Inverse problems and uncertainty quantification in remote sensing

Johanna Tamminen
Finnish Meterological Institute
johanna.tamminen@fmi.fi

ESA Earth Observation Summer School on
Earth System Monitoring and Modeling
July 30 – Aug 11, 2012, Frascati

30.7. 2012

Contents of the three lectures

- **Monday:** Introduction to inverse problems and uncertainty quantification in atmospheric remote sensing
- **Tuesday:** Introduction of Markov chain Monte Carlo method for estimating uncertainties
- **Wednesday:** More examples of how to characterize uncertainties in remote sensing (including modeling uncertainties)

Contents of this lecture

- Introduction to inverse problems in remote sensing
- Uncertainties and random variables
- Bayesian approach for solving inverse problems
- Example in atmospheric remote sensing: GOMOS

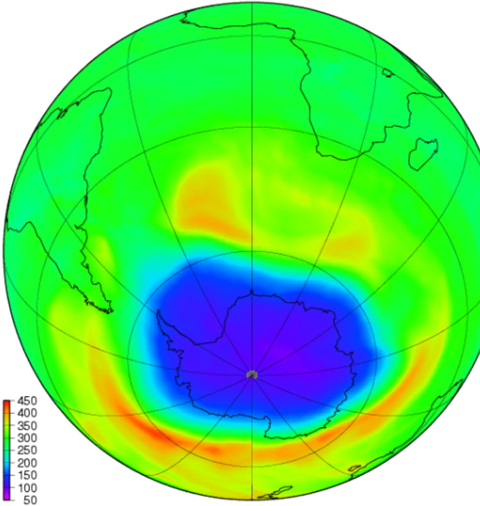
Satellite remote sensing

Satellite remote sensing has become an important way to monitor and study our environment.

- Land, vegetation, oceans, snow, ice, atmosphere, ...
- Global observations from pole to pole
- Continous observations
- Operational need for data: monitoring, forecasting, support in emergency situations, ...
- Support for monitoring the effects of international regulations, support for decision making ...
- Research: continous time series, combination of measurements, comparison with models ...

Satellite remote sensing and inverse problems

- Remote sensing measurements are typically non-direct measurements
- Data processing involves solving inverse problems (data retrieval)
- Phyiscal modeling of the measurements often complicated: typically nonlinear



Retrieval in practice

- Operative algorithms need to be fast, robust, reliable
- Often simplifications and assumptions are needed
- Typically also additional information needed for solving the problem (*ill posed* problem).
- **Important to characterize the uncertainties and validity of the simplifications and assumption – Uncertainty Quantification (UQ)**

Uncertainty Quantification (UQ)

- Uncertainty quantification: characterizing the errors and uncertainties
  reduction of the uncertainties, if possible

- UQ is becoming more and more important in environmental sciences.

- In remote sensing UQ is particularly important for:
  - Combining data from different sources
  - Assimilation
  - Comparing with models
  - Model discrepancy
  - Time series analysis, trends
  - Supporting decision making
  - Forecasting
- UQ growing area of research: theory, computational methods, simulations

This lecture series: UQ using Bayesian approach

- Bayesian formulation gives natural tools to characterize the impact of different errorr sources.
- Allowes including additional information to the retrieval in a natural way.
- In atmospheric remote sensing method called Optimal Estimation -algorithm by C. Rodgers used extensively which is based on Bayesian formulation

Forwad problem

$$y = f(x)$$

where

$y \in \mathbb{R}^m$ – unknown variable

$x \in \mathbb{R}^n$ – measurements, known paramteres

f – function describing the relationship between the unknown variable and known parameters

Inverse problem

$$y = f(x, b)$$

where

$y \in \mathbb{R}^m$ – measurements, observations, data

$x \in \mathbb{R}^n$ – unknown parameters (unknown state) which we are interested

f – function describing the relationship between the measurements and the unknown parameters (forward model)

b – known model parameters

Linear case - example

- Inverse problem: find x when y is measured and

$$y = f(x)$$

- When the dependence f is linear, we can describe the problem with matrix formulation:

$$y = Ax$$

where A is $m \times n$ matrix

- Simple solution would now be

$$x = A^{[-1]}y$$

where $A^{[-1]}$ would be 'in some sense' the inverse of matrix A.

