Lecture 2
Meteorological Data Assimilation

H. Elbern
Rhenish Institute for Environmental Research at the
University of Cologne
and
Virt. Inst. for Inverse Modelling of Atmospheric Chemical
Composition
Objectives of this lecture

Presentation of

• a review of the satellite observation data set for meteorological applications

• numerical implementation measures for spatio-temporal assimilation algorithms
Meteorological remote sensing observation suite (1)

- *Atmospheric* sounding channels from *passive* instruments: atmospheric temperature and humidity (Atmospheric sounding channels from the **HIRS** (High resolution Infrared Sounder) and **AMSU** (Advanced Microwave Sounding Unit) on board NOAA)
- *Surface* sensing channels from *passive* instruments "imaging" channels: located in atmospheric "window" regions of the infra-red and microwave spectrum: surface temperature and cloud track information
- Surface sensing channels from active instruments: scatterometer emit microwave radiation for ocean wind retrievals. Some similar-class active instruments such as altimeters and **SARS** (Synthetic Aperture Radars) provide information on wave height and spectra.
Meteorological remote sensing observation suite (2)

• visible (Lidars) or the microwave (radars) analyse the signal backscattered
  – molecules, aerosols, water droplets or ice particles.
  – penetration capability allow the derivation of information on cloud base, cloud top, wind profiles (Lidars) or cloud and rain profiles (radars).

• Radio-occulation technique using GPS (Global Positioning System). GPS receivers (e.g. METOP/GRAS) measure the Doppler shift of a GPS signal refracted along the atmospheric limb path.
Observation systems (1)

Type and number of observations used to estimate the atmosphere initial conditions in a typical day. (Buizza, 2000)

\[
\text{dim}_{\text{observation space}} = O(10^6)
\]
Observation systems (2): In-situ observations

ECMWF Data Coverage (All obs) - SYNOP/SHIP
04/NOV/2003; 00 UTC
Total number of obs = 15325

ECMWF Data Coverage (All obs) - TEMP
05/NOV/2003; 00 UTC
Total number of obs = 594
Observation systems (3): polar orbiting satellites (e.g. AMSU-A)

Data coverage for the NOAA-15 (red), NOAA-16 (cyan) and NOAA-17 (blue) AMSU-A instruments, for the four 6-hour periods centred at 00, 06, 12 and 18 UTC 12 November 2002. The plots show the data used for AMSU-A channel 5, which is a temperature-sounding channel in the mid and lower troposphere.
Data coverage provided by the GOES satellites (cyan and orange) and the METEOSAT satellites (magenta and red) for 00 UTC 10 May 2003. The total number of observations was 266,878.
Information source: numerical model

\[
\frac{\partial U}{\partial t} + \frac{1}{a \cos^2 \theta} \left\{ U \frac{\partial U}{\partial \lambda} + \nu \cos \theta \frac{\partial U}{\partial \theta} \right\} + \frac{\partial \eta}{\partial \eta} \left\{ -fU + \frac{1}{a} \left\{ \frac{\partial \phi}{\partial \lambda} + R_{\text{dry}} T_V \frac{\partial}{\partial \lambda} (\ln p) \right\} \right\} = P_U + K_U
\]

\[
\frac{\partial V}{\partial t} + \frac{1}{a \cos^2 \theta} \left\{ U \frac{\partial V}{\partial \lambda} + V \cos \theta \frac{\partial V}{\partial \theta} + \sin \theta (U^2 + V^2) \right\} + \frac{\partial \eta}{\partial \eta} \left\{ fU + \frac{\cos \theta}{a} \left\{ \frac{\partial \phi}{\partial \theta} + R_{\text{dry}} T_V \frac{\partial}{\partial \theta} (\ln p) \right\} \right\} = P_V + K_V
\]

\[
\frac{\partial T}{\partial t} + \frac{1}{a \cos^2 \theta} \left\{ U \frac{\partial T}{\partial \lambda} + V \cos \theta \frac{\partial T}{\partial \theta} \right\} + \frac{\partial \eta}{\partial \eta} \left\{ \frac{\kappa T_V \omega}{(1 + (\delta - 1)q)p} \right\} = P_T + K_T
\]

\[
\frac{\partial q}{\partial t} = \frac{1}{a \cos^2 \theta} \left\{ U \frac{\partial q}{\partial \lambda} + V \cos \theta \frac{\partial q}{\partial \theta} \right\} = \frac{\partial q}{\partial \eta} = P_q + K_q
\]

\[
\frac{\partial p_{\text{surf}}}{\partial t} = -\int_0^\eta \left\{ \nabla \cdot \left( \nu H \frac{\partial p}{\partial \eta} \right) \right\} \, d\eta
\]

\[
\eta(0, p_{\text{surf}}) = 0
\]
At which retrieval level to assimilate remote sensing data?

