Interpretation of two error statistics estimation methods: 1 - the Derozier's method 2 – the NMC method (lagged forecast)

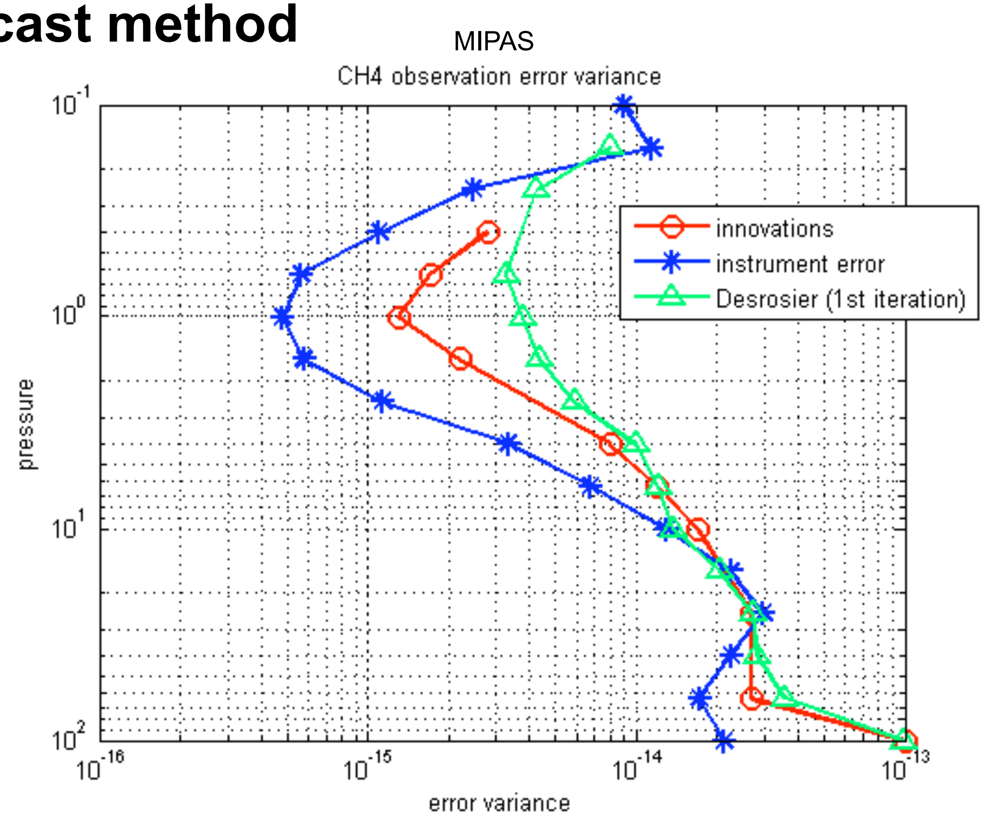
Richard Ménard, Yan Yang and Yves Rochon
Air Quality Research Division
Environment Canada

SPARC Data Assimilation Workshop
Exeter , June 21-23 2010

Many error statistics estimation approaches

- **Observation-space based methods**
  - **Innovations (ex Hollingsworth-Lonnberg)**
  - **Desroziers' approach**

- **Model-space based methods** *(no info on obs error)*
  - **NMC (NCEP) lagged-forecast method**
  - **CQC forecast differences**



Outline

1. What is the NMC method actually measuring ?

Forecast error, analysis error, something in between ?

We will show that advection has an averaging effect on error covariances and will find an interpretation in the context of assimilation of limb sounding observations of long-lived chemical species

We will get a new approach to estimate model error covariance

2. Convergence analysis of the Desroziers' method

Whether or not it converge and to which value it converges to will be examined.

Important to distinguish the constrained and unconstrained formulation of the Desroziers' method

NMC method for the chemical tracer assimilation

Evolution of mixing ratio

$$\frac{D\mu}{Dt} = \frac{\partial \mu}{\partial t} + \mathbf{V} \cdot \nabla \mu = 0$$ Mixing ratio is a conserved quantity

Evolution of errors $$\frac{D\varepsilon}{Dt} = \begin{cases} 0 \\ \varepsilon^{q} & \text{model error} \end{cases}$$

$$\varepsilon \text{ (mixing ratio error)} = \mu - \mu^{\text{true}}$$

Evolution of the error covariance

Covariance of mixing ratio error between a pair of points

$$P(x_1, x_2, t) = \langle \varepsilon_1 \varepsilon_2 \rangle = \langle \, \varepsilon^{\mu}(\mathbf{x}(t;\mathbf{X}_1), t) \;\; \varepsilon^{\mu}(\mathbf{x}(t;\mathbf{X}_2), t) \, \rangle$$

$$\nabla_1 = \left(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial y_1} \right) \; ; \; \nabla_2 = \left(\frac{\partial}{\partial x_2}, \frac{\partial}{\partial y_2} \right)$$

