ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

PAPILA
Prediction of Air Pollution
in Latin America and the Caribbean

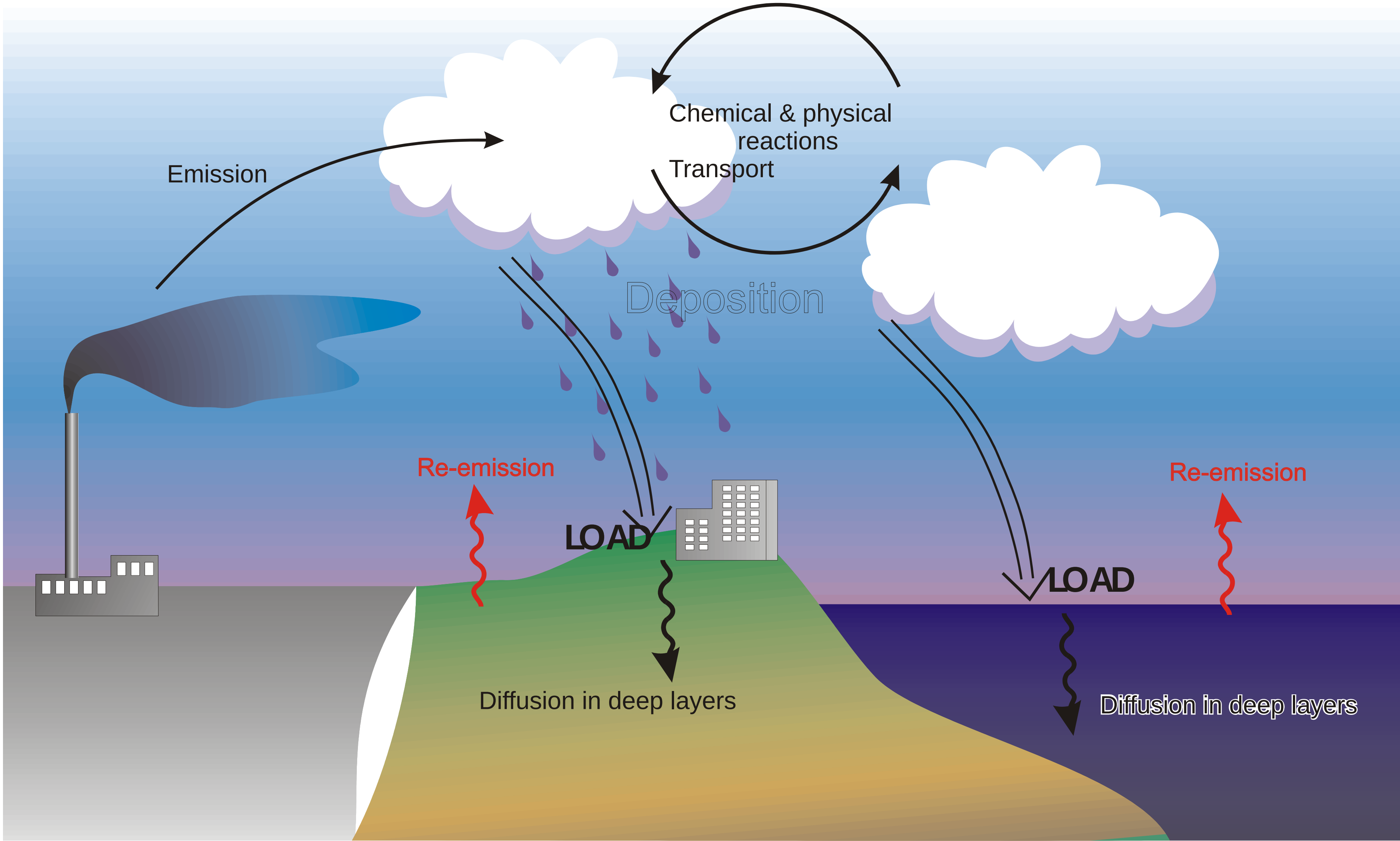
Fundamentals of air pollution modelling

M.Sofiev

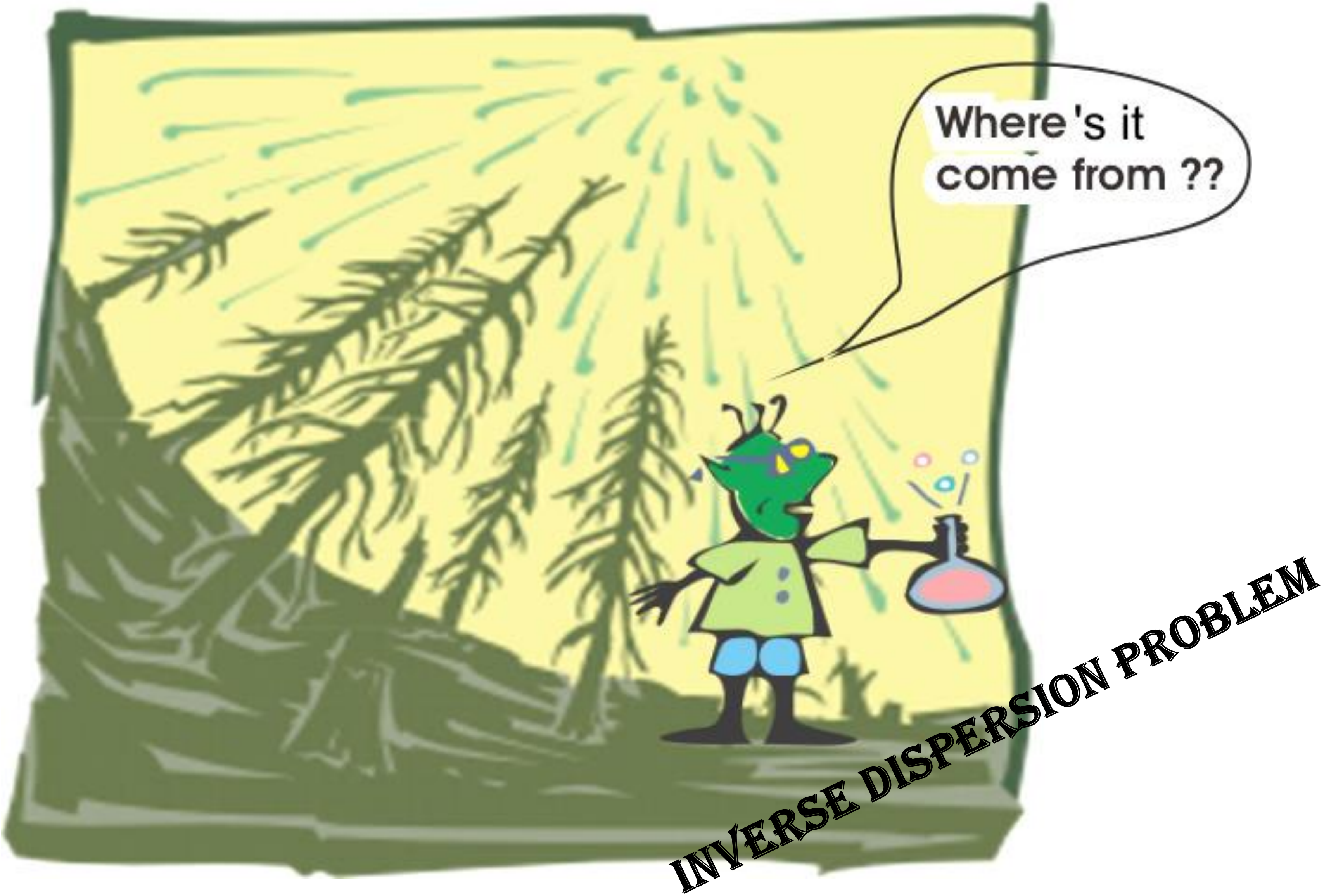
Content

- Basic terms
- What is atmospheric composition model?
- Dispersion equation
- Classifications of dispersion models
- Parts of a dispersion model
- Transport term dispersion models
- Model Quality Assurance
- Data Assimilation
- Summary

Cycle of atmospheric chemicals, troposphere



Frequently asked question

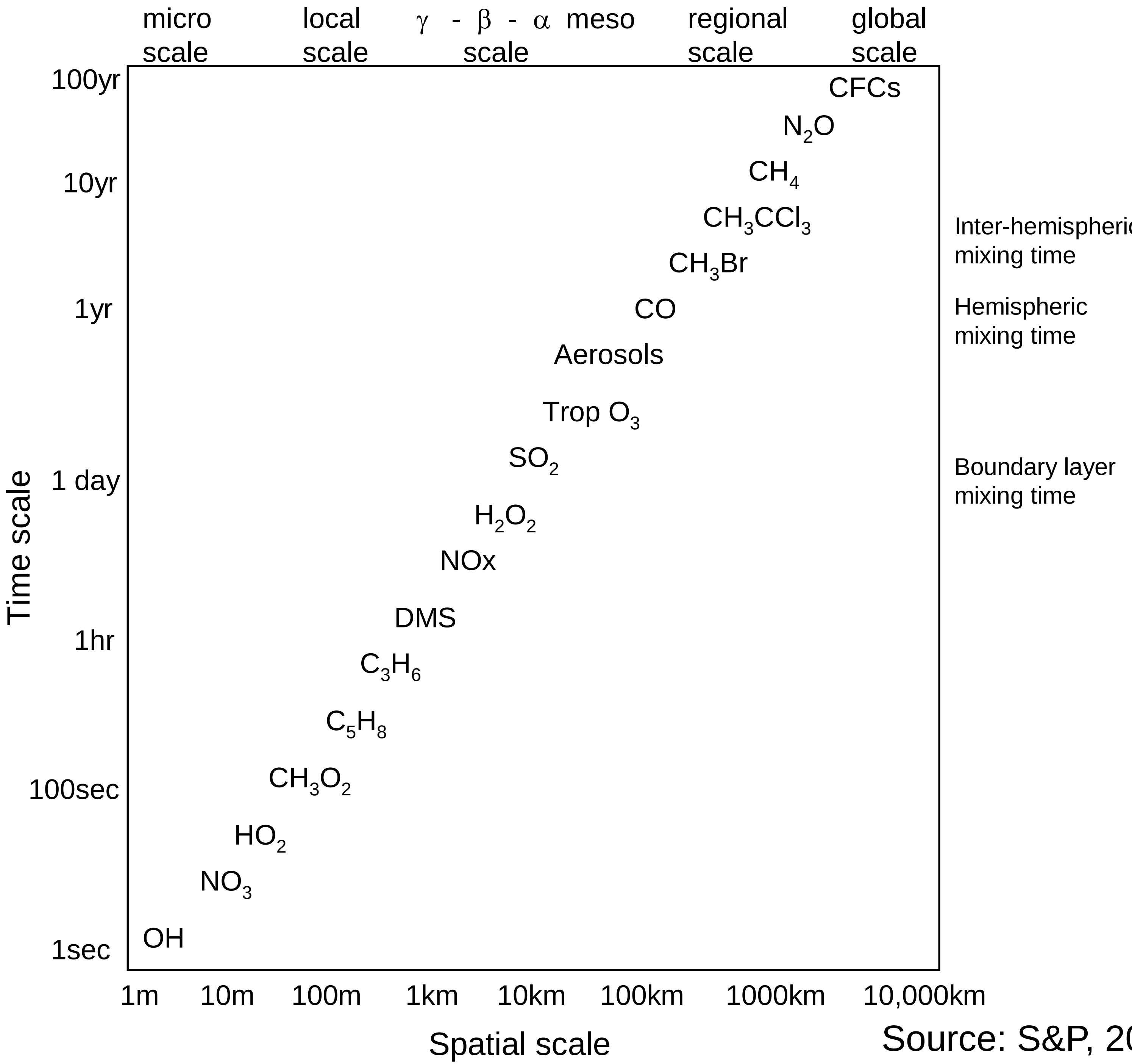


Basic terms

- Atmospheric tracer: a compound, which is transported with atmospheric flows but does not affect them (sufficiently low concentration)
  - some impact exists always: feedback to meteorological processes
- Tracer lifetime: a time period needed for 2-fold (or *e*-fold) reduction of amount of the tracer in the atmosphere
  - condition- and process-dependent: e.g. lifetime with regard to advection
  - general meaning: lifetime is related to relaxation times for any process involving this tracer (not 1:1 though)
- Spatial and temporal scales
  - related to lifetime
  - temporal scales are translated into spatial ones via wind speed
  - processes and their imporatnce are related to scales

Basi

- Atm with (suf
  - s
- Trac redu
  - c a
  - g p
- Spa
  - re
  - te
  - p



Source: S&P, 2002

Content

- Basic terms
  - atmospheric tracer
  - temporal and spatial scales
  - life time in the atmosphere
  - life cycle of atmospheric tracers
- What is atmospheric composition model?
- Dispersion equation
- Classifications of dispersion models
- Parts of a dispersion model
- Transport term dispersion models
  - Lagrangian and Eulerian models
- Model Quality Assurance
- Data Assimilation
- Summary

What is atmospheric composition model?

