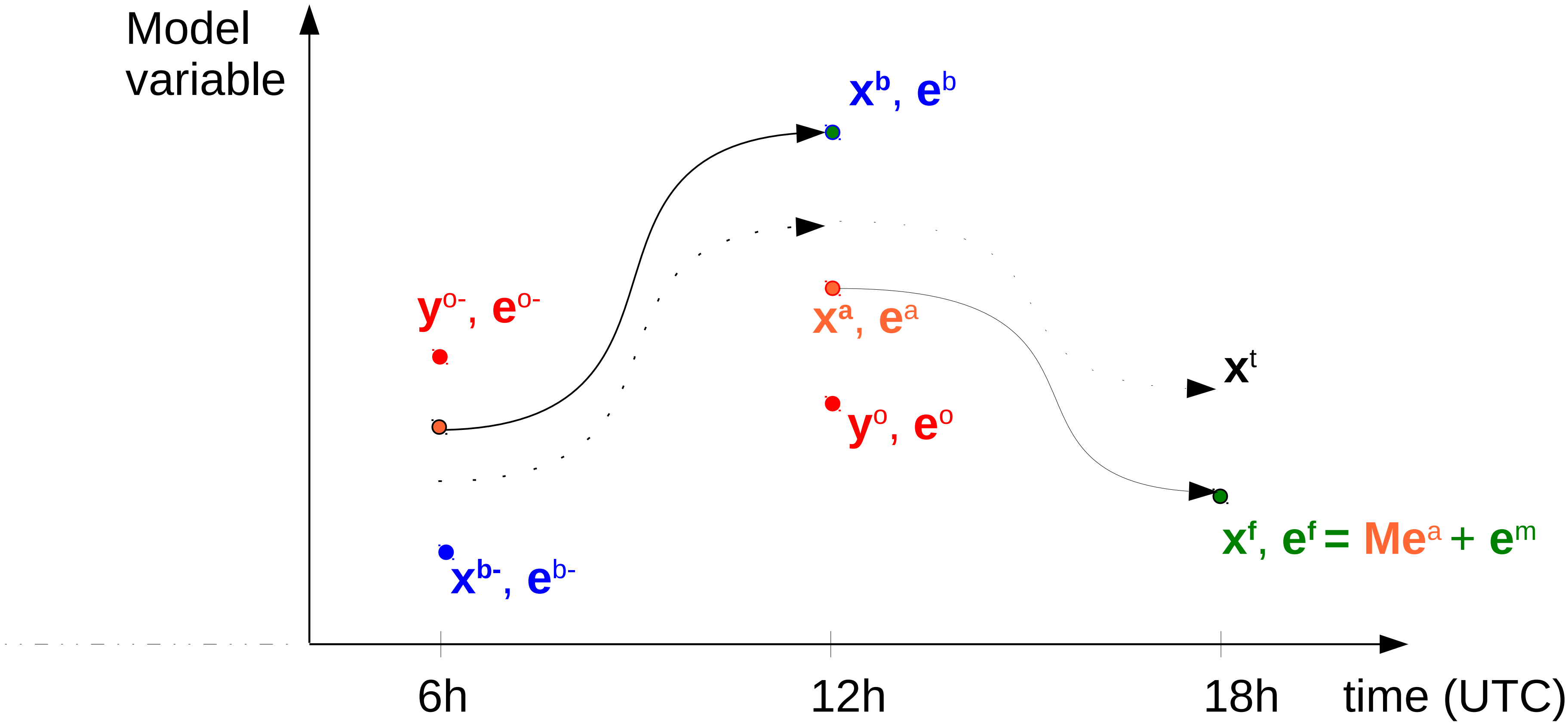
Simulation and diagnosis of error contributions in DA cycling

Loïk Berre and Benjamin Ménétrier
Workshop on Sensitivity Analysis and Data Assimilation
Aveiro, 4 July 2018

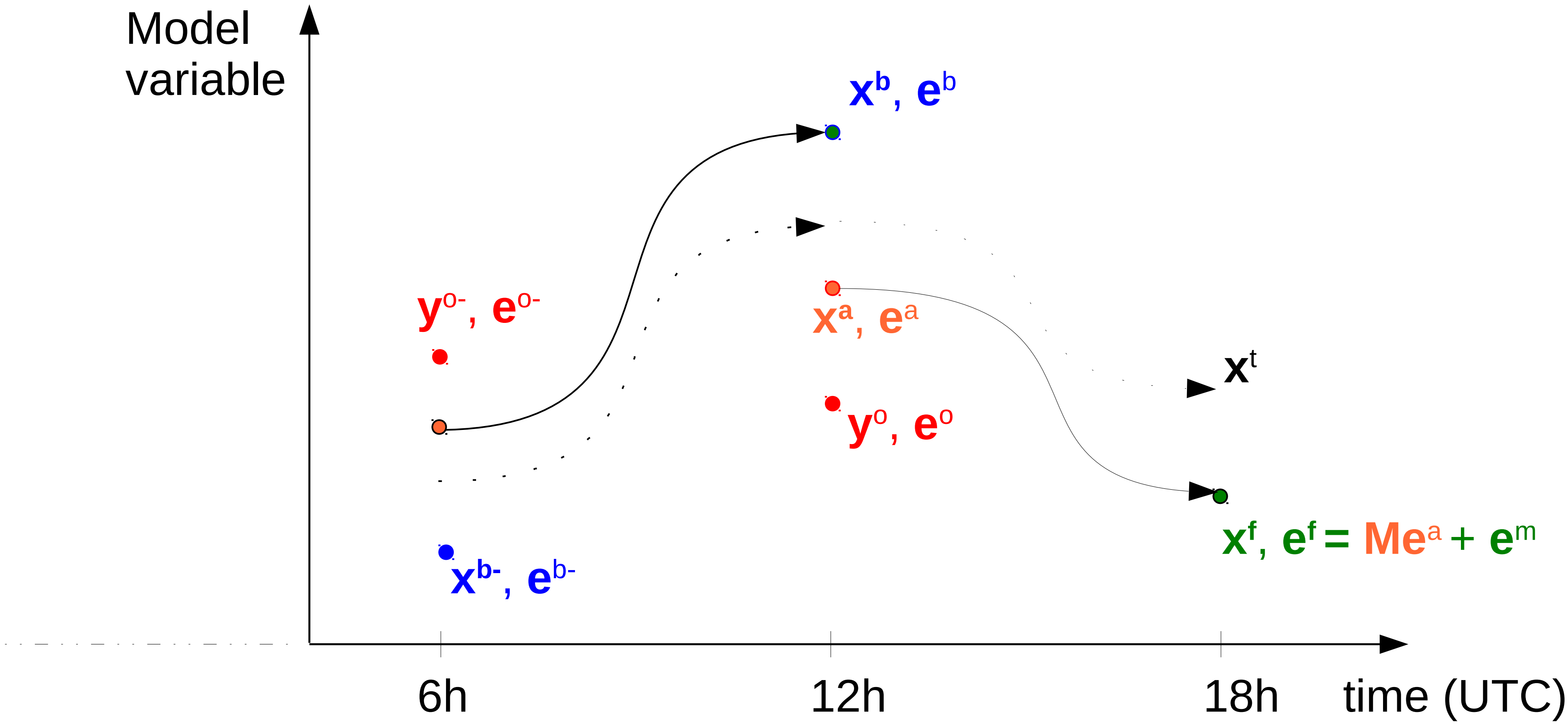
Error contributions and their propagation



Each error contribution is partly **propagated** by the successive analysis/forecast steps.