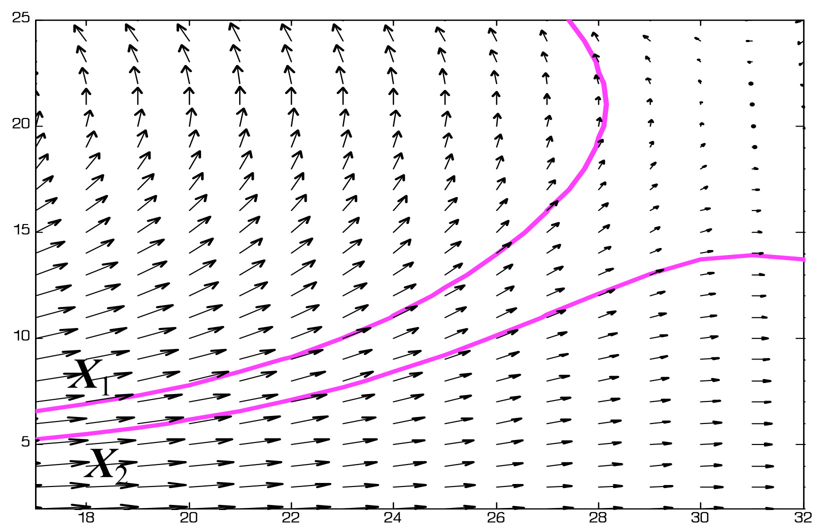
Multiplying

$$\varepsilon_2 \left(\frac{\partial \varepsilon_1}{\partial t} + \mathbf{V}_1 \cdot \nabla_1 \varepsilon_1 \right) = \varepsilon_2 \frac{\partial \varepsilon_1}{\partial t} + \mathbf{V}_1 \cdot \nabla_1 (\varepsilon_1 \varepsilon_1)$$

and likewise for the $\varepsilon_1 \partial_t \varepsilon_2$ term, and taking the expectation gives,

$$\frac{D P(\mathbf{x}_1, \mathbf{x}_2, t)}{Dt} = \frac{\partial P}{\partial t} + \mathbf{V}_1 \cdot \nabla_1 P + \mathbf{V}_2 \cdot \nabla_2 P$$

$$= \begin{cases} 0 \\ Q & \text{for white noise} \end{cases}$$



Error covariance is *conserved* (without model error)

Simplified form of NMC method

- For linear H, and taking the difference between analyses (0-hour forecast) with 6-h forecast valid at 0-hour

$$\mathbf{x}^a(0) - \mathbf{x}^f_{-6}(0) = \mathbf{K}(\mathbf{y}(0) - \mathbf{H}\mathbf{x}^f_{-6}(0)) = \mathbf{K}(\mathbf{H}\mathbf{x}^t(0) + \varepsilon^o(0) - \mathbf{H}\mathbf{x}^f_{-6}(0))$$

$$= \mathbf{K}(\varepsilon^o(0) - \mathbf{H}\varepsilon^f_{-6}(0))$$

$$\begin{aligned}
\mathbf{E}\left[\left(\mathbf{x}^f_{-6}(0) - \mathbf{x}^a(0)\right)\left(\mathbf{x}^f_{-6}(0) - \mathbf{x}^a(0)\right)^T\right] &= \mathbf{KHP}^f\mathbf{H}^T\mathbf{K}^T + \mathbf{KRK}^T \\
&= \mathbf{K}\left(\mathbf{HP}^f\mathbf{H} + \mathbf{R}\right)\mathbf{K}^T \\
&= \mathbf{K}\left(\mathbf{HP}^f\mathbf{H} + \mathbf{R}\right)\left(\mathbf{HP}^f\mathbf{H} + \mathbf{R}\right)^{-1}\mathbf{HP}^f \\
&= \mathbf{KHP}^f \\
&= \mathbf{P}^f_{-6}(0) - \mathbf{P}^a(0)
\end{aligned}$$

