Atmospheric Simulations and their Optimal Control

H. Elber, E. Friese, L. Nieradzik, K. Kasradze, J. Schwinger
Chemical Data Assimilation group
FZ Jülich, IEK-8
and
Rhenish Institute of Environmental Research at the University of Cologne
Outline

1. Optimisation: a universal scientific objective
2. Problem: How can we control a system with high degrees of freedom?
3. How can we optimally combine models with observations?
   1. a means to upgrade our knowledge basis
   2. test hypotheses

Specifically in atmospheric sciences:
1. for greenhouse gas inversion
2. for air quality and climate monitoring/reanalyses
3. Extension: “detection and attribution” algorithm for climate change
How can “good” simulations be identified?

Ozone: NO\textsubscript{x} - VOC emissions variation ensemble (21.7.1998, 15:00 UTC)

Note: not C-orthogonal, not max. sensitivities aligned

VOCs
How can “good” simulations be identified? Sequential Importance Resampling

How to select weights $w_i$?

Other techniques:
- Ensemble modelling
- (Markov Chain) Monte Carlo methods (MCMC)
- and many more versions ….
Are there more focussed approaches? e.g. quadratic optimisation

Given

- the "model" $y = M(x; a, b) := ax + b$, and
- observations $y_i, \ i = 1, \ldots, K$,

provide a best estimate of model parameters $a, b$ in a sense that

$$\min_{a, b} \left[ \sum_i (y_i - (ax_i + b))^2 \bigg| \forall \ observations(x_i, y_i) \right]$$

Writing the set of observations in column vector notation $\mathbf{y}, \mathbf{x}$, minimise

$$(\mathbf{y} - M(\mathbf{x}; a, b))^T(\mathbf{y} - M(\mathbf{x}; a, b))$$
Isopleths of the cost function and minimisation steps

Minimisation by mere gradients, quasi-Newton method L-BFGS (Large dimensional Broyden Fletcher Goldfarb Shanno),
Extensions to Generality

1. **very high degree of freedom** of highly nonlinear systems $\mathbf{M}$: “curse of dimensionality” $O(10^7\text{-}10^8)$

2. underdetermined system: **too few observations**, i.e. less than degrees of freedom

3. observations **scattered in time** and space,

4. different observation techniques: errors and representativity of observations **diverse**

5. observations often **indirectly** related to parameters of interest: remote sensing data

Methodology applicable to a wide range of problems:
Consider optimisation problem with unique optimum
1. Very high degree of freedom: “curse of dimensionality” $O(10^7 \cdot 10^8)$
(grid points x 10-100 variables)

\[
\frac{\partial U}{\partial t} + \frac{1}{a \cos^2 \theta} \left\{ U \frac{\partial U}{\partial \lambda} + v \cos \theta \frac{\partial U}{\partial \theta} \right\} + \frac{\partial U}{\partial \eta} = P_U + K_U
\]

\[
(-fu) + \frac{1}{a} \left\{ \frac{\partial \phi}{\partial \lambda} + R_{dry} T_v \frac{\partial}{\partial \lambda} (\ln p) \right\} = P_U + K_U
\]

\[
\frac{\partial V}{\partial t} + \frac{1}{a \cos^2 \theta} \left\{ U \frac{\partial V}{\partial \lambda} + V \cos \theta \frac{\partial V}{\partial \theta} + \sin \theta (U^2 + V^2) \right\} + \frac{\partial V}{\partial \eta} = P_V + K_V
\]

\[
+ fU + \frac{\cos \theta}{a} \left\{ \frac{\partial \phi}{\partial \theta} + R_{dry} T_v \frac{\partial}{\partial \theta} (\ln p) \right\} = P_V + K_V
\]

