Lecture

Objective Analysis

File: objektive_analyse.tex
Version: 0.0 (English)
Date: WS 2006/07 (September 2006)

Contents

CHAPTER 1

Introduction

/

Textbook References

- Gandin, L., Objective Analysis of meteorological fields (Leningrad: Gidromet), 1963, English translation (Jerusalem: Israel program for scientific translation), 1965.

- Daley, R., *Atmospheric Data Analysis*, Cambridge University Press pp. 457, 1991.

- Bennet, A., *Inverse Methods in Physical Oceanography*, Cambridge University Press, pp. 346, 1992.

- Bennet, A., *Inverse Modeling of the Ocean and Atmosphere*, Cambridge University Press, pp. 234, 2002.

- Bengtson, L., M. Ghil, E. Kallén, *Dynamic Meteorology: Data Assimilation Methods*, Springer, 1981.

- Kalnay, E., *Atmospheric Modeling, Data Assimilation and Predictability*, Cambridge University Press, pp. 341, 2003.

- Zou, X., F. van den Berghe, M. Pondeca, Y.H. Kuo, Introduction to Adjoint Techniques and the MM5 Adjoint Modelling System, *NCAR Technical Note TN–435–STR*, pp. 110, 1997.

Objectives of Data Assimilation and Inverse Methods

"The objective of atmospheric data assimilation 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. The regular, physically consistent aspect of the procedure comes from the use of models, and thus data assimilation is a discipline which naturally integrates theory (via models) with sampled reality (via instruments).

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 (completeness).

(Daley, 1997)

The term *inversion* or *inverse modelling* is used in a somewhat broarder sense. *The Objectives of Inversion*:

1. Estimates of dynamic fields by sparse observations, physical laws, and statistical knowledge, to analyse consistent data sets and optimal initial and boundary values (Data assimilation),
2. parameter estimates and parameter optimisation, of models and algorithms,
3. sensitivity studies and solutions of ill–posed numerical problems,
4. optimisation of observation systems,
5. test of scientific hypotheses.

(Bennet, 1992)

Given K observations $\mathbf{y}_o = ((y_o)_i, \ldots, (y_o)_K)^T$ and a regular grid with N gridpoints $\mathbf{x} = (x_1, \ldots, x_N)^T$, said the analysis or model grid. Usually, $K << N$. We seek to identify or reconstruct or "analyse" the field of the observed quantity on the analysis grid by observations, that is we try to find an estimate of the most probable, or the least flawed state in terms of all available information.

Early Data Assimilation Algorithms

Local Interpolation

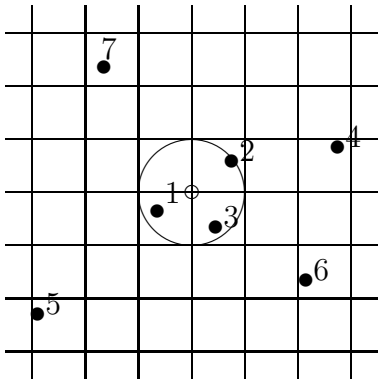
Given an x-y–plane a 2-dimensional polynomial may be adjusted to approximate the parameter field, say geopotential height, the coefficients of which are to be optimized to fit the observations $\mathbf{y}_o$. A quadratic algebraic polynomial reads

$$z(x, y) = a_{00} + a_{10}x + a_{01}y + a_{20}x^2 + a_{11}xy + a_{02}y^2$$

It may be taken as valid local approximation for an environment around a gridpoint. With coefficients $a_{vw}, v, w \in \{0, 1, 2\}$ the optimisation problem is then:
find a_{vw} such that

$$\min_{a_{vw}} \sum_{k=1}^{K_i} w_k((y_o)_k - z(x_k, y_k))^2 \quad , \tag{1.1}$$

where K_i is the number of adjacent observations, w_k the empirical weighting coefficient, dependent on the distance between observation and model grid point, and the quality of the observation.



Schematic of observation locations (solid dots), the enclosed positions 1,2, and 3 of which are used to estimate the grid point value (open circle)

Nudging

Nudging is a data assimilation method affecting the model equations, by adding a "nudging term", which denotes a weighted observation-minus-model discrepancy (Newtonian cooling formulation) at the observation location i and c the nudging coefficient:

$$\frac{\partial x_i}{\partial t} = M(\mathbf{x}) + c((y_o)_i - x_i).$$

For example, the zonal momentum equation with local nudging term reads

$$\frac{\partial u_i}{\partial t} = -\mathbf{v}_i \cdot \nabla u_i + fv - \frac{\partial \phi}{\partial x} + \frac{((u_o)_i - u_i)}{\tau_{\mathbf{v}}}$$

with $\tau := 1/c$ the relaxation constant with unit 1/time.

The Statistical Viewpoint

Even most accurate observation techniques and most complete numerical model implementations inevitably have some errors, the properties of which are only known in statistical terms. Therefore formally, any observation or model forecast can then be considered as a *realisation* of a random experiment. Data assimilation then aims to **exploit all data by weighting the contribution of each information source according to their prior error estimate**.

Let X be a random variable. The value, which is attributed to a random variable by some experiment, say an observation or a numerical integration, is designated a realisation x of X. Two estimators are frequently applied, to identify the finally sought–after estimate x_a of the always unknown true value x_t. One of the following optimality criteria is selected:

1. Minimal variance:

$$x_{a_m} = \{\tilde{x} : \min_{\tilde{\epsilon}} \mathcal{E}(\tilde{\epsilon}^2) = \mathcal{E}((\tilde{x} - x_t)^2) \,\forall\, x \in X\}, \tag{1.2}$$

2. Maximum likelyhood:

$$x_{a_l} = \{\tilde{x} : P(\tilde{x}) \geq P(x) \,\forall\, x \in X\}, \tag{1.3}$$

where $P(x) := P(x = X)$ is the probability, that the random variable X has the realisation x.

The observation

$$y_o = x_t + \epsilon_o, \tag{1.4}$$

with x_t true value and ϵ_o observational error and additionally the error of representativity for an amalysis grid resolution, is a realisation of the random variable X, as the individually unknown observation error ϵ_o is a random variable.

Any other datum, be it a second observation or a corresponding forecast or a climatological value of the measurement given above, i.e. *background or first guess value or a priori knowledge*, can also be considered as a realization of a random experiment.

$$x_b = x_t + \epsilon_b \tag{1.5}$$

Let us be given an observation y_o and a forecast x_b for the same location and time. We seek for an x_a based on y_o and x_b and their respective error variances ϵ_o and ϵ_b. with the smallest analysis error.

To proceed, we invoke the following assumptions

- the expectation (mean, average, first (statistical) moment)

$$\mathcal{E}(\epsilon_{o/b}) := \int_{-\infty}^{\infty} \epsilon_{o/b} P(\epsilon_{o/b}) d\epsilon_{o/b} = 0 \tag{1.6}$$

which demands that there is no mean or systematic error (= unbiased)

- Variance (second central moment) of error as an inverse measure of accuracy

$$\mathcal{E}(x_b^2) := \int_{-\infty}^{\infty} (x_b - \mathcal{E}(x_b))^2 P(x_b) dx_b = \int_{-\infty}^{\infty} \epsilon_b^2 \, P(\epsilon_b) \, d\epsilon_b = \sigma_b^2 =: V_b \quad (1.7)$$

and

$$\mathcal{E}({y_o}^2) := \int_{-\infty}^{\infty} (y_o - \mathcal{E}(y_o))^2 \, P(y_o) \, dy_o = \int_{-\infty}^{\infty} \epsilon_o^2 \, P(\epsilon_o) \, d\epsilon_o = \sigma_o^2 =: V_o \quad (1.8)$$

- the data y_o and x_b are independent, that is, there is no covariance between observations and the a priori information:

$$\mathcal{E}(\epsilon_o \epsilon_b) = 0 \quad (1.9)$$

We are seeking the estimator for x_a as a weighted linear combination of two competing information elements y_o and x_b, with as yet unknown weights w_o and w_b such that

$$x_a = w_o y_o + w_b x_b \quad (1.10)$$

with minimal variance V_a.

Given the assertion x_a unbiased (as y_o and x_b are unbiased), we find

$$\mathcal{E}(x_a) = \mathcal{E}(x) = \mathcal{E}(w_o y_o + w_b x_b) = (w_o + w_b)\mathcal{E}(x_t). \quad (1.11)$$

It follows that

$$w_o + w_b = 1. \quad (1.12)$$

The optimality criterion selected is the minimal error variance of the estimator X_a. Our tunable parameters are the weights w_o and w_b.

$$\begin{aligned} V_a = \sigma_a^2 &= \mathcal{E}((x_a - \mathcal{E}(x_a))^2) \\ &= \mathcal{E}((w_o y_o + w_b x_b - x_t)^2) \\ &= \mathcal{E}((w_o(y_o - x_t) + w_b(x_b - x_t))^2) \\ &= w_o^2 \mathcal{E}((y_o - x_t)^2) + w_b^2 \mathcal{E}((x_b - x_t)^2) \\ &= w_o^2 \mathcal{E}(\epsilon_o^2) + w_b^2 \mathcal{E}(\epsilon_b^2) \\ &= w_o^2 \sigma_o^2 + w_b^2 \sigma_b^2 \end{aligned} \quad (1.13)$$

where (1.9) has been used.

Which w_o and w_b provide for the analysis x_a with the smallest error variance σ_a^2?

Define a cost function $J(w_o, w_b)$ to be minimised, subject to the constraint (1.12) with the Lagrange multiplier λ

$$J(w_o, w_b) = \sigma_a^2 + \lambda(1 - w_o - w_b) = w_o^2 \sigma_o^2 + w_b^2 \sigma_b^2 + \lambda(1 - w_o - w_b), \quad (1.14)$$

given the neccessary conditions for the observational (o) and background (b) case, independently

$$\frac{\partial J}{\partial w_{o/b}} = 2w_{o/b}\sigma_{o/b}^2 - \lambda = 0, \tag{1.15}$$

from which follows

$$w_{o/b} = \frac{\lambda}{2\sigma_{o/b}^2}. \tag{1.16}$$

Again with (1.12), λ can be expressed as

$$\lambda = \frac{2\sigma_o^2\sigma_b^2}{(\sigma_o^2 + \sigma_b^2)} = \frac{2}{(\sigma_o^{-2} + \sigma_b^{-2})}. \tag{1.17}$$

From (1.13) we finally obtain the optimal weights.

$$\boxed{w_o = \frac{\sigma_o^{-2}}{\sigma_o^{-2} + \sigma_b^{-2}}, \qquad w_b = \frac{\sigma_b^{-2}}{\sigma_o^{-2} + \sigma_b^{-2}}.} \tag{1.18}$$

Hence, the minimal variance estimator reads

$$\boxed{x_a = \frac{\sigma_o^{-2}}{\sigma_o^{-2} + \sigma_b^{-2}} y_o + \frac{\sigma_b^{-2}}{\sigma_o^{-2} + \sigma_b^{-2}} x_b} \tag{1.19}$$

and is called a **BLUE**, Best Linear Unbiased Estimator.

The minimal variance of the analysis itself directly follows from (1.13)

$$\sigma_a^2 = \frac{1}{\sigma_o^{-2} + \sigma_b^{-2}} \tag{1.20}$$

and therefore

$$\frac{1}{\sigma_a^2} = \frac{1}{\sigma_o^2} + \frac{1}{\sigma_b^2} \qquad \frac{x_a}{\sigma_a^2} = \frac{y_o}{\sigma_o^2} + \frac{x_b}{\sigma_b^2}. \tag{1.21}$$

It can be stated that

$$\boxed{V_a \leq \min(V_o, V_b)} \tag{1.22}$$

Variational Approach

Result obtained by variation of a **cost function (distance function, objective function, test function)** $J(x)$

$$J(x) = (y_o - x)^2\sigma_o^{-2} + (x_b - x)^2\sigma_b^{-2} \tag{1.23}$$

$$0 = \frac{dJ(x)}{dx}|_{x=x_a} = 2(y_o - x)\sigma_o^{-2} + 2(x_b - x)\sigma_b^{-2} \tag{1.24}$$

it follows the same result as in (1.19)

$$x_a = \frac{\sigma_o^{-2}}{\sigma_o^{-2} + \sigma_b^{-2}} y_o + \frac{\sigma_b^{-2}}{\sigma_o^{-2} + \sigma_b^{-2}} x_b \tag{1.25}$$

Formulation of the inversion problem

Given a vector space $X \in \mathbf{R}^N$ suitable to describe system states $\mathbf{x} \in X$ of a fluid with sufficient accuracy. Further, a vector space Y for observations $\mathbf{y}_o \in Yc\mathbf{R}^M$. Finally, an operator H (forward interpolation, smoothing, filter, or integration operator)

$$H : \mathbf{R}^N -> \mathbf{R}^M$$

$$x -> y = H(x)$$

As we are interested $\mathbf{x}$, is there an operator H^*, with $H^*(\mathbf{y}_o) = \mathbf{x}$

(Associated phrases:
"Given the answer, what was the question?" (Foffonov)
"Can you hear the shape of the drum?" (Wunsch))

We consider the probability of a fluid to have the state $\mathbf{x}$, that is $p(x)dx := P(\mathbf{x}_0 \leq \mathbf{x} < \mathbf{x}_0 + d\mathbf{x})$ The Bayesian rule here reads (see Appendix):

$$p(\mathbf{x}|\mathbf{y}_0) = \frac{p(\mathbf{y}_0|\mathbf{x})p(\mathbf{x})}{\int_X p(\mathbf{y}_0|\mathbf{x})p(\mathbf{x})d\mathbf{x}} \tag{1.26}$$

Here, the denominator is a statistical convolution integral, as with varying fluid state $\mathbf{x}$ the conditional probability of its observations $\mathbf{y}_0$ changes.

Interpretation:

- $p(\mathbf{y}_0)$ denotes the probability that for observations as realisations of tha random variables Y_i holds $(Y_1, \ldots, Y_M)^T = \mathbf{y}_0$
- $p(\mathbf{x})$ contains our statistical knowledge on the probability of the system to take state $\mathbf{x}$ prior to the observations $\mathbf{y}_0$ (first guess, background information, a priori knowledge), usually taken from climatologies or a preceeding forecast,
- the *likelyhood function* $p(\mathbf{y}_0|\mathbf{x})$ denotes the probability to observe $\mathbf{y}_0$, if the system state is $\mathbf{x}$
- we seek the probability of the most probable sytem state $\mathbf{x}$, that is the a posteriori distribution $p(\mathbf{x}|\mathbf{y}_0)$, given observations $\mathbf{y}_0$.