- **Assimilation of retrieved products from space agencies or research institutes (“level 2”)**
  - Pro: most simple for assimilation (assimilation “like in situ observations”)
  - Con: inexact as $x_b$ and $B$ are inconsistent and mostly of poorer quality as available at NWP centres

- **Locally produced or “1D-Var” or “Averaging Kernel” retrievals**
  - Pro: $x_b$ and $B$ much better known (e.g. fronts featured and consistent with model)
  - Con: $x_b$ and $B$ used twice with the subsequent assimilation: $y$ and $x_b$ correlated

- **Direct assimilation of radiances (“level 1”)**
  - Pro: retrieval step is essentially incorporated within the main analysis by finding the model variables that minimize a cost function measuring the departure between the analysed state and both the background and available observations.
  - Con: Fast linearized radiative transfer scheme and its adjoint is needed.
Example for averaging kernels

Averaging kernels (in K/K) for AIRS (left panel) and HIRS (right panel) instruments.
Background Error Covariance Matrix

The BECM is of utmost importance, as it
• weights the model error against the competing observation errors
• spreads information from observations to the adjacent area
• influences coupled parameters: temperature ←→ wind field, chemical constituents (i.e. multivariate correlation)
• (serves for preconditioning in the case of variational assimilation)

\[
\omega(r) = \exp\left(-\frac{r^2}{2L^2}\right)
\]

\[
\omega(r) = \left(1 + \frac{1}{L}\right) \exp\left(-\frac{r}{L}\right)
\]

2 simple yet popular covariance models:

- **Gaussian**
- **Balgovind**
Background Error Covariance Matrix $\mathbf{B}$
(2 pragmatic methods to estimate $\mathbf{B}$)

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

$$B_{ij} = \frac{1}{K} \sum_{n=1}^{K} (x^n_i - \bar{x}_i)(x^n_j - \bar{x}_j)$$

2. NMC method: (e.g. Parrish and Derber, 1992)

$$\delta x(t_i) = x^{f48h}(t_i) - x^{f24h}(t_i)$$

$$\mathbf{B} = \frac{1}{K} \sum_{i=1}^{K} \delta x(t_i)\delta x^T(t_i)$$

Difference column vector of model states of two forecasts at some absolute time $t_i$, one starting 24 hours earlier than the other, thus trying to capture the most sensitive errors. Covariances calculated by outer product.
Novel multivariate covariance modelling (1)

Covariance tuning
Multivariate covariances following Derber and Bouttier, 1999

\[
\mathbf{B} = \begin{pmatrix}
    E(T'T_t) & E(T'u_t) & E(T'v_t) & E(T'q_t) \\
    E(u'T_t) & E(u'u_t) & E(u'v_t) & E(u'q_t) \\
    E(v'T_t) & E(v'u_t) & E(v'v_t) & E(v'q_t) \\
    E(q'T_t) & E(q'u_t) & E(q'v_t) & E(q'q_t)
\end{pmatrix}
\]

B composition of cross-correlation block matrices

cascadic linear relationship between parameters with separation in balanced B and unbalanced U component.

\[
\begin{align*}
T' &= T'_B \\
    &= T'_B + u'_U = K_{u_T}T' + u'_U \\
    v' &= v'_B + v'_U = K_{v_T}T' + K_{uu}u' + v'_U \\
    q' &= q'_B + q'_U = K_{q_T}T' + K_{qu}u' + K_{qv}v' + q'_U
\end{align*}
\]
Novel multivariate covariance modelling (2)

then, $B$ amenable to factorization by

$$B = KB_U K^T$$

where

$$K = \begin{pmatrix}
I & 0 & 0 & 0 \\
K_{uT} & I & 0 & 0 \\
K_{vT} & K_{vu} & I & 0 \\
K_{qT} & K_{qu} & K_{qv} & I
\end{pmatrix},
B_U = \begin{pmatrix}
B_{TT} & 0 & 0 & 0 & 0 \\
0 & B_{uu} & 0 & 0 \\
0 & 0 & B_{vv} & 0 \\
0 & 0 & 0 & B_{qq}
\end{pmatrix}$$
Construction of the adjoint code

forward model (forward differential equation) → algorithm (solver) → code → adjoint code

backward model (backward differential equation) → adjoint algorithm (adjoint solver) → code
Constructing the adjoint from forward code