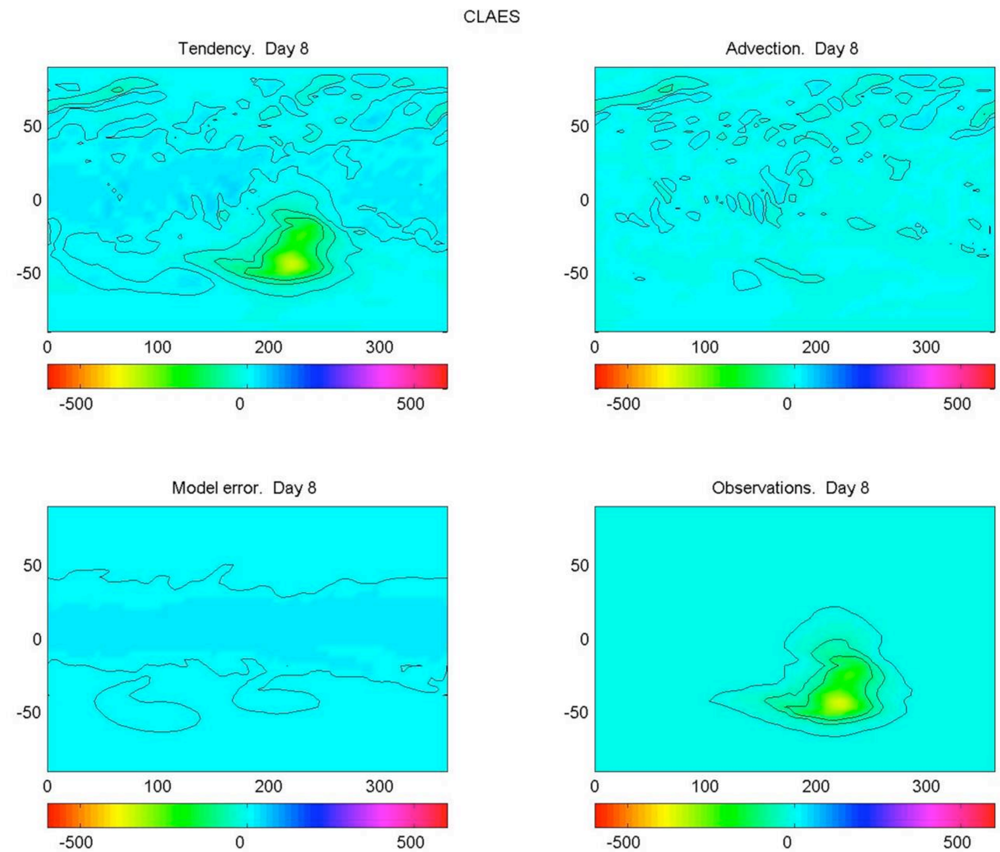
and more generally…

$$\mathbf{E}\left[\left(\mathbf{x}^f_{-6}(12) - \mathbf{x}^f_0(12)\right)\left(\mathbf{x}^f_{-6}(12) - \mathbf{x}^f_0(12)\right)^T\right] = \mathbf{M}_{0,12}[\mathbf{P}^f_{-6}(0) - \mathbf{P}^a(0)]\mathbf{M}^T_{0,12}$$

Error covariance budget : KF on isentropic coordinate limb sounding observations

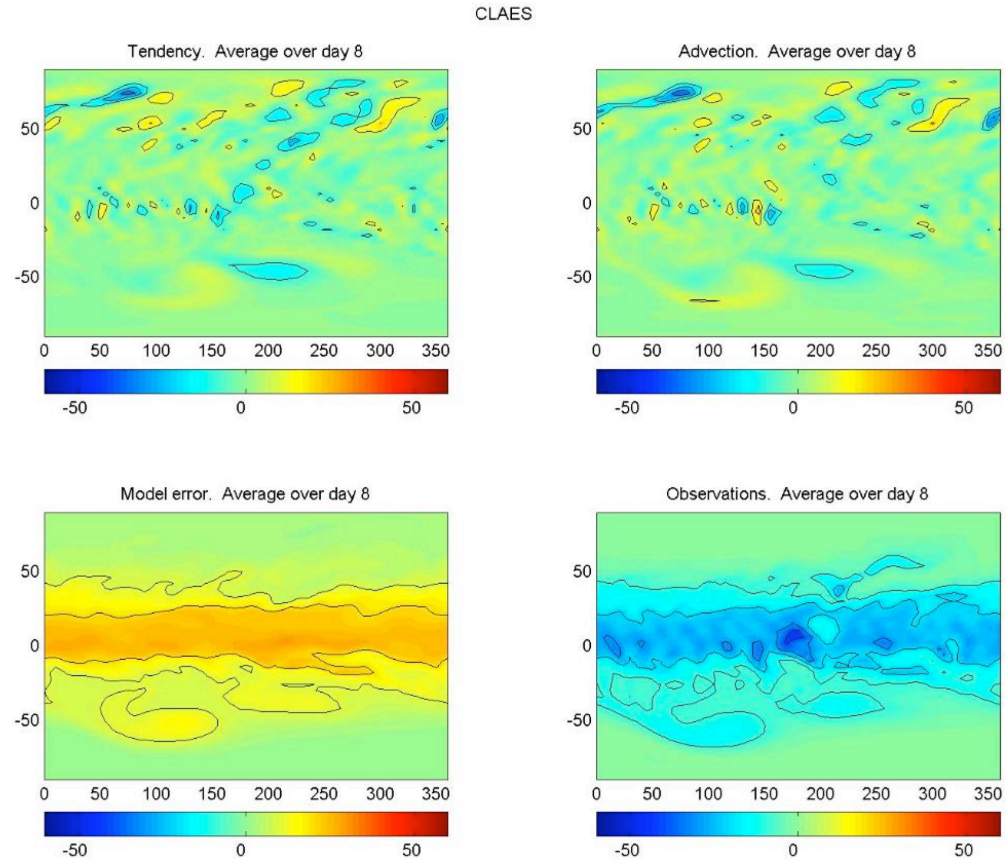
Error covariance budget (Kalman-Bucy filter, assimilation each time step)

$$\underbrace{\frac{\partial P}{\partial t}}_{tendency} + \underbrace{\mathbf{V}_1 \cdot \nabla_1 P + \mathbf{V}_2 \cdot \nabla_2 P}_{advection} = \underbrace{Q}_{\substack{model \\ error}} - \underbrace{\frac{1}{\Delta t}\mathbf{KHP}}_{\substack{observation \\ update}}$$



Taking the average over a day,
shows that the observation contribution balances
the model error contribution

$$Q\,\Delta t \approx \mathbf{KHP} = \mathbf{E}[(\mathbf{x}^f - \mathbf{x}^a)(\mathbf{x}^f - \mathbf{x}^a)^T]$$



The effect of advection on error covariances is averaged out on a time-mean (day) thus the NMC method for this particular problem is simply a measure of the model error covariance

Simple analysis

The asymptotic solution of the Kalman filter

$$\mathbf{P}_n^f = \mathbf{M}\mathbf{P}_n^a\mathbf{M}^T + \mathbf{Q}$$

$$\mathbf{P}_n^a = (\mathbf{I} - \mathbf{K}_n\mathbf{H})\mathbf{P}_n^f$$

Assuming $\mathbf{H}=\mathbf{I}$ (mimics limb sounding observations)
and $\mathbf{M}=\mathbf{I}$ (mimics advection averaging)
is

$$\mathbf{P}_\infty^f = [\mathbf{I} - \mathbf{P}_\infty^f(\mathbf{P}_\infty^f + \mathbf{R})^{-1}]\mathbf{P}_\infty^f + \mathbf{Q}$$

from which we get

$$\mathbf{K}\mathbf{P}_\infty^f = \textit{signal covariance} = \mathbf{Q}$$

Conclusion

It turns out the we have developed a method to estimate the model error covariance in dense observation network and under advection dynamics only