Forecast errors = propagation of errors with **different ages** :
* recent analysis & model errors ;
* recent background & observation errors ;
* ± old background, observation & model errors.

Error contributions and their propagation



- Goal : simulate error contributions, diagnose their amplitude and propagation.
- Motivations : knowledge about error dynamics in DA cycling, develop error simulation and estimation methods.

Outline

- Expansion of forecast error contributions
- Propagation of ***old*** versus ***recent*** error contributions
- ***Observation*** versus ***model*** error contributions
- Conclusions

Outline

- Expansion of forecast error contributions
- Propagation of *old* versus *recent* error contributions
- *Observation* versus *model* error contributions
- Conclusions

What does contribute to forecast errors ?
(linear expansion at cycling step t_i)

$$\varepsilon_i^f = \mathbf{M}_i \varepsilon_i^a + \varepsilon_i^m$$

What does contribute to forecast errors ? (linear expansion at cycling step t_i)

$$\varepsilon_i^f = \mathbf{M}_i \varepsilon_i^a + \varepsilon_i^m$$

$$= \mathbf{M}_i(\mathbf{I} - \mathbf{K}_i \mathbf{H}_i)\varepsilon_i^b + \mathbf{M}_i \mathbf{K}_i \varepsilon_i^o + \varepsilon_i^m$$

What does contribute to forecast errors ? (linear expansion at cycling step t_i)

$$
\begin{aligned}
\varepsilon_i^f &= \mathbf{M}_i \varepsilon_i^a + \varepsilon_i^m \\
&= \mathbf{M}_i(\mathbf{I} - \mathbf{K}_i\mathbf{H}_i)\varepsilon_i^b + \mathbf{M}_i\mathbf{K}_i\varepsilon_i^o + \varepsilon_i^m \\
&= \mathbf{M}_i(\mathbf{I} - \mathbf{K}_i\mathbf{H}_i)[\mathbf{M}_{i-1}(\mathbf{I} - \mathbf{K}_{i-1}\mathbf{H}_{i-1})\varepsilon_{i-1}^b + \mathbf{M}_{i-1}\mathbf{K}_{i-1}\varepsilon_{i-1}^o + \varepsilon_{i-1}^m] + \mathbf{M}_i\mathbf{K}_i\varepsilon_i^o + \varepsilon_i^m
\end{aligned}
$$

What does contribute to forecast errors ?
(linear expansion at cycling step t_i)

$$
\begin{aligned}
\varepsilon_i^f &= \mathbf{M}_i \varepsilon_i^a + \varepsilon_i^m \\
&= \mathbf{M}_i(\mathbf{I} - \mathbf{K}_i\mathbf{H}_i)\varepsilon_i^b + \mathbf{M}_i\mathbf{K}_i\varepsilon_i^o + \varepsilon_i^m \\
&= \mathbf{M}_i(\mathbf{I} - \mathbf{K}_i\mathbf{H}_i)[\mathbf{M}_{i-1}(\mathbf{I} - \mathbf{K}_{i-1}\mathbf{H}_{i-1})\varepsilon_{i-1}^b + \mathbf{M}_{i-1}\mathbf{K}_{i-1}\varepsilon_{i-1}^o + \varepsilon_{i-1}^m] + \mathbf{M}_i\mathbf{K}_i\varepsilon_i^o + \varepsilon_i^m \\
&= \mathbf{T}_2\ \varepsilon_{i-1}^b + \sum_{j=i-1}^{i} \mathbf{T}_{i-j}(\mathbf{M}_j\mathbf{K}_j\varepsilon_j^o + \varepsilon_j^m) \\
&= \ldots \\
&= \mathbf{T}_{i+1}\ \varepsilon_0^b + \sum_{j=0}^{i} \mathbf{T}_{i-j}(\mathbf{M}_j\mathbf{K}_j\varepsilon_j^o + \varepsilon_j^m)
\end{aligned}
$$

where, for $j < i$,

$$\mathbf{T}_{i-j} = \prod_{k=j+1}^{i} \mathbf{M}_k(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)$$

and t_0 is the beginning of the considered cycling period.

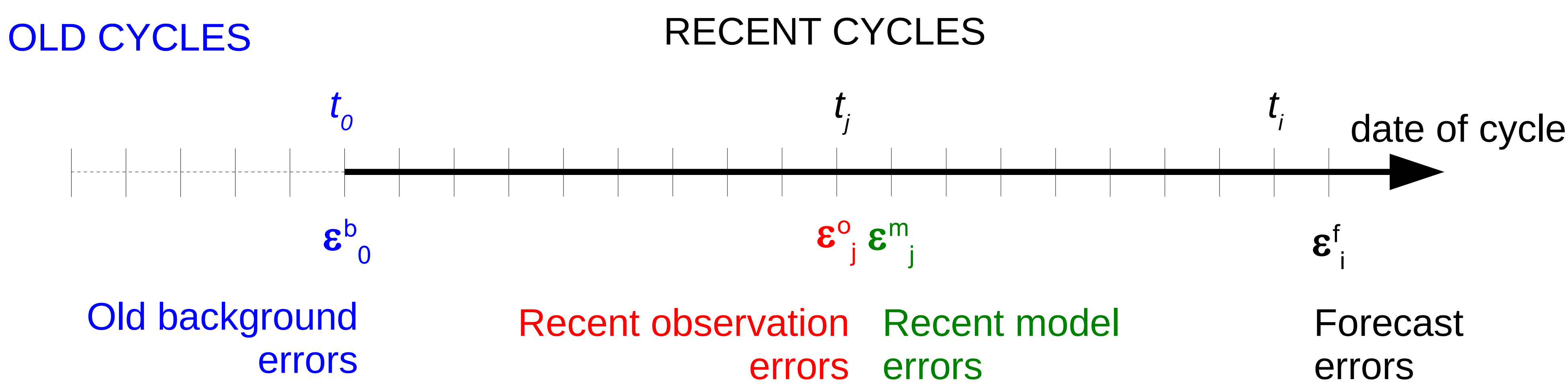
METEO FRANCE

(e.g. El Ouaraini and Berre 2011)

Old and recent error contributions

$$\varepsilon_i^f = \mathbf{T}_{i+1}\ \varepsilon_0^b + \sum_{j=0}^{i} \mathbf{T}_{i-j}(\mathbf{M}_j\mathbf{K}_j\varepsilon_j^o + \varepsilon_j^m)$$

with $\mathbf{T}_{i-j} = \prod_{k=j+1}^{i}\ \mathbf{M}_k(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\quad (=\text{ cycling operator}).$



How do these 3 error contributions compare, and how do they propagate & accumulate during the cycling ?

Outline

- Expansion of forecast error contributions
- Propagation of ***old*** versus ***recent*** error contributions
- *Observation* versus *model* error contributions
- Conclusions

Simulation of error contributions from an old background and from recent observations

- Baseline Ensemble DA experiment (EDA) : Arpege 4D-Var (global NWP), observation perturbations and multiplicative inflation, warm start on 9 January 2017 from operational EDA ; 6h cycling ; same $\mathbf{B}_j$ (provided by operational EDA) for all xp's.