- In practice it is often more complicated.

Well posed and ill-posed problems

Hadmard's definition (1902) of a well posed problem:

(i) The solution exists.

(ii) The solution is unique.

(iii) The solution hardly changes if the parameters (or initial conditions) are slightly changed.

If at least one of the criteria above is not fulfilled the problem is sayd to be ill-posed.

In remote sensing the problems are typically ill-posed as there is not enough data to result unique solutions. Therefore we need additional information to solve the problem.

Solving an ill-posed problem

Typcal ways of introducing additional information to inverse problems:

- Decreasing the size of the problem: **discretization**
- **Regularization** methods (eg. Tikhonov regularization: assuming smoothness)
- **Bayesian approach**: describing previous knowledge as *a priori* information

Measurement error, noise

- In practice measurements include always noise (ε)

$$y = f(x) + \varepsilon$$

(assuming that nose is additive)

- By repeating the measuremenets we get different answers

Measurements as random variables

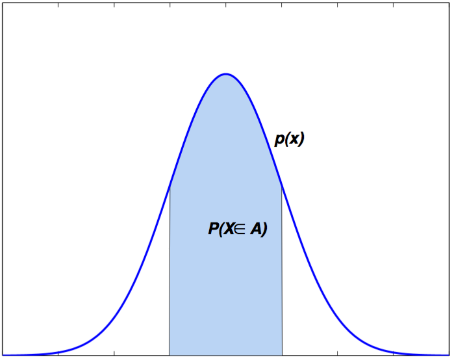
- It is natural to consider measurements as random variables.

- Intuitively Random variables get probable values often and less probable values only occasionally. These realizations form a distribution.

- Let X be a random variable. The **distribution** π of its realisations can be described as an integral π

$$\pi_X(A) = P\{X \in A\} = \int_A p(x)dx$$

where $p(x)$ is the **probablility density function (pdf)** of the random variable X.



Gaussian distributions:

- 1D Gaussian (Normal) distribution ($N(\mu, \sigma^2)$) with mean μ and covariance matrix C has pdf

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)$$

- Multivariate Gaussian distribution ($N(\mu, C)$) The n dimensional multivariate Gaussian distribution with mean μ and covariance matrix C has pdf

$$p(x) = (2\pi)^{-n/2} \left|C\right|^{-n/2} e^{-\frac{1}{2}(x-\mu)' C^{-1} (x-\mu)}$$

where C is $n \times n$ positive definite symmetric matrix.
If the components are independent, with $C = \sigma^2 I$, then the density simplifies to

$$p(x) = (2\pi)^{-n/2} \sigma^{-n} e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu_i}{\sigma}\right)^2}$$

Statistical inverse problem

- Because the measurements are random variables it is natural to consider also the unknown X as a random variable.

- Now the inverse problem is to search for the conditional distribution of X assuming that measurement $Y = y$:

$$\pi_{X\,|\,Y}(A) = P\{X \in A\,|\,Y = y\} = \int_A p(x\,|\,y)\,dx$$

where $p(x\,|\,y)$ is the pdf of the conditional distribution.

- The pdf $p(x\,|\,y)$ describes the probability that unknown $X = x$ when the measurement $Y = y$ This is what we are looking for!

Conditional probability

According to elementary probability calculation:

Joint probability that both events A and B take place

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

Now **conditional probability** is:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B \mid A)P(A)}{P(B)}$$

asuming that $P(B) \neq 0$.

Bayesian solution to an inverse problem

- The same idea of elementary probabity is used in Bayesian solution to an inverse problem.