- Model is never a copy of reality
  - ➢ It represents only those features, which are deemed important for a specific application
- The extent of their similarity is to be established in each specific case

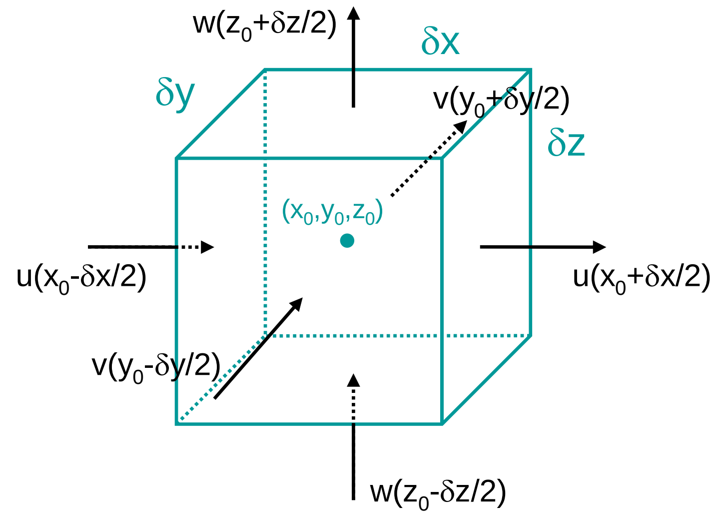
I.Repin. Zaporozhje Cossacks are writing a letter to Turkish sultan

W.Kandinski. Cossacks

Dispersion equation: vital features of nature formalized

- Mass conservation
  - ➢ transport
  - ➢ sources
  - ➢ sinks
- Scale separation
  - ➢ mean flow
  - ➢ turbulence
- Closure problem
  - ➢ K-theory $\rightarrow$ turbulent diffusion coefficient

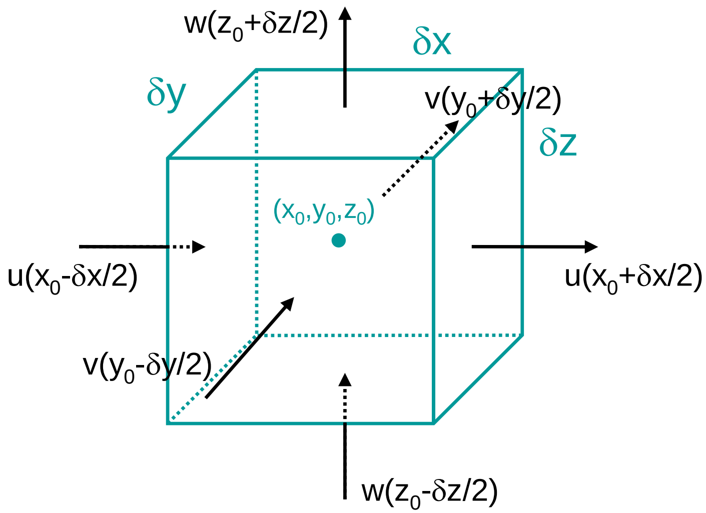
?



$$\frac{\partial c}{\partial t} = -\sum_{i=1}^{3} \frac{\partial}{\partial x_i}(u_i c) + E - R$$

Dispersion equation: vital features

- Mass conservation
  - transport
  - sources
  - sinks
- Scale separation
  - mean flow
  - turbulence
- Closure problem
  - K-theory → turbulent diffusion coefficient



$$\frac{\partial c}{\partial t} = -\sum_{i=1}^{3} \frac{\partial}{\partial x_i}(u_i c) + E - R$$

$$LC \equiv \frac{\partial C}{\partial t} + \frac{\partial}{\partial x_i}(U_i C) - \frac{\partial}{\partial x_i} \rho K_{ii} \frac{\partial (C/\rho)}{\partial x_i} + R(C) = E$$

Content

- Basic terms
  - atmospheric tracer
  - temporal and spatial scales
  - life time in the atmosphere
  - life cycle of atmospheric tracers
- What is atmospheric composition model?
- Dispersion equation
- Classifications of dispersion models
- Parts of a dispersion model
- Transport term dispersion models
  - Lagrangian and Eulerian models
- Model Quality Assurance
- Data Assimilation
- Summary

Classifications of models

- Model principles
  - Eulerian
  - Lagrangian
  - Gaussian
  - statistical Monte-Carlo
- Scales
  - global
  - continental
  - regional
  - local/urban

Classifications of models.2

- Chemicals
  - acid
  - ozone
  - greenhouse gas
  - inert aerosol/dust
  - radio-activity
  - toxic
  - persistent pollutants
- Model media
  - atmospheric
  - multi-media
  - integrated models

Classifications of models.3

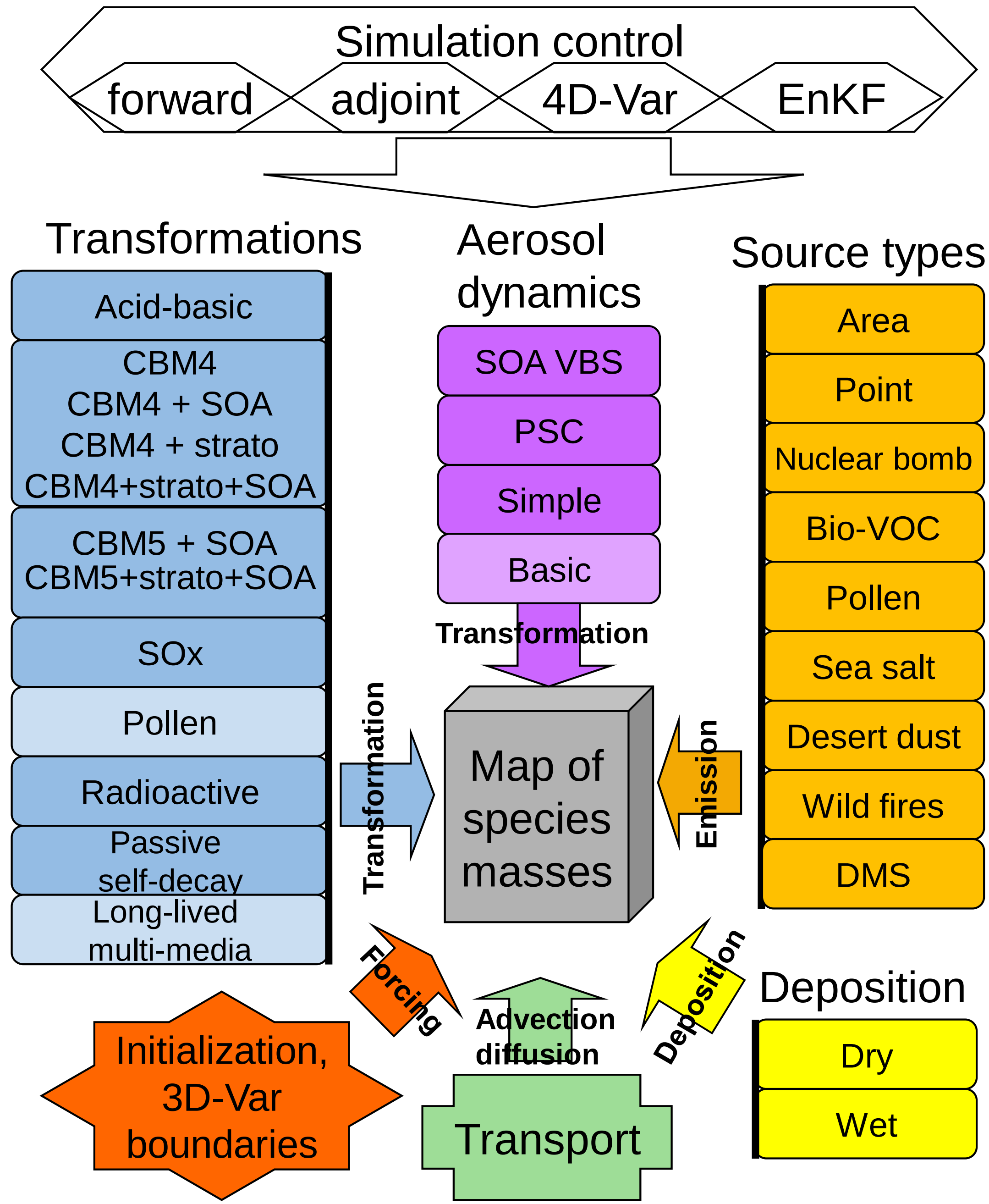
- Input data
  - climatological
  - real-time data
- Time dimension: direction, horizon
  - re-analysis
  - now-casting
  - forecasting
- Problem to solve
  - forward
  - inverse

Content

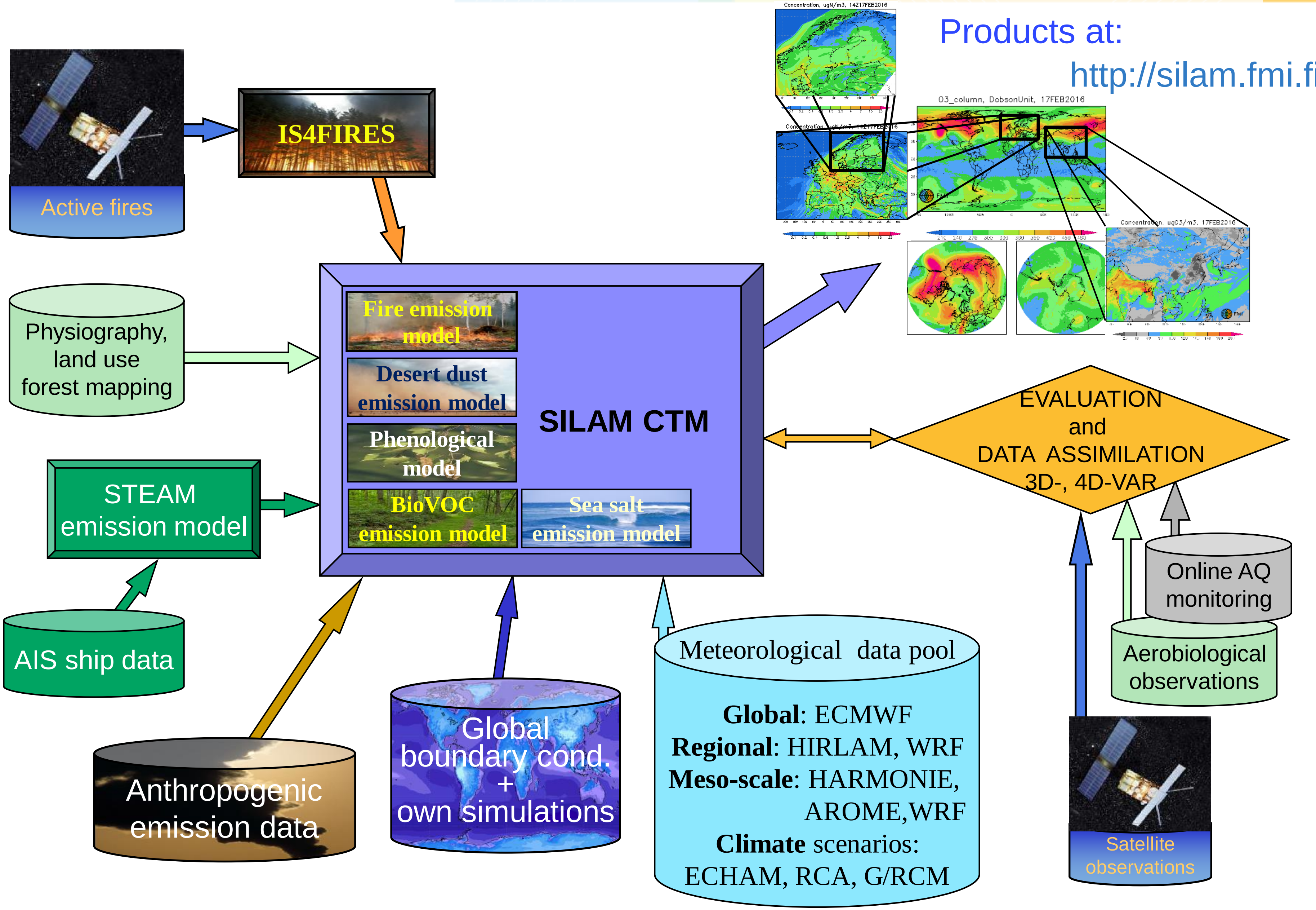
- Basic terms
  - atmospheric tracer
  - temporal and spatial scales
  - life time in the atmosphere
  - life cycle of atmospheric tracers
- What is atmospheric composition model?
- Dispersion equation
- Classifications of dispersion models
- Parts of a dispersion model
- Transport term dispersion models
  - Lagrangian and Eulerian models
- Model Quality Assurance
- Data Assimilation
- Summary

ACM components

- Modules
  - 14 transformation modules
  - 9 source terms,
  - All modern DA techniques: 3D-Var,,4D-Var, EnKF, EnKS
- Domains: from global to beta-meso scale (~1km resolution)
- Any meteo input that follows WMO standards
- Technically
  - 192 FORTRAN-2005 modules, ~250 classes, OMP+MPI parallel
  - 18 MB of code (~130,000 lines)
  - The largest FMI own model
  - Installed in a dozen of countries for research and operational purposes
  - ~10,000 lines in ~100 environment scripts (Python-Shell)