- To quantify contributions of $\boldsymbol{\varepsilon}^b_0$ and $\boldsymbol{\varepsilon}^o_j$, variants of this EDA baseline are run, from 9 to 22 January 2017 (2 weeks), using **addition and propagation of specific perturbations**.

- Propagation of states is done non linearly (4D-Var + NL model) ; propagation of perturbations is interpreted using linear formalism.

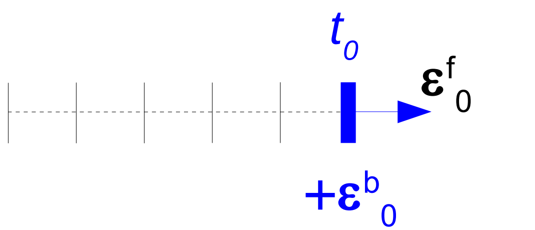
- Evolution of global variance of error contributions for temperature (500 hPa) from corresponding ensemble spread[2].

METEO FRANCE

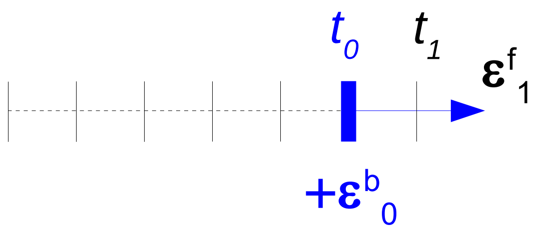
Propagation of old background error

$\boldsymbol{\varepsilon}^b_0$ = background error at t_0 simulated by warm start from operational EDA

$$\mathbf{T}_{i+1} = \prod_{k=0}^{i} \mathbf{M}_k(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)$$



Step t_0 : add $\boldsymbol{\varepsilon}^b_0$, run analysis and forecast : $\boldsymbol{\varepsilon}^f_0 = \mathbf{T}_1\, \boldsymbol{\varepsilon}^b_0$

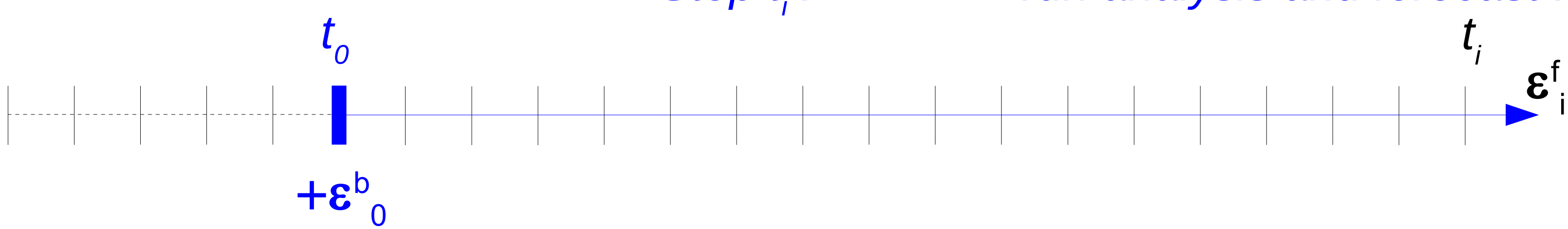


Step t_1 : run analysis and forecast : $\boldsymbol{\varepsilon}^f_1 = \mathbf{T}_2\, \boldsymbol{\varepsilon}^b_0$

...

...

Step t_i : run analysis and forecast : $\boldsymbol{\varepsilon}^f_i = \mathbf{T}_{i+1}\, \boldsymbol{\varepsilon}^b_0$