Example:
Observation and background error follow Gaussian error distribution:
Let us assume the probability $p(x)$ of the true state x deviates from a forecasted state or climatology x_b (a priori knowledge, first guess, background) is described by the normal distribution (Gaussian)

$$p(x) =: \mathcal{N}(x|x_b, \sigma_b^2) := \frac{1}{\sqrt{2\pi}\sigma_b} \exp\left(-\frac{(x - x_b)^2}{2\sigma_b^2}\right) \tag{1.27}$$

(This also includes the assumption of unbiasedness (1.6).) Likewise, for the observations, we describe

$$p(y_o|x) =: \mathcal{N}(y_o|x, \sigma_o^2) := \frac{1}{\sqrt{2\pi}\sigma_o} \exp\left(-\frac{(y_o - x)^2}{2\sigma_o^2}\right) \tag{1.28}$$

We finally need the denominator of the Bayesian rule

$$p(y_o) = \int_X p(y_0|x)p(x)dx \tag{1.29}$$

With the convolution theorem for the normal distribution

$$\mathcal{N}(\mu_1, \sigma_1^2) * \mathcal{N}(\mu_2, \sigma_2^2) = \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2) \tag{1.30}$$

(see for example Pfanzagl) we obtain

$$p(y_o) =: \mathcal{N}(y_o|x, \sigma_o^2 + \sigma_b^2) := \frac{1}{\sqrt{2\pi(\sigma_o^2 + \sigma_b^2)}} \exp\left(-\frac{(y_o - x_b)^2}{2(\sigma_o^2 + \sigma_b^2)}\right) \tag{1.31}$$

If we set $\frac{x_a}{\sigma_a^2} = \frac{y_o}{\sigma_o^2} + \frac{x_b}{\sigma_b^2}$ and again with $\frac{1}{\sigma_a^2} = \frac{1}{\sigma_o^2} + \frac{1}{\sigma_b^2}$, we arrive at

$$p(x|y_o) =: \mathcal{N}(x|x_a, \sigma_a^2) \tag{1.32}$$

The most probable state is the maximum of $p(x|y_o)$, which as it is a Gaussian distribution, is also the state with minimal variance. For convenience, we apply the negative logarithm, to obtain a simpler expression, the minimum of which is identical to the maximum of $p(x|y_o)$:

$$\begin{aligned} -\ln(p(\mathbf{x}|\mathbf{y}_0)) &= -\ln(p(\mathbf{y}_0|\mathbf{x})) - \ln(p(\mathbf{x})) + const \\ &= \frac{(x - x_b)^2}{2\sigma_b^2} + \frac{(x - y_o)^2}{2\sigma_o^2} + const \\ &= \frac{(x - x_a)^2}{2\sigma_a^2} + const \end{aligned} \tag{1.33}$$

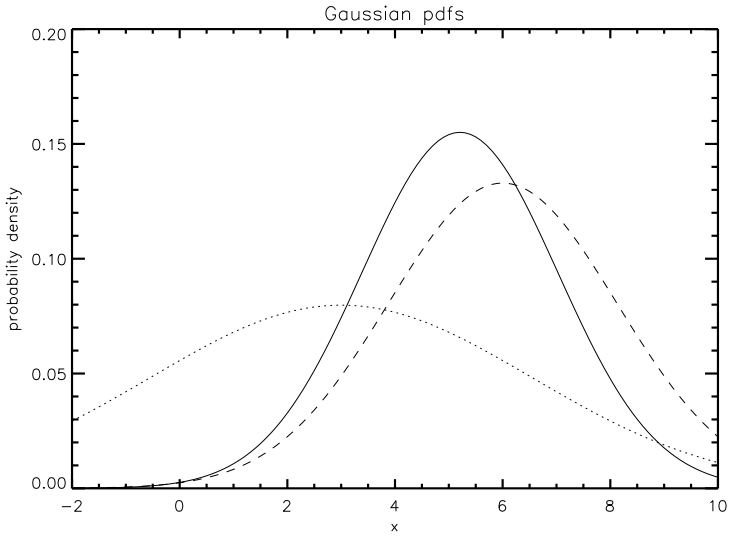


Figure 1. Gaussian PDFs for, say, the a priori distribution (1.27) with $x_b = 3, \quad \sigma_b = 5$ (dotted), the observation (1.28) with $y_o = 6, \quad \sigma_b = 3$ (dashed), and the resulting analysis (1.32) with x_a and σ_a following (1.21) (bold).

Appendix: basic statements from statistics

Given

1. a probablility space X (Grundraum) of all possible events of a random experiment
   Examples:
   - die experiment $X = \{1, 2, 3, 4, 5, 6\}$
   - all possible states of a fluid in discrete approximation
2. $\mathcal{A}$ a system of subsets from X
   Examples:
   - $\{x \in X = \mathbf{N} : x \quad \text{even}\}$
   - $\{\mathbf{x} \in X = \text{NAOflowpatterns} : x \quad \text{withNAOindex} > 0\}$
3. Function P: probability density function (PDF), which associates a probability $P(A)$ to each $A \in \mathcal{A}$

Probability theory then rests on the following axioms due to Kolmogorov

1. $P(A) \geq 0 \quad \forall A \in \mathcal{A}$
   (probability is nonnegative)
2. $P(X) = 1$
   (it is certain that the system has one of all possible states)
3. $A_n \in \mathcal{A}$ are mutually disjoint, then $P(\cup_n A_n) = \sum_{n=1}^{\infty} P(A_n)$

A statistical moment of n^{th} order is defined as

$$\int x^n P(x) dx$$

by integrating over the entire probability space. Of special interest are $n = 0$, then $\int P(x)dx = 1$, the neccesary condition for $P(x)$ being a probability density function; $n = 1$ results in the average value of x, while $n = 2$ results in the variance.

Bayes' Rule

Inverse modelling aims to identify the most probable representation of a system (mostly in terms of gridded or spectral data), given some information (that is, observations, background data, and the coded dynamical equations). The

pertinent approach to estimate a latent property X by means of a manifest property Y is Bayes' theorem (Bayes, 1763, 1764). Let $\bar{A}$ be the complement of A.

With $P(A)$ being the probability of the parameter A having a value such that $a \leq A < a + da$, and $P(A|B)$ the probability of $a \leq A < a + da$, given $b \leq B < b + db$, the probability of event A to occur, given event B, is

Let $P(A|B)$ denote the probability that A is true, given (conditional to) that B is true. Then the *Bayesian Rule* holds

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\bar{A})P(\bar{A})} \tag{1.34}$$

Proof:

$$P(A \cap B) = P(B|A)P(A) = P(A|B)P(B)$$

as, with the same reasoning

$$P(A \cap B) = P(A|B)P(B).$$

Further, from $P(B) = P(B|A)P(A) + P(B|\bar{A})P(\bar{A})$, the Bayesian rule follows.

The Bayesian Rule is helpful to solve the following class of problems:
Let there be an event B, the occurence of which can be observed. Let there be another property A (system state to be estimated), the occurence of which is latent, but is correlated to event B.

Example
with discrete probability density distribution:
Box 1 contain red and white ballots with ratio 1:2
Box 2 contain red and white ballots with ratio 1:1
A red ballot is drawn from one of the unidentified boxes.
What is the probability that the red ballot has been taken from box 1?

Solution: Estimate $P(box = 1|ballot = red)$
We know:

$$P(ballot = red|box = 1) = 1/3, \qquad P(ballot = white|box = 1) = 2/3$$

$$P(ballot = red|box = 2) = 1/2, \qquad P(ballot = white|box = 2) = 1/2$$

$$P(box = 1) = 1/2 \qquad P(box = 2) = 1/2$$

Bayesian rule:

$$P(box = 1|ballot = red) = \frac{P(ballot = red|box = 1)P(box = 1)}{P(ballot = red|box = 1)P(box = 1) + P(ballot = red|box = 2)P(box = 2)}$$

$$= \frac{1/3 * 1/2}{1/3 * 1/2 + 1/2 * 1/2} = 2/5$$

CHAPTER 2

The Spatial Analysis Equations

We strive for estimating a latent, not apparent parameter set $\mathbf{x}$, usually a geophysical field, which is to be analysed on a regular model grid, or an equivalent representation like Fourier coefficients.

We dispose of
1. indirect information on the state and processes in terms of manifest, though insufficient or inappropriate data $\mathbf{y}$. Further,
2. we dispose of a deterministic model or observation operator or forward interpolation operator $\mathcal{H}$, mapping between the state vextor $\mathbf{x}$ and the dataset $\mathbf{y}$, such that

$$\mathbf{y} = \mathcal{H}(\mathbf{x}), \tag{2.1}$$

which, for simplicity, is assumed to be linear.

Bayes' rule gives access to the probability of state $\mathbf{x}$, given $\mathbf{y}$ and $\mathcal{H}$.

$$\text{prob}(\mathbf{x}|\mathbf{y},\mathcal{H}) = \frac{\text{prob}(\mathbf{y}|\mathbf{x},\mathcal{H})\text{prob}(\mathbf{x}|\mathcal{H})}{\text{prob}(\mathbf{y}|\mathcal{H})} \tag{2.2}$$

Let us now assume that we dispose of some knowledge of the atmospheric state to be analysed, **prior** to any manifest information, which is termed background or first guess information $\mathbf{x}_b$. After making use of climatological knowledge in former times, $\mathbf{x}_b$ is now usually provided by a short range forecast. Our sought after result is an optimal estimate $\mathbf{x}_a$, the *analysis* of $\mathbf{x}_a$.

To make use of Bayes' rule we must now make assumptions on the related probabilities p. For the probability of some atmospheric state $\mathbf{x}$ to be true $p(\mathbf{x} = \mathbf{x}_t)$ prior to further information by measurements, we set

$$p(\mathbf{x} = \mathbf{x}_t) = p_b(\mathbf{x} - \mathbf{x}_b = \mathbf{x}_t) =: p_b(\mathbf{x} - \mathbf{x}_b) \tag{2.3}$$

expressing the random component of our knowledge. Likewise, any observations are affected by a random errors as well, giving

$$p(\mathbf{y}_o = \mathbf{y}_t) =: p_o(\mathbf{y}_o - \mathbf{y}_t). \tag{2.4}$$

For the conditional probability of $p(\mathbf{y} = \mathbf{y}_t)$ given $\mathbf{x} = \mathbf{x}_b$ we set

$$p(\mathbf{y} = \mathbf{y}_t|\mathbf{x} = \mathbf{x}_t) =: p_o(\mathbf{y}_o - \mathcal{H}(\mathbf{x})). \tag{2.5}$$

For explicit calculations of $p(\mathbf{x}_a|\mathbf{y}_o)$, we need to know p_b and p_o. Typically, we resort to Gaussian error characteristics for both pdf, giving

$$p(\mathbf{x} - \mathbf{x}_b) \propto \exp[-1/2(\mathbf{x} - \mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b)] \tag{2.6}$$

$$p(\mathbf{y} - \mathbf{y}_o) \propto \exp[-1/2(\mathbf{H}\mathbf{x} - \mathbf{y}_o)^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{x} - \mathbf{y}_o)] \tag{2.7}$$

where

$$\mathbf{B} := \mathcal{E}\left((\mathbf{x} - \mathbf{x}_b)(\mathbf{x} - \mathbf{x}_b)^T\right) \tag{2.8}$$

$$\mathbf{R} := \mathcal{E}\left((\mathbf{y}_t - \mathbf{y})(\mathbf{y}_t - \mathbf{y})^T\right)$$

are the background error and observation error covariance matrices, respectively. $\mathbf{H}$ is the linearized observation operator $\mathcal{H}$. We then find

$$p(\mathbf{x}|\mathbf{y}_o) \propto \exp[-1/2(\mathbf{x} - \mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{H}\mathbf{x} - \mathbf{x}_b) - 1/2(\mathbf{H}\mathbf{x} - \mathbf{y}_o)^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{x} - \mathbf{y}_o)] \tag{2.9}$$

The searched for maximum of the PDF (probability density function) is at the same location of the minimum of $J := -\ln(p(\mathbf{x}|\mathbf{y}_o))$

$$J(\mathbf{x}) = \frac{1}{2}[\mathbf{x}^b - \mathbf{x}]^T\mathbf{B}_0^{-1}[\mathbf{x}^b - \mathbf{x}] + \frac{1}{2}\left\{\mathbf{y}^0 - H[\mathbf{x}(t)]\right\}^T\mathbf{R}^{-1}\left\{\mathbf{y}^0 - H[\mathbf{x}]\right\}. \tag{2.10}$$

The gradient then reads

$$\nabla J(\mathbf{x}) = \mathbf{B}_0^{-1}[\mathbf{x}_b - \mathbf{x}] + \mathbf{H}^T\mathbf{R}^{-1}\left\{\mathbf{y}^o - \mathbf{H}[\mathbf{x} + (\mathbf{x}_b - \mathbf{x}_b)]\right\} \tag{2.11}$$

where a trivial expansion is introduced for later manipulation. At the minimum with the neccessary condition $\nabla J(\mathbf{x}) = \mathbf{0}$, where we define the analysis $\mathbf{x} =: \mathbf{x}_a$, we obtain

$$\mathbf{x}_a - \mathbf{x}_b = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}\left\{\mathbf{y}^0 - H[\mathbf{x}_b]\right\} \tag{2.12}$$

$$= \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}\left\{\mathbf{y}^0 - H[\mathbf{x}_b]\right\} \tag{2.13}$$

with the latter result obtained by the following proof of a special case of the Sherman–Morrison–Woodbury formula:

$$\begin{pmatrix} \mathbf{R}^b & \mathbf{H} \\ \mathbf{H}^T & -\mathbf{B}^{-1} \end{pmatrix}\begin{pmatrix} \mathbf{w} \\ \delta\mathbf{x}_a \end{pmatrix} = \begin{pmatrix} \delta\mathbf{y}_o \\ \mathbf{0} \end{pmatrix} \tag{2.14}$$

First, eliminate $\mathbf{w}$ to find $\delta\mathbf{x}_a = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}\delta\mathbf{y}_o$. Upon elimination of $\delta\mathbf{x}_a$, it is found that $\mathbf{w} = (\mathbf{R} + \mathbf{H}^T\mathbf{B}\mathbf{H})^{-1}\delta\mathbf{y}_o$. Using the latter expression for substitution of $\mathbf{w}$ in the second row, we finally find $\delta\mathbf{x}_a = \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}^T\mathbf{B}\mathbf{H})^{-1}\delta\mathbf{y}_o$.