- Bayes formula

$$p(x \mid y) = \frac{p_{\mathrm{lh}}(y \mid x)\, p_{\mathrm{pr}}(x)}{p(y)}$$

where

- $p(x \mid y)$ is the pdf of **a posteriori** distribution

- $p_{\mathrm{pr}}(x)$ is the pdf of **a priori** distribution. Describes prior knowledge of x.

- $p_{\mathrm{lh}}(y \mid x)$ is the pdf of the **likelihood** distribution. Characterizes the dependence of the measurements on the unknown.

- $p(y) \neq 0$ is a scaling factor (constant)

- The scaling factor is obtained by integrating over the state space:

$$p(y) = \int p_{lh}(y \mid x) p_{pr}(x) dx$$

- It is typically not considered, but we will come back to this later in the lectures.

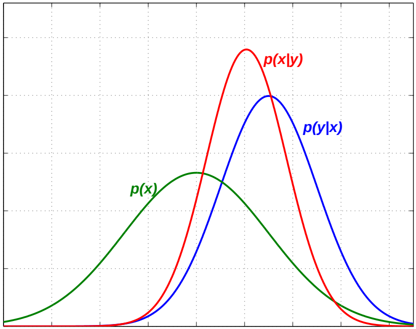
A posteriori distribution

- The Bayes formula

$$p(x \mid y) = \frac{p_{\mathrm{lh}}(y \mid x)\, p_{\mathrm{pr}}(x)}{p(y)}$$

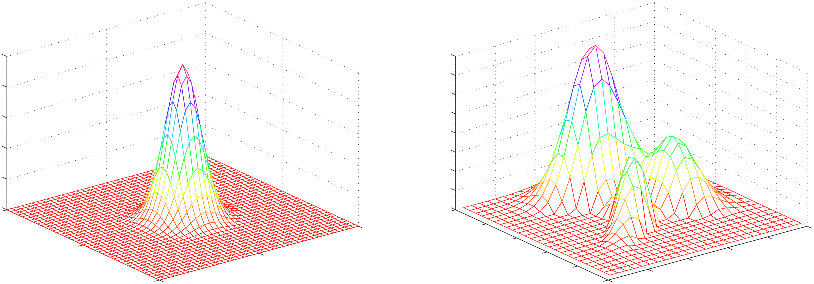
describes the solution of an inverse problem

Natural way for using all available information: how to combine new measurements with our old prior knowledge $(p(x) \longrightarrow p(x \mid y))$



- The solution is a distribution

Toy examples of 2-dimensional posterior distributions



How to characterize solution which is a distribution?

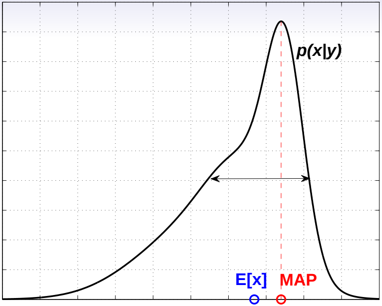
- Maximum a posteriori (MAP) estimate - most probable value

$$\text{MAP} = \widehat{x} = \underset{x}{\text{argmax}}\{p(x \mid y)\}$$

- Expectation

$$\overline{x} = \mathbb{E}_{X \mid Y}[x] = \int_{R^n} x \, p(x \mid y) \, dx$$

- The uncertainty of the estimates is described as the 'width' of the distribution



- Shape of the distribution
  kth moment: $\mathbb{E}_{X \mid Y}[(x - \overline{x})^k]$
  $k = 2$, variace
  $k = 3$, skewness
  $k = 4$, kurtosis

Linear inverse problem

Linear problem:

$$y = Ax + \varepsilon$$

- Assume Gaussian noise $\varepsilon \sim \mathrm{N}(0, C_y)$ and prior information $p_{\mathrm{pr}}(x) = \mathrm{N}(x_0, C_{x_0})$.
- In this special case the posterior distribution is also normally distributed

$$p(x \mid y) \propto e^{-\frac{1}{2}(x-\overline{x})^T Q (x-\overline{x})},$$

where the the expectation and the covariance matrix are:

$$\overline{x} = Q^{-1}(C_{x_0}^{-1} x_0 + A^T C_y^{-1} y), \quad Q^{-1} = (C_{x_0}^{-1} + A^T C_y^{-1} A)^{-1}.$$

- In this case the posterior estimate $\widehat{x} = \overline{x}$ and the posterior covariance matrix $C_{\widehat{x}} = Q^{-1}$ fully describe the posterior distribution.