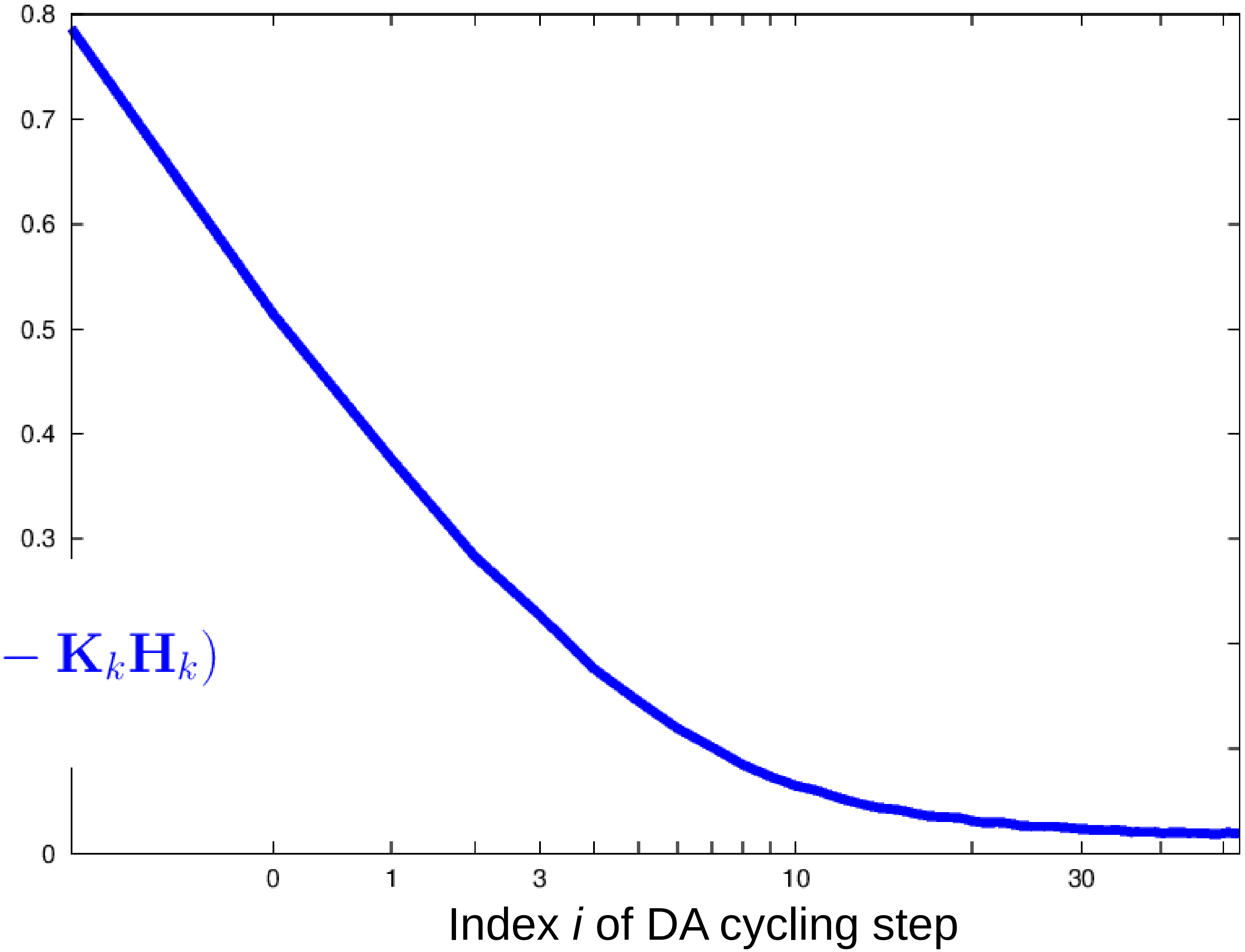


Propagation of old background error

$$\mathrm{Var}(\mathbf{T}_{i+1}\ \boldsymbol{\varepsilon}_0^b)$$

with

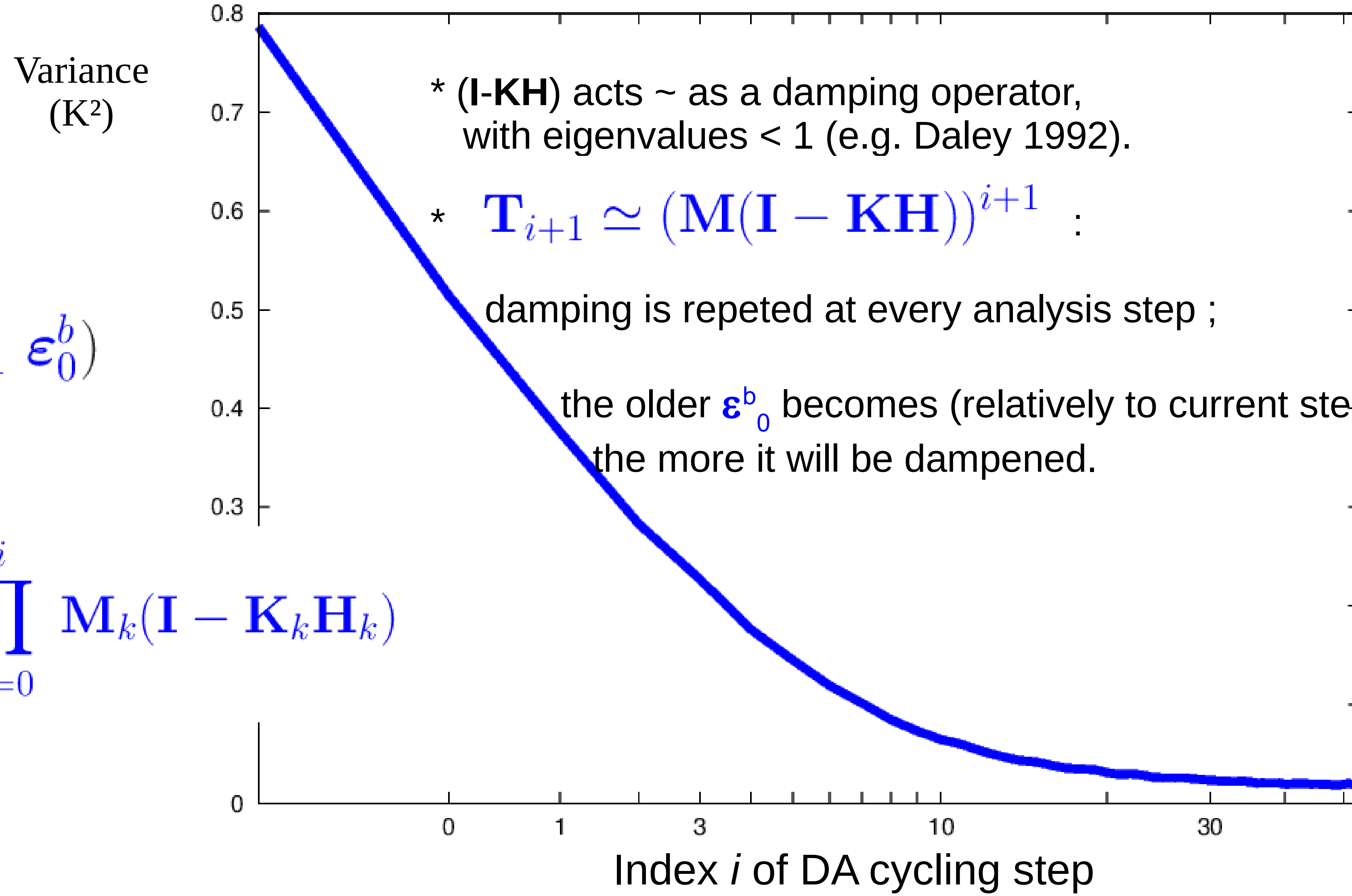
$$\mathbf{T}_{i+1} = \prod_{k=0}^{i} \mathbf{M}_k(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)$$



Old background errors are dampened
by successive DA steps (~ 4-day period).

METEO
FRANCE

Propagation of old background error



* (**I-KH**) acts ~ as a damping operator, with eigenvalues < 1 (e.g. Daley 1992).

* $\mathbf{T}_{i+1} \simeq (\mathbf{M}(\mathbf{I} - \mathbf{KH}))^{i+1}$:

damping is repeted at every analysis step ;

the older $\boldsymbol{\varepsilon}^b_0$ becomes (relatively to current step t_i), the more it will be dampened.

$$\mathrm{Var}(\mathbf{T}_{i+1}\ \boldsymbol{\varepsilon}_0^b)$$

with

$$\mathbf{T}_{i+1} = \prod_{k=0}^{i} \mathbf{M}_k(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)$$

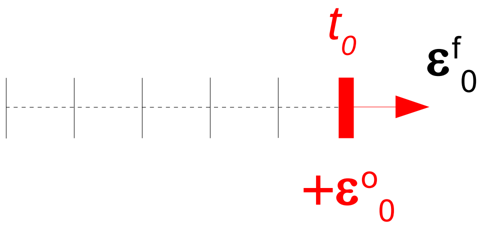
Old background errors are dampened by successive DA steps (~ 4-day period).

METEO FRANCE

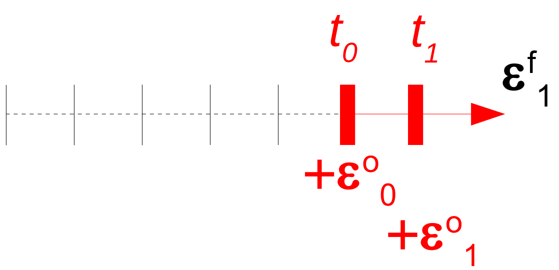
Accumulation and propagation of recent observation errors

$$\mathrm{Var}\left(\sum_{j=0}^{i} \mathbf{T}_{i-j}\, \mathbf{M}_j \mathbf{K}_j\, \boldsymbol{\varepsilon}^o_j \right)$$

$\boldsymbol{\varepsilon}^o_j = \mathbf{R}^{1/2}\, \boldsymbol{\eta}$



Step t_0 : add $\boldsymbol{\varepsilon}^o_0$, run analysis and forecast : $\boldsymbol{\varepsilon}^f_0 = \mathbf{M}_0\mathbf{K}_0\boldsymbol{\varepsilon}^o_0$

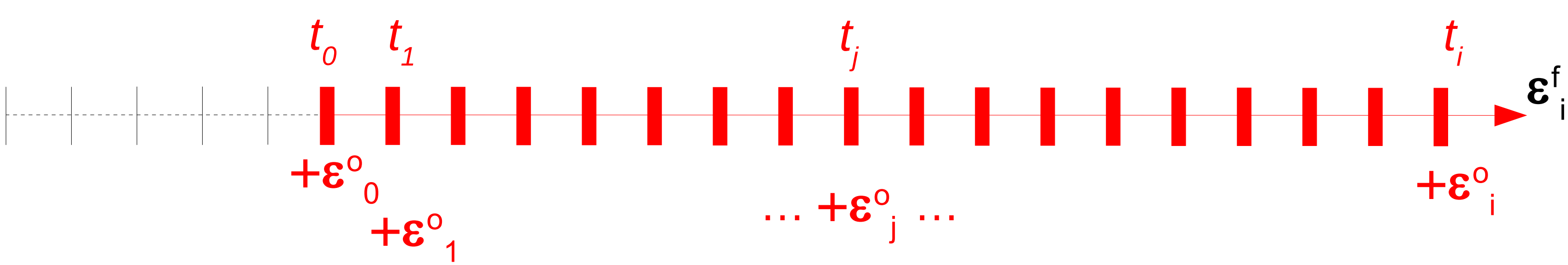


Step t_1 : add $\boldsymbol{\varepsilon}^o_1$, run analysis and forecast : $\boldsymbol{\varepsilon}^f_1 = \mathbf{T}_1\mathbf{M}_0\mathbf{K}_0\boldsymbol{\varepsilon}^o_0 + \mathbf{M}_1\mathbf{K}_1\boldsymbol{\varepsilon}^o_1$

...