The Desrozier's method

$$\overline{\mathbf{R}}_{n+1} = (\mathbf{I} - \mathbf{HK}_n)(\mathbf{HBH} + \mathbf{R})$$

$$= \overline{\mathbf{R}}_n (\mathbf{H}\overline{\mathbf{B}}_n \mathbf{H}^T + \overline{\mathbf{R}}_n)^{-1} (\mathbf{HBH} + \mathbf{R})$$

$$\mathbf{H}\overline{\mathbf{B}}_{n+1}\mathbf{H}^T = \mathbf{HK}_n (\mathbf{HBH} + \mathbf{R})$$

$$= \mathbf{H}\overline{\mathbf{B}}_n \mathbf{H}^T (\mathbf{H}\overline{\mathbf{B}}_n \mathbf{H}^T + \overline{\mathbf{R}}_n)^{-1} (\mathbf{HBH} + \mathbf{R})$$

where assimilation residuals are used to provide the information about the error statistics

$$\left\langle (O - F)(O - F)^T \right\rangle = \mathbf{HBH}^T + \mathbf{R}$$

$$\left\langle (O - A_{n+1})(O - F)^T \right\rangle = \overline{\mathbf{R}}_{n+1}$$

$$\left\langle (A_{n+1} - F)(O - F)^T \right\rangle = \mathbf{H}\overline{\mathbf{B}}_{n+1}\mathbf{H}^T$$

Illustration - scalar case

Iteration on observation error variance

$$\left\langle (O-A)(O-F)^T \right\rangle = \overline{\mathbf{R}}(\mathbf{H}\overline{\mathbf{B}}\mathbf{H}^T + \overline{\mathbf{R}})^{-1}(\mathbf{HBH} + \mathbf{R})$$

where $\left\langle (O-F)(O-F)^T \right\rangle = \mathbf{HBH}^T + \mathbf{R}$ is obtained from assimilation residuals and overbar denotes *prescribed* error covariances

i)- Correctly prescribed forecast error variance

$$\overline{\mathbf{B}} = \mathbf{B} = \sigma_f^2 \qquad \overline{\mathbf{R}} = \alpha\mathbf{R} = \alpha\sigma_o^2$$

optimal value $\alpha = 1$

$$\left\langle (O-A)(O-F) \right\rangle = \frac{\alpha\,\sigma_o^2}{\alpha\,\sigma_o^2 + \sigma_f^2}(\sigma_o^2 + \sigma_f^2) = \alpha\,\sigma_o^2\left(\frac{\gamma+1}{\alpha\,\gamma+1}\right)$$

where $\gamma = \dfrac{\sigma_o^2}{\sigma_f^2}$

Convergence – scalar case

let $\langle (O-A)(O-F) \rangle = \alpha_{n+1}\sigma_o^2$ be the next iterate

so the iteration on α_n takes the form

$$\alpha_{n+1} = \alpha_n \left(\frac{\gamma + 1}{\alpha_n \gamma + 1} \right) = G(\alpha_n)$$

Define a *mapping* G

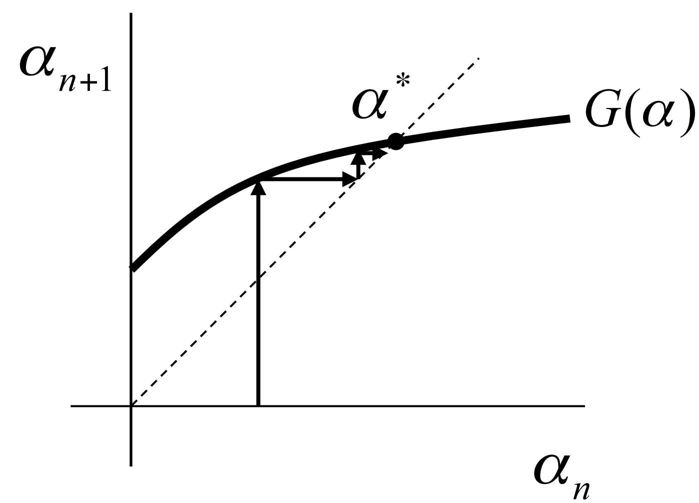
$$G(\alpha) = \alpha \left(\frac{\gamma + 1}{\alpha \gamma + 1} \right)$$

The fixed-point is

$$\alpha^* = G(\alpha^*)$$

condition for convergence

$$\left| G'(\alpha^*) \right| < 1$$



Convergence – scalar case

and so for this case we get $\alpha^* = 1$

$$G'(\alpha^*) = \frac{1}{\gamma + 1} = \frac{\sigma_f^2}{\sigma_o^2 + \sigma_f^2} = K \le 1$$

the scheme is always convergent and converges to the true value, $\alpha = 1$

ii)- Incorrectly prescribed forecast error variance

$$\overline{\mathbf{B}} = \beta \mathbf{B} = \beta \sigma_f^2 \qquad \overline{\mathbf{R}} = \alpha \mathbf{R} = \alpha \sigma_o^2$$

the mapping is now different

$$\alpha_{n+1} = \alpha_n \left(\frac{\gamma + 1}{\alpha_n \gamma + \beta} \right) = G(\alpha_n)$$

Convergence – scalar case

The fixed-point is

$$\alpha^* = 1 + \frac{1-\beta}{\gamma} = 1 + \frac{\left(\sigma_f^2 - \overline{\sigma}_f^2\right)}{\sigma_o^2}$$

that is *not the true observation error* value.