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

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

\[
\frac{\partial p_{surf}}{\partial t} = \int_0^1 \nabla \cdot \left( n_H \frac{\partial p}{\partial \eta} \right) \, d\eta
\]

\[
\frac{\partial p}{\partial \eta} = -\frac{\partial p}{\partial \eta} - \int_0^\eta \nabla \cdot \left( n_H \frac{\partial p}{\partial \eta} \right) \, d\eta
\]

\[
\eta(0, p_{surf}) = 0
\]
1. Very high degree of freedom: “curse of dimensionality” $O(10^7 - 10^8)$
(grid points $\times$ 10-100 variables)

SACADA icosahedral grid
(see presentation K. Kasradze)
2. Underdetermined system: too few observations, i.e. less than degrees of freedom

Type and number of observations used to estimate the atmosphere initial conditions during a typical day. (Buizza, 2000)
Satellite data sources in 2007+, but only a fraction can be used

Number of satellite sources used at ECMWF

Source: ECMWF
Take the model for over-determination: Synergy of information sources

Bayes’ rule:

\[ p(x|y_0) \propto p(y_0|x)p(x) \]

Analysis (=estimation) **BLUE**
Best Linear Unbiased Estimate

\[
\frac{x_a}{\sigma_a^2} = \frac{y_0}{\sigma_o^2} + \frac{x_b}{\sigma_b^2} \\
\frac{1}{\sigma_a^2} = \frac{1}{\sigma_o^2} + \frac{1}{\sigma_b^2} \\
p(x|y_0) =: \mathcal{N}(x|x_a, \sigma_a^2) \\
= \frac{1}{\sqrt{2\pi\sigma_a}} \exp\left(-\frac{(x_a - x)^2}{2\sigma_a^2}\right)
\]

\[
p(y_0|x) =: \mathcal{N}(y_0|x, \sigma_o^2) = \frac{1}{\sqrt{2\pi\sigma_o}} \exp\left(-\frac{(y_o - x)^2}{2\sigma_o^2}\right)
\]

**Example: inconsistent data**

\[
p(x) =: \mathcal{N}(x|x_b, \sigma_b^2) := \frac{1}{\sqrt{2\pi\sigma_b}} \exp\left(-\frac{(x - x_b)^2}{2\sigma_b^2}\right)
\]

**a priori (=prediction or climatology)**

**observation**
3. Observations scattered in time (and space)

MOZAIC - JAGOS

Aeronet LIDAR

MetOp-1 IASI, GOME-2

AIREP, AMDAR, ACAR

00 UTC
Types of assimilation algorithms: “smoother” and filter

$y = ax + b$
Tendency Equations

direct chemistry transport equation

\[
\frac{\partial c_i}{\partial t} + \nabla \cdot (\mathbf{v} c_i) - \nabla \cdot (\rho \mathbf{K} \nabla \overline{c_i}) - \sum_{r=1}^{R} \left( k(r) \left( s_i(r_+) - s_i(r_-) \right) \Pi_{j=1}^{U} c_j^{s_j(r_-)} \right) = E_i + D_i
\]

- \( c_i \) concentration of species \( i \)
- \( \mathbf{v} \) wind velocity
- \( k(r) \) reaction rate of reaction \( r \)
- \( U \) number of species in the mechanism
- \( E_i \) emission rate of species \( i \) (source)

adjoint chemistry transport equation

\[
- \frac{\partial \delta c_i^*}{\partial t} - \mathbf{v} \nabla \delta c_i^* - \frac{1}{\rho} \nabla \cdot (\rho \mathbf{K} \nabla \delta c_i^*) + \sum_{r=1}^{R} \left( k(r) \frac{s_i(r_-)}{s_i} \Pi_{j=1}^{U} c_j^{s_j(r_-)} \sum_{n=1}^{U} \left( s_n(r_+) - s_n(r_-) \right) \delta c_n^* \right) = 0
\]

