TUNING OF
OBSERVATION ERROR PARAMETERS
IN A VARIATIONAL DATA ASSIMILATION

G. Desroziers*, B. Chapnik*, S. Ivanov*, F. Rabier*
and O. Talagrand**

*Météo-France, CNRM, Toulouse
**Ecole Normale Supérieure, LMD, Paris
Introduction

- Assimilation schemes used for Weather Prediction basically rely on linear estimation theory or on an extension of this formalism.
- In such an approach, the final analysis is very dependent on the specification of the relative weights given to each source of information through the error covariances.
- Observation errors are not perfectly known: a large potential for improvement on analyses is offered by methods allowing their tuning.
- Large operational centers are now using or have planned to use assimilation schemes based on a 3D or a 4D variational approach.
- Diagnoses based on statistics of departures between information (including background) and the minimizing solution have been proposed, that can be applied in a variational framework.
- We present a method, based on those diagnostics, that aims to perform a diagnostics and tuning of information errors from a given set of observations and background fields.
Contents

- The French Weather Prediction System
- A posteriori diagnostics
- Derivation of a method to tune observation error parameters
- Test of the tuning procedure
- Relationship with Maximum Likelihood
- Discussion and conclusion
The French Arpège Weather Prediction System

- Unique global spectral model with a stretched geometry.
- Roughly 20 km resolution near the pole of interest.
- 41 levels in the vertical.
- Prognostic variables: temperature, specific humidity, vorticity, divergence + surface pressure.
Variational formulation

3D/4D variational (3D/4D-Var) incremental algorithms:

- \( x^a = x^b + \delta x \)
- one seeks the increment \( \delta x \) that minimizes

\[
J(\delta x) = J^b(\delta x) + J^o(\delta x) \\
= \frac{1}{2} \delta x^T B^{-1} \delta x \\
+ \frac{1}{2} (d - H \delta x)^T R^{-1} (d - H \delta x).
\]

\( J^b \) : background term, with \( B \) the covariance matrix of forecast errors.

\( H \) : linearized observation operator,
\( R \) : covariance matrix of observation errors,
\( d = y^o - H(x^b) \) : innovation vector.

Solution is given by

\[
\delta x^a = Kd = K(y - H(x^b)),
\]

where \( K \) is the gain matrix.

\( K \) is not explicitly determined.

⇒ Operational system based on a 4D-Var with a 6h assimilation period (4 assimilations per day).
Observations

Radiosondes

\begin{center}
\includegraphics[width=\textwidth]{radiosonde_map}
\end{center}

\begin{center}
\textit{NOAA radiances}
\end{center}

\begin{center}
\includegraphics[width=\textwidth]{noaa_radiance_map}
\end{center}

\approx 1M data / 6-hour assimilation period.

\approx 200 000 after quality control and ”thinning”.
Statistical expectation of $J$ terms

Measurements $y^o$, of dimension $p$
background $x^b$, of dimension $n$

→ information vector $z$ (Talagrand 1997):

$$z = \begin{cases} \begin{array}{c} I_{n \times n} \quad (x^t) + \begin{pmatrix} \epsilon^b \\ \epsilon^y \end{pmatrix} \\ H \end{array} \end{cases} = \Gamma(x^t) + \epsilon,$$

where $x^t$ is the truth, $\Gamma$ the full observation operator and $\epsilon$ the vector of errors.

$J_j$ term of $J$, sum of $m_j$ elements, then

$$E[J_j(x^a)] = \frac{1}{2}[m_j - Tr(\Gamma_j^T S_j^{-1} \Gamma_j P^a)],$$

with $\Gamma_j$ the linearized observation operator and $S_j$ the associated observation error covariance matrix, $P^a$ the full estimation error covariance matrix.

In particular

$$\begin{cases} E(J^b(x^a)) = \frac{1}{2}Tr(KH) \\ E(J^o(x^a)) = \frac{1}{2}[p - Tr(HK)] \end{cases}$$

and thus

$$E(J(x^a)) = E(J^b(x^a)) + E(J^o(x^a)) = \frac{p}{2}. $$
Randomized estimation of

\[ Tr(HK) \]

If \( \delta y^o = \mathcal{N}(0, I_{p \times p}) \), then

\[
\delta y^{oT} HK \delta y^o = Tr(HK \delta y^o \delta y^{oT}) \approx Tr(HK).
\]

(Girard, 1987).