- If forecast error variance is underestimated, obs error is overestimated
- If forecast error variance is overestimated, obs error is underestimated

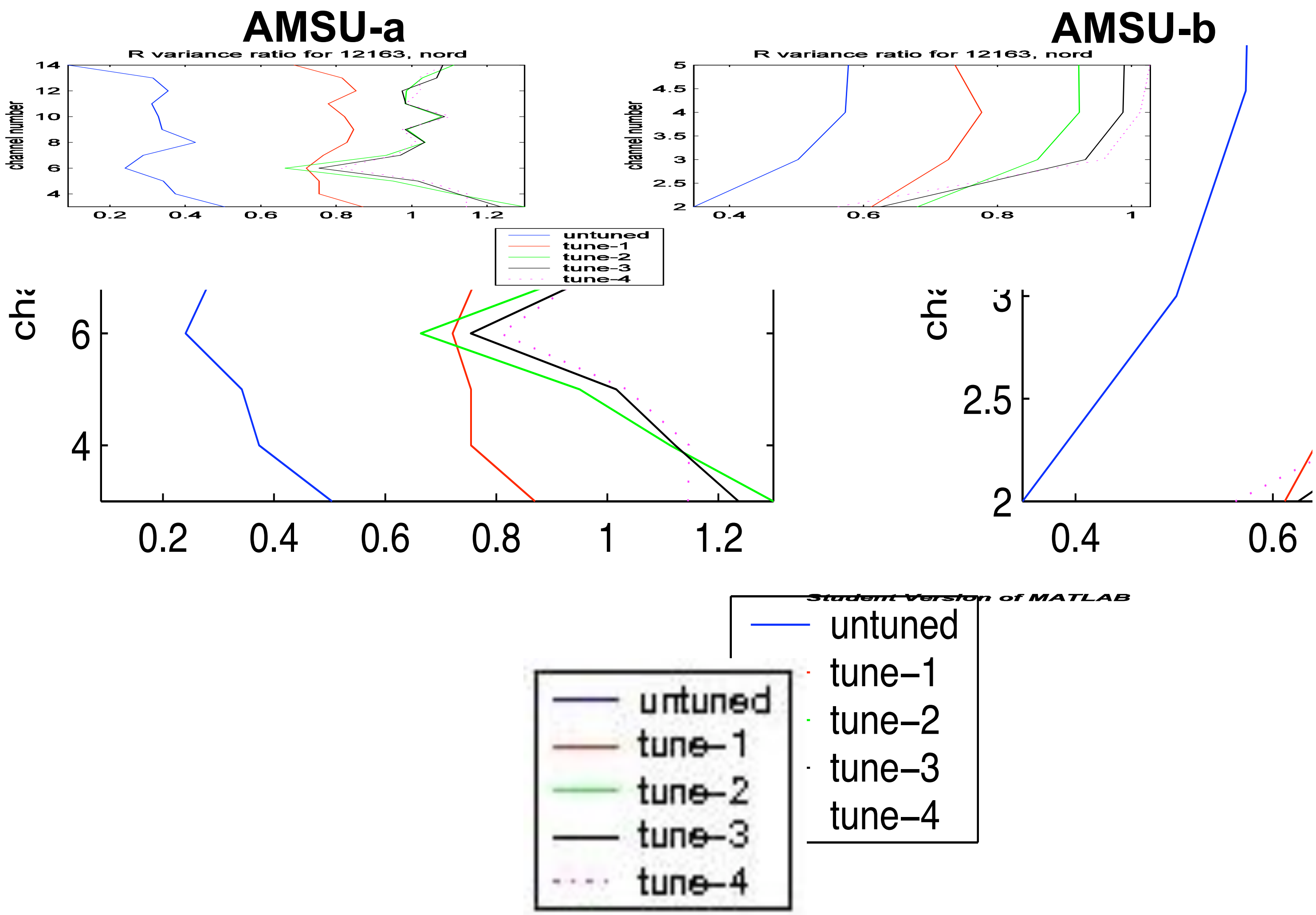
$$G'(\alpha^*) = \frac{\beta}{\gamma + 1} = \frac{\beta\sigma_f^2}{\sigma_o^2 + \sigma_f^2}$$

Will not converge if: $\beta\sigma_f^2 = \overline{\sigma}_f^2 > \sigma_o^2 + \sigma_f^2$

In practice the estimated forecast error variance will never be larger than the innovation error variance, so for all practical cases the scheme converges.

Iteration on observation error variance

Error variance (n) / reference error variance



Iteration on both observation and background error

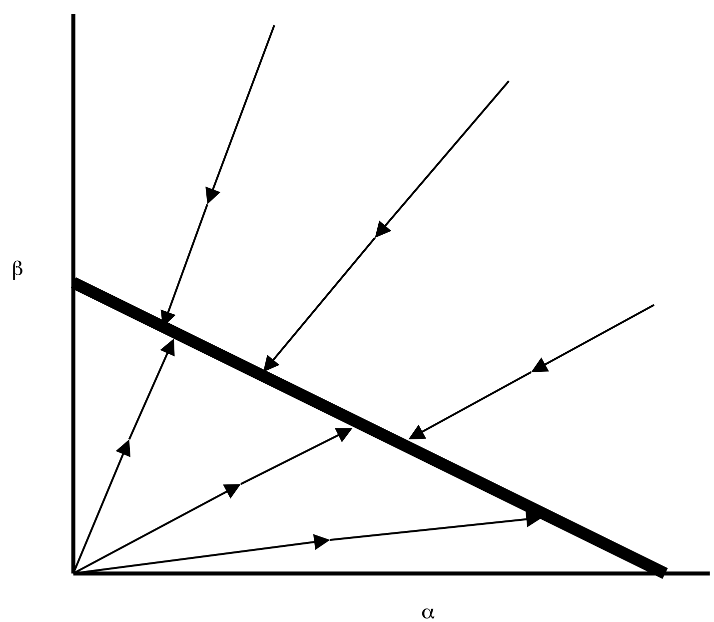
Consider the case of tuning together α and β

$$\alpha_{n+1} = \alpha_n \left(\frac{\gamma + 1}{\alpha_n \gamma + \beta_n} \right) = G(\alpha_n, \beta_n) \qquad \beta_{n+1} = \beta_n \left(\frac{\gamma + 1}{\alpha_n \gamma + \beta_n} \right) = F(\alpha_n, \beta_n)$$