Program line

\[ x_{i}^{k+1} = f(x_{1}^{k}, \ldots, x_{i}^{k}, \ldots, x_{n}^{k}) \]

Tangent linear form

\[ \delta x_{i}^{k+1} = \frac{\partial f}{\partial x_{1}^{k}} \delta x_{1}^{k} + \cdots + \frac{\partial f}{\partial x_{i}^{k}} \delta x_{i}^{k} + \cdots + \frac{\partial f}{\partial x_{n}^{k}} \delta x_{n}^{k} \]

Blown-up equivalent notation as ”transformation“

\[
\begin{pmatrix}
\delta x_{1}^{k} \\
\vdots \\
\delta x_{i}^{k} \\
\vdots \\
\delta x_{n}^{k} \\
\delta x_{i}^{k+1}
\end{pmatrix} =
\begin{pmatrix}
1 & 0 & \cdots & \cdots & 0 & 0 \\
\vdots & \ddots & \ddots & \cdots & \vdots & \vdots \\
0 & \cdots & 1 & \cdots & 0 & 0 \\
\frac{\partial f}{\partial x_{1}^{k}} & \cdots & \frac{\partial f}{\partial x_{i}^{k}} & \cdots & \frac{\partial f}{\partial x_{n}^{k}} & 0 \\
\delta x_{1}^{k} \\
\vdots \\
\delta x_{i}^{k} \\
\delta x_{n}^{k} \\
\delta x_{i}^{k+1}
\end{pmatrix}
\]
Constructing the adjoint from forward code

adjoint (transposition)

\[
\begin{pmatrix}
\lambda x_i^k \\
\vdots \\
\lambda x_i^k \\
\lambda x_i^k \\
\lambda x_i^k + 1
\end{pmatrix}
= 
\begin{pmatrix}
1 & \frac{\partial f}{\partial x_i^k} \\
0 & \vdots \\
0 & 1 \\
0 & 0 \\
0 & 0
\end{pmatrix}
\begin{pmatrix}
\lambda x_1^k \\
\vdots \\
\lambda x_i^k \\
\vdots \\
\lambda x_n^k \\
\lambda x_{n+1}^k
\end{pmatrix}
\]

Hence

\[
\begin{align*}
\lambda x_1^k &= \lambda x_1^k + \frac{\partial f}{\partial x_1^k} \lambda x_1^{k+1} \\
\vdots &= \vdots \\
\lambda x_i^k &= \lambda x_i^k + \frac{\partial f}{\partial x_i^k} \lambda x_i^{k+1} \\
\vdots &= \vdots \\
\lambda x_n^k &= \lambda x_n^k + \frac{\partial f}{\partial x_n^k} \lambda x_i^{k+1} \\
\lambda x_{n+1}^k &= 0
\end{align*}
\]
Adjoint code verification

Let $F$ be a coded function, the tangent–linear (TL) and adjoint (AJ) of which be under development. Test the correctness of both the TL and AJ code!

\[ y = F(x), \quad F : \mathcal{R}^n \rightarrow \mathcal{R}^m, \quad z = D(y), \quad D : \mathcal{R}^m \rightarrow \mathcal{R} \]

\[ \delta z = \left( \frac{\partial D}{\partial y} \right) \left( \frac{\partial F}{\partial x} \right) \delta x = \langle \left( \frac{\partial D}{\partial y} \right)^T, H \delta x \rangle \]

\[ = \langle H^T \left( \frac{\partial D}{\partial y} \right)^T, \delta x \rangle = \left( \left( \frac{\partial F}{\partial x} \right)^T \left( \frac{\partial D}{\partial y} \right)^T \right)^T \delta x \]

Application of $H$ and $H^T$ means running the TL and AJ code, respectively. With suitably selected input $\frac{\partial D}{\partial y_i} = (0, \ldots, 0, 1, 0, \ldots, 0)^T$ and $\delta x_i := (0, \ldots, 0, 1, 0, \ldots, x_n)^T$ the $i^{th}$ row of $H$ can be compared with the $i^{th}$ column of $H^T$. Further, the TL code linearized at input $x$ be approximated by $\lim_{\delta x_i \rightarrow 0} \frac{1}{\delta x_i} H((x + \delta x_i) - x)$, where $\delta x_i = (0, \ldots, 0, \delta x_i, 0, \ldots, x_n)^T$. 
Adjoint compiler