On the other hand,

\[ H \delta x(y^o + \delta y^o) - H \delta x(y^o) = HK \delta y^o \]

and then

\[
\delta y^{oT} [H \delta x(y^o + \delta y^o) - H \delta x(y^o)] \approx \delta y^{oT} HK \delta y^o \\
\approx Tr(HK).
\]

\( \Rightarrow Tr(HK) \) can be obtained with 2 analyses with perturbed and un-perturbed observations.
Randomized estimation of sub-parts of $J^o$ or $J^b$

- If $J^o_j$ is a term of $J^o$ with $p_j$ observations, then

$$E(J^o_j) = \frac{1}{2} [p_j - \text{Tr}(\Pi_j (HK) \Pi_j^T)],$$

where $\Pi_j$ is a projection operator: $y - \Pi_j \rightarrow \{y\}_j$.

An estimator of $E(J^o_j)$ can be obtained: $E(J^o_j) \simeq 1/2[p_j - (\Pi_j \delta y^o)^T \Pi_j (H \delta x(y^o + \delta y^o) - H \delta x(y^o))].$

$\Rightarrow E(J^o_j)$ can be obtained for an ensemble of sub-parts of $J^o$ with only 2 analyses with perturbed and un-perturbed observations.

- Similar expression for parts of $J^b$, but with a perturbation of the background fields:

$$E(J^b_j) \simeq - (\Pi_j \delta x^b)^T \Pi_j [\delta x(x^b + \delta x^b) - \delta x(x^b)].$$

$\Rightarrow E(J^b_j)$ can be obtained for an ensemble of sub-parts of $J^b$ with 2 analyses with perturbed and un-perturbed background.
Tuning
background and observation errors

New expression of the cost function:

\[ J(\delta \mathbf{x}) = \sum_j \frac{1}{s_j^b} J_j^b(\delta \mathbf{x}) + \sum_j \frac{1}{s_j^o} J_j^o(\delta \mathbf{x}). \]

Rationale:

find the set of parameters \((s_j^b)^2\) and \((s_j^o)^2\) such as

\[
\begin{align*}
  s_j^b &= 2J_j^b(\delta \mathbf{x})/E(J_j^b) \\
  &= 2J_j^b(\delta \mathbf{x})/\text{Tr}(\Pi_j(KH)\Pi_j^T) \\
  s_j^o &= 2J_j^o(\delta \mathbf{x})/E(J_j^o) \\
  &= 2J_j^o(\delta \mathbf{x})/[p_j - \text{Tr}(\Pi_j(HK)\Pi_j^T)].
\end{align*}
\]

Use of a fixed-point algorithm.
Test in the French Arpège 3D-Var

Simulation of background and observations with known error covariances.

The pseudo-truth $x^t$ is given by a 6h Arpège forecast.

**Background:** $\epsilon^b = B^{1/2} \zeta^b$, where $\zeta^b = \mathcal{N}(0, I_{n \times n})$.

**Observations:** evaluations of $x^t$ at actual TEMP locations ($H(x^t)$), perturbed with operational observation error profiles (for geopotential, temperature, wind and specific humidity).
Optimization of observation error profiles for TEMP

Observation error profiles: (solid) true; (dotted) imposed at the beginning of the iterative procedure; (dashdot) after 1 iteration; (dashed) after 5 iterations.
Observation error profiles: (solid) true; (dotted) imposed at the beginning of the iterative procedure; (dashdot) after 1 iteration; (dashed) after 5 iterations.
Application to
NOAA TOVS radiances

\[ R_\alpha = e B_\alpha(T_0) \tau_\alpha(p_0) \]
\[ + \int_{p_0}^{0} B_\alpha(T(p)) \frac{\delta \tau_\alpha}{\delta p} dp \]
\[ + (1 - e) \int_{0}^{p_0} B_\alpha(T(p)) \frac{\delta \tau^*_\alpha}{\delta p} dp, \]

where
\[ \alpha \] is the frequency,
\[ p \] is the pressure,
\[ p_0 \] is the surface pressure,
\[ T \] is the temperature,
\[ T_0 \] is the surface temperature,
\[ e \] is the surface emissivity,
\[ B_\alpha(T(p)) \] is the Planck function,
\[ \tau_\alpha \] is the transmittance from pressure \( p \) to space,
\[ \tau^*_\alpha \] is the transmittance from pressure \( p \) down to the surface and then to space.
TOVS weighting functions \( \frac{\partial \tau_{\text{OLR}}}{\partial \ln p} \) (taken from Smith et al 1979).
Tuning of observation errors for TOVS brightness temperatures (real dataset) (NOAA14 / 21 February 1997) (Clear observations)
Tuning of observation errors for **TOVS** brightness temperatures