- \( \delta c_i^* \) adjoint of concentration of species \( i \)
- \( \delta \) stoichiometric coefficient
- \( \mathbf{K} \) diffusion coefficient
- \( R \) number of reactions in the mechanism
- \( D_i \) deposition rate of species \( i \) (sink)
Adjoint integration “backward in time”

\[ \frac{dx}{dt} = \mathcal{M}(x) + e(t), \quad \frac{d\delta x}{dt} = \mathcal{M}'(\delta x) + \delta e(t) \] (1)

\[ \delta x(t_n) = M(t_n, t_0)\delta x(t_0) = \prod_{i=n}^{1} M(t_i, t_{i-1})\delta x(t_0) \] (2)

\[ -\frac{d\delta x^*(t)}{dt} - \mathcal{M}^T(\delta x^*(t)) = R^{-1}(y_0(t) - H[x(t)]) \] (3)

Gradient of the cost function

\[ \nabla_{[x(t_0), e]} J = -B_0^{-1}(x^b(t_0) - x(t_0)) - K^{-1}(e^b(t) - e(t)) - \sum_{m=0}^{N} \prod_{i=1}^{m} M^T(t_{i-1}, t_i)R^{-1}(y_0(t_m) - H[x(t_m)]) \]

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 and Impact of Chemistry
Coord.: H. Huntrieser, DLR flight path Nov. 14, 2001
day = 318 ; hour = 16.0
CONTRACE
Nov. 14, 2001 north (= home) bound

$\text{O}_3$

$\text{H}_2\text{O}_2$

1. guess

assimilation result

observations

flight height [km]
4. Observation errors are varying and representativity of observations are diverse

Example: Analysis increments using the novel background error covariance matrix formulation

Background (forecast) error correlation can bridge gaps according to subgrid scale land use information:

e.g. from forest to forest
5. Observations often indirectly related to parameters of interest: Remote sensing

level 0: detector tensions: digital data

level 1: calculate spectra

Retrieval:
Solve the model equivalent:
radiative transfer equation $H(x)$

Calculate difference ($y - H(x)$), and assimilate

Example: SCIAMACHY

level 2: calculate located geo data (say: $y=[NO_2]$)

level 3: “analysis fields” (say NO$_2$)
Main observational data (spaceborne)

**ENVISAT** (2002-2012) MIPAS, SCIAMACHY, GOMOS; temperature, ozone, water vapour and other atmospheric constituents

(ii) AATSR, MERIS aerosol, MERIS sea colour, ASAR land and ocean images; RA-2 land, ice and ocean monitoring; MWR water vapour column and land surface parameters; DORIS cryosphere and land surface parameters

**AQUA** (2002) AMSR/E: clouds, radiation and precipitation; MODIS: clouds, radiation, aerosol and vegetation parameters; AMSU, AIRS, HSB temperature and humidity; CERES radiation

**AQUA** (2004) MLS trace gases of the upper troposphere to upper stratosphere; + water HIRDLS, temperature and trace gases in the upper troposphere, stratosphere and mesosphere; TES trop. ozone and some photochemical precursors; OMI total column ozone and NO2 and UV-B radiation.

**TERRA** (1999). ASTER, land surface, water and ice; CERES radiation, MISR radiation and biosphere parameters; MODIS biological and physical processes on land and the ocean; MOPITT CO and CH4 in the troposphere.

**MetOp-1**

IASI ozone, NO2, GOME-2 ozone; SO2 NO2 formaldehyde

**A train**

CloudSat, Aqua, PARASOL, CALIPSO
Satellite information:
ESA UV-VIS satellite footprints Ruhr area comparison

Minimal areas:

- **GOME 1** (special mode): 320 x 40 km²
- **SCIAMACHY**: 80 x 40 "
- **GOME 2**: 80 x 40 "
- **OMI**: 24 x 13 "

Ruhr area domain: 90 x 80 km²
1 km resolution
(~12 000 000 inhabitants)
We started with: \((y - M(x; a, b))^T(y - M(x; a, b))\)

**Generalized cost function to be minimized**

Minimize \(J\) by variation of \(x(t_0)\):

\[
J(x(t_0)) = \frac{1}{2} (x^b(t_0) - x(t_0))^T B_0^{-1} (x^b(t_0) - x(t_0)) + \\
\frac{1}{2} \int_{t_0}^{t_N} (y^0(t) - HM[x(t)])^T R^{-1} (y^0(t) - HM[x(t)]) dt
\]