Some examples

• Odyssee
• TAMC
• Tapenade
• ADOL-C and ADOL-F (includes ability for higher order derivatives)
• IMAS

For much more comprehensive list see:
www.autodiff.org/?module=Tools&submenu=&language=ALL

Also codes from adjoint compilers must be tested!!
Algorithm and performance bugs occur!
Adjoint integration “backward in time”

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

\[
\frac{dx}{dt} = M(x)
\]

tangent linear model

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

adjoint model

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

gradient of the cost function

\[
\nabla_{[x(t_0), e]} J = -B_0^{-1}(x^*(t_c) - x(t_0)); -K^{-1}(e^*(t) - e(t)); -\sum_{m=0}^{N} M'^T(t_{i-1}, t_i)R^{-1}(y^0(t_m) - H[x(t_m)])
\]

Find minimum of $J(x(t_0), e)$ with $\nabla_{[x(t_0), e]} J$ by use of a minimization routine
CONTRACE
Convective Transport of Trace Gases into the upper Troposphere over Europe: Budget and Impact of Chemistry
Coord.: H. Huntrieser, DLR
flight path Nov. 14, 2001
CONTRACE
Nov. 14, 2001 north (= home) bound

1. guess assimilation result observations flight height [km]
Isopleths of the cost function and transformed cost function and minimisation steps

Minimisation by mere gradients, quasi-Newton method L-BFGS (Large dimensional Broyden Fletcher Goldfarb Shanno), and preconditioned (transformed) L-BFGS application
The Preconditioning Problem

Hessian matrix = (analysis error covariance matrix)^{-1}

\[ \nabla^2 J = B_0^{-1} + H^T R^{-1} H \]

the optimum is (linear case)

\[ x_{opt} - x_b = (B_0^{-1} + H^T R^{-1} H)^{-1} \nabla J \]

practical problem:
with increasing L, B_0 is increasingly ill-conditioned

Example:
2D grid 25 × 25
influence radius L = 6
gives

\[ \text{cond}(B_0) \sim 10^9 \]
Transformed cost function

define

\[ d_i := y_i^0 - HM(t_i, t_0) x(t_0) \]

\[ \delta x(t_0) := x^b - x(t_0) \]

transformation

\[ v := B^{-1/2} \delta x_b \]

transformed cost function

\[ J(v) = \frac{1}{2} v^T v + \frac{1}{2} \sum_{m=0}^{N} d_m^T R^{-1} d_m \]

transformed gradient of the cost function

\[ \nabla_v J(v) = v + B^{1/2} \sum_{m=0}^{N} M^T(t_m, t_0) H^T R^{-1} d_m \]

Pro transformation:

minimisation problem is better conditioned

Contra:

strictly positive definite approximation to \( B \) required

Computation of inverse \( B \), square root \( B \), inverse square root \( B \)

by (Sca)LAPACK eigenpair decomposition, but better choices available.
Kalman filter: basic equations

Forecast equations

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

\[ P^b_i = M(t_i, t_{i-1})P^a_iM^T(t_i, t_{i-1}) + Q \]

Analysis equations

\[ x^a(t_i) = x^b(t_i) + K_i d_i, \quad (1) \]

\[ K_i := P^b_i H^T_i (H_i P^b_i H^T_i + R_i)^{-1} \quad \in \mathcal{R}^{n \times p_i} \quad (2) \]

and

\[ P^a_i = (I - K_i H_i)P^b_i. \quad (3) \]
Reduced rank Kalman filter (basic idea)

Approximate covariance matrices $P^{b,a}$ ($n \times n$) by a product of suitably low ranked matrix $S^{b,a}$ ($n \times p$), and $p \ll n$. Same procedure for the system noise matrix $Q$ with $T$ ($n \times r$).

The forecast step remains unchanged.

The forecast error covariance matrix rests on $2 \times p$ model integrations only!

The enlargement of $p$ by $r$ enforces periodic reductions of columns (with lowest ranked eigenvalues)
Reduced rank Kalman filter (calculus)

\[ \psi := HS \]
\[ K = S^f \psi^T (\psi \psi^T + R)^{-1} \]
\[ x^o = x^f + K(y^o - Hx^f) \]

In practice, all calculations can be performed without actually calculating matrices P!