SILAM AQ assessment and forecasting platform



Content

Model core: transport algorithm (advection scheme)

- A key part of every chemistry transport model
- Derived from mass conservation law
  - Expresses and guarantees mass conservation in the model
- Interacts with all other modules
- Very simple basic form…
- … but strikingly difficult to solve numerically
- Two key approaches:
  - Eulerian
  - Lagrangian

Eulerian:

$$\frac{\partial c}{\partial t} = -\sum_{i=1}^{3} \frac{\partial}{\partial x_i}(u_i c)$$

Lagrangian:

$$\frac{dr_i}{dt} = u_i$$

Lagrangian model principles

- A tracers release is represented via many discrete volumes of air enriched with the corresponding tracer: Lagrangian particles
- Lagrangian particle is
  - NOT a particle (in the common meaning of the word)
  - a finite-size volume of air completely isolated from other volumes
- LPs move along wind streamlines: advection term
- LPs get randomly relocated: diffusion term
- LPs can contain reacting chemical admixtures
- LPs do not interact with each other
- Final concentration is the sum of mass of many LPs that appeared to be in a large volume, divided by that volume

LP motions

- Transport with wind: $$\frac{dr_i}{dt} = u_i$$

- Diffusion: Langevin equation (drift *a*, diffusion *b*, Wiener zero-mean *dt*-variance process)

$$\frac{dv_i}{dt} = a_i(\vec{x},\vec{u},t)dt + b_{ij}(\vec{x},\vec{u},t)dW_j \qquad \sigma_{r_i} \sim \sqrt{\gamma\, K_i\, \Delta t}$$

- Connection to dispersion equation and Eulerian formalism

$$\frac{dc}{dt} = \frac{\partial c}{\partial t} + \left(\frac{\partial c}{\partial x_1}u_1 + \frac{\partial c}{\partial x_2}u_2 + \frac{\partial c}{\partial x_3}u_3 \right) = \frac{\partial c}{\partial t} + u_i\,\frac{\partial c}{\partial x_i}$$

Chemical transformations

- Lagrangian models can support linear chemical calculations
  - the reactions are of the first order, i.e. the rates are independent from the concentrations of the species
  - radioactive decay
- Non-linear transformations can also be imagined… in theory

Problems of Lagrangian models

- Non-linear processes are all but impossible
- Boundary conditions are all but impossible
- Representativeness problem: limited number of particles forced a single LP to represent large volume of air
    - ➢ Increase of the LP number leads to prohibiting computational demand

- All-in-all, lagrangian models are good for point source of point receptor, for limited time period and region size. Example: forward and inverse emergency applications

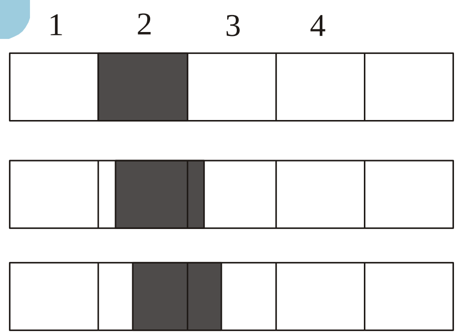
Eulerian modelling

- Directly approaching the dispersion equation
- Discretization of every term using finite-size meshes: (3D) Eulerian grid
- Non-linearities are naturally included
- Boundary conditions are simple

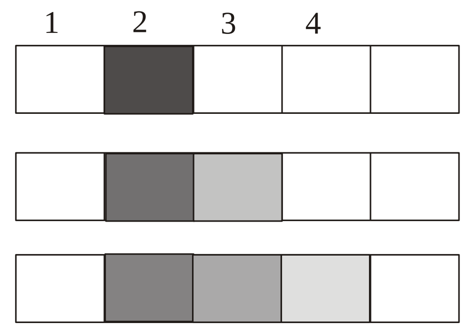
Numerical solution: why a problem?

$u = 0.2\, \Delta x\, \tau^{-1}$ →

In “reality”



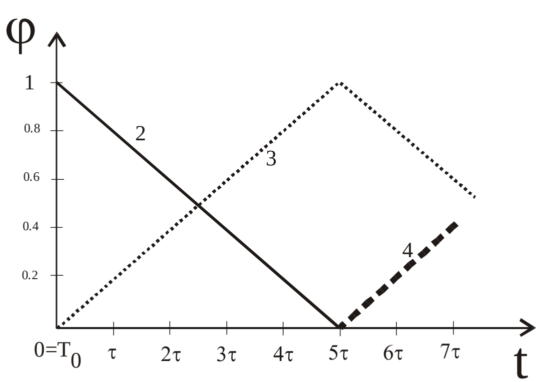
In idealized grid scheme

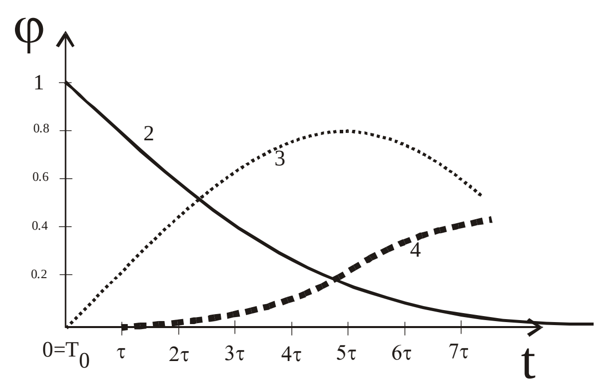


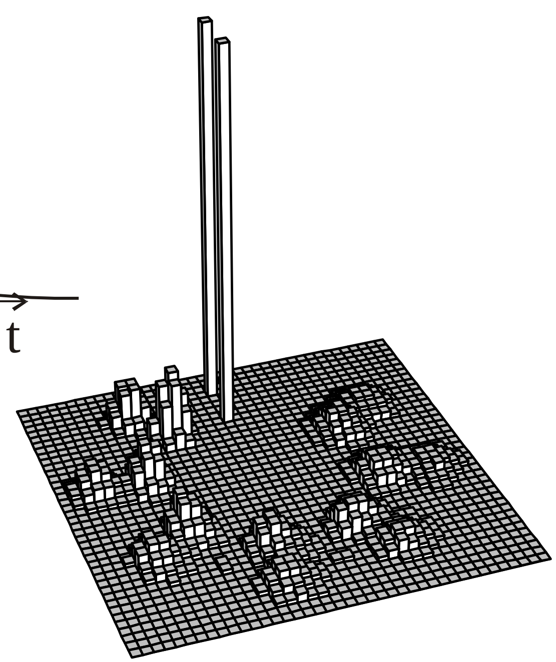
$t = T_0$

$t = T_0 + \tau$

$t = T_0 + 2\tau$



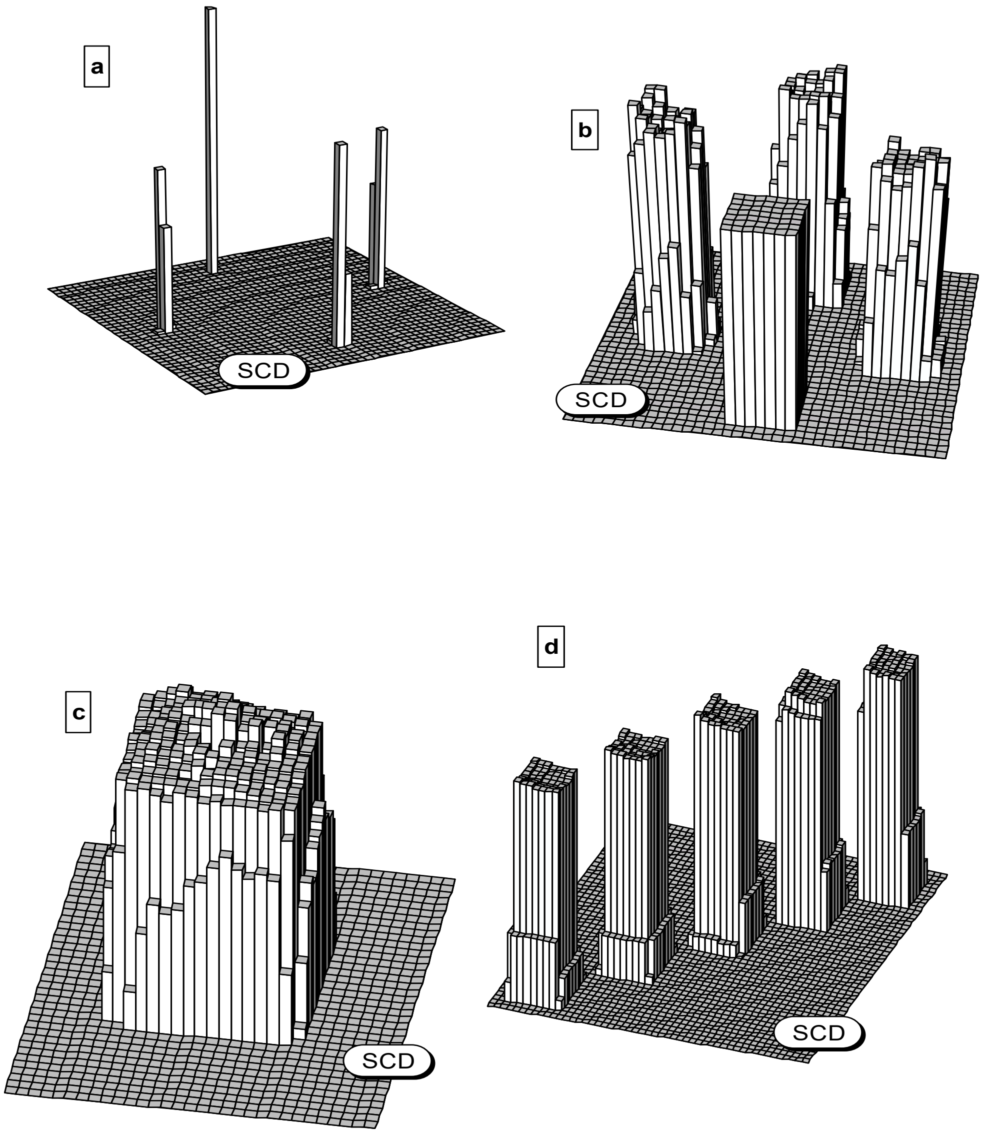




Courtesy: M.Galperin

Numerical diffusion: solutions exist

- Modern approaches are capable of reducing or nearly eliminating numerical diffusion
- … but nothing comes for free: higher order non-linear distortions



Courtesy: M.Galperin

Problem of Eulerian models

- Complexity
- For “trajectory” adepts: no such term
  - Might complicate solutions for strictly time-driven processes, such as radioactive decay

Content

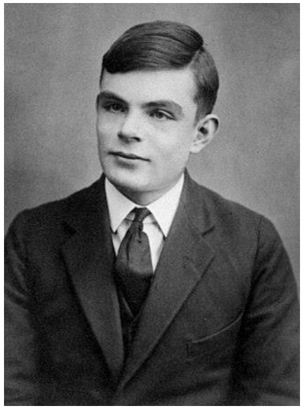
- Basic terms
  - atmospheric tracer
  - temporal and spatial scales
  - life time in the atmosphere
  - life cycle of atmospheric tracers
- What is atmospheric composition model?
- Dispersion equation
- Classifications of dispersion models
- Parts of a dispersion model
- Transport term dispersion models
  - Lagrangian and Eulerian models
- Model Quality Assurance
- Data Assimilation
- Summary

Model Quality Assurance

- **Verification**: confirmation that the model is correctly implemented with respect to the conceptual model
  - mission impossible
- **Validation**: demonstration that a model shows a satisfactory range of accuracy sufficient for the intended application.
  - Requires a specific purpose and objectives, against which the validity is determined
  - mission impossible
- **Evaluation**: quantification of the model performance against the metrics deemed important for the possible model applications
  - largely resembles random search: many metrics, case studies, …

Mission impossible?