An alternative way to obtain $\mathbf{K}$. As *observation increment* we define

$$\begin{aligned} \mathbf{d} &:= \mathbf{y}_o - H(\mathbf{x}_b) && (2.15)\\ &= \mathbf{y}_o - H(\mathbf{x}_t) - H(\mathbf{x}_b - \mathbf{x}_t) \\ &= \vec{\epsilon}_o - \mathbf{H}\vec{\epsilon}_b && (2.16) \end{aligned}$$

We are searching for an optimal weight matrix $\mathbf{K}$, also known as *gain matrix*, mapping the observation increment $\mathbf{d}$ linearly on the *analysis increment* $\delta\mathbf{x}_a = \mathbf{x}_a - \mathbf{x}_b$

$$\mathbf{x}_a - \mathbf{x}_b = \mathbf{K}\ \mathbf{d} \tag{2.17}$$

with the error

$$-\vec{\epsilon}_b = \mathbf{x}_t - \mathbf{x}_b = \mathbf{K}\ \mathbf{d} - \vec{\epsilon}_a \tag{2.18}$$

or

$$\vec{\epsilon}_a = \mathbf{K}\ \mathbf{d} + \vec{\epsilon}_b \tag{2.19}$$

$$\vec{\epsilon}_a\vec{\epsilon}_a^T = (\mathbf{K}\ \mathbf{d} + \vec{\epsilon}_b)(\mathbf{K}\ \mathbf{d} + \epsilon_b)^T \tag{2.20}$$

Seek for minimal $|\vec{\epsilon_a}|$ by differentiation with respect to the elements K_{ij} of $\mathbf{K}$ and equating with $\mathbf{0}$

$$0 = (\mathbf{K}\ \mathbf{d} + \vec{\epsilon}_b)\ \mathbf{d}^T \tag{2.21}$$

Substituting $\mathbf{d}$ by (2.16) and taking expectations $\mathcal{E}\{\}$ gives

$$\mathbf{K}\mathcal{E}\left\{(\vec{\epsilon}_o - \mathbf{H}\vec{\epsilon}_b)(\vec{\epsilon}_o - \mathbf{H}\vec{\epsilon}_b)^T\right\} = \mathcal{E}\left\{\vec{\epsilon}_b(\vec{\epsilon}_o - \mathbf{H}\vec{\epsilon}_b)^T\right\} \tag{2.22}$$

Gain matrix $\mathbf{K}$ then is

$$\begin{aligned} \mathbf{K} &= \mathcal{E}\left\{\vec{\epsilon}_b(\vec{\epsilon}_o - \mathbf{H}\vec{\epsilon}_b)^T\right\}\left(\mathcal{E}\left\{(\vec{\epsilon}_o - \mathbf{H}\vec{\epsilon}_b)(\vec{\epsilon}_o - \mathbf{H}\vec{\epsilon}_b)^T\right\}\right)^{-1} \\ &= \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1} && (2.23) \end{aligned}$$

Upon resorting to (2.19), the analysis error covariance matrix $\mathbf{A}$ or $\mathbf{P}_a$ can be obtained by

$$\begin{aligned} \mathbf{A} &= \mathcal{E}\left\{\vec{\epsilon}_a(\vec{\epsilon}_a)^T\right\} \\ &= \mathcal{E}\left\{\vec{\epsilon}_b(\vec{\epsilon}_b^T + \vec{\epsilon}_b(\vec{\epsilon}_o - \mathbf{H}\vec{\epsilon}_b)^T\mathbf{K}^T + \mathbf{K}(\vec{\epsilon}_o - \mathbf{H}\vec{\epsilon}_b)\vec{\epsilon}_b^T + \mathbf{K}(\vec{\epsilon}_o - \mathbf{H}\vec{\epsilon}_b)(\vec{\epsilon}_o - \mathbf{H}\vec{\epsilon}_b)^T\mathbf{K}^T\right\} \\ &= \mathbf{B} - \mathbf{B}\mathbf{H}^T\mathbf{K}^T - \mathbf{K}\mathbf{H}\mathbf{B} + \mathbf{K}\mathbf{R}\mathbf{K}^T + \mathbf{K}\mathbf{H}\mathbf{B}\mathbf{H}^T\mathbf{K}^T, && (2.24) \end{aligned}$$

Introducing the explicit formula of $\mathbf{K}$ (2.23), the analysis error covariance matrix A then reads

$$\mathbf{A} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B} \tag{2.25}$$

Alternative form

With the expression (2.12), $\vec{\epsilon}_a$ can be written as

$$\begin{aligned}\vec{\epsilon}_a &= \vec{\epsilon}_b + \left(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{R}^{-1}(\vec{\epsilon}_o - \mathbf{H}\vec{\epsilon}_b) \qquad (2.26)\\ &= \left(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\left(\mathbf{H}^T\mathbf{R}^{-1}(\vec{\epsilon}_o - \mathbf{H}\vec{\epsilon}_b) + \left(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)\vec{\epsilon}_b\right)\\ &= \left(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\left(\mathbf{B}^{-1}\vec{\epsilon}_b + \mathbf{H}^T\mathbf{R}^{-1}\vec{\epsilon}_o\right) \qquad (2.27)\end{aligned}$$

To obtain an alternative formulation of the analysis error covariance matrix, use is made from

$$\begin{aligned}\mathbf{A} &= \mathbf{P}_a = \mathcal{E}\left\{\vec{\epsilon}_a(\vec{\epsilon}_a)^T\right\}\\ &= \mathcal{E}\left\{\left(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\left(\mathbf{B}^{-1}\vec{\epsilon}_b + \mathbf{H}^T\mathbf{R}^{-1}\vec{\epsilon}_o\right)\left(\mathbf{B}^{-1}\vec{\epsilon}_b + \mathbf{H}^T\mathbf{R}^{-1}\vec{\epsilon}_o\right)^T\left(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)^{-T}\right\}\\ &= \left(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)^{-1} \qquad (2.28)\end{aligned}$$

Frequently, the expression is needed in its inverse form

$$\mathbf{A}^{-1} = \left(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right) =: \mathbf{P}_a^{-1} \qquad (2.29)$$

Practical implementation

Procurement of a priori information

The a priori knowledge, required to process the OI-formula (ref2.11) includes the background field $\mathbf{x}_b$ and the background error covariance matrix $\mathbf{B}$. The former is usually obtained from a preceeding short term forecast, in earlier days from climatologies. As $\mathbf{B}$ is a purely statistical quantity, it must be estimated in some way. The following parameters are usually of special importance:

- The variance of the quantity to be analysed, denoting the diagonal elements B_{ii} of $\mathbf{B}$.
- The distance to which correlations exist, or a suitable parameterisation thereof L, is usually termed *radius of influence* or *(de)correlation length*. $L > 0$ introduces off-diagonal elements of $\mathbf{B}$. If L is not azimuth (direction) dependent, it is termed *isotropic*. If L is not location dependent, it is termed *homogeneous*.
- If $\mathbf{B}$ comprises only a single atmospheric field parameter, say, temperature, or several parameters, which are not connected by correlations, the formulation is termed *monovariate*, otherwise *multivariate*.

A classical way to estimate a monovariate $\mathbf{B}$ proceeds as follows: Compute correlations R_{kl} between locations k and l, given time series of observations $y^o_{k|l}(t)$ and model values $x^b_{k|l}(t)$, with the assumption that observation errors

do not correlate $\mathcal{E}\left\{\epsilon_k^o \epsilon_l^o\right\} = 0$, and observation errors do not correlate with any model error $\mathcal{E}\left\{\epsilon_k^o \epsilon_l^b\right\} = 0$. Then

$$\begin{aligned} R_{kl} &= \frac{\mathcal{E}\left\{(y_k^o - x_k^b)(y_l^o - x_l^b)\right\}}{\sqrt{\mathcal{E}\left\{(y_k^o - x_k^b)^2\right\}\mathcal{E}\left\{(y_l^o - x_l^b)^2\right\}}} \\ &= \frac{\mathcal{E}\left\{((y_k^o - x_k^t) - (x_k^b - x_k^t))((y_l^o - x_l^t) - (x_l^b - x_l^t))\right\}}{\sqrt{\mathcal{E}\left\{((y_k^o - x_k^t)) - (x_k^b - x_k^t))^2\right\}\mathcal{E}\left\{((y_l^o - x_l^t) - (x_l^b - x_l^t))^2\right\}}} \\ &= \frac{\mathcal{E}\left\{\epsilon_k^b \epsilon_l^b\right\}}{\sqrt{\left(\mathcal{E}\left\{(\epsilon_k^o)^2\right\} + \mathcal{E}\left\{(\epsilon_k^b)^2\right\}\right)\left(\mathcal{E}\left\{(\epsilon_l^o)^2\right\} + \mathcal{E}\left\{(\epsilon_l^b)^2\right\}\right)}} \\ &= \frac{\mathcal{E}\left\{\epsilon_k^b \epsilon_l^b\right\}}{\sqrt{((\sigma_k^o)^2 + (\sigma_k^b)^2)((\sigma_l^o)^2 + (\sigma_l^b)^2)}} \end{aligned} \tag{2.30}$$

In the limit, for any location k, l

$$\lim_{k \leftarrow l} R_{kl} = \frac{(\sigma^b)^2}{(\sigma^o)^2 + (\sigma^b)^2}$$

Multivariate covariance modelling

Given some observation increment $y_o(\mathbf{r}_i) - Hx(\mathbf{r}_i) \neq 0$, other parameters at the same location can be correlated sufficiently well to be assumed also deviating from the background. For example, a temperature observation increment δT is likely to be correlated with a wind perturbation $\delta\mathbf{u}(\mathbf{r})$ with respect to the background. Hence, we need covariances between temperature /geopotential fields and wind fields, resulting in six auto– and cross–covariances.

Using Helmholtz's theorem, can be used to express the wind fields as by streamfunction ψ and velocity potential χ, and the related perturbations $\delta\psi, \delta\chi$, respectively

$$\delta u = -\frac{\partial\delta\psi}{\partial y} + \frac{\partial\delta\chi}{\partial x}, \qquad \delta v = \frac{\partial\delta\psi}{\partial x} + \frac{\partial\delta\chi}{\partial y} \tag{2.31}$$

For correlations at locations $\mathbf{r}_i = (x_i, y_i)$ we are then able to compute covariances formally. For example, for the covariance of $u(\mathbf{r}_i) = u(x_i, y_i)$ and $v(\mathbf{r}_j) = v(x_j, y_j)$ it follows

$$\begin{aligned}\mathcal{E}\left\{\delta u(\mathbf{r}_1)\delta\mathbf{v}(r_2)\right\} =: \mathcal{E}\left\{\delta u_1\delta v_2\right\}\\ = -\mathcal{E}\left\{\frac{\partial\delta\psi_1\partial\delta\psi_2}{\partial y_1\partial x_2}\right\} + \mathcal{E}\left\{\frac{\partial\delta\psi_1\partial\delta\psi_2}{\partial x_1\partial x_2}\right\} - \mathcal{E}\left\{\frac{\partial\delta\psi_1\partial\delta\chi_2}{\partial y_1\partial x_2}\right\} + \mathcal{E}\left\{\frac{\partial\delta\chi_1\partial\delta\chi_2}{\partial x_1\partial y_2}\right\}\end{aligned} \tag{2.32}$$

We now assume homogeneous flow, that is, covariances are independent from the individual location, only distance $|\tilde{\mathbf{r}}_{12}| := |\mathbf{r}_1 - \mathbf{r}_2|$ matters. Then, based on arguments from turbulence theory applied to the determination of two-point velocity correlations (Batchelor, 1953; Panchev, 1971, Monin and Yaglom, 1975), cross-correlation can be formally derived. Taking the first term of (2.32), it can be set

$$\begin{aligned}\mathcal{E}\left\{\frac{\partial\delta\psi_1\partial\delta\psi_2}{\partial y_1\partial x_2}\right\} &= \mathcal{E}\left\{\lim_{\Delta y_1\to 0}\left[\frac{\delta\psi_1(x_1, y_1+\Delta y) - \delta\psi_1(x_1, y_1)}{\Delta y_1}\right]\frac{\partial\delta\psi_2}{\partial x_2}\right\}\\ &= \lim_{\Delta y_1\to 0}\left[\frac{\mathcal{E}\left\{\delta\psi_1(x_1, y_1+\Delta y)\frac{\partial\delta\psi_2}{\partial x_2}\right\} - \mathcal{E}\left\{\delta\psi_1(x_1, y_1)\frac{\partial\delta\psi_2}{\partial x_2}\right\}}{\Delta y_1}\right]\\ &= \frac{\partial}{\partial y_1}\mathcal{E}\left\{\delta\psi_1\frac{\partial\delta\psi_2}{\partial x_2}\right\}\end{aligned} \tag{2.33}$$

With the analog procedure applied to the still infinitesimal term $\frac{\delta\psi_2}{\partial x_2}$, and defining $\tilde{y} = y_1 - y_2$ and $\tilde{x} := x_2 - x_1$ it follows that

$$\begin{aligned}\mathcal{E}\left\{\frac{\partial\delta\psi_1\partial\delta\psi_2}{\partial y_1\partial x_2}\right\} &= \frac{\partial^2}{\partial y_1\partial x_2}\mathcal{E}\left\{\delta\psi_1\delta\psi_2\right\}\\ &= -\frac{\partial^2}{\partial\tilde{y}_1\partial\tilde{x}_2}\mathcal{E}\left\{\delta\psi_1\delta\psi_2\right\}.\end{aligned} \tag{2.34}$$