Positive semidefiniteness is maintained!

\[ S^a S^a T = (I - KH)S^f S^f T \]
\[ = S^f [I - \psi^T (\psi \psi^T + R)^{-1} \psi^T] S^f T \]
\[ S^a T = S^f [I - \psi^T (\psi \psi^T + R)^{-1} \psi^T]^{1/2} \]
Ensemble Kalman Filter

Assimilation system with different background fields and perturbed observations, reflecting estimated error covariance matrices

\[ P^a = \mathcal{E}(x_a x_a^T) \]

After short prediction:

\[ P^b = \mathcal{E}(x_b x_b^T) \]

- Rank of matrices = number of elements orders of magnitude smaller (O(100)) than dimension of phase space
  => analysis increment only within the sub space spanned by the ensemble
Further measures: suppression of spurious covariances

Random vectors do not provide optimal subspaces
Implementing ensemble Kalman filter in practice

Setting

\[ \overline{\mathcal{H}(x_b)} = \frac{1}{K} \sum_{i=1}^{K} \mathcal{H}(x_b^{(i)}) \]

Approximation of \( PH^T \) and \( PHPH^T \) by

\[ \tilde{PH}^T = \frac{1}{K - 1} \sum_{i=1}^{K} (x_b^{(i)} - \overline{x_b})(\mathcal{H}(x_b^{(i)}) - \overline{\mathcal{H}(x_b)})^T \]

and

\[ \tilde{HPH}^T = \frac{1}{K - 1} \sum_{i=1}^{K} (\mathcal{H}(x_b^{(i)}) - \overline{\mathcal{H}(x_b)})(\mathcal{H}(x_b^{(i)}) - \overline{\mathcal{H}(x_b)})^T \]
**Some advanced auxiliary calculations (1)**

**singular value analysis**

unit constraint (scalar product):

\[ (\epsilon, C\epsilon) = 1 \]

maximise

Raleigh quotient:

\[ \max_i \left( \frac{(PM(t_1, t_0)\epsilon_i, EPM(t_1, t_0)\epsilon_i)}{\epsilon_i, C\epsilon_i} \right) \]

\[ J(\epsilon) = (PM(t_1, t_0)\epsilon_i, EPM(t_1, t_0)\epsilon_i) - \lambda(\epsilon^T C\epsilon - 1) \]

\[ \nabla_\epsilon J(\epsilon) = 0 \equiv M^T(t_1, t_0)P^T EPM(t_1, t_0)\epsilon = \lambda C\epsilon \]

\( \epsilon_i \) perturbation vector of potential optimisation parameters:
initial values boundary values, emission rates, deposition velocities

C norm inducing pos. def., sym. operator at initial time \( t_0 \) (Mahalanobis)

M tangent linear model

E norm inducing pos. def., sym. operator at optimisation time \( t_1 \)

(P projection operator, extinguishing areas or species outside focus)

\( \lambda \) Lagrange parameter and generalised eigenvalues
Some advanced auxiliary calculations (2)
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:

Expectation

\[ \mathbb{E}[J_{\text{min}}] = \frac{p}{2} \]

Variance

\[ \mathbb{V}[J_{\text{min}}] = \frac{p}{2} \]
Some advanced auxiliary calculations (3)
a posteriori validation of data assimilation results

The partition of costs in terms of all observations and
the background at the final analysis

\[ J(x^a) = J^o(x^a) + J^b(x^a) \]

How do the partial costs in the
cost function divide?

partial costs of observations

\[ \mathcal{E}[J^o(x^a)] = \frac{1}{2} \text{tr}(I_p - HK) \]

partial costs of background:

\[ \mathcal{E}[J^b(x^a)] = \frac{1}{2} \text{tr}(KH) \]
The partition of costs in terms of individual classes of information sources $z_j$

objective function as a sum of various data / information sources

$$J_j(x) = \sum_{j=1}^{k} J_j(x)$$

with

$$J_j(x^a) = 1/2 (G_j x - z_j)^T S_j^{-1} (G_j x - z_j)$$

$G_j = H_j M(t_j, t_0)$ combined observation - model operator

$S_j$ information error covariance of class $j$ with $m_j$ elements of information

$$\mathcal{E}[J(x^a)] = 1/2 \left( m_j - \text{tr}(G_j^T S_j^{-1} G_j P_a) \right)$$
Summary

- Satellite data still not fully exploited, but within the scope of variational calculus
- 4Dvar is presently regarded as the most rewarding data assimilation method
- Kalman Filtering with suitably reduced complexity (square root and/or ensemble) feasible, but superiority yet to be proved