- No number of case studies, model tests or alike constitute a formal proof
  - ➢ Turing's halting problem: determining, from a description of an arbitrary computer program and an input, whether the program will finish running or continue to run forever.
  - ➢ Alan Turing, 1936: there is no generic algorithm to solve the halting problem for all possible program-input pairs



Alan Turing, 1928
age 16

- All numerical systems can only be DISproven – by a negative example
- Increasing number of tests and history of the model applications reduces the chances that the next case will fail (but never zero it)

Incomplete verification: a real-life case

- Example: supercooled ternary solution mixture parameterization
  - Carslaw et al, 1995, JRL (F-77 code is in the paper body)
  - Applied in a number of models, used for decades
  - Failed within 3 days in SILAM stratospheric simulations
- FORTRAN operator:
  xsb = (-ks(1)-ks(2)/T-sqrt((ks(1)+ks(2)/T)**2 - 4.*(ks(3)+ks(4)/T)*(ks(5)+ks(6)/T+ks(7)*ln(T)-ln(pw)))) / (2.*(ks(3)+ks(4)/T))
- Breaks down if T = 211.92 ± 0.01K
- Equation:

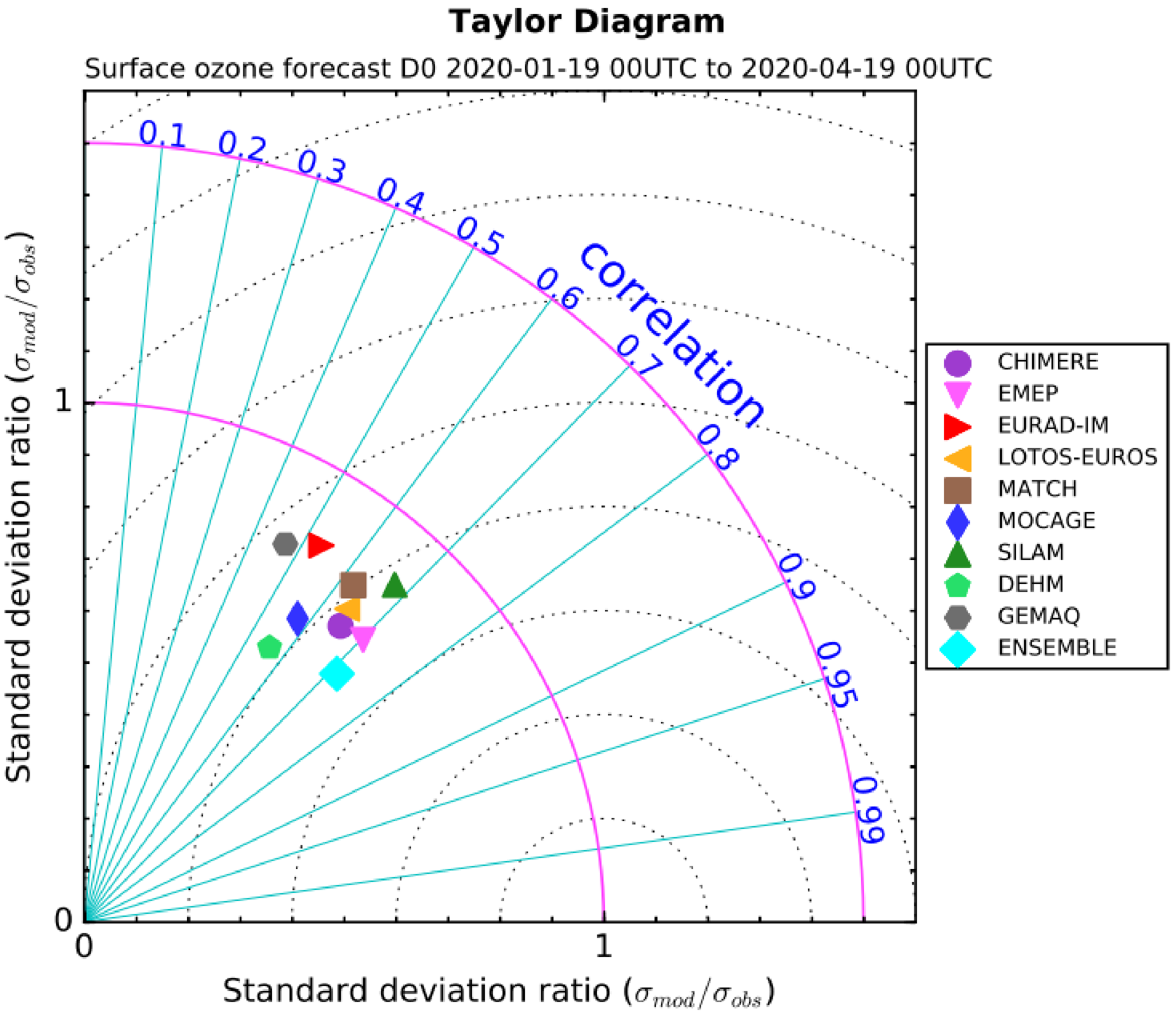
$$x_{\mathrm{Nb}}, x_{\mathrm{Sb}} = \{-k_1 - k_2/T - [(k_1+k_2/T)^2 - 4(k_3+k_4/T)\times (k_5+k_6/T+k_7\ln(T)-\ln(p_w))]^{1/2}\} \;/\; 2(k_3+k_4/T).$$

Model-measurement comparison: evaluation

- The only connection between model and reality
- Data sets from different origin
  - point observations vs grid-mean model results
  - representativeness error
  - instrumental errors
- Observations are expensive => sparse
- Limited number of observed variables
- Specific statistical methods are required to obtain non-trivial conclusions

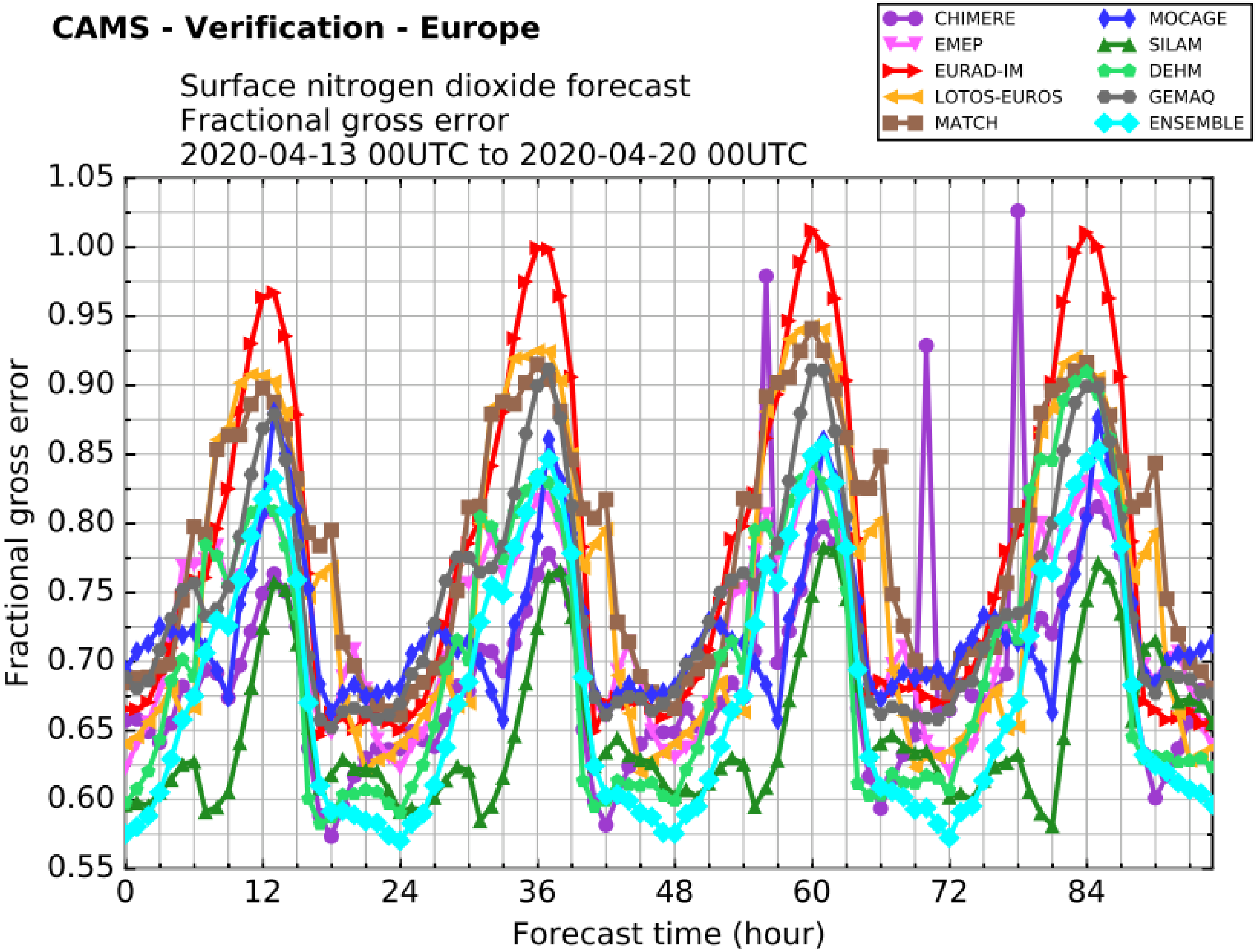
Examples of evaluation statistics

- http://atmosphere.copernicus.eu: Copernicus Atmospheric Monitoring Service



Examples of evaluation statistics

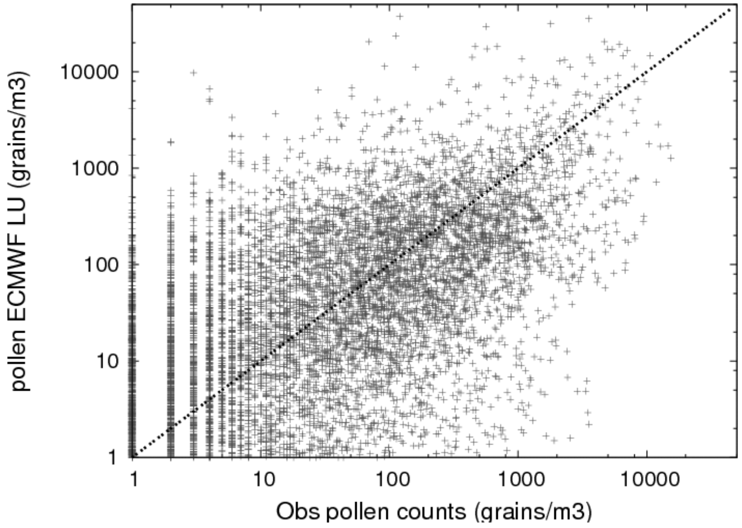
- http://atmosphere.copernicus.eu: Copernicus Atmospheric Monitoring Service

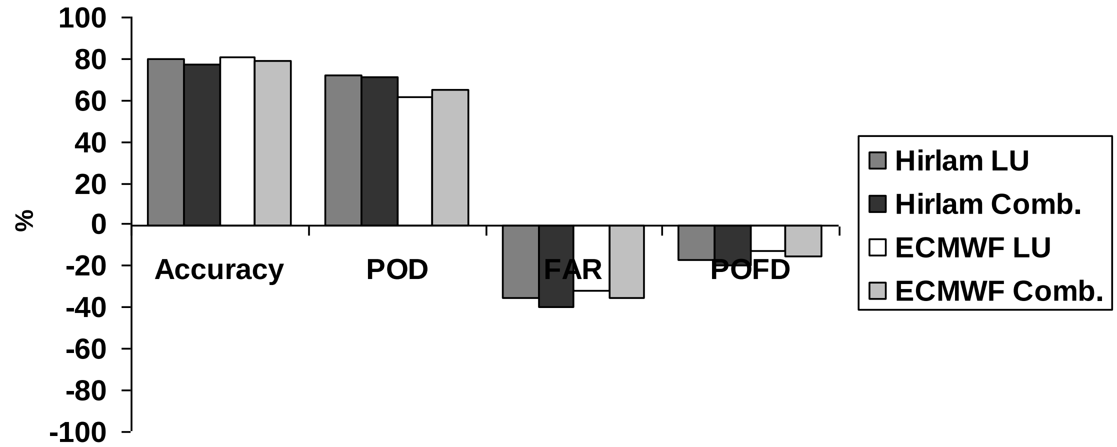


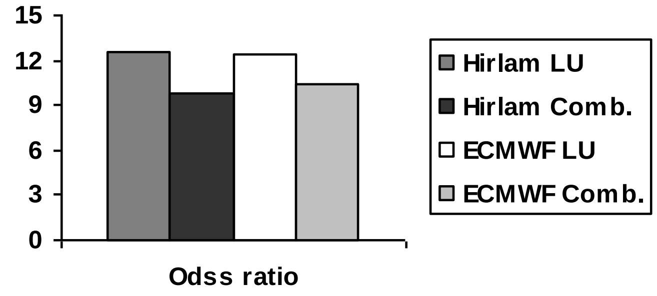
Example: health impact (validation attempt)

- SILAM birch pollen
  - scatter plot

relevant statistical measures







Accuracy=low-low + high-high
POD = high-high
FAR = high-low / all_obs
POFD = high-low / low_obs
OR = chances of high vs low
if predicted high

Siljamo et al, 2012

Model inter-comparison

- Useful if lacking measurements
- Similar features of data
- Large data sets => high precision
- Wide variety of analytical methods
- Ensemble model