After proceeding with all covariances in the manner described above, the following relations result:

$$\begin{aligned}
\mathcal{E}\{\delta u_i \delta u_j\} &= -\frac{\partial^2}{\partial \tilde{y}^2}\mathcal{E}\{\delta\psi_i\delta\psi_j\} - \frac{\partial^2}{\partial \tilde{x}^2}\mathcal{E}\{\delta\chi_i\delta\chi_j\} + \frac{\partial^2}{\partial \tilde{x}^2\partial\tilde{y}^2}\left(\mathcal{E}\{\delta\psi_i\delta\chi_j\} + \mathcal{E}\{\delta\chi_i\delta\psi_j\}\right)\\
\mathcal{E}\{\delta v_i \delta v_j\} &= -\frac{\partial^2}{\partial \tilde{x}^2}\mathcal{E}\{\delta\psi_i\delta\psi_j\} - \frac{\partial^2}{\partial \tilde{y}^2}\mathcal{E}\{\delta\chi_i\delta\chi_j\} + - - -\frac{\partial^2}{\partial\tilde{x}\partial\tilde{y}}\left(\mathcal{E}\{\delta\psi_i\delta\chi_j\} + \mathcal{E}\{\delta\chi_i\delta\psi_j\}\right)\\
\mathcal{E}\{\delta u_i \delta v_j\} &= \frac{\partial^2}{\partial\tilde{x}\partial\tilde{y}}\mathcal{E}\{\delta\psi_i\delta\chi_j\} - \frac{\partial^2}{\partial \tilde{x}^2}\mathcal{E}\{\delta\chi_i\delta\psi_j\} + \frac{\partial^2}{\partial\tilde{x}\partial\tilde{y}}\left(\mathcal{E}\{\delta\psi_i\delta\psi_j\} - \mathcal{E}\{\delta\chi_i\delta\chi_j\}\right)\\
\mathcal{E}\{\delta v_i \delta u_j\} &= \frac{\partial^2}{\partial \tilde{y}^2}\mathcal{E}\{\delta\psi_i\delta\chi_j\} - \frac{\partial^2}{\partial \tilde{x}^2}\mathcal{E}\{\delta\chi_i\delta\psi_j\} + \frac{\partial^2}{\partial\tilde{x}\partial\tilde{y}}\left(\mathcal{E}\{\delta\psi_i\delta\psi_j\} - \mathcal{E}\{\delta\chi_i\delta\chi_j\}\right)\\
\mathcal{E}\{\delta\psi_i \delta u_j\} &= -\frac{\partial}{\partial\tilde{y}}\mathcal{E}\{\delta\psi_i\delta\psi_j\} + \frac{\partial}{\partial\tilde{x}}\mathcal{E}\{\delta\psi_i\delta\chi_j\}\\
\mathcal{E}\{\delta\psi_i \delta v_j\} &= +\frac{\partial}{\partial\tilde{x}}\mathcal{E}\{\delta\psi_i\delta\psi_j\} + \frac{\partial}{\partial\tilde{y}}\mathcal{E}\{\delta\psi_i\delta\chi_j\}\\
\mathcal{E}\{\delta u_i \delta\psi_j\} &= +\frac{\partial}{\partial\tilde{y}}\mathcal{E}\{\delta\psi_i\delta\psi_j\} - \frac{\partial}{\partial\tilde{x}}\mathcal{E}\{\delta\chi_i\delta\psi_j\}\\
\mathcal{E}\{\delta v_i \delta\psi_j\} &= -\frac{\partial}{\partial\tilde{x}}\mathcal{E}\{\delta\psi_i\delta\psi_j\} - \frac{\partial}{\partial\tilde{y}}\mathcal{E}\{\delta\chi_i\delta\psi_j\}
\end{aligned} \tag{2.35}$$

It should be noted, that $\mathcal{E}\{\psi_i u_j\} = -\mathcal{E}\{u_i\psi_j\}$. This is due to the sign dependence of direction of derivative: given a positive increment ψ_i at some location i, a positive increment u_j will be engendered to its north. Conversely, a positive increment u_j will engender a negative increment ψ_i to its north.

Now we restrict attention to the purely geostrophic case, that is, only the stream function ψ is further needed . Then the geostrophic relations $f\delta u = g\,\partial\delta h/\partial y$ and $f\delta v = -g\,\partial\delta h/\partial x$ hold. In addition, we define the gepotential covariance model

$$\mu_{ij} := \exp -\frac{r_{ij}^2}{2L_\psi^2}. \tag{2.36}$$

Then

$$\mathcal{E}\{\delta u_i \delta v_j\} = -\frac{g^2}{f_i f_j}\mathcal{E}\left\{\frac{\partial\delta h_i}{\partial y_i}\frac{\partial\delta h_j}{\partial x_j}\right\}, \tag{2.37}$$

and

$$\begin{aligned}
\mathcal{E}\{\delta u_i \delta v_j\} &= -\frac{g^2}{f_i f_j}\frac{\partial^2\mathcal{E}\{\partial\delta h_i\partial\delta h_j\}}{\partial y_i\partial x_j}\\
&= -\frac{g^2}{f_i f_j}\frac{\sigma_h^2\partial^2\mu_{ij}}{\partial y_i\partial x_j}
\end{aligned} \tag{2.38}$$

Variational approach and PSAS

We set again:

$$\mathbf{x}_a = \mathbf{x}_b + \delta\mathbf{x}_a, \qquad \mathbf{d} := \mathbf{y} - \mathbf{H}\mathbf{x}_b$$

Then

$$J(\delta\mathbf{x}) = 1/2\delta\mathbf{x}^T\mathbf{B}^{-1}\delta\mathbf{x} + 1/2(\mathbf{H}\delta\mathbf{x} - \mathbf{d})^T\mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x} - \mathbf{d}) \tag{2.39}$$

Covariance matrices, like $\mathbf{B}$, can be factorized by square roots (which are not unique!).

$$\mathbf{B} = \sqrt{\mathbf{B}}\sqrt{\mathbf{B}}, \qquad \mathbf{B}^{-1} = \sqrt{\mathbf{B}^{-1}}\sqrt{\mathbf{B}^{-1}} \tag{2.40}$$

The cost function is minimized with respect to $\delta\mathbf{x}$, that is a quantity in the model or analysis space. The minimisation problem is however ill–posed, that is, $\mathbf{B}$ is likely to have a poor condition number, namely if the radii of influence are large. Condition numbers, that is the quotent of smallest and largest eigenvalue, can be easily lower than 10^{-9}. The minimisation problem is then easily amenable for preconditioning with

$$\mathbf{v} := \sqrt{\mathbf{B}^{-1}}\delta\mathbf{x} \tag{2.41}$$

Then, the cost function reads

$$J(\mathbf{v}) = 1/2\mathbf{v}^T\mathbf{v} + 1/2(\mathbf{H}\sqrt{\mathbf{B}}\mathbf{v} - \mathbf{d})^T\mathbf{R}^{-1}(\mathbf{H}\sqrt{\mathbf{B}}\mathbf{v} - \mathbf{d}) \tag{2.42}$$

The gradient with respect to $\mathbf{v}$ then is

$$\nabla_{\mathbf{v}}J(\mathbf{v}) = (\mathbf{I} + \sqrt{\mathbf{B}}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\sqrt{\mathbf{B}})\mathbf{v} - \sqrt{\mathbf{B}}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{d} \tag{2.43}$$

The Hessian matrix is

$$\nabla^2_{\mathbf{v}}J(\mathbf{v}) = (\mathbf{I} + \sqrt{\mathbf{B}}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\sqrt{\mathbf{B}}) \tag{2.44}$$

Clearly, at the minimum, the transformed equation holds

$$(\mathbf{I} + \sqrt{\mathbf{B}}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\sqrt{\mathbf{B}})\mathbf{v} = \sqrt{\mathbf{B}}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{d} \tag{2.45}$$

which is obviously fully equivalent to the result

$$(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})\delta\mathbf{x} = \mathbf{H}^T\mathbf{R}^{-1}\mathbf{d} \tag{2.46}$$

The Physical–space Statistical Analysis System (PSAS). Equation (2.46) has been shown to be equivalent to

$$\delta\mathbf{x}_a = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{d} \tag{2.47}$$

The Physical–space Statistical Analysis System (PSAS) aims to exploit the frequently occuring situation, that there are considerably less observations

than degrees of freedom in the model space. The critical step to solve (2.47) is the the inversion of $(\mathbf{HBH}^T + \mathbf{R})$, that is the solution of the equation

$$(\mathbf{HBH}^T + \mathbf{R})\mathbf{w} = \mathbf{d}, \tag{2.48}$$

as $\mathbf{w}$ is the vector of analysis increments at the observation locations

$$\delta\mathbf{x}_a = \mathbf{BH}^T\mathbf{w} \tag{2.49}$$

Solving (2.48) is equivalent with solving the classical quadratic minimisation problem with a new objective function F

$$F(\mathbf{w}) = 1/2\mathbf{w}^T(\mathbf{HBH}^T + \mathbf{R})\mathbf{w} - \mathbf{w}^T\mathbf{d}. \tag{2.50}$$

The gradient of F is then

$$\nabla_{\mathbf{w}}F(\mathbf{w}) = (\mathbf{HBH}^T + \mathbf{R})\mathbf{w} - \mathbf{d}. \tag{2.51}$$

Again, in practice, solving equation (2.53) needs to be preconditioned. This has to be done with the variable in observation space $\mathbf{w} =: \sqrt{\mathbf{R}^{-1}}\mathbf{u}$. Equation (2.53) then becomes

$$F(\mathbf{u}) = 1/2\mathbf{u}^T(\mathbf{I} + \sqrt{\mathbf{R}^{-1}}\mathbf{HBH}^T\sqrt{\mathbf{R}^{-1}})\mathbf{u} - \mathbf{u}^T\sqrt{\mathbf{R}^{-1}}\mathbf{d} \tag{2.52}$$

and the gradient

$$\nabla_{\mathbf{u}}F(\mathbf{u}) = (\mathbf{I} + \sqrt{\mathbf{R}^{-1}}\mathbf{HBH}^T\sqrt{\mathbf{R}^{-1}})\mathbf{u} - \sqrt{\mathbf{R}^{-1}}\mathbf{d} \tag{2.53}$$

and Hessian matrix

$$\nabla^2_{\mathbf{u}}F(\mathbf{u}) = \mathbf{I} + \sqrt{\mathbf{R}^{-1}}\mathbf{HBH}^T\sqrt{\mathbf{R}^{-1}}. \tag{2.54}$$

CHAPTER 3

A posteriori validation and Degree of Freedom of Signal

A posteriori validation in observation space

(to be delivered in electronic form)

A posteriori validation in model space

(to be delivered in electronic form)

Degree of Freedom of Signal

(to be delivered in electronic form)

CHAPTER 4

Space–time data assimilation

Theory of Advanced Data Assimilation

Basically, today's applications can be traced from two principal mathematical disciplines, stochastic differential equations and control theory. Main references for these area are for example, Jazwinski (1970) for the former, and Lions (1971) for the latter. The objectives of the present section is to provide a broader basis for the statistical foundation of advanced data assimilation algorithms as it is used in atmospheric sciences, while placing special emphasis on the problem formulation as an overdetermined system and to point out the relations to other advanced data assimilation algorithms. The present theoretical exposition follows Lorenc (1986, 1988), Talagrand (1998), and Courtier (1997). As far as possible the mathematical notation adopts the recommendations expressed in Ide et al. (1997).

Maximisation of Probability

Let $\mathbf{z} \in \mathbf{R}^m$ be the the available declarative information base, that is some background knowledge and observations, and $\mathbf{x} \in \mathbf{R}^n$ be the searched after estimate of the state, conditional to the information contained in $\mathbf{z}$. We then have

$$P(\mathbf{x}|\mathbf{z}) = \frac{P(\mathbf{z}|\mathbf{x})P(\mathbf{x})}{\int P(\mathbf{z}|\mathbf{x})P(\mathbf{x})d\mathbf{x}}. \tag{4.1}$$

To proceed, two models of different types are required: 1. a numerical model, which maps the system state $\mathbf{x}$ to the available data $\mathbf{z}$ in the information space by applying coded geophysical relations, and 2. a statistical model, describing the probability density function $P(\mathbf{x}|\mathbf{z})$.

1. For expository purposes we assume the numerical model to be linear. The general case is considered below. The data assimilation problem may be formulated as searching for the solution of an overdetermined linear system:

$$\mathbf{z} = \mathbf{\Gamma}\mathbf{x}^t + \zeta, \tag{4.2}$$

where $\mathbf{z} \in \mathrm{R}^m$ denotes any declarative information. For most cases, this is a merger of a priori knowledge and observations. Let the number of information elements be m, while n be the dimension of the analysis or model space, matrix $\mathbf{\Gamma} \in \mathrm{R}^{m \times n}, \quad m > n$, be one–to–one defining the observability condition, $\mathbf{x}^t \in \mathrm{R}^n$ be the "true" system state, and $\zeta \in \mathrm{R}^n$ be the unbiased and unknown error vector of $\mathbf{z}$.

2. Assuming a Gaussian distribution to be sufficiently accurate, the probability density function $P(\mathbf{x}|\mathbf{z})$ in (4.1) of an analysis $\mathbf{x}^a$, given an information basis $\mathbf{z}$ can be modelled as

$$P(\mathbf{x}|\mathbf{z}) \propto \exp[-1/2(\Gamma\mathbf{x} - \mathbf{z})^T \mathbf{S}^{-1}(\Gamma\mathbf{x} - \mathbf{z})], \tag{4.3}$$

where $\mathbf{S} := \mathcal{E}(\zeta\zeta^T) \in \mathrm{R}^{m \times m}$ is the known "information" error covariance matrix.

The assumption of a Gaussian density distribution is generally made for convenience. However, it is widely justified in most cases, where it describes the measurement error characteristics of the dynamical meteorological parameters and gas phase constituents. Severe problems must be envisaged however for the assimilation of humidity observations, and likewise for chemical species under the impact of heterogeneous chemistry processes due to phase transitions. In atmospheric chemistry data assimilation this problem has not yet been considered. Since the present study is concerned with gas phase chemistry only, the potential error introduced by assuming a Gaussian error distribution is esteemed to be within the error margins other model and information uncertainties.