MAP and ML estimate

$$p(x \mid y) \propto p(y \mid x)p(x)$$

- Assume 'non-informative' prior distribution $p(x) = c$

- Now maximum a posteriori (MAP) estimate is the same as Maximum likelihood estimate (ML)

$$\mathrm{MAP} = \operatorname*{argmax}_x \{p(x \mid y)\} = \operatorname*{argmax}_x \{p(y \mid x)\} = \mathrm{ML}$$

- In natural sciences 'non-informative' prior is often attractive as it allows solutions that are purely based on measurements.

- However, there are often non-physical solutions that shoud not be taken into account, like positivity.

Special case: Gaussian noise

- Assume additive, normally distributed measurement noise

$\varepsilon \sim \mathrm{N}(0, C_y)$

- The likelihood function

$$p_{\mathrm{lh}}(y \,|\, x) = \frac{1}{(2\pi)^{\frac{m}{2}}\sqrt{|C_y|}} \times \exp(-\frac{1}{2}(f(x) - y)^T C_y^{-1}(f(x) - y)).$$

- Assuming non-informative prior density the posterior

distribution is proportional to the likelihood function only:

$$p(x \,|\, y) \propto p(y \,|\, x),$$

and the MAP estimate $\hat{x}$ equals with ML estimate which further equals with the one that minimizes the **sum of squared residuals** function

$$\mathrm{SSR}(x) = (f(x) - y)^T C_y^{-1}(f(x) - y).$$

- This formula has been the basis for the traditional parameter

estimation, which concentrates simply on minimizing the SSR function (least squares solution).

In practice ...

- **Prior:** Typically, prior information is defined as Gaussian for simplifying computations.

- In **linear Gaussian** case the problem reduces to simply solving weighted least squares problem.

- **Non-linear and/or non-Gaussian** problems. Typically assumed that solution is 'close to Gaussian'.

Different techniques to solve nonlinear/non-Gaussian problems

Linearization and Gaussian assumption:

- Linearize the problem. If noise after linarization is close to Gaussian then linear Gaussian theory and simple matrix inversions can be applied (assuming that prior is also Gaussian).

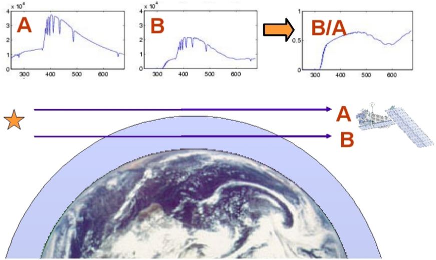
Non-linear optimization - seaching for maximum a posteriori (MAP) or maximum likelihood (ML) estimate:

- Iterative methods to search for MAP (ML) and assume linearity around estimate. Approximate uncertainty with covariance matrix. Some commonly used methods:
  - Levenberg-Marquardt iterative algorithm.
    (*See eg. Press et al. Numerical Recipes. The art of scientific computing*).
    - Combination of steepst descent (gradient) and Newtonian iteration (approximation with quadratic function).
    - Ready made algorithms are available.
    - MAP (ML) estimate $\hat{x}$ is computed. Posterior is assumed to be Gaussian close to estimate $\hat{x}$ and posterior covariance matrix $C_{\hat{x}}$ is computed.
  - In atmospheric research typically used iteratively method called Optimal estimation which is based on Bayes theorem.
    (*See Rodgers 2001: Inverse methods for atmospheric sounding*)

GOMOS/Envisat: Stellar occultation instrument

GOMOS - Global Ozone Monitoring by Occultation of Stars

- One of the three atmospheric instruments on-board ESA's Envisat satellite
- Launched 2002
- Measurements till April 2012 when Envisat lost connection to Earth.