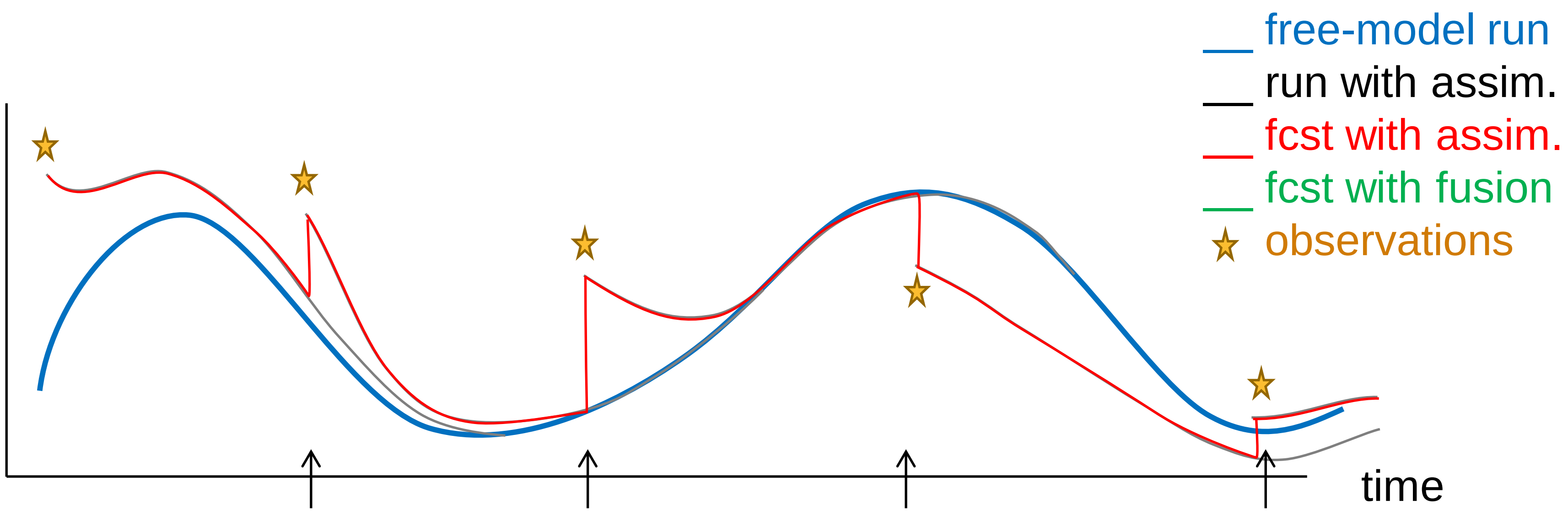
- Possibly, no connection to reality

Content

- Basic terms
  - atmospheric tracer
  - temporal and spatial scales
  - life time in the atmosphere
  - life cycle of atmospheric tracers
- What is atmospheric composition model?
- Dispersion equation
- Classifications of dispersion models
- Parts of a dispersion model
- Transport term dispersion models
  - Lagrangian and Eulerian models
- Model Quality Assurance
- Data Assimilation
- Summary

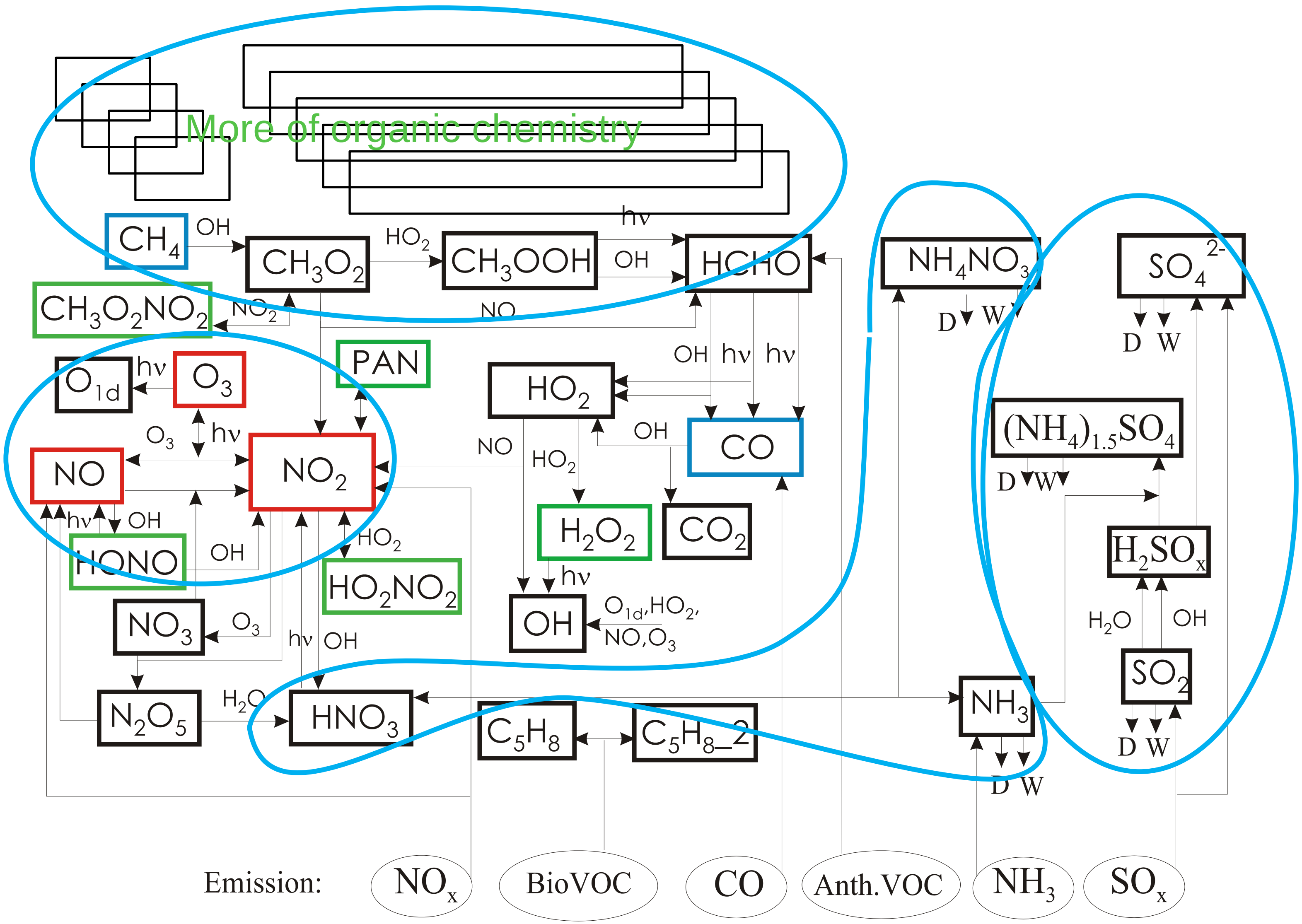
Data assimilation in a classical form

- The approach designed for meteorological forecasting
  - Corrects the model state, i.e. the predicted variables (T,q,U,V,p,...)
  - Works there



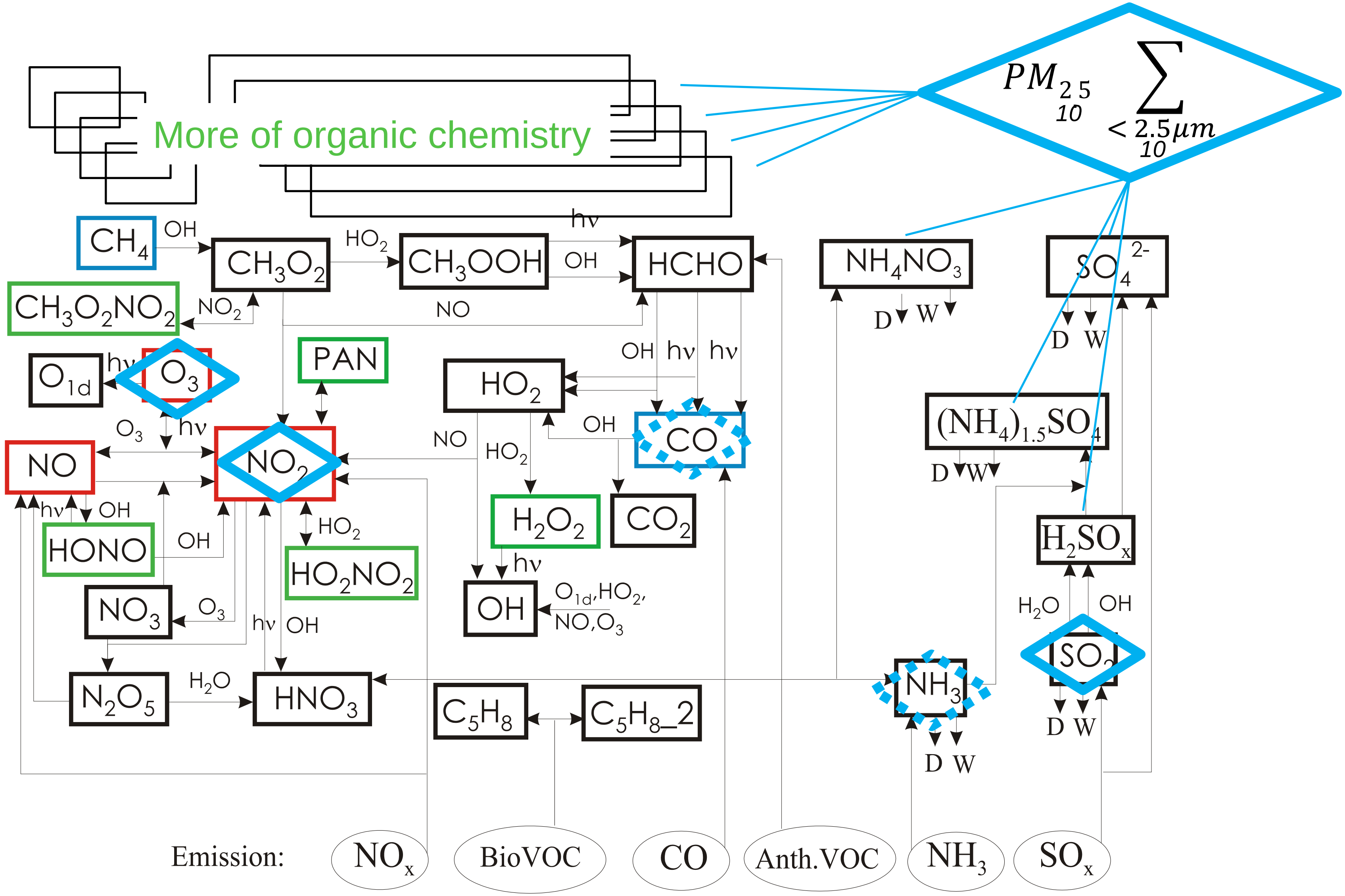
What to assimilate?

chemistry scheme for SOx/NOx/NHx



Soifev, 2000

What do we observe routinely?



Soifev, 2000

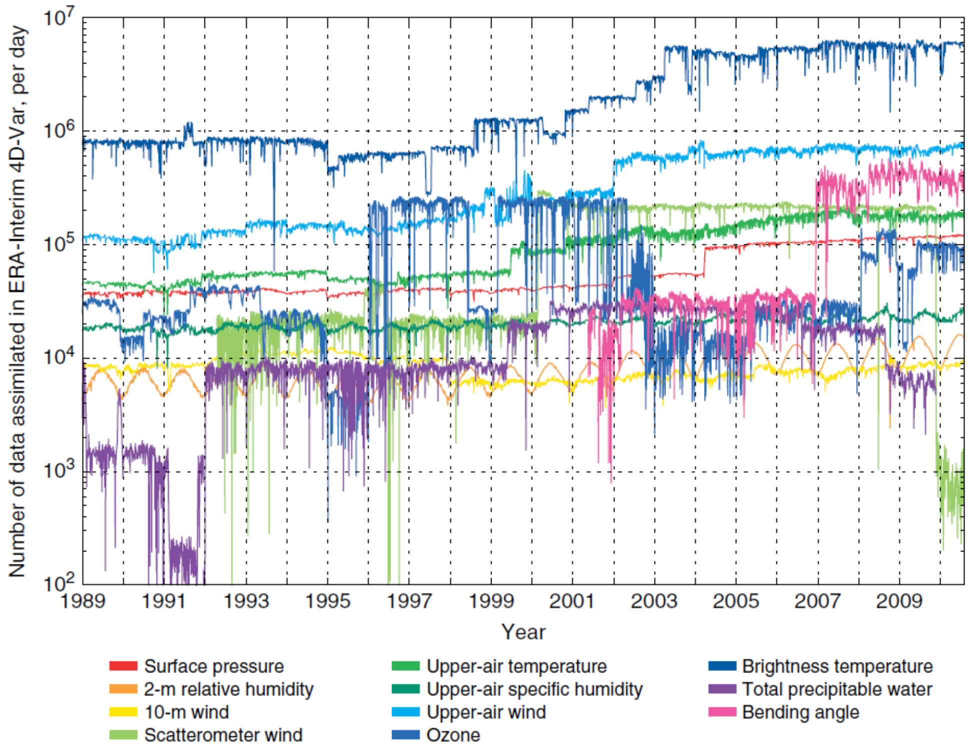
AC observations is a problem

Model variables

model state concentrations ~100	c_1 … c_{nobs}	observed (~10)
	… c_n	non-observed (~100)
emission fluxes ~10	e_1 … e_m	emission (10-20)
model parameters meteorology ~10	p_1 … p_k	physical state (meteorology, ~10)