*(real dataset)*

*(NOAA14 / 21 February 1997)*

*(Partly cloudy observations)*

---

**Tuning coefficients** of operational errors as determined by the iterative procedure

(r00: 00Z; r06: 06Z; r12: 12Z; r18: 18Z).
Relationship with
Maximum-likelihood estimation

We have

\[ d = y^o - H(x^b) \]
\[ = y^o - H(x^t) + H(x^t) - H(x^b) \]
\[ \approx \epsilon^o - H\epsilon^b. \]

Then, the covariance of the innovation vector \( d \) can be written

\[
E(dd^T) = R^t + HBH^T
\]
\[ = s^o R^t + s^b HB^T H^T \]
\[ = D(s^o, s^b). \]

Parameters \( s^o \) and \( s^b \) can be determined by maximizing the likelihood that the actual data arise from this model.

Equivalently, this can be obtained by minimizing the negative log-likelihood:

\[
\mathcal{L}(s^o, s^b) = \log(\text{det}(D)) + d^T D^{-1} d.
\]

This can also be written

\[
\mathcal{L}(s^o, s^b) = \text{Tr}(\log(D)) + d^T D^{-1} d.
\]
The optimal values of $s^o$ and $s^b$ are found for
\[
\frac{\partial}{\partial s^o} \mathcal{L}(s^o, s^b) = \frac{\partial}{\partial s^o} Tr(Log(D)) + \frac{\partial}{\partial s^o}(d^T D^{-1} d)
\]
\[
= Tr(RD^{-1}) - d^T D^{-1} RD^{-1} d
\]
\[
= \frac{1}{s^o}(p - Tr(HK)) - \frac{2}{(s^o)^2} J^o(x^a)
\]
\[
= \frac{1}{s^o}[(p - Tr(HK)) - \frac{2}{s^o} J^o(x^a)]
\]
\[
= 0.
\]

and
\[
\frac{\partial}{\partial s^b} \mathcal{L}(s^o, s^b) = \frac{\partial}{\partial s^b} Tr(Log(D)) + \frac{\partial}{\partial s^b}(d^T D^{-1} d)
\]
\[
= \frac{1}{s^b}[Tr(KH) - \frac{2}{s^b} J^b(x^a)]
\]
\[
= 0.
\]

Thus, the procedure introduced above solves
\[
\begin{align*}
\frac{\partial}{\partial s^o} \mathcal{L}(s^o, s^b) &= 0 \\
\frac{\partial}{\partial s^b} \mathcal{L}(s^o, s^b) &= 0.
\end{align*}
\]
Discussion

- The tuning procedure proposed here is equivalent to the Maximum Likelihood approach for determining error variances.
- It avoids the minimization of the log-likelihood function and in particular the computation of $Tr(\log(D))$.
- Quality of the estimators:
  - this method is based on a statistics (on a single realization) of the innovation vector.
  - a sufficiently large number of observations is needed,
  - $HBH^T$ must be different from $R$,
  - role of the lengthscales of the background error covariances found in $B$ and of the form of the observation operator.
  - impact of a mis-specification of matrix $B$, if only observation errors are tuned?
Conclusion

• A practical computation of the statistical expectation of sub-parts of the cost function is feasible for the cost of an additional analysis with perturbed observations or perturbed background.

• This allows the computation of a posteriori diagnostics of the consistency of the analysis scheme.

• It also permits the adaptive tuning of observation (or background) errors from a single (or possibly a set) of innovation vector(s).

• Results are very good with simulated background and observations.

• A first application to a real satellite dataset is also encouraging: results have, in particular, shown a good time consistency.

• This method is equivalent to the Maximum Likelihood approach for the determination of error variances, but with a lower cost and an easier implementation.