Maximum likelihood estimation

Given two temperature estimates, T_1 and T_2, valid at same location and same time. Assume temperatures are drawn from Gaussian probability distributions, with uncertainties σ_1 and σ_2. The probability of the true temperature being T is,

$$P(T|T_1, T_2, \sigma_1, \sigma_2) = \frac{1}{\sqrt{2\sigma_1^2}} \exp(-\frac{(T - T_1)^2}{2\sigma_1^2}) \frac{1}{\sqrt{2\sigma_2^2}} \exp(-\frac{(T - T_2)^2}{2\sigma_2^2}) \quad (1)$$

$$\log P = const - \frac{1}{2}(\frac{(T - T_1)^2}{\sigma_1^2} + \frac{(T - T_2)^2}{\sigma_2^2}) \quad (2)$$

Differentiate and set to zero to find most likely value of true T, the so-called **maximum likelihood estimate**,

$$T = T_1 \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} + T_2 \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \quad (3)$$

$$= T_1 + (T_2 - T_1) \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \quad (4)$$

3 underlines it is a symmetrical combination of two values, with weights depending on their uncertainty. 4 shows it as a "correction" of one estimates, with the correction term depending the "distance" and the uncertainties.

In meteorological data assimilation T_1 would be a measurement, T_2 would be a model estimate. σ_2 would be the uncertainty of the model. σ_1 would include both the uncertainty of the observation and the *error of representativeness*.

The most likely value for the uncertainty of the temperature estimate:

$$\frac{1}{\sigma^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \quad (5)$$

σ is *always* smaller than both σ_1 and σ_2!

3 dimensional variational data assimilation

In 3DVar the most likely model state, given the observations and an old forecast state, is found by minimising a cost function.

$$J = J_1 + J_2 = \frac{1}{2}\frac{(T - T_1)^2}{\sigma_1^2} + \frac{1}{2}\frac{(T - T_2)^2}{\sigma_2^2}. \tag{6}$$

This is generalised to a situation with many observations, observations of different types, and information coming also from a NWP model forecast,

$$\begin{aligned}
J &= J_b + J_o &(7)\\
&= \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathrm{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + \frac{1}{2}(H(\mathbf{x}) - \mathbf{y})^T \mathrm{R}^{-1}(H(\mathbf{x}) - \mathbf{y}) &(8)\\
&= \frac{1}{2}(\delta\mathbf{x})^T \mathrm{B}^{-1}\delta\mathbf{x} + \frac{1}{2}(\mathrm{H}\delta\mathbf{x} + H(\mathbf{x}_b) - \mathbf{y})^T \mathrm{R}^{-1}(\mathrm{H}\delta\mathbf{x} + H(\mathbf{x}_b) - \mathbf{y}) &(9)
\end{aligned}$$

The vector $\delta\mathbf{x}$, which minimises the cost function J, is the analysis increment, $\delta\mathbf{x}_a$. The analysis itself is then $\mathbf{x}_a = \mathbf{x}_b + \delta\mathbf{x}_a$. The first term, J_b measures the deviation between the first guess model state and the trial state, $\mathbf{x}$, weighted by the statistical errors of the nwp model. The second term, J_o, measures the deviation between the observations and the prediction of the observations given trial state x, weighted by the statistical errors of the observations, including errors of representativeness.

In meteorology the *errors of the observations* are in general assumed *uncorrelated*. That corresponds to R being a diagonal matrix. The errors of the model are heavily correlated. Since the dimensions of B are enormous, being smart about B is essential. This is a big difference between NWP data assimilation and general data assimilation. At DMI B is found by looking at the differences in O-B offset statistics for different forecast lengths, and projecting it on a set of functions that enable inversion of B.

In principle the minimum of the cost function could be found by solving,

$$\nabla J = \mathrm{B}^{-1}\delta\mathbf{x} + \mathrm{H}^T\mathrm{R}^{-1}(\mathrm{H}\delta\mathbf{x} + H(\mathbf{x}_b) - \mathbf{y}) = 0 \qquad (10)$$

In practise an approximation to $\delta\mathbf{x}_a$ is found by iteration, using the expression for the gradient and a minimisation algorithm until "proper" convergence of the cost function is obtained.

4 dimensional variational data assimilation

In 3DVar the observations are compared to a first guess model state valid the time of the analysis. The observations included in the data analysis come from a time interval centred on the time of the analysis. If a property is measured frequently at a given observing site the observation valid closest in time to the time of the analysis is selected for the data analysis, the remainder are neglected. Key meteorological observations, like radiosondings, are performed simultaneously, in line with the time of the data analyses. However, other important observations, in particular observations performed from non stationary satellites may appear at any time during the data assimilation time window. Using them as if the were obtained at the time of the analysis will introduce errors due to the evolution of the atmospheric state during the time between performing the observation and the analysis.