Each variable is a map of 10^6 - 10^8 grid cells

AC observations is a problem



Dee et al, 2011

Daily count of observations in ERA-Interim

	o3	no2	pm25	pm10	so2
20161101	9839	11424	3746	8628	5826

Ea

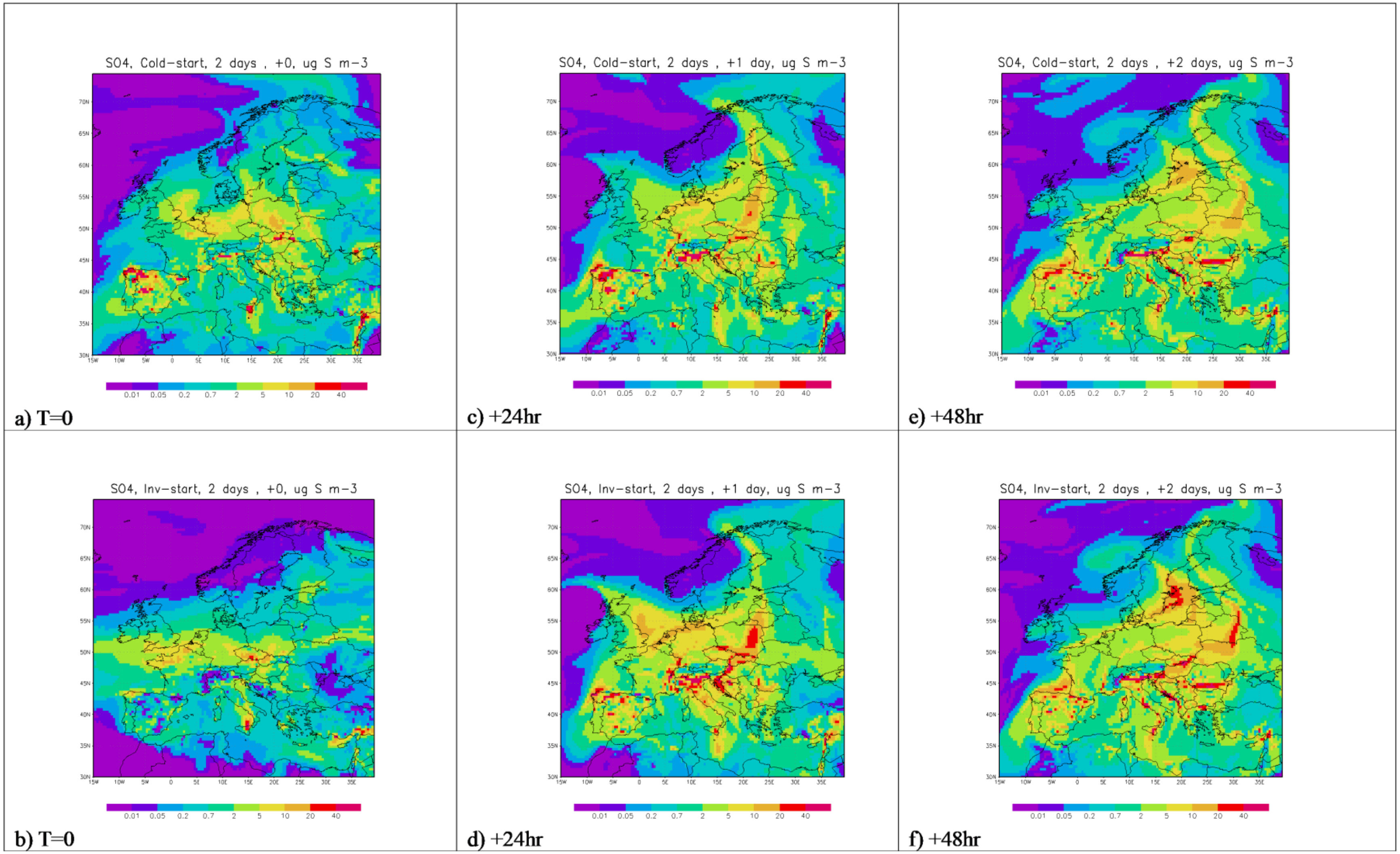
Daily count of observations in CAMS: example of 1.11.2016

More bad news

- Chemical-system state vector contains concentrations of numerous species…
- … and adjusting this vector is not enough:
  - forced motion of this non-autonomous non-linear system may be (and often is) the most significant driver
  - the own system relaxation is often fast and quickly eliminates the effect of DA

Memory of the troposphere

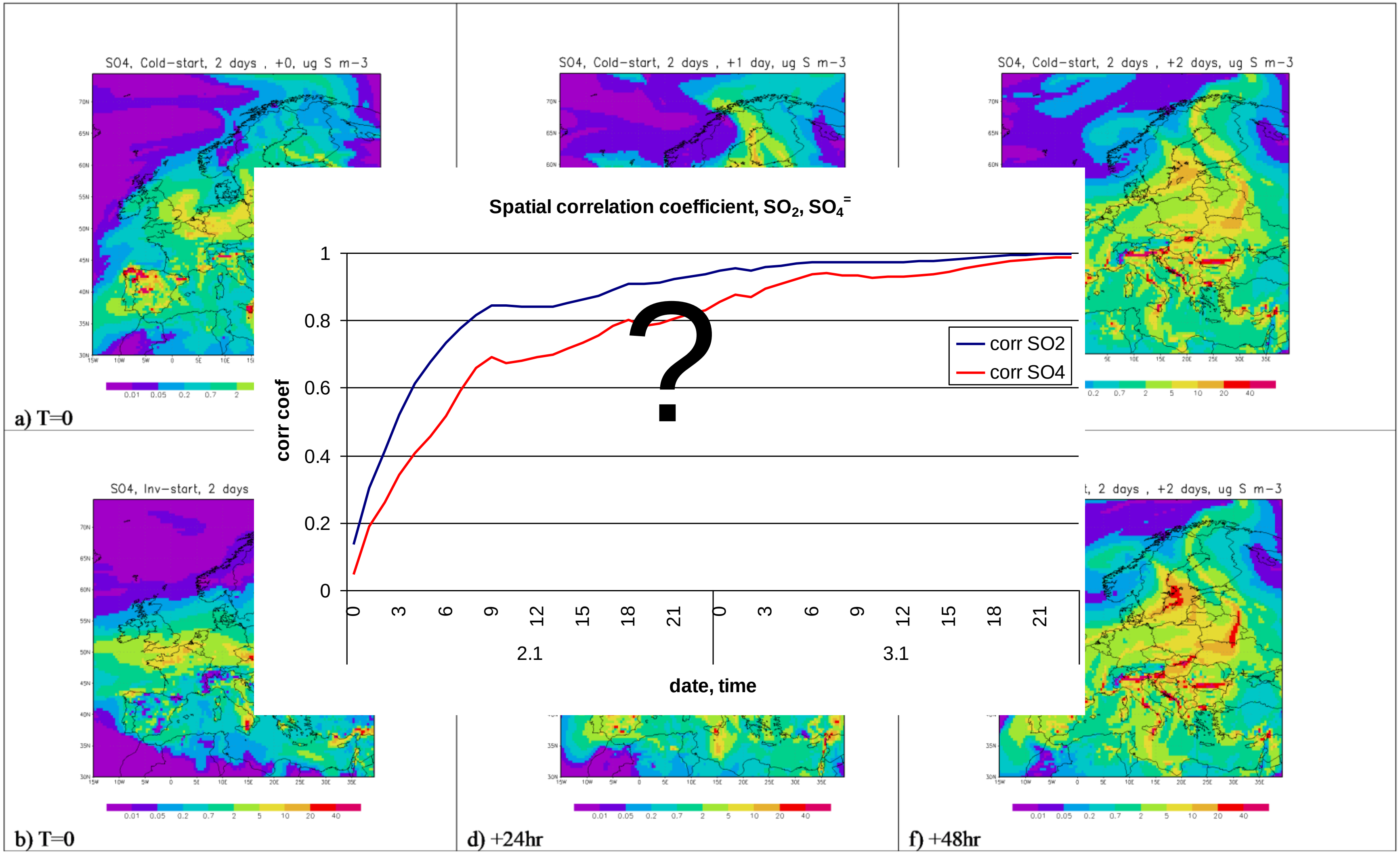
SO_2, hourly mean, µg m^{-3}



Vira & Sofiev, 2012

Memory of the troposphere

SO_2, hourly mean, µg m^{-3}



Vira & Sofiev, 2012

Atmospheric composition DA's grand challenge

- Classical assimilation of concentrations makes little sense: the model forgets the impact much too fast
- Reason: mathematically, the system has short relaxation time, thus being driven by external forcing rather than by initial conditions

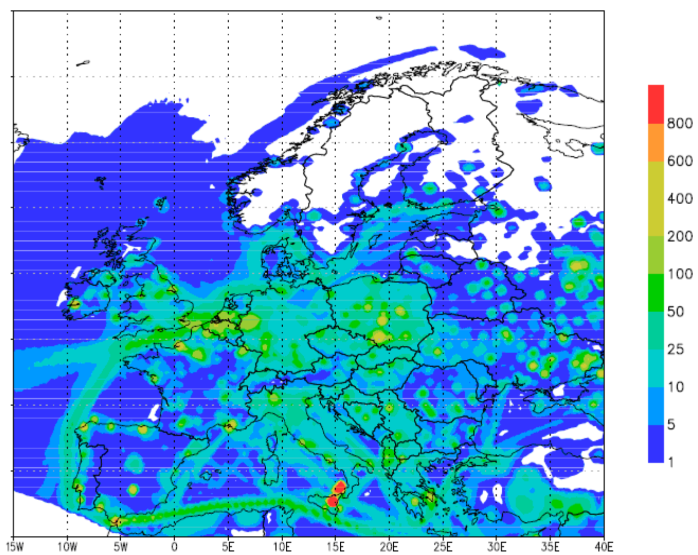
How to handle such system?

- Ignore the difficulties and system constraints and apply known techniques with available observations.
  State estimation with
  - OI / 3D-VAR
- Account for the system constraints and chemical links.
  State estimation with
  - 4D-VAR / EnKF
- Expand the control variable
  - include emission fluxes
  - include meteorology
- Consider non-classical forms of “DA-looking” techniques
  - data fusion
  - optimised ensemble