Upon applying the negative logarithm to (4.3), a weighted cost function $\mathcal{J}(\psi)$ is found, which quantifies the discrepancy between two vectors of the information space

$$-\ln(P(\psi - \mathbf{z})) \propto \mathcal{J}(\psi) = (\psi - \mathbf{z})^T \mathbf{S}^{-1}(\psi - \mathbf{z}).$$

With the information error covariance matrix $\mathbf{S}$ being known, and given $\mathbf{z} \in \mathrm{R}^m$, we solve for $\mathbf{x}^a \in \mathrm{R}^n$, minimising

$$\min_{\mathbf{x}} \parallel \mathbf{z} - \mathbf{\Gamma}\mathbf{x} \parallel_{\mathbf{S}^{-1}} \tag{4.4}$$

with $\parallel \cdot \parallel_{\mathbf{S}^{-1}}$ the Mahalanobis scalar product $(\cdot,\cdot)_{\mathbf{S}^{-1}} := \langle \cdot, \mathbf{S}^{-1} \cdot \rangle$, subject to the extremal condition

$$1/2 \frac{\partial}{\partial \mathbf{x}} \langle \mathbf{z} - \mathbf{\Gamma}\mathbf{x}, \mathbf{S}^{-1}(\mathbf{z} - \mathbf{\Gamma}\mathbf{x}) \rangle = -\mathbf{\Gamma}^T \mathbf{S}^{-1}(\mathbf{z} - \mathbf{\Gamma}\mathbf{x}) = 0. \tag{4.5}$$

The solution to this problem is given by

$$\mathbf{x}^a = (\mathbf{\Gamma}^T\mathbf{S}^{-1}\mathbf{\Gamma})^{-1}\mathbf{\Gamma}^T\mathbf{S}^{-1}\mathbf{z}, \tag{4.6}$$

where $\mathbf{x}^a$ has the property to be a Best Linear Unbiased Estimate (BLUE), that is, $\mathcal{E}(\mathbf{x}^a - \mathbf{x}^t) = 0$.

With (4.6) and (4.2), the associated analysis error is found to be $\mathbf{x}^a - \mathbf{x}^t = (\mathbf{\Gamma}^T\mathbf{S}^{-1}\mathbf{\Gamma})^{-1}\mathbf{\Gamma}^T\mathbf{S}^{-1}(\zeta)$.

Making use of $\mathcal{E}(\zeta\zeta^T) = \mathbf{S}$ it is easy to see that the analysis error covariance matrix is

$$\mathbf{P}^a := \mathcal{E}\left((\mathbf{x}^a - \mathbf{x})(\mathbf{x}^a - \mathbf{x})^T\right) = (\mathbf{\Gamma}^T\mathbf{S}^{-1}\mathbf{\Gamma})^{-1}. \tag{4.7}$$

The Prognostic Model

An advanced spatio–temporal data assimilation procedure involves a model $\mathcal{M}$, which can formally written as a stochastic differential equation

$$\frac{d\mathbf{x}^t}{dt} = \mathcal{M}(\mathbf{x}^t) + \eta, \tag{4.8}$$

with model error η. In data assimilation parlance, model $\mathcal{M}$ is often referred to as the 'forward' or 'direct' model. Upon differentiation with respect to $\mathbf{x}$ we obtain

$$\frac{d\delta\mathbf{x}}{dt} = \mathcal{M}(\mathbf{x}^t + \delta\mathbf{x}) - \mathcal{M}(\mathbf{x}) = \mathbf{M}'\delta\mathbf{x} + \mathcal{O}(\parallel \delta\mathbf{x} \parallel^2), \tag{4.9}$$

where $\mathbf{M}'$ is the model tangent–linear to $\mathcal{M}$. Introducing the integration operator or resolvent $\mathbf{M}(t_j, t_i)$, which propagates a perturbation $\delta\mathbf{x}(t)$ of the state variable $\mathbf{x}(t)$ from time t_i to time t_j, a stepwise tangent–linear model integration gives

$$\delta\mathbf{x}(t_n) = \mathbf{M}(t_n, t_{n-1})\mathbf{M}(t_{n-1}, t_{n-2})\ldots\mathbf{M}(t_1, t_0)\delta\mathbf{x}(t_0). \tag{4.10}$$

For a direct inference of the observation–minus–model discrepancy $\mathbf{y}_i^o - \mathbf{H}_i\mathbf{x}(t_i)$ at time t_i we calculate $\mathbf{H}_i\mathbf{x}(t_i)$ as the model equivalent to the observation set $\mathbf{y}_i^o$ such that

$$\mathbf{y}_i^o - \mathbf{H}_i\mathbf{x}(t_i) = \mathbf{H}_i\delta\mathbf{x}(t_i) = \mathbf{H}_i\mathbf{M}(t_i, t_0)\delta\mathbf{x}(t_0), \tag{4.11}$$

where $\mathbf{y}_i^o \in \mathrm{R}^{p_i}$ is the vector of observations with $p_i = m_i - n$ being the number of available observations, and $\mathbf{H}_i \in \mathrm{R}^{p_i \times n}$ is the forward interpolator or observation operator.

The information vector $\mathbf{z}$ is now divided into two categories of information sources, that is $\mathbf{x}^b$, as the a priori state, which can be obtained from climatologies or antecedent forecasts, and the vector of observations $\mathbf{y}^o$. Likewise, the operator $\mathbf{\Gamma}$ can be composed from an identity operator $\mathbf{I}$ and the model resolvents (4.30) for each time step $[t_0, \ldots, t_N]$, $\mathbf{G} :=$

$(\mathbf{H}_0, \mathbf{H}_1\mathbf{M}(t_1,t_0), \ldots, \mathbf{H}_N\mathbf{M}(t_N,t_0))^T$, with a concatenation of forward interpolators and the tangent linear model resolvent. The time interval $[t_0, \ldots, t_N]$ is called the assimilation interval or assimilation window. Observations outside this time span are ignored. It can be seen that Γ and $\mathbf{z}$ may be split as

$$\mathbf{\Gamma} := \begin{pmatrix} \mathbf{I} \\ \mathbf{G} \end{pmatrix} \qquad \mathbf{z} := \begin{pmatrix} \mathbf{x}^b \\ \mathbf{y}^0 \end{pmatrix} = \begin{pmatrix} \mathbf{x} & + & \zeta^b \\ \mathbf{Gx} & + & \epsilon \end{pmatrix}. \tag{4.12}$$

The error of the background estimate and the observations are ζ^b and ϵ, respectively. Assuming the errors of the background estimate and the observations being uncorrelated, we set $\mathcal{E}(\zeta^b\epsilon^T) = \mathbf{0}$. The background error covariance matrix is then $\mathbf{B} = \mathbf{P}^b := \mathcal{E}(\zeta^b\zeta^{bT}) \in \mathrm{R}^{n\times n}$ and the observation error covariance matrix is $\mathbf{R} := \mathcal{E}(\epsilon\epsilon^T) \in \mathrm{R}^{p\times p}$, with $p = m - n$ the number of available observations. Then the information error covariance matrix can be written as

$$\mathbf{S} := \begin{pmatrix} \mathbf{P}^b & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{pmatrix}. \tag{4.13}$$

As a matter of convenience we assume that no linear combination of observations is perfect, that is, that $\mathbf{R}$ is nonsingular. Likewise, the background information is assumed to be redundant, assuring that $\mathbf{P}^b$ is nonsingular. However, as described in (Elbern and Schmidt, 2001), $\mathbf{P}^b$ has often a poor condition number.

With (4.7), (4.12), and (4.13) and the application of the Woodbury formula[1] we have for the analogy to the analysis error covariance matrix $\mathbf{P}^a \in \mathrm{R}^{n\times n}$

$$\mathbf{P}^a := (\mathbf{P}^{b^{-1}} + \mathbf{G}^T\mathbf{R}^{-1}\mathbf{G})^{-1} = \mathbf{P}^b - \mathbf{P}^b\mathbf{G}^T(\mathbf{G}\mathbf{P}^b\mathbf{G}^T + \mathbf{R})^{-1}\mathbf{G}\mathbf{P}^b = (\mathbf{I} - \mathbf{K}\mathbf{G})\mathbf{P}^b. \tag{4.14}$$

where

$$\mathbf{K} := \mathbf{P}^b\mathbf{G}^T(\mathbf{G}\mathbf{P}^b\mathbf{G}^T + \mathbf{R})^{-1} \quad \in \mathrm{R}^{n\times p} \tag{4.15}$$

is the Kalman gain matrix.

In the case of split and uncorrelated information sources (4.13), the analogy to equation (4.6) can be derived, again by use of the Woodbury formula, as follows (see also (2.14) :

$$\begin{aligned}
\mathbf{x}^a &= (\mathbf{P}^{b^{-1}} + \mathbf{G}^T\mathbf{R}^{-1}\mathbf{G})^{-1}(\mathbf{I}, \mathbf{G}^T) \begin{pmatrix} \mathbf{P}^b & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{pmatrix} \begin{pmatrix} \mathbf{x}^b \\ \mathbf{y}^0 \end{pmatrix} \\
&= (\mathbf{P}^{b^{-1}} + \mathbf{G}^T\mathbf{R}^{-1}\mathbf{G})^{-1}\mathbf{P}^{b^{-1}}\mathbf{x}^b + (\mathbf{P}^{b^{-1}} + \mathbf{G}^T\mathbf{R}^{-1}\mathbf{G})^{-1}\mathbf{G}^T\mathbf{R}^{-1}\mathbf{y}^0 \\
&= (\mathbf{I} - \mathbf{K}\mathbf{G})\mathbf{x}^b + \\
&\quad (\mathbf{P}^{b^{-1}} + \mathbf{G}^T\mathbf{R}^{-1}\mathbf{G})^{-1}(\mathbf{G}^T\mathbf{R}^{-1}\mathbf{G}\mathbf{P}^b\mathbf{G}^T + \mathbf{G}^T)(\mathbf{G}\mathbf{P}^b\mathbf{G}^T + \mathbf{R})^{-1}\mathbf{y}^0 \\
&= (\mathbf{I} - \mathbf{K}\mathbf{G})\mathbf{x}^b + \mathbf{K}\mathbf{y}^0.
\end{aligned}$$

Hence we find

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}\mathbf{d} \tag{4.16}$$

with $\mathbf{d} := \mathbf{y}^0 - \mathbf{G}\mathbf{x}^b$ being the innovation vector.

[1] A familiar form of the Woodbury formula reads
$(\mathbf{A} + \mathbf{CBD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{C}(\mathbf{B}^{-1} + \mathbf{D}\mathbf{A}^{-1}\mathbf{C})^{-1}\mathbf{D}\mathbf{A}^{-1}$

Smoother and Filter Algorithms

In the framework of the variational calculus, problem (4.4) is reformulated as an iterative minimisation of a cost function

$$\mathcal{J}(\xi) := 1/2(\mathbf{\Gamma}\xi - \mathbf{z})^T \mathbf{S}^{-1}(\mathbf{\Gamma}\xi - \mathbf{z}).$$

After transfer into the form of two uncorrelated information sources (4.12) and (4.13), the cost function then is

$$J(\xi(t_0)) = \frac{1}{2}(\mathbf{x}^b(t_0) - \xi(t_0))^T \mathbf{P^b}^{-1}(\mathbf{x}^b(t_0) - \xi(t_0)) + \frac{1}{2}\sum_0^N \left(\mathbf{y}^0(t_i) - \mathbf{H}\xi(t_i)\right)^T \mathbf{R}^{-1}(\mathbf{y}^0(t_i) - \mathbf{H}\xi(t_i)). \tag{4.17}$$

The minimisation of this quadratic expression is treated by classical minimisation methods, say, quasi-Newton algorithms, which require the gradient of J with respect to the optimisation parameter $\xi(t_0)$. This is easily found to be

$$\nabla_{\xi(t_0)} J = -\mathbf{P^b}^{-1}(\mathbf{x}^b(t_0) - \xi(t_0)) - \sum_{m=0}^{N} \mathbf{M}^T(t_m, t_0)\mathbf{H}_m^T \mathbf{R}^{-1}(\mathbf{y}^0(t_m) - \mathbf{H}_m\xi(t_m)). \tag{4.18}$$

As described above, the information contents introduced by the background knowledge renders the optimisation problem formally overdetermined and acts as a measure to reduce the ill-conditioning. On the other hand the background error covariance matrix is likely to be ill–conditioned itself and may hamper the minimisation procedure.

Formally, the preconditioning requires the calculation of the inverse, its square root, and the inverse of the square root of the background error covariance matrix (Lorenc, 1988). Thus, the general approach is given by the transformation $\mathbf{v} := \sqrt{\mathbf{P^b}^{-1}}\delta\mathbf{x}(t_0)$, and the associated gradient reads $\nabla_{\mathbf{v}} J(\mathbf{v}) = \sqrt{\mathbf{P^b}_0}\nabla_{\xi(t_0)} J$. The transformed state $\mathbf{v}$ and $\nabla_{\mathbf{v}} J(\mathbf{v})$ is the input for the minimisation procedure, with the total costs $\mathcal{J}$ remaining unaffected by the coordinate transformations.
Equation (4.25) is a smoothing data assimilation algorithm, as it provides for a smooth adaption between the observations and the background values within an assimilation interval. The constraint maintaining the smoothing property is given by the dynamic model (4.8).

The minimisation by iteration is only a technical issue and a simultaneous minimisation for all time steps within the assimilation interval is feasible, but inefficient for very high dimensional problems. Formally, 4D–var and 3D–var can be cast into the same algorithm, differing only in the inclusion of the time dimension by model evolution.