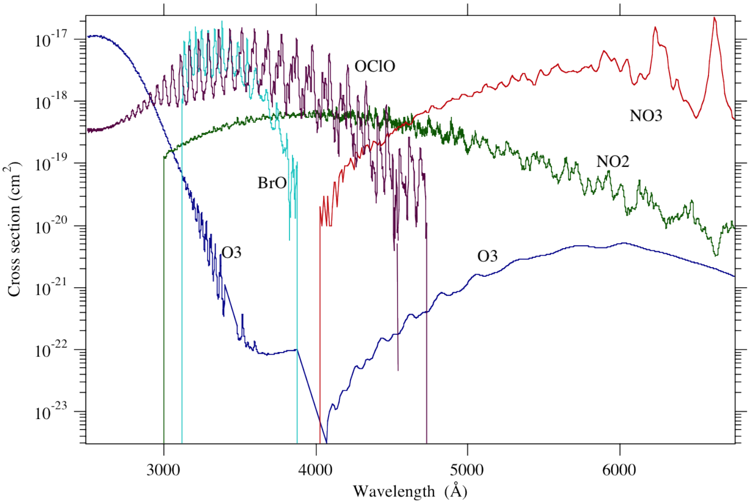
'Self-calibrating'

$$T_{\lambda,\ell} = \frac{I_{\lambda,\ell}}{I_{\lambda}^{\mathrm{star}}}$$

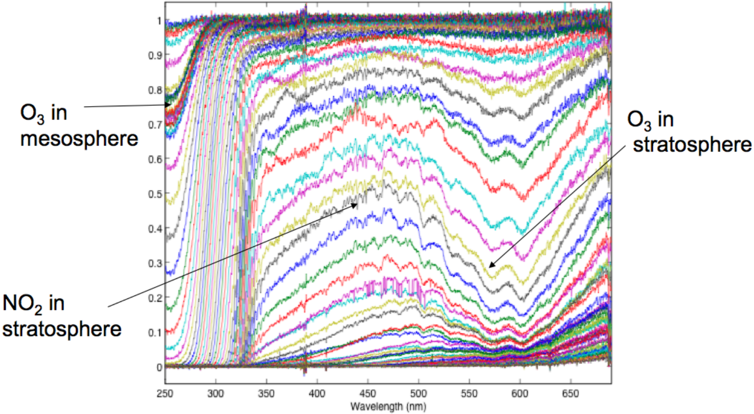
Measurement principle

- 'Fingerprints' of atmospheric gases in transmission spectra
- 20–40 stellar occultations/orbit.
- 50–100 ray path measurements/star from 100 km down to 10 km.

Cross sections of gases relevant to GOMOS (UV-VIS)

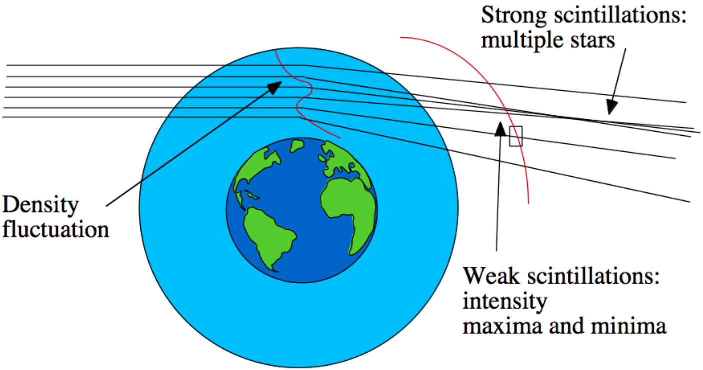


Absorption by gases



Transmission spectra measured by GOMOS at descinding altitudes from 100 km down to 5 km.