- **x^b(t_0)**: background state at \(t = 0\)
- **x(t)**: model state at time \(t\)
- **e_b(t_0)**: background emission rate at \(t = 0\)
- **e(t)**: emission rate field at time \(t\)
- **K**: emission rate error covariance matrix
- **H[ ]**: forward interpolator
- **y^0(t)**: observation at time \(t\)
- **B_0**: background error covariance matrix

1. Model constraint and time propagator (resolvent)
2. Background term for artificial over-determination
3. Forecast errors
4. Observation errors and errors of representativity
5. Observation operator
Question: Which parameter to be optimized?

Hypothesis: initial state and emission rates are least known

- only emission rate opt.
- only initial value opt.
- joint opt.
- emission biased model state
- true state
- observations
- only emission rate opt.
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)
Data Assimilation in general

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)
Objective of atmospheric data assimilation

"is to produce a regular, physically consistent four dimensional representation of the state of the system 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)
Information sources and theories

Information set

*declarative information:*
- observations/retrievals
- forecasts
- “Climate” statistics,
- error statistics

*procedural Information*
 differential equations
 models

4D-consistent process description

- control theory
- statist. filter theory
- classical numerics
- optimisation algorithm.

Interpolation
in space and time

Filter
error affected
data

Completion
non-observed
parameters

Advection

micro-physics

aqueous phase

chem.

gas-phase

aerosols

radiation

Emissionen
Deposition
How can I select “good” simulations?

Ozone: NO$_x$-VOC emissions variation ensemble (21.7.1998, 15:00 UTC)

Note: not C-orthogonal, not max. sensitivities aligned

NO$_x$ 1.5 $\times$

1 $\times$

0.6 $\times$

VOCs
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)

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.
Emission source estimates by inverse modelling
Optimised emission factors for Nest 3

height layer ~32~70m

NO\textsubscript{2}, (xylene (bottom), CO (top))

SO\textsubscript{2}.

without assimilation

with assimilation
2. Analyses, Example (ii):
Zepter 2: 4D-var assimilation of particle number densities

Flight 14 assimilation of PND (0.005-3.0 μm) 02.11.2008 (11-15 UTC)

(from Lars Nieradzik, PhD thesis 2011)
water solubles include Friese and Ebel, 2010:

$$\begin{align*}
H^+ & \quad NH_4^+ & \quad Na^+ & \quad SO_4^{2-} & \quad NO_3^- & \quad Cl^- & \quad H_2O
\end{align*}$$
2. Analyses, Example (ii cntd.):
Flight 14 assimilation of PND (0.005-3.0 µm), Nov. 2
Analysis increment (Analysis – Background)

Aitken mode $\Delta$PND
4 hour assimilation
5 km grid

Acc. mode $\Delta$PND
(~ 450m a.g.)

from Lars Nieradzik, PhD thesis 2011
Thank you for your attention!
Minimize $J$ by variation of $x(t_0)$:

$$J(x(t_0)) = \frac{1}{2}(x^b(t_0) - x(t_0))^T B_0^{-1}(x^b(t_0) - x(t_0)) + \frac{1}{2} \int_{t_0}^{t_N} (y^0(t) - HM[x(t)])^T R^{-1}(y^0(t) - HM[x(t)]) dt$$

- $x^b(t_0)$ background state at $t = 0$
- $x(t)$ model state at time $t$
- $e_b(t_0)$ background emission rate at $t = 0$
- $e(t)$ emission rate field at time $t$
- $K$ emission rate error covariance matrix
- $H[ ]$ forward interpolator
- $y^0(t)$ observation at time $t$
- $B_0$ background error covariance matrix

2 Questions:

1. **Spatial optimisation only (no time evolution involved):** How does a closed formula for the optimum read?

2. **Temporal optimisation:** How can we integrate “backward in time” with an adjoint model $M^T$ for an optimum at initial time?