The mapping is an attractor

$$\frac{\partial(G,F)}{\partial(\alpha_n, \beta_n)} = \frac{\gamma + 1}{(\alpha_n \gamma + \beta_n)} \begin{pmatrix} \beta_n & -\alpha_n \\ -\beta_n \gamma & \alpha_n \gamma \end{pmatrix}$$

is rank deficient ! and its determinant is 0
So the scheme is strongly convergent



The fixed-point solution (thick black line)

$$(\alpha^* - 1)\sigma_o^2 + (\beta^* - 1)\sigma_f^2 = 0$$

corresponds to where the estimated total variance (obs + background) is equal to that of the innovation variance

Accounting for spatial correlation - Spectral Analysis

Case where the background error covariance is *spatially correlated* and the observation error covariance is *spatially uncorrelated*

Assume an homogeneous **B** in a 1D periodic domain with observations at each grid points, **H = I** .
We can write the Fourier transform as a matrix **F**, and its inverse as $\mathbf{F}^T$

Then in the system

$$\mathbf{R}_{n+1} = \mathbf{R}_n (\mathbf{B}_n + \mathbf{R}_n)^{-1} (\mathbf{B} + \mathbf{R})$$

$$\mathbf{B}_{n+1} = \mathbf{B}_n (\mathbf{B}_n + \mathbf{R}_n)^{-1} (\mathbf{B} + \mathbf{R})$$

All matrices can be simultaneously diagonalized giving a N systems of scalar (variance) equations (one for each wavenumber *k*)

$$\hat{\mathbf{R}}_{n+1}(k) = \hat{\mathbf{R}}_n(k)(\hat{\mathbf{B}}_n(k) + \hat{\mathbf{R}}_n(k))^{-1}(\hat{\mathbf{B}}(k) + \hat{\mathbf{R}}(k))$$

$$\hat{\mathbf{B}}_{n+1}(k) = \hat{\mathbf{B}}_n(k)\,(\hat{\mathbf{B}}_n(k) + \hat{\mathbf{R}}_n(k))^{-1}(\hat{\mathbf{B}}(k) + \hat{\mathbf{R}}(k))$$

$$\Rightarrow$$

$$\hat{\alpha}_{n+1} = \hat{\alpha}_n \left(\frac{\gamma + 1}{\hat{\alpha}_n \gamma + \hat{\beta}_n} \right) = G(\hat{\alpha}_n, \hat{\beta}_n)$$

$$\hat{\beta}_{n+1} = \hat{\beta}_n \left(\frac{\gamma + 1}{\hat{\alpha}_n \gamma + \hat{\beta}_n} \right) = F(\hat{\alpha}_n, \hat{\beta}_n)$$

but for each wavenumber we have the same ill-conditioned system as before
additional information is therefore needed

Constrained Desroziers' method

Let's introduce a correlation model

$$\hat{\mathbf{R}}_n(k) = \alpha_n \sigma_o^2 \, r_k(L_o)$$

$$\hat{\mathbf{B}}_n(k) = \beta_n \sigma_f^2 \, b_k(L_B)$$

so there is only a total of four parameters α, β, L_o and L_B
And the iteration equations then take the form

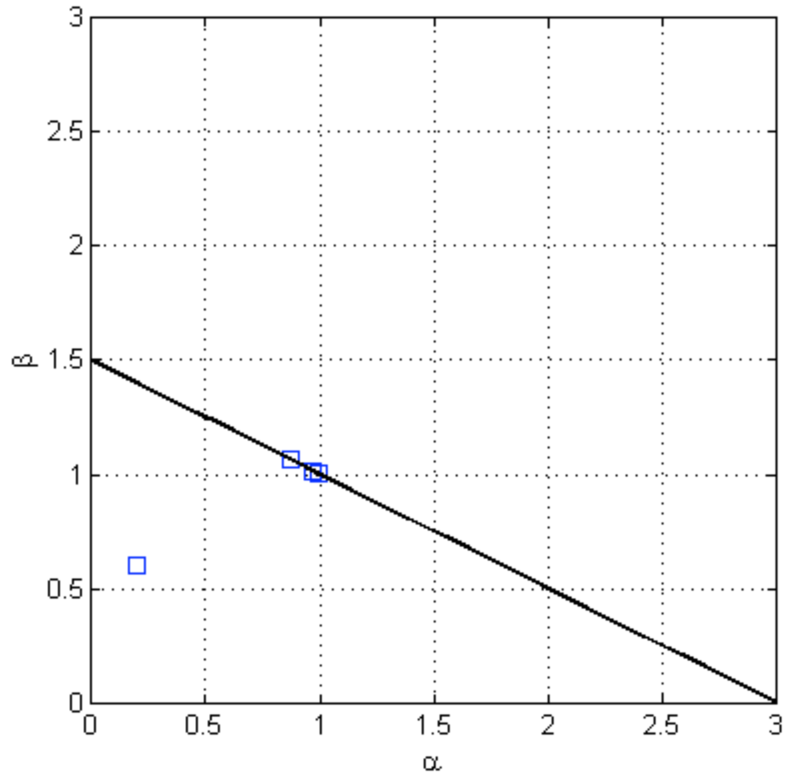
$$\alpha_{n+1} = \alpha_n \sum_{k=1}^{N} \frac{r_k(L_o)[\gamma\, r_k(L_o^t) + b_k(L_B^t)]}{\alpha_n \gamma\, r_k(L_o) + \beta_n b_k(L_b)}$$