...

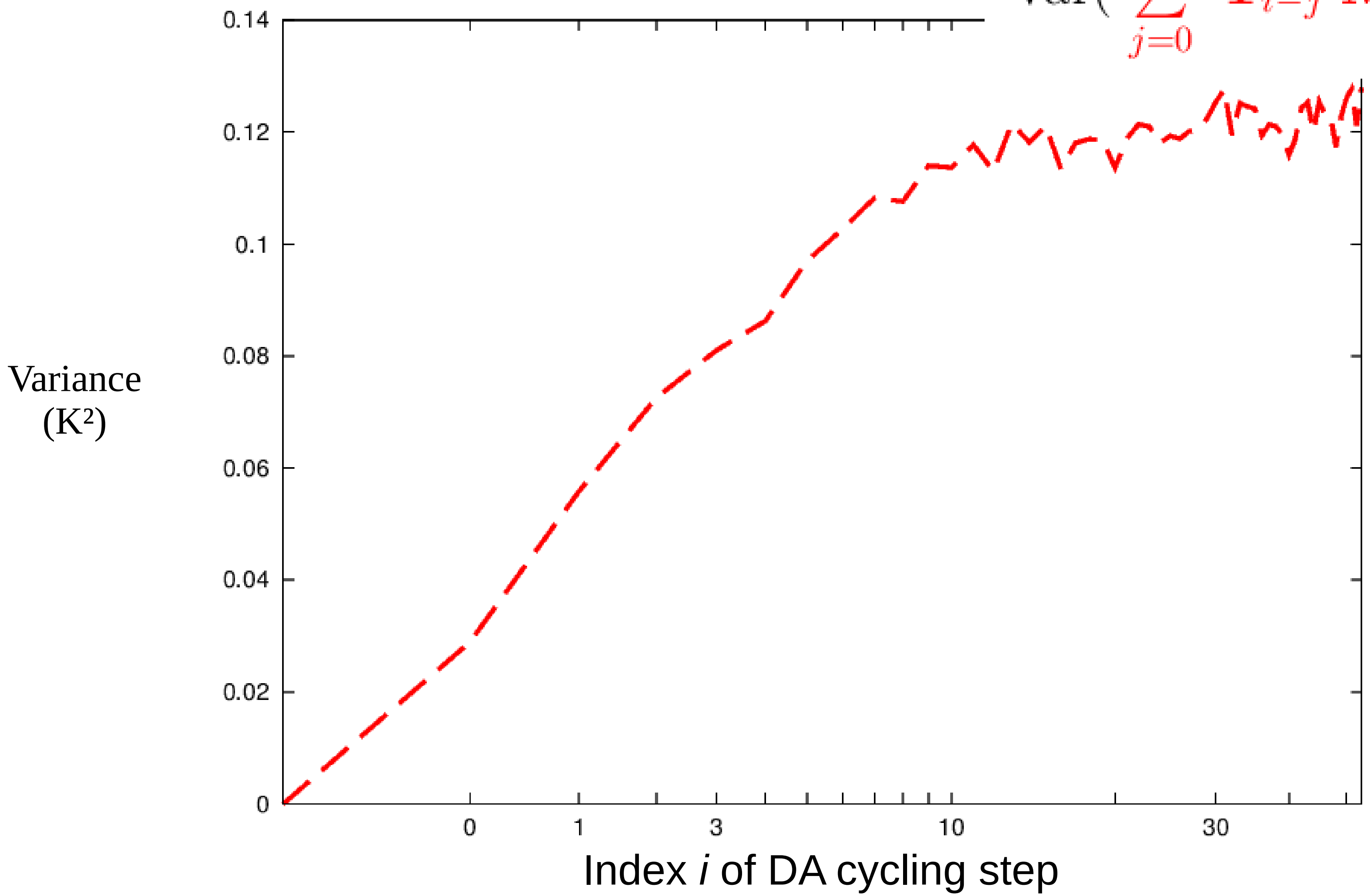
Step t_i : add $\boldsymbol{\varepsilon}^o_i$, run analysis and forecast : $\boldsymbol{\varepsilon}^f_i = \Sigma^i_{j=0}\mathbf{T}_{i-j}\mathbf{M}_j\mathbf{K}_j\boldsymbol{\varepsilon}^o_j$



METEO FRANCE

Accumulation and propagation of recent observation errors

$$\mathrm{Var}\left(\sum_{j=0}^{i} \mathbf{T}_{i-j}\, \mathbf{M}_j \mathbf{K}_j\, \varepsilon_j^o \right)$$



Recent observation errors are accumulated and dampened by successive DA steps.

Why does it converge like this ?

Spectral interpretation of the convergence of $\mathrm{Var}\left(\sum_{j=0}^{i} \mathbf{T}_{i-j}\, \mathbf{M}_j \mathbf{K}_j\, \boldsymbol{\varepsilon}_j^o\right)$
as a power series

Spectral variance (for a given wave vector $\vec{m}$) :
for homogeneous and static covariances, etc.

$$V[e_{i,\vec{m}}^{f,o_{rct}}] = \sum_{j=0}^{i} (|T_{\vec{m}}|^2)^{i-j}\, V[\tilde{e}_{\vec{m}}^{f,o_j}] = V[\tilde{e}_{\vec{m}}^{f,o}] \sum_{j=0}^{i} (|T_{\vec{m}}|^2)^{i-j}$$