This can be mitigated somewhat in the 3DVar scheme, by using in equation $\mathrm{H}\delta\mathbf{x}+H(\mathbf{x_{bi}})-\mathbf{y}_i$, where $\mathbf{x}_{bi}$ is the model forecast for time i, operating, for example with forecasts separated by 1 hour through the data analysis time window, rather than just $\mathbf{x}_b$ valid at the center of the time window. This method can be considered an intermediate step between 3DVar and 4DVar.

In real 4DVar the observations are compared to the model prediction at (or close to) the time of the observations. The cost function now takes the form,

$$J = J_b + J_o \tag{11}$$

$$\frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T\mathrm{B}^{-1}(\mathbf{x}-\mathbf{x}_b)+\frac{1}{2}(H_i(\mathbf{x}_i)-\mathbf{y}_i)^T\mathrm{R_i}^{-1}(H_i(\mathbf{x}_i)-\mathbf{y}_i) \tag{12}$$

where $\mathbf{x}$ is the state vector at time t_0 and $\mathbf{x}_i$ the state vector at time t_i as forecast by the model, when starting the model from

$\mathbf{x}$ at t_0.

For the *forecast operator* we have:

$$\mathbf{x}_i = M_{0\to i}(\mathbf{x}) = M_i M_{i-1}...(M_1(\mathbf{x})), \quad \text{where} \ \ \mathbf{x}_i = M_i(\mathbf{x}_{i-1}) \tag{13}$$

Assume that for realistic deviations, $\delta\mathbf{x} = \mathbf{x} - \mathbf{x}_b$, the following holds:

$$H_i(M_{0\to i}(\mathbf{x})) - \mathbf{y}_i \approx \mathrm{H_i M}_{0\to i}\delta\mathbf{x} + H_i(M_{0\to i}(\mathbf{x}_b)) - \mathbf{y}_i, \tag{14}$$

where M is the *tangent linear model*, corresponding to the differential of the forecast model M, and H_i is the observation operator for the observations at time t_i and $\mathrm{H_i}$ is the tangent linear of that. Then,

$$\nabla J_o = \sum_{i=0}^{n} \mathrm{M}_1^T...\mathrm{M}_i^T \mathrm{R}_i^{-1}(H_i(\mathbf{x}_i) - \mathbf{y}_i) \tag{15}$$

$$= \mathrm{H}_0^T d_0 + \mathrm{M}_1^T\ [\mathrm{H}_1^T d_1\ +\ \mathrm{M}_2^T\ [\mathrm{H}_2^T d_2\ +\ ...\ +\ \mathrm{M}_n^T \mathrm{H}_n^T d_n\ ]\ ...\ ], \tag{16}$$

$$\text{where} \quad d_i = \mathrm{R}_i^{-1}(H_i(\mathbf{x}_i) - \mathbf{y}_i) \tag{17}$$

M_i^T is called the *adjoint* model and $\mathrm{H_i}^T$ the adjoint forecast operator.

Again the minimum of J with respect to $\delta\mathbf{x}$ defines the analysis increment.

Both 3 and 4DVar are based on static error covariance matrices. More advanced assimilation, such as Kalman filtering, enable dynamic error estimation of the error correlations, exit. Attempts are being made to include part of the benefits from Kalman filtering in NWP data assimilation, but the dimension of the problem is to large to enable a genuine Kalman filtering approach.

Nudging

Instead of doing a statistical analysis, a fictive term forcing the model towards an observation in introduced in the set of differential equation.

Example, the standard continuity equation,

$$\frac{d\rho}{dt} = \rho \nabla \cdot \mathbf{V} x \tag{18}$$

can be expanded into,

$$\frac{d\rho}{dt} = \rho(\nabla \cdot \mathbf{V} + \epsilon) \tag{19}$$

This is used at DMI, to include precipitation information (from the DMI radar network) and cloud information (from geostationary satellites) in the nowcasting system we are developing,

$$\epsilon = \frac{\text{precip}_{\text{radar}} - \text{precip}_{\text{nwp}}}{\tau} \tag{20}$$

τ is a timescale determining the strength of the fictive forcing. Done in such a way that mass is added low in the column and removed high, to keep it constant. Leading to a lift, formation of precipitation, and stronger convergence at the bottum.

While it is complicated and costly to build a 3 or 4DVar data assimilation system, it is in certain cases very quick to introduce nudging terms. On the other hand the stringent statistical approach used in Var is laking in nudging, and it can be difficult to utilise observations that are not directly related to the model variables.