$$\beta_{n+1} = \beta_n \sum_{k=1}^{N} \frac{b_k(L_B)[\gamma\, r_k(L_o^t) + b_k(L_B^t)]}{\alpha_n \gamma\, r_k(L_o) + \beta_n b_k(L_b)}$$

This system has also the innovation attractor solution

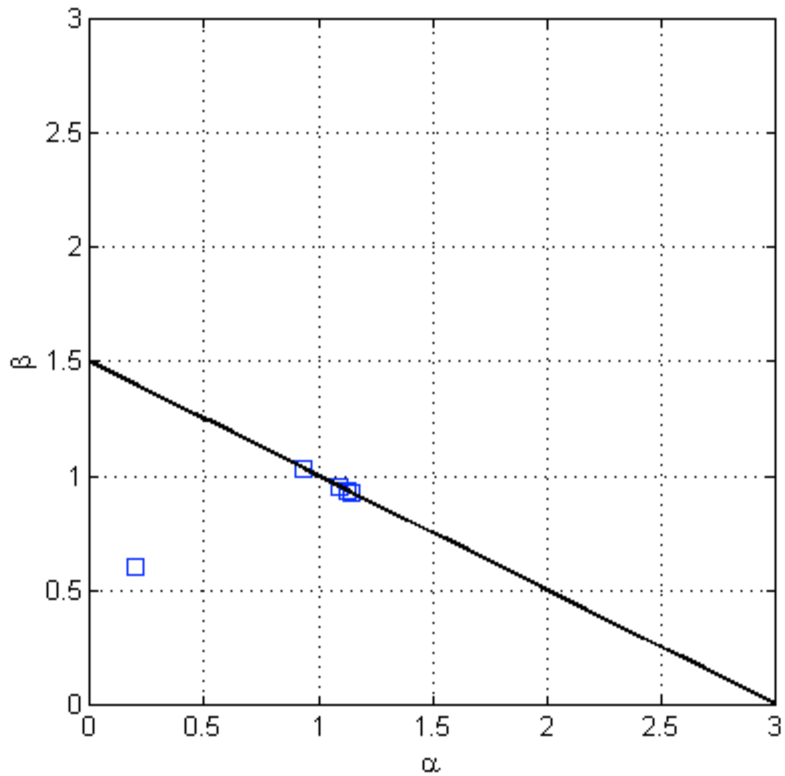
$$\alpha_{n+1}\gamma + \beta_{n+1} = \alpha_n \gamma + \beta_n = \gamma + 1$$

If $L_o = L_o^t$ and $L_B = L_B^t$ then it can be shown that $\overline{\alpha} = \overline{\beta} = 1$
the estimated variances and the true variances



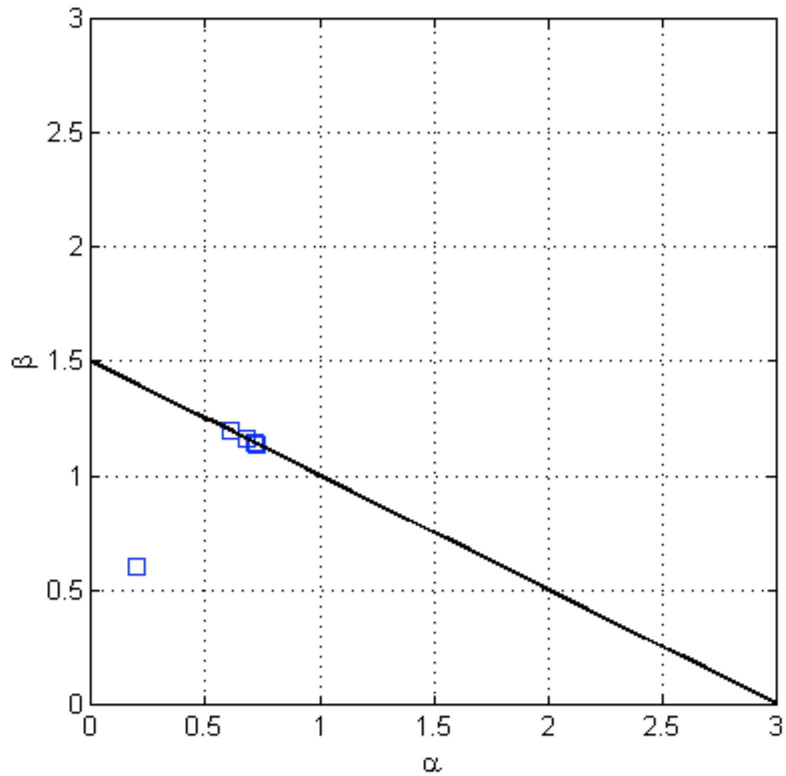
If $L_B > L_B^t$ while $L_o = L_o^t$ then $\overline{\alpha} > 1$ and $\overline{\beta} < 1$

i.e. the estimated background error variances is underestimated, and the estimated observation error is overestimated



case where background error correlation length 50% larger than truth

If $L_o < L_o^t$ while $L_B = L_B^t$ then $\overline{\alpha} < 1$ and $\overline{\beta} > 1$
i.e. the estimated observation error variance is underestimated and and the background error variance is overestimated



case where the true observation error correlation length is 50 km, but is prescribed as spatially uncorrelated