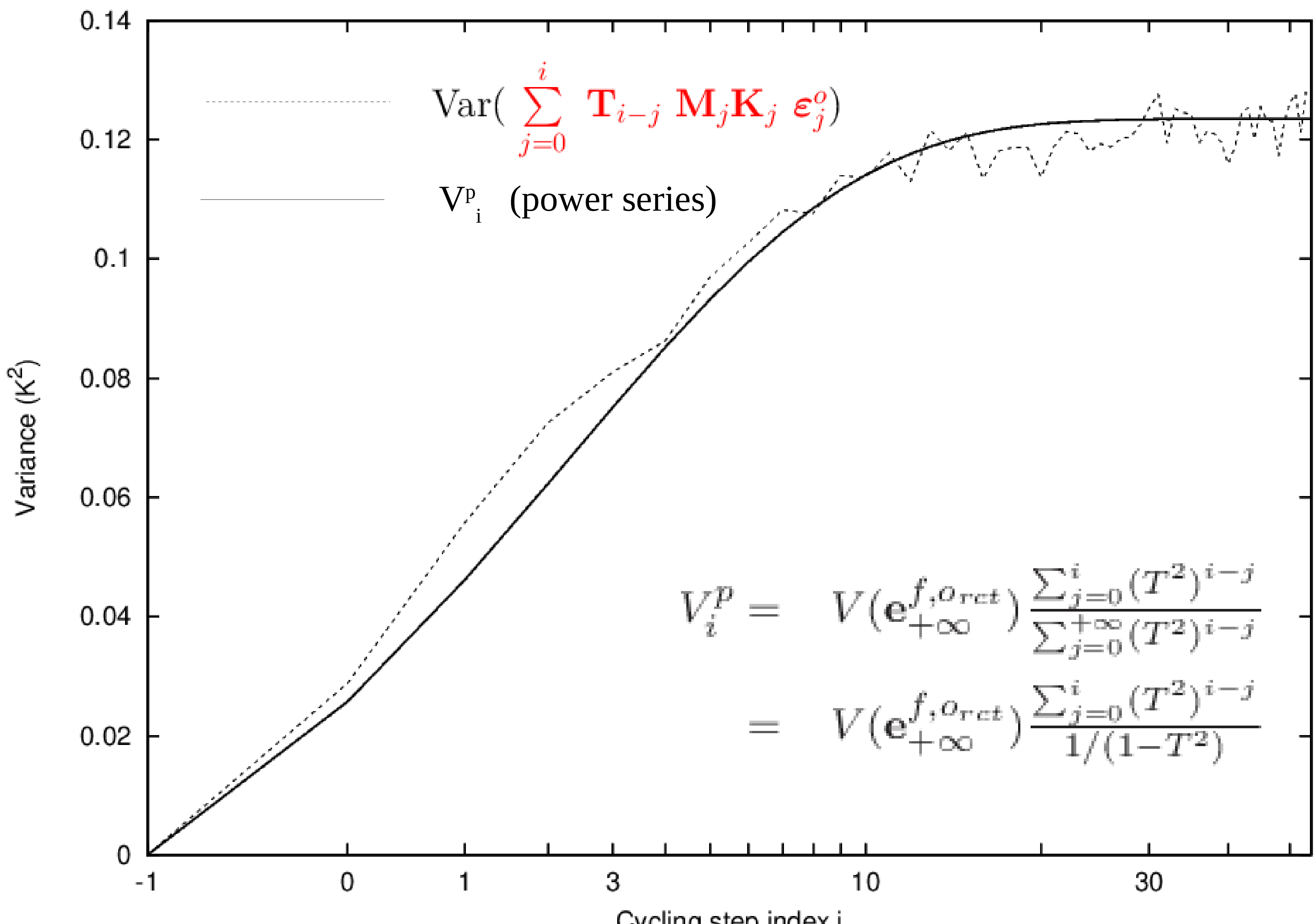
Convergence of power series :
(if $|T_m|^2 < 1$)

$$\lim_{i \to +\infty} \sum_{j=0}^{i} (|T_{\vec{m}}|^2)^{i-j} = \sum_{j'=0}^{+\infty} (|T_{\vec{m}}|^2)^{j'} = \frac{1}{1 - |T_{\vec{m}}|^2}$$

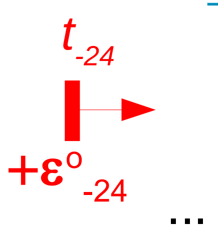
Weighted sum of power series :
(with **K**' = **I** – **KH**)

$$V(e_{+\infty}^{f,o_{rct}}) = \sum_{\vec{m}} \frac{|(MK)_{\vec{m}}|^2}{1 - |(MK')_{\vec{m}}|^2} V(e_{\vec{m}}^o)$$

Spectral interpretation of the convergence as a power series



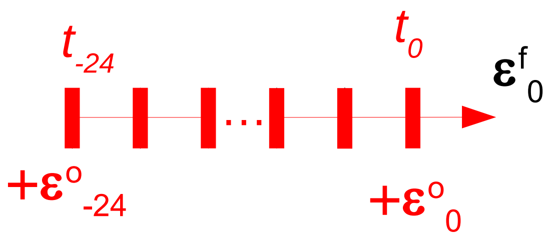
Accumulation and propagation of old+recent observation errors



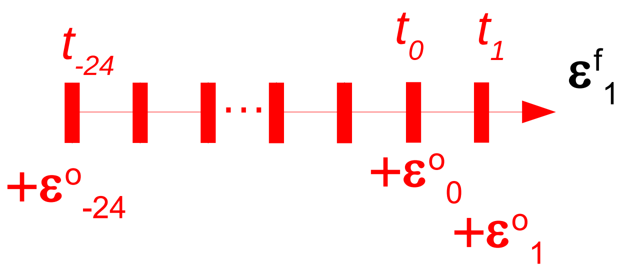
Step t_{-24} : add $\boldsymbol{\varepsilon}^o_{-24}$, run analysis and forecast
(6 days before t_0)

...

...



Step t_0 : add $\boldsymbol{\varepsilon}^o_0$, run analysis and forecast

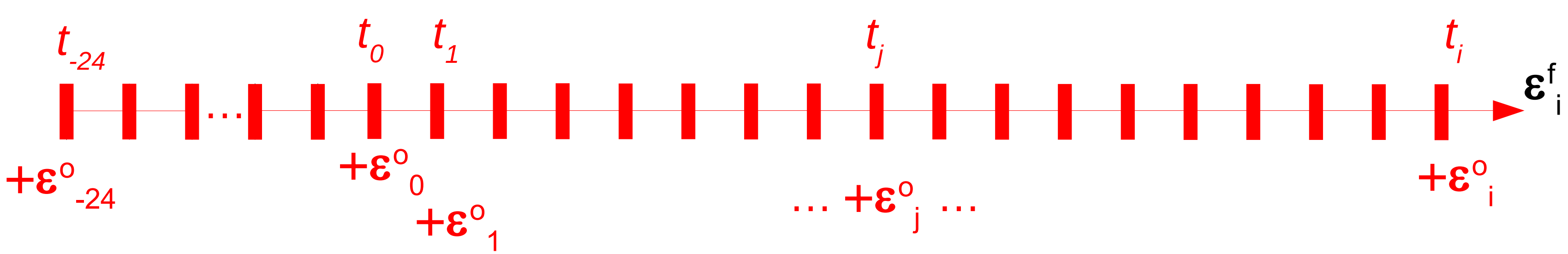


Step t_1 : add $\boldsymbol{\varepsilon}^o_1$, run analysis and forecast

...