In contrast ro smoother algorithms, sequential data assimilation algorithm propagates in time and updates the system state at each instant of observation

supply. Thus far, the background error covariance matrix $\mathbf{P}^b$ has been assumed to be known. The Kalman filter (Kalman, 1960; Kalman and Bucy, 1962) relaxes this requirement with elevated sophistication, as it not only estimates the system state, but also the then purely spatial analysis error covariance matrix $\mathbf{P}_i^a$ and the background error covariance matrix $\mathbf{P}_i^b$ for each time step i. $\mathbf{P}_i^b$ is commonly referred to as the forecast error covariance matrix in this context. Assuming a linear model integrator $\mathbf{M}(t_i, t_{i-1})$ for simplicity, the analysis based forecast is then

$$\mathbf{x}^f(t_i) = \mathbf{M}(t_i, t_{i-1})\mathbf{x}^a(t_{i-1}) \tag{4.19}$$

to obtain the forecasted model state $\mathbf{x}^f(t_{i+1})$. Preserving the unknown model error η and $\epsilon^a := \mathbf{x}^a(t_{i-1}) - \mathbf{x}^t(t_{i-1})$, equation (4.8) then modifies to

$$\epsilon^f := \mathbf{x}^f(t_i) - \mathbf{x}^t(t_i) = \mathbf{M}(t_i, t_{i-1})\epsilon^a - \eta \tag{4.20}$$

Multiplying (4.20) by its transposed from the right hand side, applying the expectation operator $\mathcal{E}$ and observing the analysis and model errors to be uncorrelated, it is seen that the update of the forecast error covariance matrix $\mathbf{P}_i^b$ is (Jaszwinski, 1970; Daley, 1991)

$$\mathbf{P}_i^b = \mathbf{M}(t_i, t_{i-1})\mathbf{P}_i^a\mathbf{M}^T(t_i, t_{i-1}) + \mathbf{Q} \tag{4.21}$$

with $\mathbf{Q} := \mathcal{E}(\eta\eta^T)$ being the model error covariance matrix, assumed to be known. Each time step i, which introduces new observations, results in a state analysis $\mathbf{x}^a(t_{i+1})$, that rests on the forecasted state $\mathbf{x}^f(t_i)$ and the innovation vector $\mathbf{d}_i := \mathbf{y}_i^0 - \mathbf{H}_i\mathbf{x}^f(t_{i+1})$, weighted by the Kalman gain matrix. Observe that now operator $\mathbf{G}$ is replaced by the mere forward interpolation operator for time i, $\mathbf{H}_i$. With obvious subscripts for time indexing the remaining Kalman filter equations are readily available from prior derivations. We find

$$\mathbf{x}^a(t_i) = \mathbf{x}^b(t_i) + \mathbf{K}_i\mathbf{d}_i, \tag{4.22}$$

$$\mathbf{K}_i := \mathbf{P}_i^b\mathbf{H}_i^T(\mathbf{H}_i\mathbf{P}_i^b\mathbf{H}_i^T + \mathbf{R}_i)^{-1} \quad \in \mathrm{R}^{n\times p_i} \tag{4.23}$$

and

$$\mathbf{P}_i^a = (\mathbf{I} - \mathbf{K}_i\mathbf{G}_i)\mathbf{P}_i^b. \tag{4.24}$$

Equations (4.20) and (4.21) are termed the forecast equations of the Kalman filter, while equations (4.22), (4.23), and (4.24) are the analysis equations. It is equation (4.21) which incurs the excessively high computational demands, as it formally requires two model integrations per model dimension.

Four–dimenional variational approach

As with 3D–var, the optimization problem may be condensed in a suitably defined distance function $\mathcal{J}(\mathbf{x}(t))$ which is a measure of the model's predictive skill with respect to observations. The scalar function $\mathcal{J}(\mathbf{x}(t))$ is to be minimized, where $\mathbf{x}(t)$ is the vector valued state variable of the model. A coarse outline of the procedure reads as follows: (i) find the gradient $\nabla_{\mathbf{x}(t_0)}\mathcal{J}$ of the distance function, and (ii) find $\mathbf{x}^{n+1}(t_0), \quad n = 0, 1, \ldots$ such that $\mathcal{J}(\mathbf{x}^{n+1}(t_0)) < \mathcal{J}(\mathbf{x}^n(t_0))$, with the aid of $\nabla_{\mathbf{x}(t_0)}\mathcal{J}$ and previously calculated gradients. (iii) Repeat this sequence until $\mathcal{J}$ becomes as small as a prescribed threshold value χ^2, to be inferred later.

Adjoint computation of $\nabla_{\mathbf{x}(t_0)}\mathcal{J}$

For the sake of completeness and later notational reference, we will briefly outline the approach that provides the gradient of the distance function by means of the adjoint calculus. A mathematically strict and comprehensive derivation is dealt with in Talagrand and Courtier (1987) and references therein. The exposition here mainly follows this study.

The distance function $\mathcal{J}$ may be defined as follows:

$$\mathcal{J}(\mathbf{x}(t)) = \frac{1}{2}(\mathbf{x}_b - \mathbf{x}(t_0))^T \mathbf{B}^{-1}(\mathbf{x}_b - \mathbf{x}(t_0)) + \frac{1}{2}\int_{t_0}^{t_N} (\mathbf{y}(t) - \mathbf{H}\mathbf{x}(t))^T \mathbf{R}^{-1}(\mathbf{y}(t) - \mathbf{H}\mathbf{x}(t))dt \tag{4.25}$$

where $\mathcal{J}$ is a scalar functional defined on the time interval $t_0 \leq t \leq t_N$ dependent on the vector valued state variable $\mathbf{x} \in \mathcal{H}$ with $\mathcal{H}$ denoting a Hilbert space. The first guess or background values $\mathbf{x}_b$ are defined at $t = t_0$, and $\mathbf{B}$ is the covariance matrix of the estimated background error. The observations are denoted $\mathbf{y}$ and the observation and representativeness errors are included in the covariance matrix $\mathbf{R}^{-1}$. For simplicity we assume direct observations of the model's state variables.

Denoting the inner product of $\mathcal{H}$ by bracketed parentheses $\langle \quad , \quad \rangle$, the operator $\mathbf{M}'$, mapping from $\mathcal{H}$ into $\mathcal{H}$ itself, has the adjoint operator $\mathbf{M}^T$, which is defined by $\langle \mathbf{y}, \mathbf{M}'\mathbf{z} \rangle = \langle \mathbf{M}^T\mathbf{y}, \mathbf{z} \rangle$ for all $\mathbf{y}, \mathbf{z} \in \mathcal{H}$. For the remainder we drop the background term in (4.25), however introduce the model equation (**??**) as a strong constraint with Lagrange multipliers $\lambda(t)$. Setting $1/2\,(\mathbf{y}(t) - \mathbf{H}\mathbf{x}(t))^T \mathbf{R}^{-1}(\mathbf{y}(t) - \mathbf{H}\mathbf{x}(t)) = \mathcal{O}(t)$ for notational convenience, equation (4.25) then reads

$$\mathcal{J}^o(\mathbf{x}(t)) = \int_{t_0}^{t_N} \left(\langle \lambda, \frac{d\mathbf{x}(t)}{dt} - \mathcal{M}\mathbf{x}(t) \rangle \right) dt \tag{4.26}$$

Then, the variation of the distance function $\mathcal{J}^o(\mathbf{x}(t_0))$ can be expressed as

$$\delta\mathcal{J}^o = \int_{t_0}^{t_N} \left(\langle \nabla_{\mathbf{x}}\mathcal{O}(t), \delta\mathbf{x}(t) \rangle + \langle \delta\lambda, \frac{d\mathbf{x}(t)}{dt} - \mathcal{M}\mathbf{x}(t) \rangle + \langle \lambda, \frac{d\delta\mathbf{x}(t)}{dt} - \mathbf{M}'\delta\mathbf{x}(t) \rangle \right) dt \tag{4.27}$$

$$= \int_{t_0}^{t_N} \left(\langle \nabla_{\mathbf{x}} \mathcal{O}(t) - \frac{d\lambda(t)}{dt} - \mathbf{M}^T \lambda(t), \delta\mathbf{x}(t) \rangle + \langle \delta\lambda, \frac{d\mathbf{x}(t)}{dt} - \mathcal{M}\mathbf{x}(t) \rangle \right) dt + \lambda(t_N)\delta\mathbf{x}(t_N) - \lambda(t_0)\delta\mathbf{x}(t_0),$$

where integration by parts was applied. Introducing the extremal principle $\delta\mathcal{J}^o = 0$, the integrand includes the inhomogeneous adjoint equation, which is forced by the observation increment $\mathbf{R^{-1}}(\mathbf{y}(t) - \mathbf{H}\mathbf{x}(t))$ at time t

$$-\frac{d\lambda(t)}{dt} - \mathbf{M}^T \lambda(t) = \mathbf{R^{-1}}(\mathbf{y}(t) - \mathbf{H}\mathbf{x}(t)). \tag{4.28}$$

Again, integration of the tangent linear equation (4.9) gives the evolution of an initial perturbation $\delta\mathbf{x}(t_0)$ at later times t_n and can formally be expressed as

$$\delta\mathbf{x}(t_n) = \mathbf{L}^T(t_n, t_0)\delta\mathbf{x}(t_0), \tag{4.29}$$

where the operator $\mathbf{M}(t_n, t_0)$ denotes the resolvent of $\mathbf{M}'$ for the time interval $[t_0, t_n]$ acting on the initial state $\mathbf{x}(t_0)$.

Designating $\mathbf{L}^T$ as the resolvent of $\mathbf{M}^T$, it can be shown that $\mathbf{L}^T(t_i, t_j)$, is the adjoint of $\mathbf{M}(t_j, t_i)$, or, for any $\delta\mathbf{x}, \lambda$, we have $\langle \lambda(t_N), \mathbf{M}(t_N, t_0)\delta\mathbf{x}(t_0) \rangle = \langle \mathbf{L}^T(t_0, t_N)\lambda(t_N), \delta\mathbf{x}(t_0) \rangle$. This is readily seen from the conservation of the inner product of a perturbed state variable $\delta\mathbf{x}(t)$ and its adjoint $\lambda(t)$, if external forcing in (4.28) is excluded:

$$\frac{d}{dt}\langle \lambda(t), \delta\mathbf{x}(t) \rangle = \langle \lambda(t), \mathbf{M}'\delta\mathbf{x}(t) \rangle - \langle \mathbf{M}^T\lambda(t), \delta\mathbf{x}(t) \rangle = 0. \tag{4.30}$$

With the aid of (4.27) and (4.29) it is now possible to express $\delta\mathcal{J}^o$ as a function of $\mathbf{x}(t_0)$ alone

$$\delta\mathcal{J}^o(\mathbf{x}(t_0)) = \int_{t_0}^{t_N} \langle \nabla_{\mathbf{x}} \mathcal{O}(t), \mathbf{L}(t, t_0)\delta\mathbf{x}(t_0) \rangle dt = \langle \int_{t_0}^{t_N} \mathbf{L}^T(t_0, t)\nabla_{\mathbf{x}} \mathcal{O}(t) dt, \delta\mathbf{x}(t_0) \rangle, \tag{4.31}$$

which finally leads to the desired gradient of $\mathcal{J}^o$, given in terms of a discretized expression of the adjoint operator $\mathbf{L}^T(t_0, t)$

$$\nabla_{\mathbf{x}(t_0)} \mathcal{J}^o = \sum_{m=0}^{N} \mathbf{L}^T(t_0, t_1)\mathbf{L}^T(t_1, t_2)...\mathbf{L}^T(t_{m-1}, t_m)\nabla_{\mathbf{x}} \mathcal{O}(t_m). \tag{4.32}$$

It remains to be shown that $\nabla_{\mathbf{x}(t_0)} \mathcal{J}^o = \lambda(t_0)$. Defining the backward initial condition $\lambda(t_N) = 0$ and given a single instantaneous forcing $\lambda(t') = -\nabla_{\mathbf{x}(t')} \mathcal{O}(t')$ at any time $t', \quad t \leq t' < t_N$, equation (4.28) may be rewritten, again in exact form for convenience, as

$$\frac{\partial}{\partial t}\mathbf{L}^T(t, t')\nabla_{\mathbf{x}(t')} \mathcal{O}(t') = -\frac{d\lambda(t)}{dt} = \mathbf{M}^T\mathbf{L}^T(t, t')\nabla_{\mathbf{x}(t')} \mathcal{O}(t'), \tag{4.33}$$

to reveal that $\frac{\partial}{\partial t}\mathbf{L}^T(t, t') = \mathbf{M}^T\mathbf{L}^T(t, t')$.

Integration of (4.33) then gives

$$\lambda(t) = \mathbf{L}^T(t,t')\nabla_{\mathbf{x}(t')}\mathcal{O}(t'). \tag{4.34}$$

Since the resolvent $\mathbf{L}$ is a linear operator, all observations can be combined by adding up the corresponding equations (4.34), which demonstrates the preposition for $t = t_0$.

The computational procedure to calculate the gradient $\nabla_{\mathbf{x}(t_0)}\mathcal{J}^o$ is first to integrate 'forward in time' the 'forward' model $\mathbf{M}$ by equation (4.29), followed by an integration 'backward in time' of the adjoint ('backward') equation (4.32). The forward integration provides the distance function value $\mathcal{J}^o(\mathbf{x}(t_0))$ and intermediate values for retrieving the operators $\tilde{\mathbf{L}}^T$ for the backward integration, while the backward integration gives the gradient $\nabla_{\mathbf{x}(t_0)}\mathcal{J}^o$. Both $\mathcal{J}^o$ and $\nabla_{\mathbf{x}(t_0)}\mathcal{J}^o$ values then enter the minimization routine.