Refraction


Strong scintillations:
multiple stars
Density
fluctuation
Weak scintillations:
intensity
maxima and minima

Modelling

- The transmission at wavelength λ, along the ray path ℓ, includes $T_{\lambda,\ell}^{\text{abs}}$ due to absorption and scattering by gases and $T_{\lambda,\ell}^{\text{ref}}$ due to refractive attenuation and scintillation.
- $T_{\lambda,\ell}^{\text{abs}}$ is given by the Beer's law,

$$T_{\lambda,\ell}^{\text{abs}} = \exp\left[-\int_{\ell} \sum_{\text{gas}} \alpha_{\lambda}^{\text{gas}}(z(s))\rho^{\text{gas}}(z(s))ds \right]$$

where the integral is over the ray path ℓ.

The temperature dependent cross sections α^{gas} are assumed to be known from laboratory measurements.

The inversion problem is to estimate the gas profiles $\rho^{\text{gas}}(z)$ from the measurements

$$y_{\lambda,\ell} = T_{\lambda,\ell}^{\text{abs}} T_{\lambda,\ell}^{\text{ref}} + \epsilon_{\lambda,\ell}.$$

Operational algorithm: Assumptions

Model:

$$T^{\text{abs}}_{\lambda,\ell}(\rho) = \exp\left[-\int_{\ell} \sum_{\text{gas}} \alpha^{\text{gas}}_{\lambda}(z(s))\rho^{\text{gas}}(z(s))ds\right]$$

From now on we consider data to be simply:

$$y_{\lambda,\ell} = T^{\text{abs}}_{\lambda,\ell} + \epsilon_{\lambda,\ell}.$$

where $\lambda = \lambda_1, \ldots, \lambda_\Lambda, \quad \ell = \ell_1, \ldots \ell_M$

- Noise Gaussian, uncorrelated between different altitudes and wavelengths.
- $T^{\text{ref}}_{\lambda,\ell}$, scintillation and dilution, obtained from separate measurements (scintillation/turbulence effects not fully corrected).
- Temperature dependence of cross–sections can be modelled with 'effective' cross sections (only wavelength dependence)
- Temperature obtained from ECMWF

Operational two step algorithm

Since the cross sections are assumed constant on each ray path and noise uncorrelated, the inversion separates into

$$T^{\mathrm{abs}}_{\lambda,\ell} = \exp\left[-\sum_{\mathrm{gas}} \alpha^{\mathrm{gas}}_{\lambda,\ell} N^{\mathrm{gas}}_{\ell}\right], \quad \lambda = \lambda_1, \dots, \lambda_\Lambda,$$

with $N^{\mathrm{gas}}_{\ell} = \int_{\ell} \rho^{\mathrm{gas}}(z(s))ds, \quad \ell = \ell_1, \dots, \ell_M.$

- Two step approach:
  - **Spectral inversion** - retrieval of horizontally integrated densities of several constituents but each altitude separately
  - **Vertical inversion** - retrieval of full profile for each constituent separately

Spectral inversion

The line density vector $N_\ell = (N_\ell^{\mathrm{gas}})$, $\mathrm{gas} = 1, \ldots, n_{\mathrm{gas}}$ with the posterior density

$$P(N_\ell|y_\ell) \propto e^{-\frac{1}{2}(T_\ell(N_\ell)-y_\ell)C_\ell^{-1}(T_\ell(N_\ell)-y_\ell)}p(N_\ell),$$

is fitted to the spectral data $y_\ell = (y_{\lambda,\ell})$, $C_\ell = \mathrm{diag}(\sigma^2_{\lambda,\ell})$, $\lambda = 1, \ldots, \Lambda$, separately for each ray path ℓ.

- Prior: Fixed prior for neutral density from ECMWF. For gases and aerosols non-informative prior

- Non-linear problem solved iterativeely using Levenber-Marquard algorithm.
- MAP (=ML) estimate $\widehat{N_\ell}$ obtained with uncertainty described by its covariance matrix $C_{\widehat{N_\ell}}$ for all ray paths (altitudes) ℓ

Pointestimate demo

BLUE - GOMOS
measurement
RED - GOMOS iterative fit