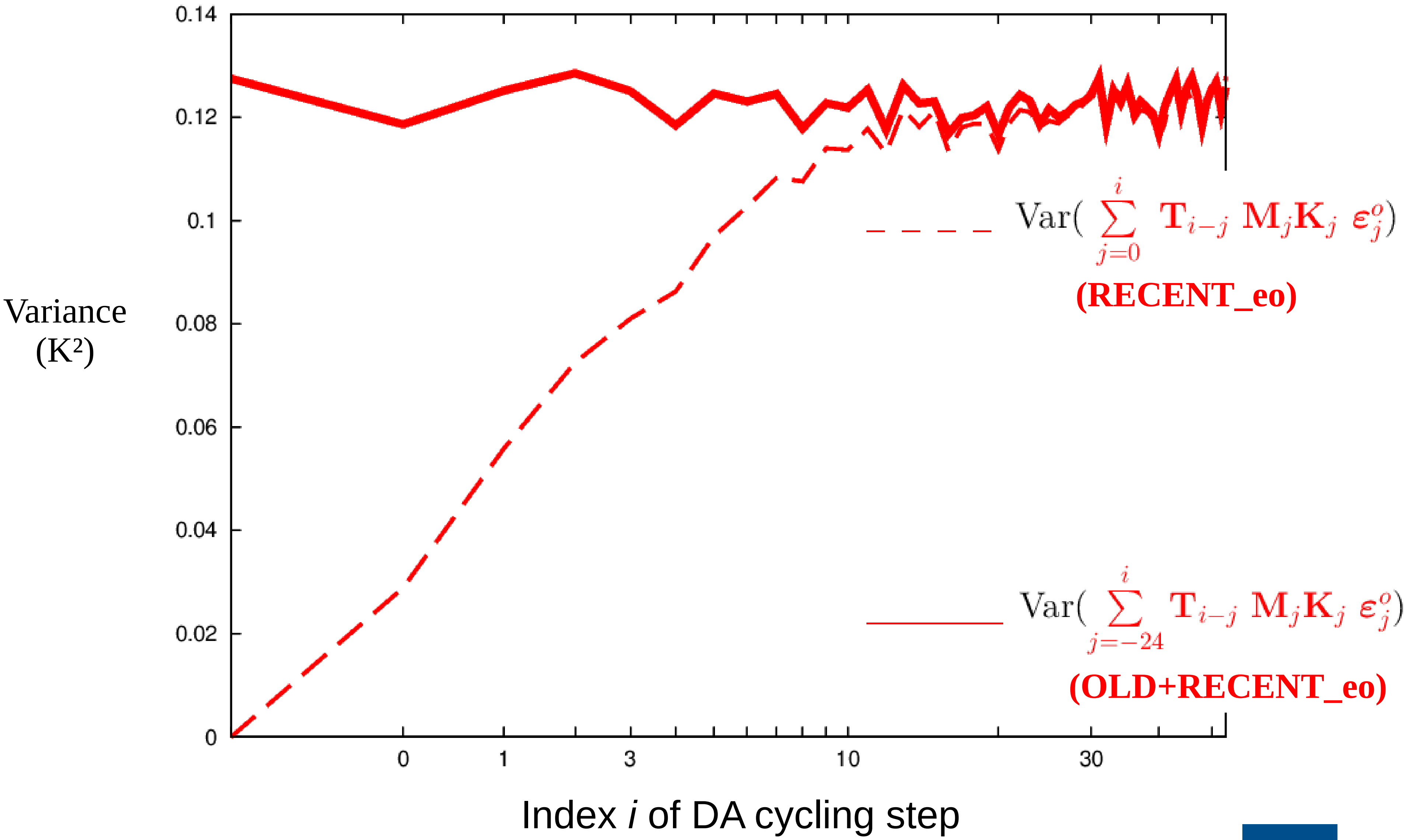
...

Step t_i : add $\boldsymbol{\varepsilon}^o_i$, run analysis and forecast : $\boldsymbol{\varepsilon}^f_i = \Sigma^i_{j=-24} \mathbf{T}_{i-j} \mathbf{M}_j \mathbf{K}_j \boldsymbol{\varepsilon}^o_j$



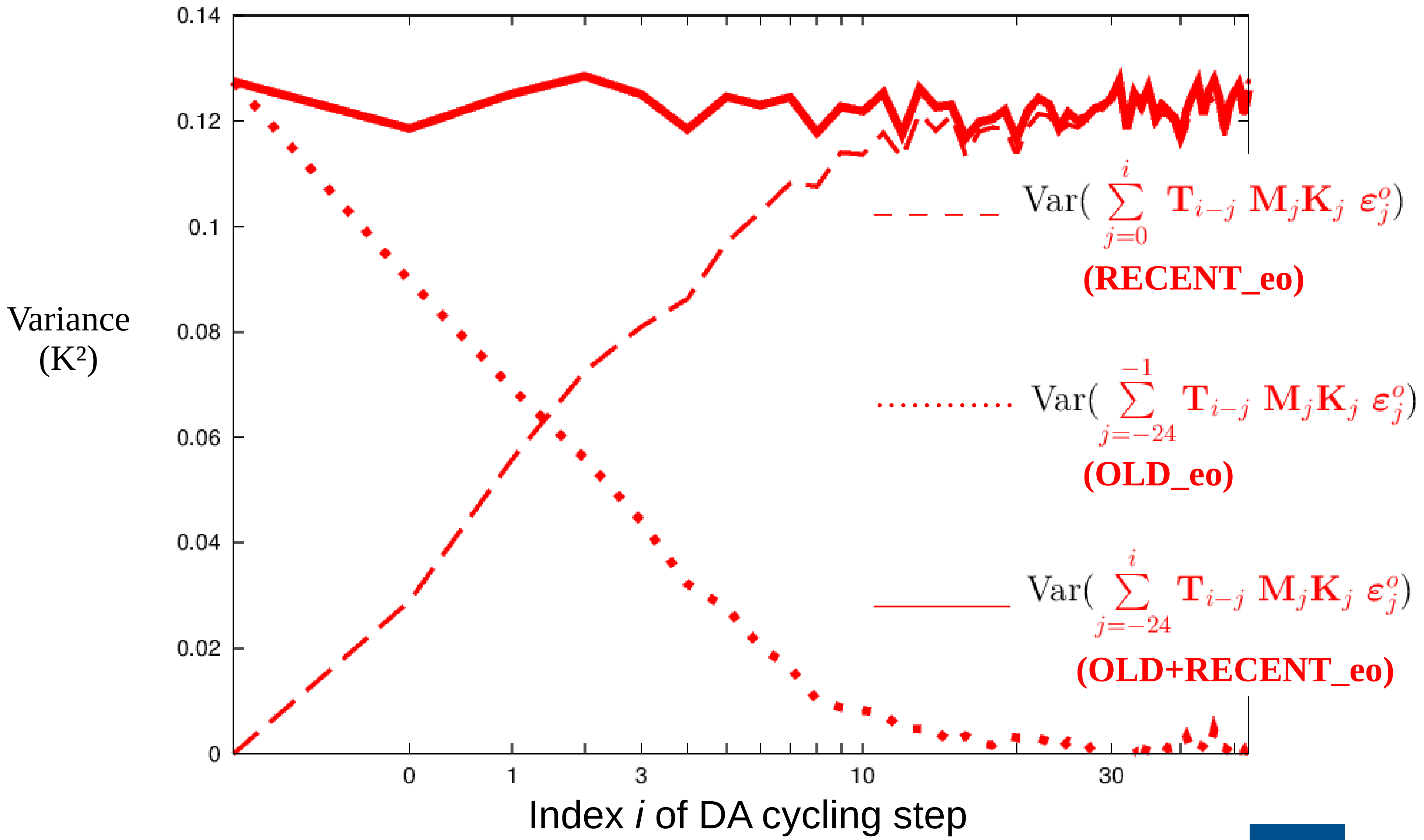
METEO FRANCE

Accumulation and propagation of old+recent observation errors



Total (old+recent) contribution is stable.

Evolution of old & recent observation error contributions



Total (old+recent) contribution is stable : compensation between damping of old errors and accumulation of recent errors.