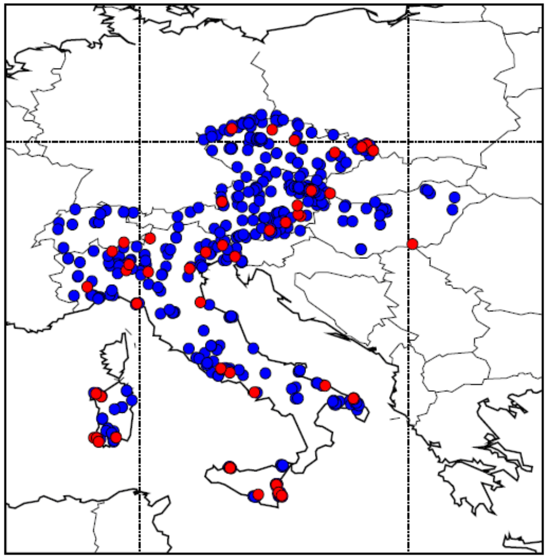
Computational experiment: SOx in Europe

- SILAM experiment 8-22.02.2006
- 3D-VAR, 4D-VAR
- state estimation problem

SO_2 emission



SO_2 observations



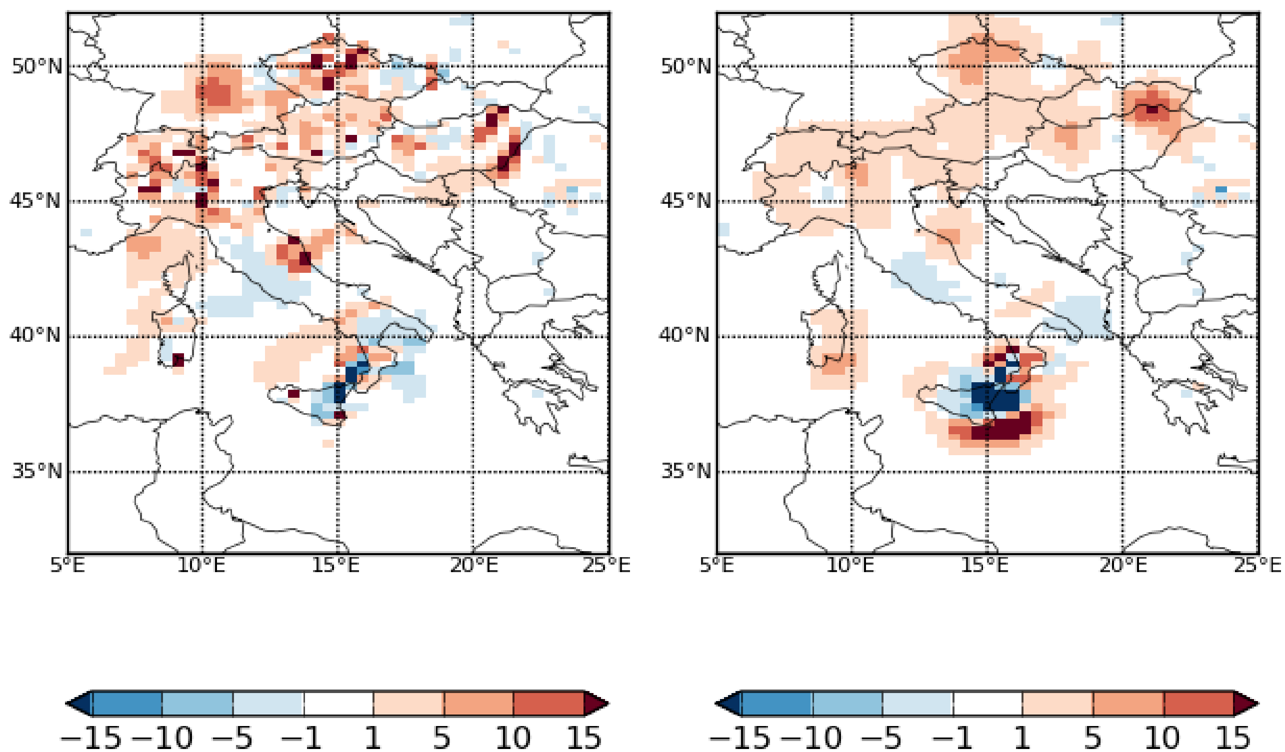
Vira & Sofiev, 2012

Effect of complexity: 3D-VAR vs 4D-VAR

- SO_2 near-surface concentration, changes due to DA

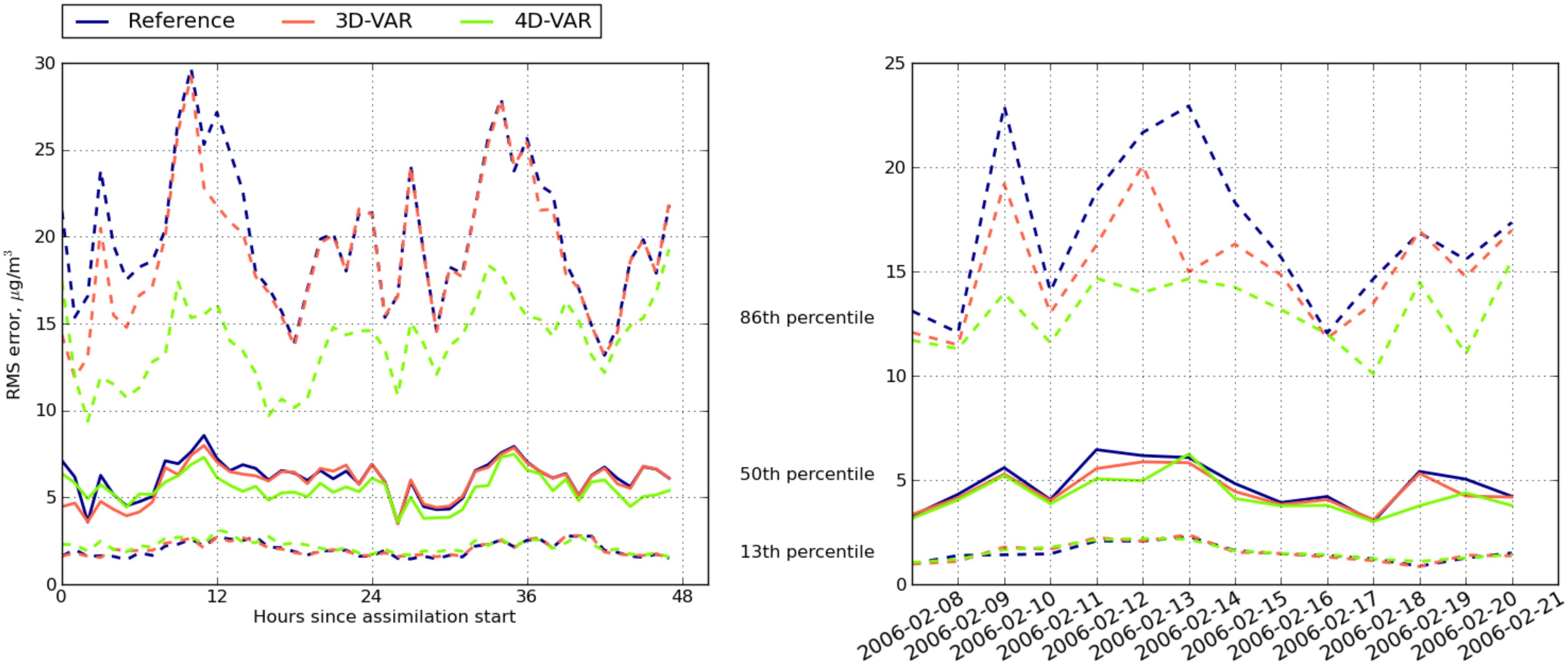
4D-VAR
diagonal error covariance

3D-VAR
non-diagonal error covariance



Vira & Sofiev, 2012

Effect on scores

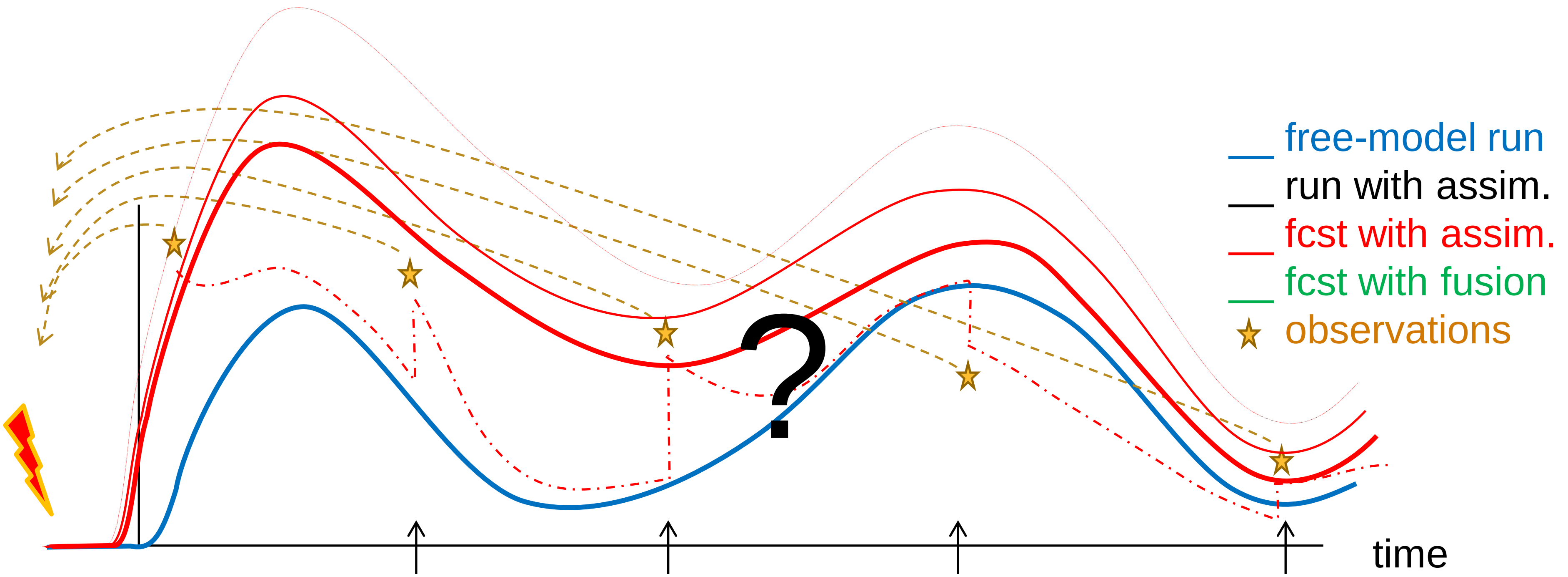


Vira & Sofiev, 2012

How to handle such system?

- Ignore the difficulties and apply known techniques with available observations. State estimation with
  - OI / 3D-VAR
- Account for the system constraints and chemical links. State estimation with
  - 4D-VAR / EnKF
- Expand the control variable: find what has longer impact
  - include emission fluxes
  - include meteorology
- Consider non-classical forms of “DA-looking” techniques
  - data fusion
  - optimised ensemble