Practical construction of adjoint code

A program line, which modifies a state variable x_i^{k+1} at a model step $k+1$, (which is not neccessarily time)

$$x_i^{k+1} = f(x_1^k, \cdots, x_i^k, \cdots, x_n^k)$$

has the tangent linear form

$$\delta x_i^{k+1} = \frac{\partial f}{\partial x_1^k}\delta x_1^k + \cdots + \frac{\partial f}{\partial x_i^k}\delta x_i^k + \cdots + \frac{\partial f}{\partial x_n^k}\delta x_n^k$$

A blown–up equivalent notation as "transformation" reads

$$\begin{pmatrix} \delta x_1^k \\ \vdots \\ \delta x_i^k \\ \vdots \\ \delta x_n^k \\ \delta x_i^{k+1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 & 0 \\ & \ddots & & & & \vdots \\ & & 1 & & & \vdots \\ & & & \ddots & & \vdots \\ 0 & \cdots & \cdots & 0 & 1 & 0 \\ \frac{\partial f}{\partial x_1^k} & \cdots & \frac{\partial f}{\partial x_i^k} & \cdots & \frac{\partial f}{\partial x_n^k} & 0 \end{pmatrix} \begin{pmatrix} \delta x_1^k \\ \vdots \\ \delta x_i^k \\ \vdots \\ \delta x_n^k \\ \delta x_i^{k+1} \end{pmatrix}$$

adjoint (transposition)

$$\begin{pmatrix} \lambda x_1^k \\ \vdots \\ \lambda x_i^k \\ \vdots \\ \lambda x_n^k \\ \lambda x_i^{k+1} \end{pmatrix} = \begin{pmatrix} 1 & & & & \frac{\partial f}{\partial x_1^k} \\ 0 & \ddots & & & \vdots \\ \vdots & & 1 & & \frac{\partial f}{\partial x_i^k} \\ 0 & & & \ddots & \vdots \\ \vdots & & & 1 & \frac{\partial f}{\partial x_n^k} \\ 0 & & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \lambda x_1^k \\ \vdots \\ \lambda x_i^k \\ \vdots \\ \lambda x_n^k \\ \lambda x_i^{k+1} \end{pmatrix}$$

Hence

$$\begin{array}{rcl} \lambda x_1^k & = & \lambda x_1^k + \frac{\partial f}{\partial x_1^k}\lambda x_i^{k+1} \\ \vdots & & \vdots \\ \lambda x_i^k & = & \lambda x_i^k + \frac{\partial f}{\partial x_i^k}\lambda x_i^{k+1} \\ \vdots & & \vdots \\ \lambda x_n^k & = & \lambda x_n^k + \frac{\partial f}{\partial x_n^k}\lambda x_i^{k+1} \\ \lambda x_i^{k+1} & = & 0 \end{array}$$

Example

$$\frac{\partial c(t,x)}{\partial t} = \kappa \frac{\partial^2 c(t,x)}{\partial^2 x}$$

Program line inside loop over $i = 2, \ldots, N-1$

$$c_i^{k+1} = c_i^k + \gamma(c_{i-1}^k - 2c_i^k + c_{i+1}^k),$$

where

$$\gamma := \frac{\kappa \Delta t}{\Delta x^2}$$

The tangent linear form reads

$$\delta c_i^{k+1} = (1-2\gamma)\delta c_i^k + \gamma \delta c_{i-1}^k + \gamma \delta c_{i+1}^k$$

equivalent notation as "transformation"

$$\begin{pmatrix} \delta c_{i-1}^k \\ \delta c_i^k \\ \delta c_{i+1}^k \\ \delta c_i^{k+1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \gamma & 1-2\gamma & \gamma & 0 \end{pmatrix} \begin{pmatrix} \delta c_{1-1}^k \\ \delta c_i^k \\ \delta c_{i+1}^k \\ \delta c_i^{k+1} \end{pmatrix}$$

adjoint (transposition)

$$\begin{pmatrix} \lambda c_{i-1}^k \\ \lambda c_i^k \\ \lambda c_{i+1}^k \\ \lambda c_i^{k+1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \gamma \\ 0 & 1 & 0 & 1-2\gamma \\ 0 & 0 & 1 & \gamma \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \lambda c_{i-1}^k \\ \lambda c_i^k \\ \lambda c_{i+1}^k \\ \lambda c_i^{k+1} \end{pmatrix}$$

Hence, the code to be placed in the loop over $i = N-1, \ldots, 2$ reads

$$\begin{array}{rcl} \lambda c_{i-1}^k & = & \lambda c_{i-1}^k + \gamma \lambda c_i^{k+1} \\ \lambda c_i^k & = & \lambda c_i^k + (1-2\gamma)\lambda c_i^{k+1} \\ \lambda c_{i+1}^k & = & \lambda c_{i+1}^k + \gamma \lambda c_i^{k+1} \\ \lambda c_i^{k+1} & = & 0 \end{array}$$

Practical check of correctness of adjoint code

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

$$\mathbf{y} = \mathbf{F}(\mathbf{x}) \quad \mathbf{F} : \mathcal{R}^n \to \mathcal{R}^m, \qquad z = D(\mathbf{y}) \quad D : \mathcal{R}^m \to \mathcal{R}$$

$$\delta z = \nabla_y D \mathbf{F}' \delta \mathbf{x} = \langle (\nabla_y D)^T, \mathbf{H}\delta\mathbf{x} \rangle$$

$$= \langle \mathbf{H}^T (\nabla_y D)^T, \delta\mathbf{x} \rangle = \left((\mathbf{F}')^T (\nabla_y D)^T \right)^T \delta\mathbf{x}$$

Application of $\mathbf{H}$ and $\mathbf{H}^T$ 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 $\delta\mathbf{x}_i := (0, \ldots, 0, 1, 0, \ldots, x_n)^T$ the i^{th} row of $\mathbf{H}$ can be compared with the i^{th} column of $\mathbf{H}^T$. Further, the TL code linearized at input $\mathbf{x}$ be approximated by $\lim_{\delta x_i \to 0} \frac{1}{|\delta x_i|} \mathbf{H}((\mathbf{x} + \delta\mathbf{x}_i) - \mathbf{x})$, where $\delta\mathbf{x}_i = (0, \ldots, 0, \delta x_i, 0, \ldots, x_n)^T$.

Ensembe Transform Kalman Filter

Ensemble Kalman Filter

This section follows Hamill, Th., Ensemble nased atmospheric data assimilation"' in Palmer, T. and Hagedorn, R., eds., 2006.

ensemble matrix composed by ensemble members

$$\mathbf{X}^b := (\mathbf{x}_1^b, \ldots, \mathbf{x}_m^b) \tag{4.35}$$

ensemble mean

$$\bar{\mathbf{x}}^b := \frac{1}{m} \sum_{i=1}^{m} \mathbf{x}_i^b \tag{4.36}$$

The perturbation of the i^{th} member is $\mathbf{x'}_i^b := \mathbf{x}_i^b - \bar{\mathbf{x}}^b$. The matrix of the ensemble perturbations then reads

$$\mathbf{X'}^b := (\mathbf{x'}_1^b, \ldots, \mathbf{x'}_m^b) \tag{4.37}$$

Let $\tilde{\mathbf{P}}^b$ denote the ensemble estimate of the forecast error covariance matrix $\mathbf{P}^b$. This is then calculated by

$$\tilde{\mathbf{P}}^b = \frac{1}{m-1} \mathbf{X}'^b \mathbf{X}'^{bT} \tag{4.38}$$

The question how to generate $\mathbf{X}^b$ has not been considered so far. Several methods are available: Optimal perturbations are used for the ensemble transformed Kalman filter (EnTKF) (Bishop et al., 2001). Another option is to generate m perturbed observations. A stochastic update algorithm reads

$$\mathbf{x}_i^a = \mathbf{x}_i^b + \mathbf{K}_{i+1}(\mathbf{y}_i - H(\mathbf{x}_i^b), \tag{4.39}$$

we generate m sets of "'perturbed observations"'

$$\mathbf{y}_i = \mathbf{y} + \mathbf{y'}_i, \quad i = 1, \ldots, m, \tag{4.40}$$

where $\mathbf{y'}_i \propto N(0, \mathbf{R})$.

The elements of the Kalman gain matrix can then be calculated as follows

$$\overline{H(\mathbf{x}^b)} := \frac{1}{m} \sum_{i=1}^{m} H(\mathbf{x}_i^b) \tag{4.41}$$

$$\tilde{\mathbf{P}}^b \mathbf{H}^T := \frac{1}{m-1} \sum_{i=1}^{m} (\mathbf{x}_i^b - \bar{\mathbf{x}}^b)(H(\mathbf{x}_i^b) - \overline{H(\mathbf{x}^b)})^T \tag{4.42}$$

and

$$\mathbf{H}\tilde{\mathbf{P}}^b \mathbf{H}^T := \frac{1}{m-1} \sum_{i=1}^{m} (H(\mathbf{x}_i^b) - \overline{H(\mathbf{x}^b)})(H(\mathbf{x}_i^b) - \overline{H(\mathbf{x}^b)})^T \tag{4.43}$$

We obtain for the analysis mean

$$\bar{\mathbf{x}}^a(t_{i+1}) = \bar{\mathbf{x}}^b(t_{i+1}) + \mathbf{K}_{i+1}(\mathbf{y} - H(\bar{\mathbf{x}}_i^b), \tag{4.44}$$

and for the individual perturbations

$$\mathbf{x'}^a_{(t+1)} = \mathbf{x'}^b_{(t+1)} + \mathbf{K}_{t+1} H(\mathbf{x'}_i^b) \tag{4.45}$$

Processing of an ensemble Kalman filter

Sequence for an ensemble square root Kalman filter (EnSqRF) with serial processing

1. construct an ensemble of initial values
2. integrate all ensemble members forward in time until the next observation time (4.8)
3. perform update step
   (a) add a random model error $\breve{\eta}$ to the forecast $\breve{\mathbf{x}}^b_{i,(t+1)} = \mathbf{x}^b_{i,(t+1)} + \breve{\eta}$ for all available observations at time $(t+1)$
   (b) determine the forecast mean $\bar{\mathbf{x}}^b$ by (4.36) and the matrix of ensemble perturbations $\mathbf{X}'^b$ by (4.37),
   (c) determine (4.42) and (4.43)
   (d) determine (4.23)
   (e) calculate
4. Add the updated mean and the perturbations together to form the analysis ensemble
5. Go to step 2

CHAPTER 5

The Initialisation Problem

So far we have treated two information sources for data assimilation: observations and a priori information like forecasts, which provide background knowledge. Despite the fact that this set of information provides an overdetermined, (yet ill–posed) problem, it is still insufficient for the inference of initial values for the integration of numerical forecast models. Why is this the case? Theoretical meteorology teaches the solutions of the primitive equations, which is only amenable to numerical techniques. Upon simlification as far as shallow water equations on a sphere, Rossby (or rotational) modes are a family of solutions, which are well known to sustain the general flow pattern in nature. In addition, gravity modes figure as a further set of solutions. In a multilevel shallow water model, internal gravity modes are an additional family of solutions. In the case of non–hydrostatic models, sound waves are yet another set of solutions. While these gravity modes can be observed, they are neither relevant for weather, nor of such high amplitudes that they can be directly observed in the troposphere in the undisturbed atmosphere or ocean. (In the middle and upper atmosphere conditions are different). Rare occasions, in which gravity waves are observed, include volcano erutions, and tsunamis and seiches in oceans, seas and lakes. In most cases, *atmsopheric dynamics are captured by slow modes like Rossby waves*, and not by fast modes, like gravity waves. However, our data assimilation algorithms so far treated here do not account of the clear predominance of slow modes, as there are no constraints imposed to suppress the fast ones.

In the case of hydrostatic models, procedures were introduced to suppress high frequency waves, assuming a cut–off frequency as acceptable limit. Techniques used so far, were first the normal mode initialisation, later the non–linear normal mode initialisation. With the advent of 4D–dimensional variational data assimilation, high frequency waves were inherently hampered to

survive until the occurence of an observation, and the initialisation problem is mitigated. Therefore these techniques are not included in this lecture. However the initialisation problem prevails for non-hydrostatic models. An efficient technique is digital filtering, introduced by Lynch and Huang (1992).

Digital filtering

We consider a function of time $f(t)$, composed of low and high frequency components. The implementation of alow–passfiltes includes:

1. calculate the Fourier transform $F(\omega)$ of $f(t)$
2. set the high frequency components $= 0$,

$$H(\omega) := \begin{cases} 1, \text{if} |\omega| \leq 2|\omega_c| \\ 0, \text{if} |\omega| > |\omega_c| \end{cases} \tag{5.1}$$

3. retrieve the low pass filter function by inversion.

Let $h(t) = \sin(\omega_c t)/(\pi t)$ be the inverse Fourier transform of $H(\omega)$. Then the three steps are equivalent og the convolution of $f(t)$ with $h(t)$. Digression on convolution

Convolution theorem: The Fourier transform $\mathcal{F}(f * g)$ of a convolution product of two functions $(f * g)(x) = \int_{-infty}^{\infty} f(y)g(x-y)dy$ equals the product of the two Fourier transform of its factors $\mathcal{F}(f) = F(y), \mathcal{F}(g) = G(y)$ times $\sqrt{2\pi}$

$$\frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} (f * g)(x) \exp(ixy)dx = \sqrt{(2\pi)} F(y) \cdot G(y) \tag{5.2}$$

To filter $f(t)$ we have

$$f^*(t) = (h * f)(t) = \int_{-\infty}^{\infty} h(\tau) f(t-\tau) d\tau. \tag{5.3}$$

Given a time series $\{f_n\} := \{\ldots, f_{-2}, f_{-1}, f_0, f_1, f_2, \ldots\}$ at discrete moments $t_n = n\Delta t$, the shortest period which can be identified is $\tau_N = 2\Delta t$, which corresponds to the Nyquist frequency $\omega_N = \pi/\Delta t$. We can find the Fourier transform

$$F(\theta) = \sum_{n=-\infty}^{\infty} f_n \exp(-in\theta) \tag{5.4}$$

The Fourier transform of $H(\theta) = \sum_{n=-\infty}^{\infty} h_n \exp(-in\theta)$ is

$$h_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} H(\theta) \exp(in\theta) d\theta \tag{5.5}$$

The coefficients are

$$h_n = \frac{\sin n\theta_c}{n\pi}. \tag{5.6}$$

With $\{f_n^*\}$ denoting the interesting low frequency part of $\{f_n\}$,

$$H(\theta) \cdot F(\theta) = \sum_{n=-\infty}^{\infty} f_n^* \exp(-in\theta) \tag{5.7}$$

The convolution theorem implies that $H(\theta) \cdot F(\theta)$ is the convolution of $\{h_n\}$ with $\{f_n\}$

$$f_n^* = (h * f)_n = \sum_{n=-\infty}^{\infty} h_k f_{n-k}. \tag{5.8}$$

An approximation to the low frequency part reads

$$f_n^* = \sum_{k=-N}^{N} h_k f_{n-k} \tag{5.9}$$

To avoid Gibbs' oscillations, filter windows are introduced, like Dolph-Tchebychef or Lanczos window

$$w_n = \frac{\sin\left(n\pi/(N+1)\right)}{n\pi/(N+1)} \tag{5.10}$$

Transfer functions $T(\theta)$ of a filter are functions, which act as filter for pure sinusoidal waves upon multiplication, that is, for $f_n = \exp(in\theta)$ we have $f_n^* = T(\theta) \cdot f_n$

$$T(\theta) = \sum_{k=-N}^{N} h_k \exp(-ik\theta) = \left(h_0 + 2\sum_{k=1}^{N} h_k \cos(k\theta) \right) \tag{5.11}$$

In general we find formulations like equation (5.9) equivalent to non–rekursive digital filters

$$y_n^* = \sum_{k=-N}^{N} a_k x_{n-k} \tag{5.12}$$

More elaborate froms are rekursive digital filters

$$y_n^* = \sum_{k=0}^{N} a_k x_{n-k} + \sum_{k=0}^{M} b_k y_{n-k} \tag{5.13}$$

Application for initialisation

$$f_F^*(0) = 1/2 h_0 f_0 + \sum_{n=1}^{N} h_{-n} f_n \tag{5.14}$$

hindcast

$$f_B^*(0) = 1/2 h_0 f_0 + \sum_{n=-1}^{-N} h_{-n} f_n \tag{5.15}$$

Summation of both sums to find

$$f^*(0) = f_F^*(0) + f_B^*(0) \tag{5.16}$$

Filter coefficients

$$h_n = \frac{\sin\left(n\pi/(N+1)\right)}{n\pi/(N+1)} \left\{ \frac{\sin(n\theta_c)}{n\pi} \right\} \tag{5.17}$$

References

References

Armstrong, M., Problems with universal kriging. *Math. Geol.*, *16*, 101–108, 1984.