Summary and future work

- The convergence analysis of the Desroziers' method to included the estimation of the correlation length scales can be made by taking the second moments of the spectral equations. The problem of estimating three parameters (out of four) and all four parameters will be investigated
- The estimation of the error variances is sensitive to the misspecification of observations error correlation length (more than what they are from misspecification of the background error correlation length scale)

- In its simplest form the NMC method (forecast minus analysis) and when applied for chemical tracers with a dense observation network, provide an estimate of the model error covariance. In the meteorological context and with longer assimilation windows, these conclusions have to be revised

- It is not clear what the CQC forecast difference method is actually providing. The estimate is of course dependent on the model error covariance, but but depend also on the correlation of the advection terms

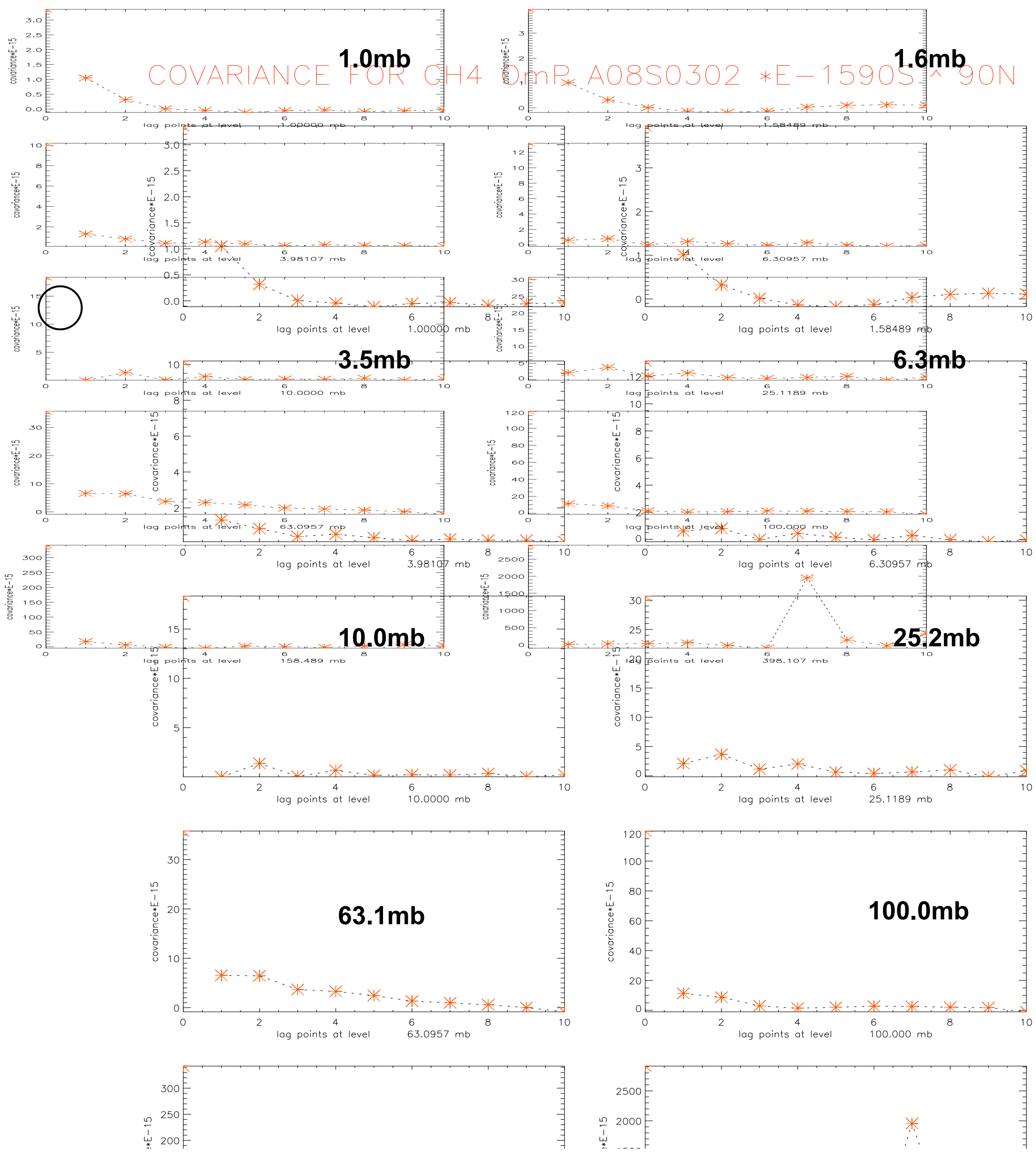
$$\left\langle (\mathbf{V}_1 \cdot \nabla \mu_1)(\mathbf{V}_2 \cdot \nabla \mu_2) \right\rangle$$

that may easily introduce smaller scales variances and correlations

Merci

Thank you

CH_4



COVARIANCE FOR CH4 0mP A08S0302 *E−15 90S ^ 90N
1.0mb
1.6mb
3.5mb
6.3mb
10.0mb
25.2mb
63.1mb
100.0mb
covariance*E−15
lag points at level
1.00000 mb
1.58489 mb
3.98107 mb
6.30957 mb
10.0000 mb
25.1189 mb
63.0957 mb
100.000 mb

Tuning in alternance –CH4