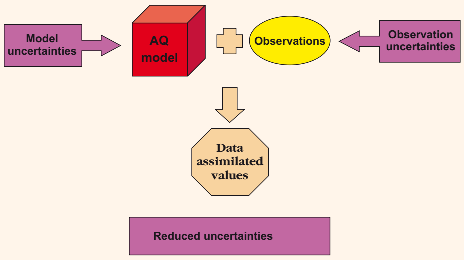
Air4EU

Cross Cutting Issue 5: Data Assimilation

Sam-Erik Walker (NILU), **Martijn Schaap** (TNO), **Lora Slini** (AUTH)

Benefits of using data assimilation

- Assessing the air quality (AQ) at local, urban and regional scales can potentially be improved by combining models and observations, i.e. by using data assimilation
- Combining models and observations reduces uncertainties inherent in both and is strongly recommended in order to achieve a better depiction of the real situation in an area of interest
- The most important areas where data assimilation can be used to improve AQ assessments are in:
  - Estimating concentration levels and their uncertainties
  - Estimating different source and scale contributions to the AQ levels
  - Performing and evaluating design and optimization of new or existing monitoring networks

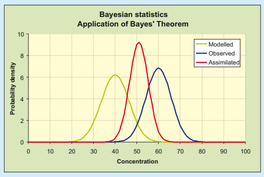


Combining models and monitoring leads to assimilated results with reduced uncertainties.

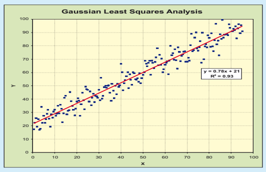
Historical Background

The idea of data assimilation dates back to work in the 18th and 19th century by Thomas Bayes and Carl Friedrich Gauss.

Bayes developed Bayes' theorem, which describes how to combine prior knowledge (model) with observations.



Gauss developed the method of least squares, which describes how to fit a model (e.g. a straight line) to a set of observations.



Applications in meteorology

In the past few decades data assimilation has been used with great success in meteorology, and especially in weather forecasting.

Development of assimilation methodologies are a major activity at different national and international meteorological research centres in Europe such as e.g. DARC (Data Assimilation Research Centre), ECMWF, KNMI, Meteo-France, UK Met Office etc.

The importance of data assimilation is also highlighted in international earth observation strategies such as the recent GMES (Global Monitoring for Environment and Security) and GEOSS (Global Earth Observation System of Systems).

Applications in air quality modelling

Application of data assimilation for improved air pollution dispersion modelling and AQ assessment is still in its early stages, though it has been applied already to several atmospheric chemical transport models on urban and regional scales

In the Air4EU project data assimilation has been used in a number of case studies

See references below for more details

The data assimilation cycle

Data assimilation always involves two phases:

1. Before new observations are used
2. After the observations have been used to correct the model

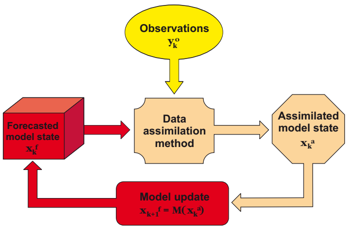
Model calculated values are usually produced on a timely basis, e.g. every hour.

When observations becomes available, data assimilation can be used to update the model calculations.

There are two types of updates: Active and Passive

Active: The updated model values are actively used in further calculations with the model.

Passive: The observations are only used to adjust the model output at each time step, but does not influence the model further.



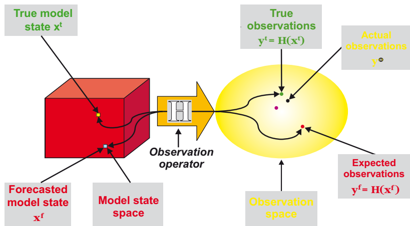
How to link model variables with observations

Model variables are either:

- Input to models, such as emissions, meteorological parameters, boundary conditions etc., or
- Output from models, typically concentration values in regular grids, irregular receptor points etc.

An observation operator maps the current model variables (a point in the model state space) onto a set of expected observations (a point in the observation space).

The actual observations leads to adjustments of the model state through the observation operator, using estimates of model and observations uncertainties.