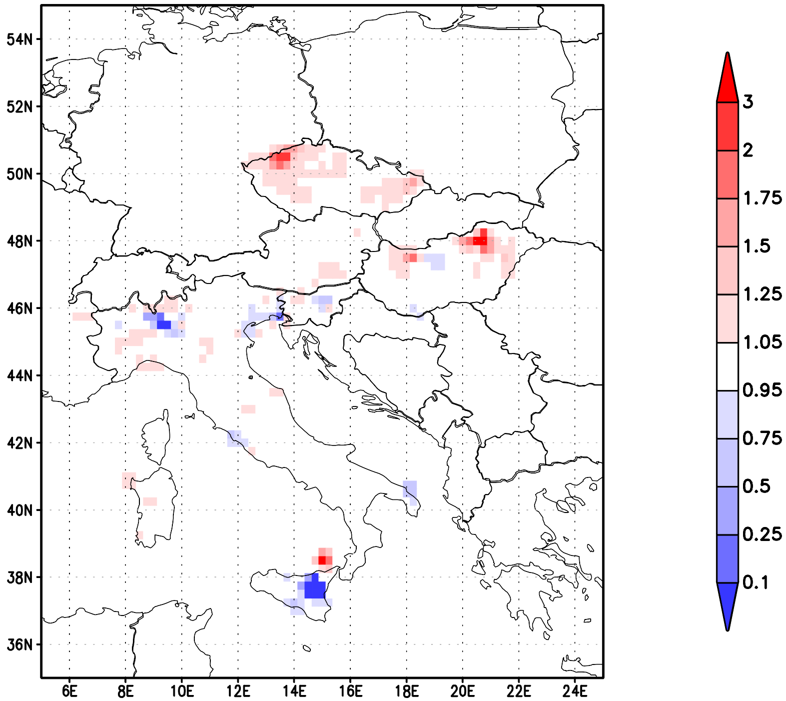
Source term inversion



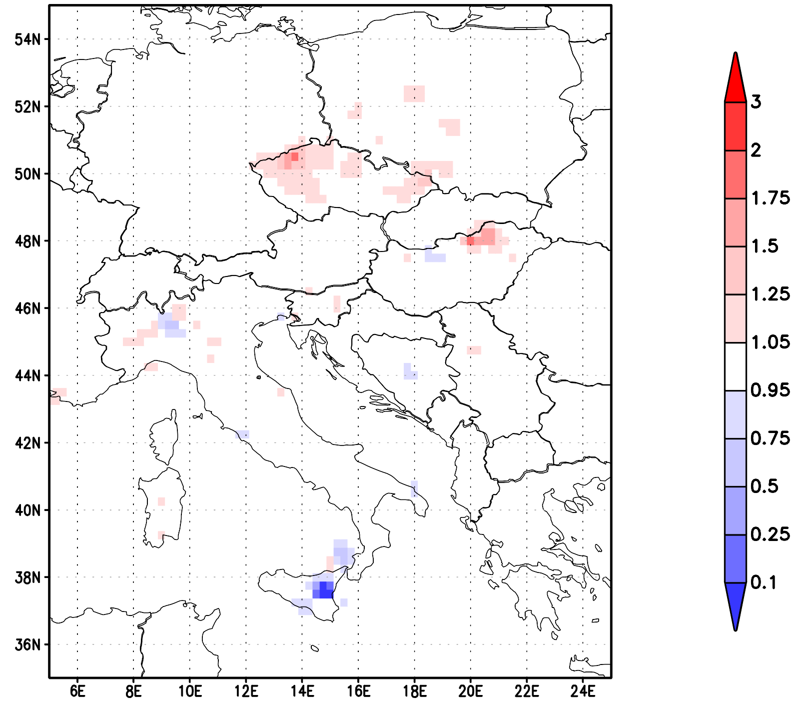
Emission correction factor

- Same SOx experiment, now with 4D-VAR towards emission

Day 1 correction



Weeks 1-2 mean correction



Vira & Sofiev, 2012

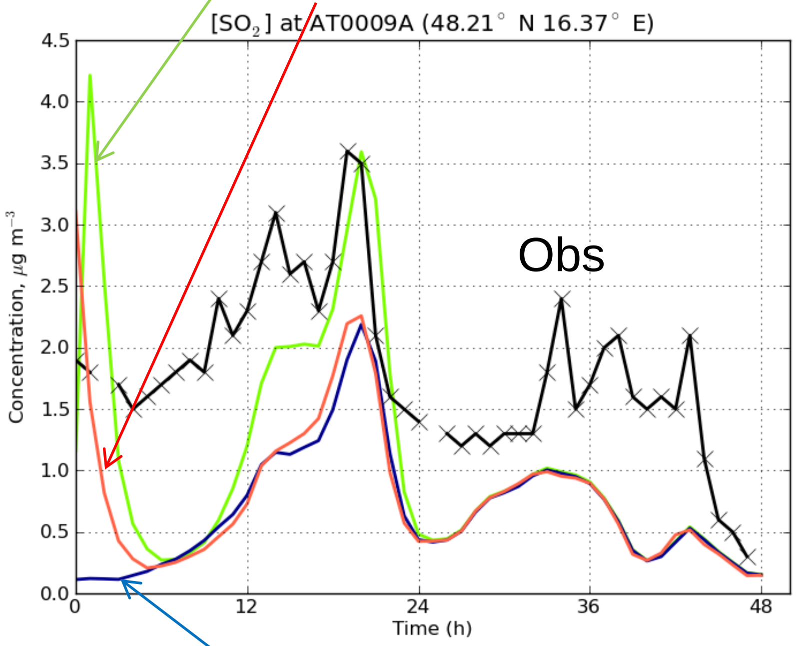
Comparison of the approaches

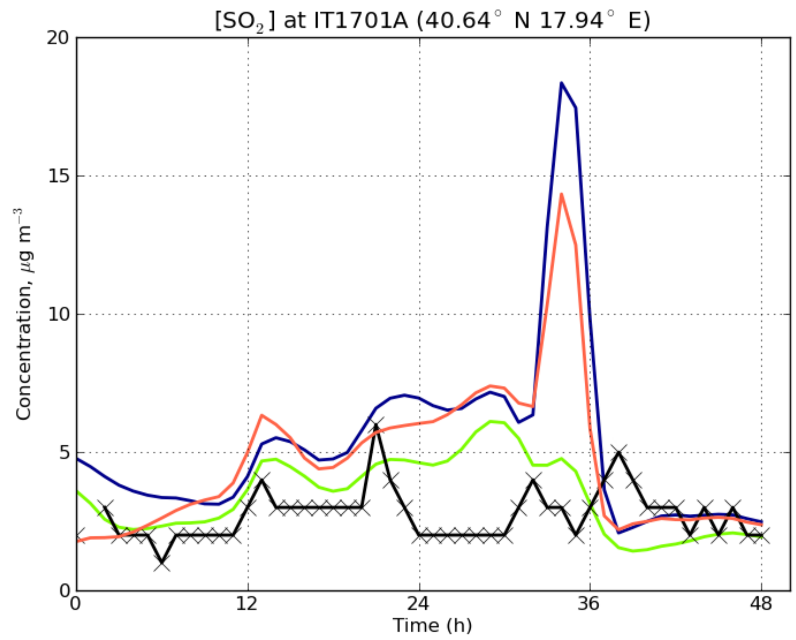
4D-VAR state+emissoin

Site: AT 9, Austria

Site: IT 17, Italy

3D-VAR





reference run

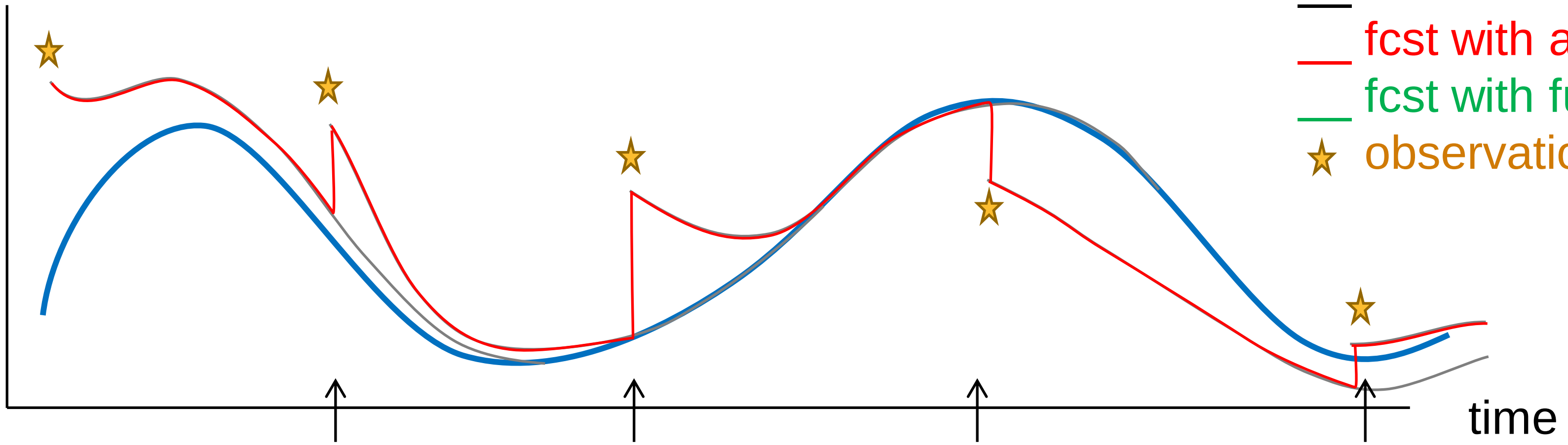
Vira & Sofiev, 2012

How to handle such system?

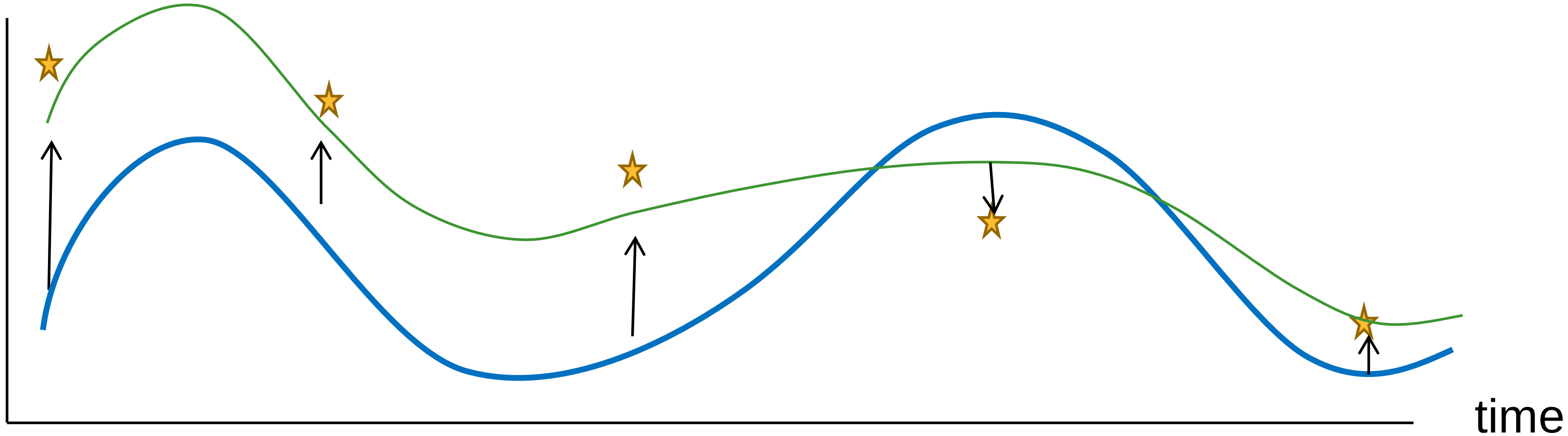
- Ignore the difficulties and apply known techniques with available observations. State estimation with
  - OI / 3D-VAR
- Account for the system constraints and chemical links. State estimation with
  - 4D-VAR / EnKF
- Expand the control variable: find what has longer impact
  - include emission fluxes
  - include meteorology
- Consider non-classical forms of “DA-looking” techniques
  - data fusion
  - optimised ensemble

Data Assimilation vs Data Fusion

- Data assimilation



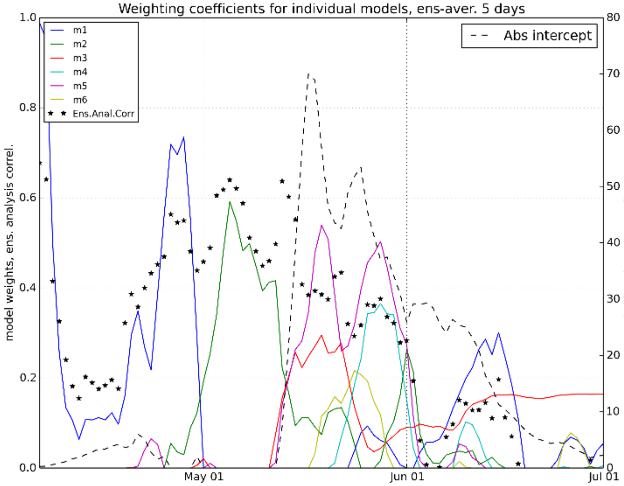
- Data fusion



Data assimilation vs data fusion

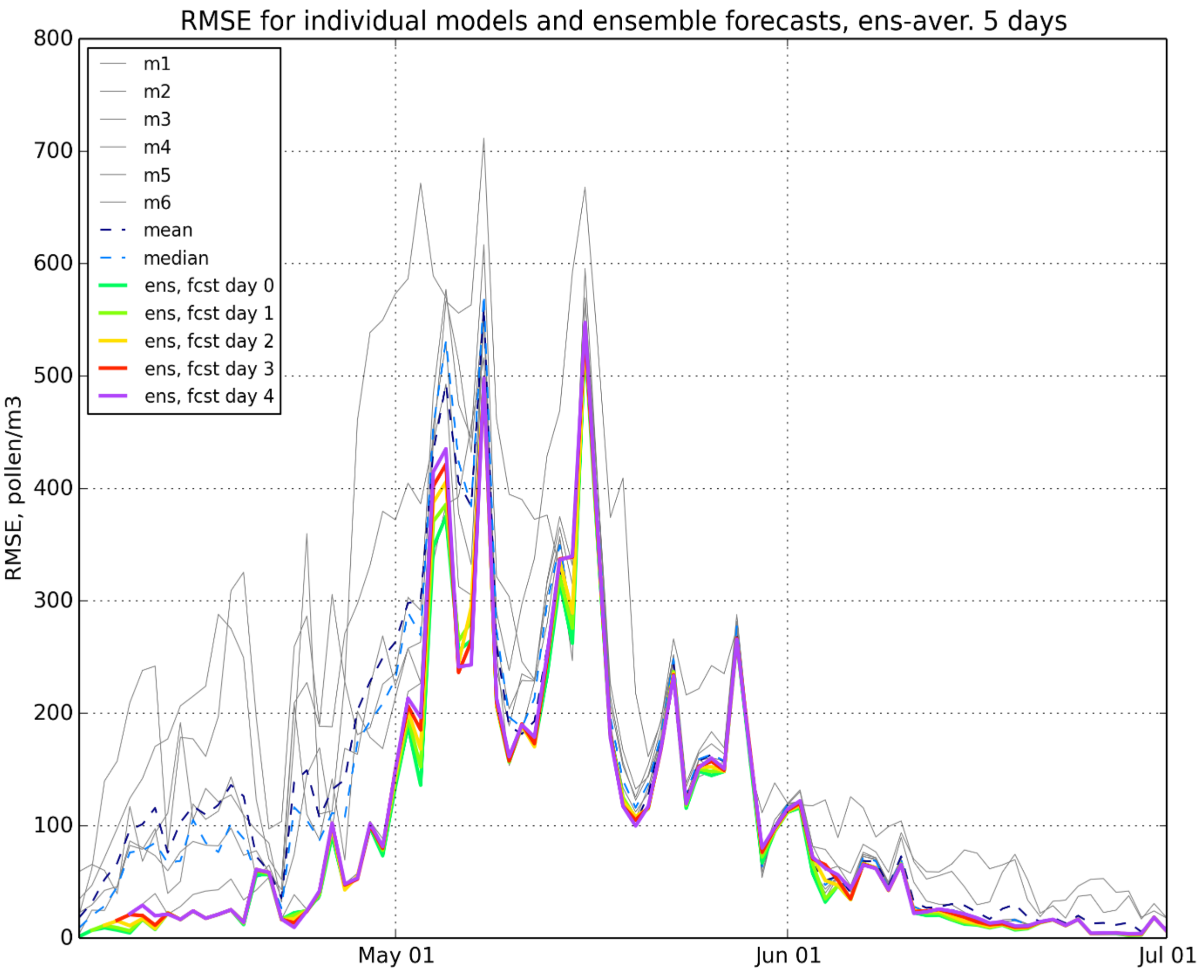
- DA: data are used to adjust model internal variables, parameters or forcing
  - model is "informed" about deviation from the observations and asked to behave better
- DF: data are used to adjust model output after the simulations are finished
  - model has no clue about its errors, it runs without feedback from observations
  - all corrections are applied as post-processing of the model predictions
- A simple example: bias correction
- Promising: error of model predictions (e.g., bias) can be less varying than the predictions themselves

Ensemble-based data fusion: works!



Weights of individual models

Error of individual models, simple and optimised ensemble



Sofiev et al, 2017

Summary model development / application

- **Atmospheric Composition model is a numerical realization of the mass conservation law**
- No universal solution: model is not an image of the real world but an image of a few processes of the real world
- Two main types of transport kernels: Lagrangian and Eulerian
  - Solve the same dispersion equation
  - There is rigorous (well, mostly) transformation from one formalism to another
  - Lagrangian: simple things are simple, complicated things are not possible
  - Eulerian: simple things are complicated, complicated things are possible
- Take the model, which fits the task
  - Emergency response: Lagrangian is popular but Eulerian solutions are challenging this status-quo
  - Atmospheric composition: exclusively Eulerian

Summary model evaluation

- Model verification / validation / evaluation
  - evaluation is the most-meaningful objective for the model as a whole, can (and should) be performed by the model user
  - **The principle of known quantified quality**
- Verification is applied for model sub-systems, development stage
- Validation is the ultimate goal in model applications, rarely reached
  - Each model application requires tailored evaluation, followed by decision of the model applicability for the task
  - Statistical evaluation measures are task-specific

Summary for Data Assimilation

- Atmospheric composition is tough for data assimilation: violates almost all assumptions behind DA methods
- Classic methods give ~20% of improvement for the analysis, next to nothing for the follow-up forecast
- Expansion of control variable is among the most-promising albeit complicated approaches for improving the forecast
- Data fusion technology shows very promising first results