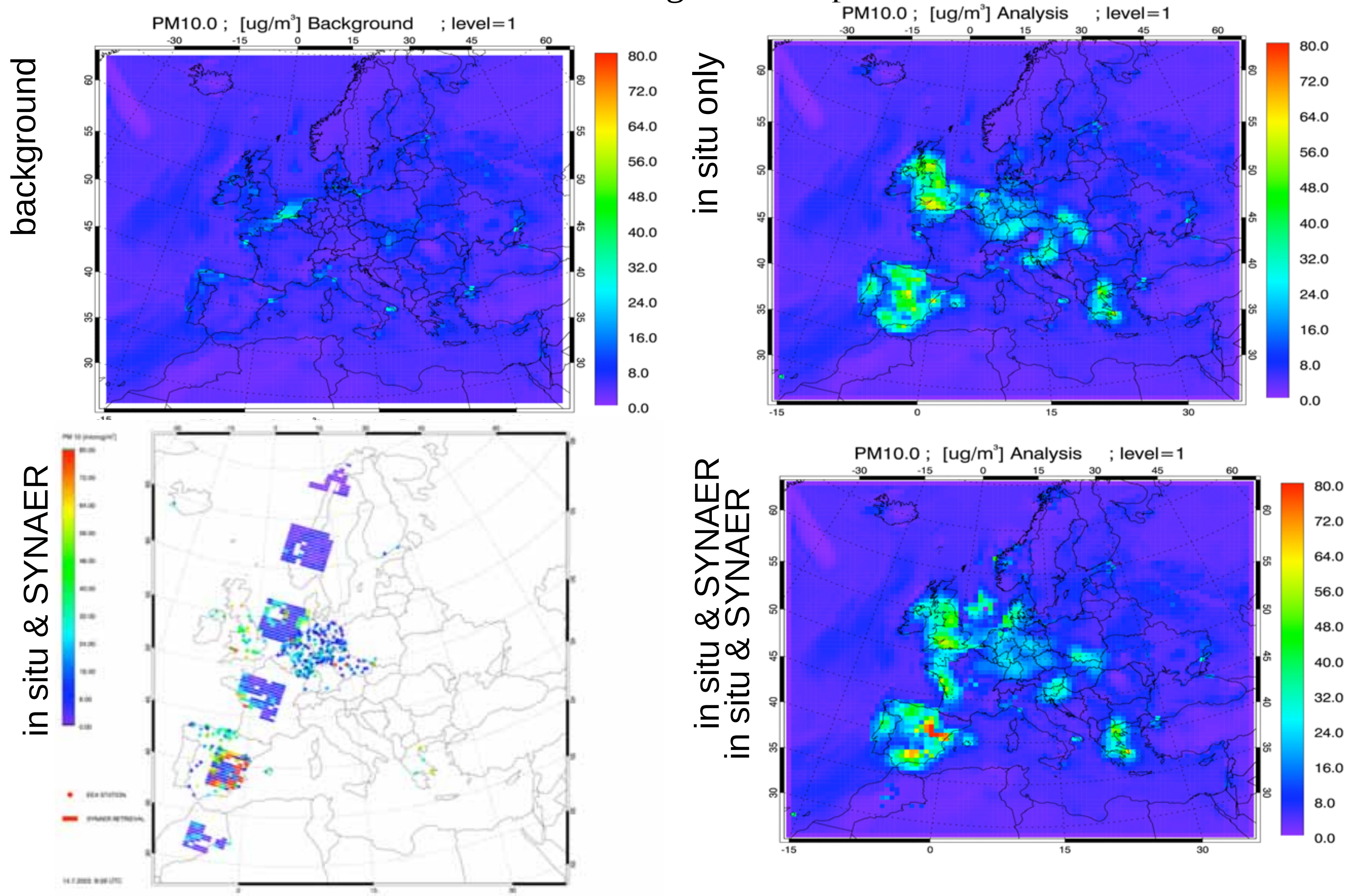
SYNAER PM_{10} retrievals

EEA PM_{10} stations

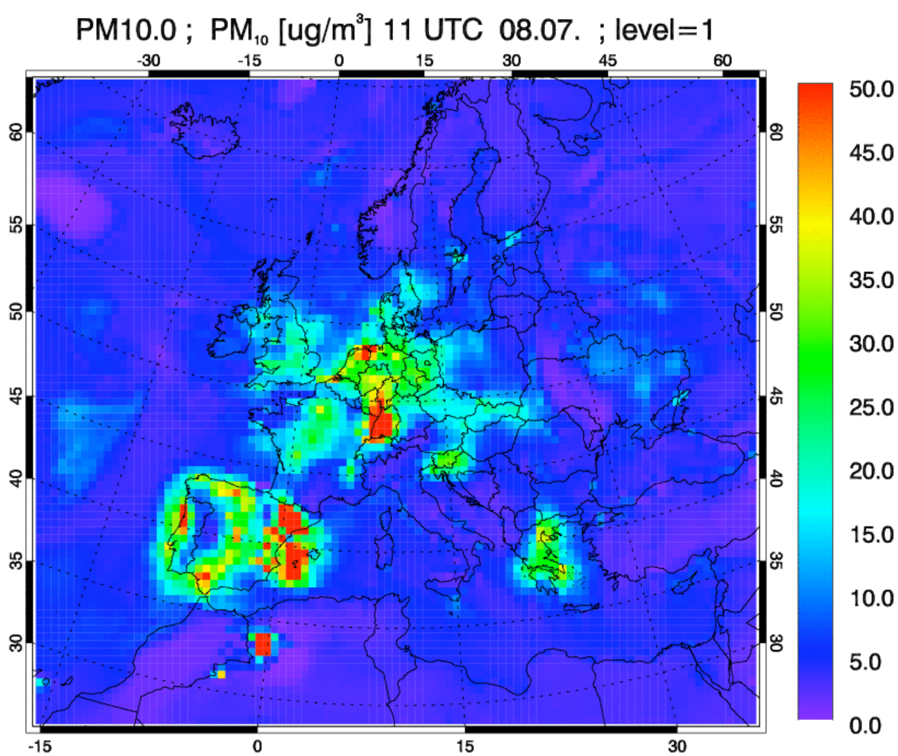
3Dvar aerosol assimilation (13.7.2003)



3Dvar aerosol assimilation (14.7.2003)
biomass burning case in Spain