General recommendations

- Define the main objective(s) of using data assimilation for improved AQ assessment
- Aquire the necessary modelling and assimilation expertise to perform the task
- Conduct a proper evaluation of the model
- Select variables to estimate based on e.g. sensitivity analysis
- Select an assimilation methodology
- Perform assimilation using real observations
- Evaluate assimilation systems using simulated observations

See below and Walker et al. (2006) for more details.

Defining the main objectives

One should provide a clear and unambiguous statement of the main objective(s) and purpose(s) of using data assimilation for improved AQ assessment.

Examples of such objectives could be e.g.:

- To obtain a more accurate modelling of the air pollution in an area, e.g. in order to check compliance with limit value(s).
- To obtain improved estimates of contributions from different sources and to check compliance with emission regulations.

Selecting model variables to estimate

The purpose of data assimilation is to adjust the values of some or all of a set of uncertain model variables in order to better fit the model to observed concentrations.

When selecting variables to be used one should generally focus on variables which are most uncertain, and for which observations will be most useful.

This can be based on e.g. model sensitivity analysis.

Variables selected could e.g. either be:

- Input data to models such as e.g. emissions, meteorology etc.
- Output data such as grid concentrations etc.

Data assimilation methodologies:

- Optimal Interpolation (OI)
- Statistical Interpolation (Kriging)
- Variational methods: 3D-Var, 4D-Var
- Reduced rank and Ensemble Kalman filters
- Sequential Monte Carlo methods (particle filters)
- Other statistical methods such as regression analysis, receptor modelling, principal component analysis, time series analysis, neural networks etc.

A description of these methods with references is given in Walker et al. (2006)

References

Kalnay, E. (2003), Atmospheric Modeling, Data Assimilation and Predictability, *Cambridge University Press, Cambridge UK.*

Swinbank R., et al. (2003) Data Assimilation for the Earth System, *Nato Sciences Series, IV. Earth and Environmental Sciences, Vol. 26.*

Walker, S.E., Schaap M., Slini L. (2006) Data assimilation, Air4EU WP6 synthesis, *Milestone report 6.8.*

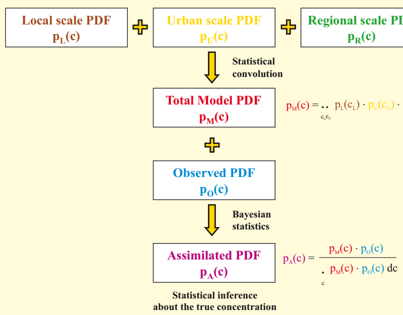
Air4EU data assimilation tool SAM

A PC-based software tool SAM (Simple Statistical System for Assimilation of Models and Monitoring data) has been made in order to enable users to perform data assimilation in a simple way.

The tool is based on specifying concentration contributions from each of the spatial scales, local, urban and regional, together with a specification of their uncertainties in the form of probability density functions (PDFs).

Combining the total model (local + urban + regional) concentration with an observed concentration, the tool calculates an assimilated concentration with its associated uncertainty (PDF) using Bayesian statistics.

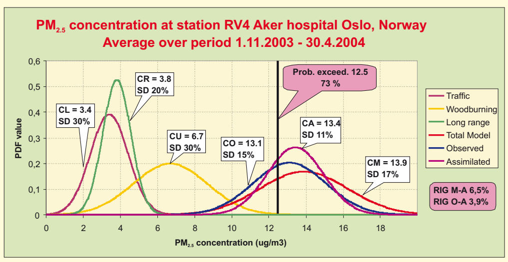
For more information please contact Sam-Erik Walker (sew@nilu.no)



$$p_M(c) = \iint_{c_L c_U} p_L(c_L) \cdot p_U(c_U) \cdot p_R(c - c_L - c_U)\, dc_L\, dc_U$$

$$p_A(c) = \frac{p_M(c) \cdot p_O(c)}{\int_c p_M(c) \cdot p_O(c)\, dc}$$

Example of using the SAM tool



NILU PP 29/2006

For more information about Air4EU please visit www.air4eu.nl