Barkmeijer, J., M. van Gijzen, and F. Bouttier, Singular vectors and estimates of the analysis error covariance matrix, Q. J. R. Met. Soc, 124, 1695-1713, 1998.

Barnes, M. G., The use of kriging for estimating the spatial distribution of radionuclides and other spatial phenomena. Batelle Memorial Institute, Pacific Northwest Laboratory, Richland, Washington, PNL–SA–9051, 1980.

Bayes, T., An assay towards solving a problem in the doctrine of chances, *Phil. Trans. Roy. Soc. London*, *53*, 370–418, 1763.

Bayes, T., An assay towards solving a problem in the doctrine of chances, *Phil. Trans. Roy. Soc. London*, *54*, 298–310, 1764.

Bengtson, L., Ghil, M., Källén, E., Dynamic meteorology: data assimilation methods, Springer–Verlag, New York, 1981.

Bennet, A., *Inverse methods in physical oceanography*, Cambridge University Press, pp. 346, 1992.

Bergthorsson, P., and B. R. Döös, Numerical weather map analysis, *Tellus*, *7*, 329–340, 1955.

Bott, A., A positive definite advection scheme obtained by nonlinear renormalization of advective fluxes. *Mon. Wea. Rev.*, *117*, 1006–1015, 1989.

Charney, J.G., M. Halem, R. Jastrow, Use of incomplete historical data to infer the present state of the atmosphere, *J. Atmos. Sci.*, *26*, 1160–1163, 1969.

Clark, T. L., Voldner, E. C., Olson, M. P., Seilkop, S. K., and Alvo, M., International sulfur deposition model evaluation (SDME). Report, 1986.

Courtier, Ph., and O. Talagrand, Variational assimilation of meteorological observations with the adjoint vorticity equation. Part II. Numerical results. Q. J. R. Meteorol. Soc.,, *113*, 1329–1347, 1987.

Courtier, Ph., J.–N. Thépaut, and A. Hollingthworth, A strategy for operational implementation of 4D–Var, using an incremental approach, Q. J. R.

Meteorol. Soc.,, *120*, 1367–1387, 1994.

Courtier, Ph., E. Andersson, W. Heckley, J. Pailleux, D. Vasiljevic, M. Hamrud, A. Hollingthworth, F. Rabier, and M. Fisher, The ECMWF implementation of the three dimensional variational assimilation (3D–Var). I: Formulation, Q. J. R. Meteorol. Soc.,, *124*, 1738–1807, 1998.

Daley, R., *Atmospheric Data Analysis*, Cambridge University Press, Cambridge, 1991.

Daley, R., The lagged innovation covariance: a performance diagnostic for atmospheric data assimilation. *Mon. Wea. Rev.*, *120*, 178–196, 1992a.

Daley, R., Forecast-error statistics for homogeneous and inhomogeneous observation networks. *Mon. Wea. Rev.*, *120*, 627–643, 1992b.

Daley, R., Estimating model–error covariances for application to atmospheric data assimilation. *Mon. Wea. Rev.*, *120*, 1735–1746, 1992c.

Dee, D. and A. da Silva, Data assimilation in the presence of forecast bias, *Quart. J. Roy. Met. Soc.*, *124*, 269–295, 1998.

Dennis, J, and J. More, Quasi-Newton methods, motivation and theory, SIAM review, 19, 46-89, 1977.

Elbern, H., Parallelization and load balancing of a comprehensive atmospheric chemistry transport model. *Atmos. Env.*, *31*, 3561–3574, 1997.

Elbern H., H. Schmidt, and A. Ebel; Variational data assimilation for tropospheric chemistry modeling. *J. Geophys. Res.*, *102*, D13, 15,967–15,985, 1997.

Elbern, H., and H. Schmidt, A four–dimensional variational chemistry data assimilation scheme for Eulerian chemistry transport modeling. *J. Geophys. Res.*, *104*, 18 583–18 598, 1999.

Elbern, H., and H. Schmidt, Ozone episode analysis by four–dimensional variational chemistry data assimilation, accepted for publication in *J. Geophys. Res.*, 2001.

Elbern, H., H. Schmidt, O. Talagrand, A. Ebel, 4D–variational data assimilation with an adjoint air quality model for emission analysis, *Environ. Mod. Softw.*, *15*, 539-548, 2000.

Eliassen, A., A provisional report on calculation of spatial covariance and autocorrelation of the pressure field, Inst. Weather Clim. Res., Rep. No. 5, Academy of Science, Oslo, (reprinted in Bengtson et al., (1981)), 1954.

Endlich, R. M., Eynon, B. P., Ferek, R. J., Valdes, A. D., and Maxwell, C., Statistical analysis of precipitation chemistry measurements over the ecosystem United States. UAPSP–112, EPRI, Palo Alto, California, 1986.

Fisher, M., and D. J. Lary, Lagrangian four-dimensional variational data assimilation of chemical species, *Q. J. R. Meteorol. Soc., 121*, 1681-1704, 1995.

Fedorov, V., Kriging and other estimators of spatial field characteristics (with special reference to environmental studies). *Atmos. Env.*, *23*, 174–184, 1989.

Gandin, L. S., Objective analysis of meteorological fields. Gidrometeorol. Izd., Leningrad, (English translation by Israel program for scientific translations, Jerusalem,) 1965.

Gauß, K.,F., *Theoria motus corporum coelestium*, english translation: Theory of motion of the heavenly bodies, Dover, New York, 1809.

Ghil, M., S. Cohn, J. Tavantzis, K. Bube, and E. Isaacson, Applications of estimation theory to numerical weather prediction. In *Dynamic meteorology: data assimilation methods*, Eds.: L. Bengtson, M. Ghil, and E. Kallen, Springer Verlag, New York, 139–224, 1981.

Ghil, M., and P. Malanotte–Rizzoli, Data assimilation in meteorology and oceanography. *Advances in Geophysics*, *33*, 141–266, 1991.

Golub, G., and J. M. Ortega, *Scientific Computing. An Introduction with Parallel Computing*, Academic Press, 1993.

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.

Jazwinski, A., *Stochastic Processes and Filtering Theory*, Academic Press, San Diego, pp. 376, 1970.

Journel, A. G., Huijbregts, C., *Mining Geostatistics*, Academic Press, NY, 1978.

Kalman, R. E., A new approach to linear filtering and prediction problems. *J. Basic Eng.*, *82D*, 35–45, 1960.

Kalman, R. E., and R. S. Bucy, New results in linear filtering and prediction theory. *ASME t J. Basic Eng.*, *83D*, 95–108, 1961.

Kirsch, A., *An introduction to the mathjematical theory of inverse problems*, Springer Verlag, 1996.

Kolmogorov, A., Interpolated and extrapolated stationary random sequences, *Izvestia SSSR, seriya mathem.*, *5*, 85–95, 1941.

Lanczos, C., 1950; An iteration method for the solution of the eigenvalue problem of linear differential and integral operators, *J. Res. Nat. Bur. Stand.*, 45, 225-282.

Le Dimet, F. X., O. Talagrand, Variational algorithms for the analysis and assimilation of meteorological observations: theoretical aspects, *Tellus*, *38A*, 97–110, 1986.

Lions, J.–L., *Optimal control of systems governed by partial differential equations*. Springer Verlag, Berlin, 1971.

Liu, D. C., and J. Nocedal, On the limited memory BFGS method for large scale minimization. *Math. Prog.*, *45*, 503–528, 1989.

Lorenc, A. C., Analysis methods for numerical weather predictions, *Q. J. R. Met. Soc.*, *112*, 1177–1194, 1986.

Lorenc, A. C., Optimal nonlinear objective analysis,*Q. J. R. Met. Soc.*, *114*, 205–240, 1988.

Lyster, P., S. Cohn, Menard, R., L.–P. Chang, and S.–J. Lin, Parallel implementation of a Kalman filter for constituent data assimilation, *Mon. Wea. Rev.*, *125*, 1674–1686, 1997.

Mahfouf, J.–F., Analysis of soil moisture from neart surface parameters: A feasibility study, *J. Appl. Meteorol.*, *30*, 1534–1547, 1991.

Marchuk, Y., *Numerical solution of the problems of the dynamics of the atmosphere and ocean*, in Russian, Gidrometeorol. Izd., Leningrad, 1974.

Menard, R., S. Cohn, L.–P. Chang, and P. Lyster, Assimilation of stratospheric chemical tracer observations usin a Kalman filter. Part I: Formulation, *Mon. Wea. Rev.*, *128*, 2654–2671, 2000.

Menard, R., L.–P. Chang, Assimilation of stratospheric chemical tracer observations usin a Kalman filter. Part II: χ^2–validated results and analysis of variances and correlation dynamics, *Mon. Wea. Rev.*, *128*, 2672–2686, 2000.

Menke, W., *Geophysical Data Analysis*, Academic Press, Orlando, 260 pp, 1984.

Mhamdi, A., and W. Marquardt, An inversion approach to the estimation of reaction rates in chemical reactors. In: *Proceedings of European control Conference*, Karlsruhe, 1999.

Panofsky, H, Objective weather map analysis, *J. Meteorol.*, *6*, 386–392, 1949.

Penenko, V., and N. N. Obraztsov, A variational initialization method for the fields of the meteorological elements, *Meteorol. Hydrol. (Engl. Transl.)*, *11*, 1-11, 1976.

Peters, L., C. Berkowitz, G. Carmichael, R. Easter, G. Fairweather, S. Ghan, J. Hales, R. Leung, W. Pennel, F. Potra, R. Saylor, and T. Tsang, The current state and future direction of Eulerian models in simulating the tropospheric chemistry and transport of trace species: a review. *Atmos. Env.*, *28*, 189–222, 1995.

Rabier, F. and Ph. Courtier, Four-dimensional assimilation in the presence of baroclinic instability, *Q. J. R. Met. Soc.*, *118*, 649-672, 1992.

Rhodin, A., F. Kucharski, U. Callies, D.P. Eppel, W. Wergen, Variational analysis of effective soil moisture from screen–level atmospheric parameters: applications to a short range forecast model, Q. J. R. Met. Soc., *125*, 2427–2448, 1999.

Robertson, L., and C. Persson, On the application of four dimensional data assimilation of air pollution data using the adjoint technique, in *Air Pollution Modeling and Its Application*, vol. 9, edited by H. van Dop and G. Kallos, 365–373, Plenum, New York, 1992.

Rutherford, I. D., Data assimilation by statistical interpolation of forecast error fields, *J. Atmos. Sci.*, **29**, 809–815, 1972.

Sadokov, V., and Schteinbok, D., Application of conjugate functions in analysis and forecasting of the temperatur anomaly, english translation, *Soviet Meteor. Hydrol.*, **10**, 16–21, 1977.

Salby, M. L., Sampling theory for asynoptic satellite observations, I: Space time spectra, resolution, and aliasing. *J. Atmos. Sci.*, **39**, 2577–2600, 1982a.

– , Sampling theory for asynoptic satellite observations, II: Fast Fourier synoptic mapping. *J. Atmos. Sci.*, **39**, 2601–2614, 1982b.

Smagorinsky, J., K. Miyakoda, R. Strickler, The relative importance of variables in initial conditions for dynamical weather prediction. *Tellus*, **122**, 141–157, 1970.

Stockwell, W. R., P. Middleton, J. S. Chang, and X. Tang, The second generation regional acid deposition model chemical mechanism for regional air quality modeling, *J. Geophys. Res., 95*, 16,343-16,367, 1990.

Tadjbakhsh, I., Utilization of time dependent data in running solution of initial value problems, *J. Appl. Meteor.*, **8**, 389–391, 1969.

Talagrand, O., A posteriori evaluation and verification of analysis and assimilation algorithms, in *Proceeding of the ECMWF seminar "Recent developments in numerical methods for atmospheric modelling"*, Reading, UK, 17–28, 1998.

Talagrand, O., and Ph. Courtier, Variational assimilation of meteorological observations with the adjoint vorticity equation. Part I. Theory, Q. J. R. Meteorol. Soc.,, **113**, 1321–1328, 1987.

van Loon, M., P. J. H. Builtjes, and A. J. Segers data assimilation of ozone in the atmospheric transport chemistry model LOTOS, *Environ. Mod. Softw.*, **15**, 603-609, 2000.

Wiener, N., *Extrapolation, interpolation and smoothing of stationary time series*, New York, J. Wiley, 1949.

Yanenko, N. N., *The method of fractional steps.*, Springer–Verlag, New York, pp. 160, 1971.

Zou, X., F. van den Berghe, M. Pondeca, Y.–H. Kuo, Introduction to adjoint techniques and the MM5 adjoint modelling system, NCAR/TN–435+STR,

1997.

Zou, X., Y.–H. Kuo, Y.–R. Guo, Assimilation of atmospheric refractivity using a nonhydrostatic mesoscale model, *Mon. Wea. Rev.*, **123**, 2229–2249, 1995.