Outline

- Expansion of forecast error contributions
- Propagation of *old* versus *recent* error contributions
- ***Observation*** versus ***model*** error contributions
- Conclusions

Contributions to forecast error variance

$$\varepsilon_i^f = \mathbf{T}_{i+1}\,\varepsilon_0^b + \sum_{j=0}^{i} \mathbf{T}_{i-j}\,\mathbf{M}_j\mathbf{K}_j\,\varepsilon_j^o + \sum_{j=0}^{i} \mathbf{T}_{i-j}\,\varepsilon_j^m$$

with

$\mathrm{Var}(\mathbf{T}_{i+1}\,\varepsilon_0^b) \simeq 0 \quad$ for $i \gtrsim \tau^T$,
where $\tau^T \simeq 4$ days is the timescale over which old errors vanish,

$\mathrm{Cov}(\varepsilon_0^b,\ \varepsilon_j^o) = 0 \quad$ for time uncorrelated random observation errors,

$\mathrm{Cov}(\mathbf{T}_{i+1}\,\varepsilon_0^b,\ \mathbf{T}_{i-j}\,\varepsilon_j^m) \simeq 0 \quad$ when $i \gtrsim max(\tau^T, \tau^m)$
where τ^m is the correlation timescale of random model errors.

for $i \gtrsim max(\tau^T, \tau^m)$:

$$\mathrm{Var}(\varepsilon_i^f) = \mathrm{Var}\left(\sum_{j=0}^{i} \mathbf{T}_{i-j}\,\mathbf{M}_j\mathbf{K}_j\,\varepsilon_j^o\right) + \mathrm{Var}\left(\sum_{j=0}^{i} \mathbf{T}_{i-j}\,\varepsilon_j^m\right)$$

METEO FRANCE

Diagnosis of recent model error contributions

Contributions to forecast error variance :

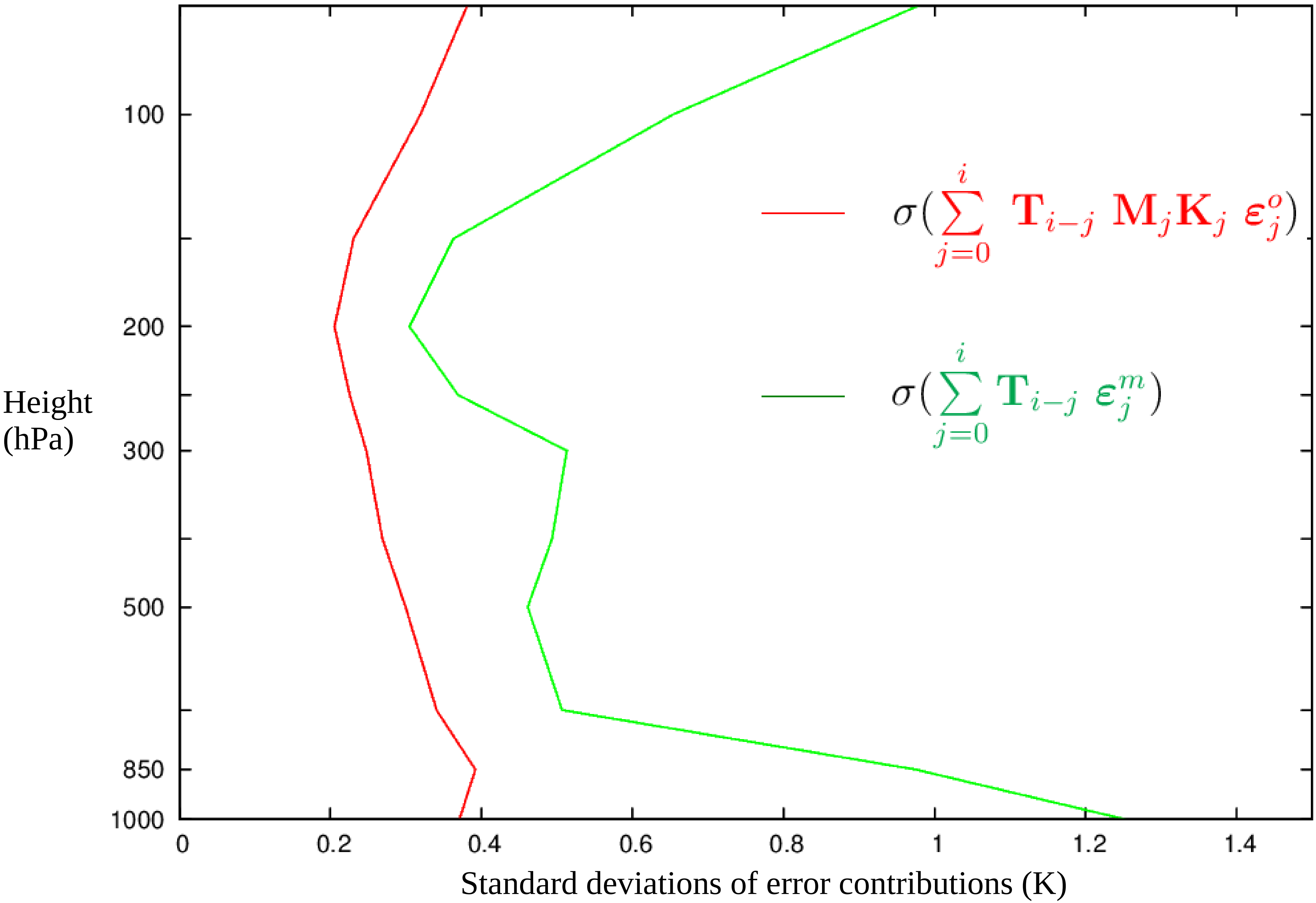
$$\text{Var}(\varepsilon_i^f) = \text{Var}(\sum_{j=0}^{i} \mathbf{T}_{i-j}\ \mathbf{M}_j\mathbf{K}_j\ \varepsilon_j^o) + \text{Var}(\sum_{j=0}^{i} \mathbf{T}_{i-j}\ \varepsilon_j^m)$$

which leads to the following estimation approach
(e.g. at day 4, considering $max(\tau^T, \tau^m) \simeq 4$ days) :

- $\text{Var}(\varepsilon_i^f)$ estimated by innovation-based diagnostics (e.g. Desroziers et al 2005);
- $\text{Var}(\sum_{j=0}^{i} \mathbf{T}_{i-j}\ \mathbf{M}_j\mathbf{K}_j\ \varepsilon_j^o)$ estimated by EDA with observation perturbations only ;
- $\text{Var}(\sum_{j=0}^{i} \mathbf{T}_{i-j}\ \varepsilon_j^m) = \text{Var}(\varepsilon_i^f) - \text{Var}(\sum_{j=0}^{i} \mathbf{T}_{i-j}\ \mathbf{M}_j\mathbf{K}_j\ \varepsilon_j^o)$

METEO FRANCE

Diagnosis of model error contributions versus observation error contributions



Conclusions

- Linear forecast error expansion to diagnose background, observation and model error contributions, with different ages.

- Global forecast error variance tends to be stable : compensation between damping of old errors (by successive analyses, within 4 days) and accumulation of recent errors (like a power series).

- Observation error contributions are significant, and model error contributions seem to be even larger.

- Extend this study to spatial correlation aspects, regional variations, etc. Possible use for calibration of model error representations.

Thank you for your attention