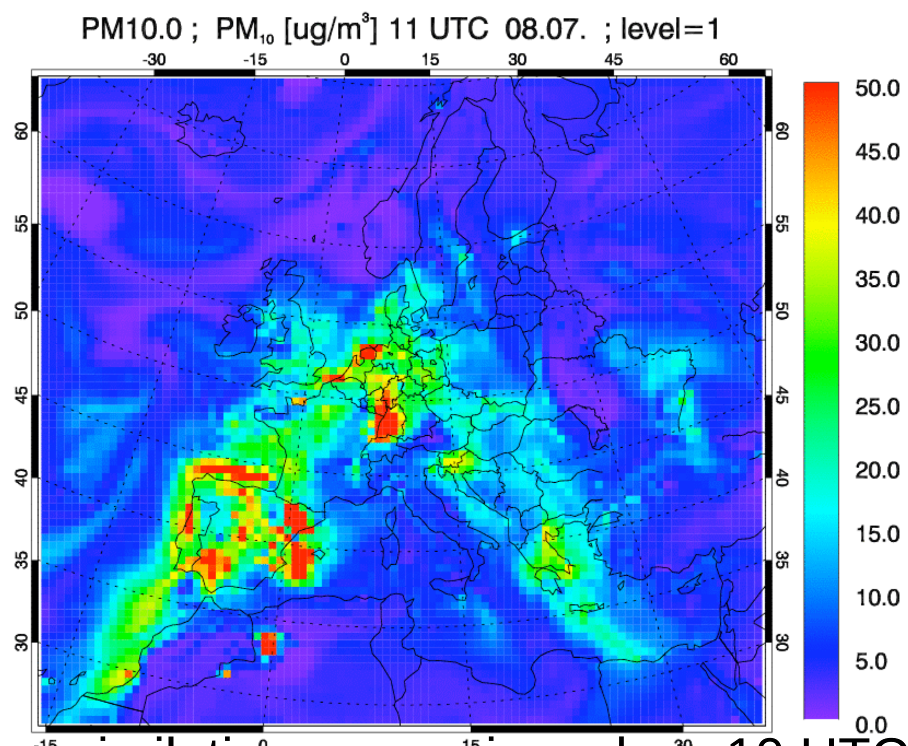


Do aerosol data assimilation effects accumulate? (14. July 2003)



No previous assimilation

only 14. July 2003



assimilation on previous days 10 UTC

Accumulation of retrieval information over 14 days