12th International EnKF workshop

Estimating model evidence using ensemble-based data assimilation with localization

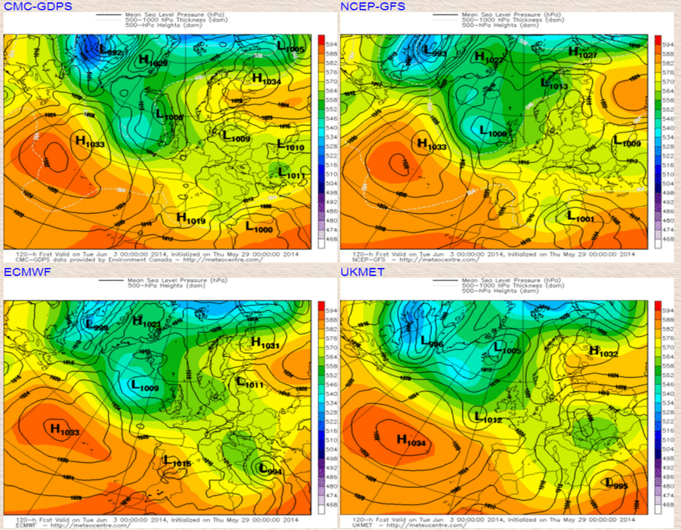
The model selection problem

Sammy Metref, Juan Ruiz, Alexis Hannart, Alberto Carrassi,
Marc Bocquet and Michael Ghil

Project DADA

June 12th, 2017

A comparison of geopotential heights at 500hPa for 4 short range models



Outline

Model evidence

For a model $\mathcal{M}$ simulating an unknown process such that:

$$\mathbf{x}_k = \mathcal{M}(\mathbf{x}_{k-1}), \tag{1}$$

where $\mathcal{M} : \mathbb{R}^M \to \mathbb{R}^M$.

And for an ideal infinite set of observations of the same process,

$$\mathbf{y}_{K:} = \{\mathbf{y}_K, \mathbf{y}_{K-1}, ..., \mathbf{y}_1, \mathbf{y}_0, ..., \mathbf{y}_{-\infty}\},$$

such that:

$$\mathbf{y}_k = \mathcal{H}_k(\mathbf{x}_k) + \epsilon_k, \tag{2}$$

where $\mathcal{H}_k : \mathbb{R}^M \to \mathbb{R}^d$ and ϵ_k represents observation error.

Model evidence (marginal likelihood of the observations)

$$p(\mathbf{y}_{K:}|\mathcal{M}) = \int \mathrm{d}\mathbf{x}\; p(\mathbf{y}_{K:}|\mathbf{x})p(\mathbf{x}). \tag{3}$$

Model evidence

For a model $\mathcal{M}$ simulating an unknown process such that:

$$\mathbf{x}_k = \mathcal{M}(\mathbf{x}_{k-1}), \tag{1}$$

where $\mathcal{M} : \mathbb{R}^M \to \mathbb{R}^M$.

And for an ideal infinite set of observations of the same process,

$$\mathbf{y}_{K:} = \{\mathbf{y}_K, \mathbf{y}_{K-1}, ..., \mathbf{y}_1, \mathbf{y}_0, ..., \mathbf{y}_{-\infty}\},$$

such that:

$$\mathbf{y}_k = \mathcal{H}_k(\mathbf{x}_k) + \epsilon_k, \tag{2}$$

where $\mathcal{H}_k : \mathbb{R}^M \to \mathbb{R}^d$ and ϵ_k represents observation error.

Model evidence (marginal likelihood of the observations)

$$p(\mathbf{y}_{K:}|\mathcal{M}) = \int \mathrm{d}\mathbf{x}\; p(\mathbf{y}_{K:}|\mathbf{x})p(\mathbf{x}). \tag{3}$$

Defined as a “climatological” model evidence

Model evidence using data assimilation

We rather define a contextual model evidence i.e. conditioned on the present

- $p(\mathbf{y}_{K:}|\mathcal{M}) \to p(\mathbf{y}_{K:1}|\mathbf{y}_{0:})$ [$\mathcal{M}$ is dropped for clarity]

In the context of present time, we marginalize over $\mathbf{x}_0$ and not over $\mathbf{x}$

The Contextual Model Evidence (CME)

$$p(\mathbf{y}_{K:1}|\mathbf{y}_{0:}) = \int \mathrm{d}\mathbf{x}_0 \; p(\mathbf{y}_{K:1}|\mathbf{x}_0)p(\mathbf{x}_0|\mathbf{y}_{0:}) \tag{4}$$

Model evidence using data assimilation

We rather define a contextual model evidence i.e. conditioned on the present

- $p(\mathbf{y}_{K:}|\mathcal{M}) \rightarrow p(\mathbf{y}_{K:1}|\mathbf{y}_{0:})$ [$\mathcal{M}$ is dropped for clarity]

In the context of present time, we marginalize over $\mathbf{x}_0$ and not over $\mathbf{x}$

The Contextual Model Evidence (CME)

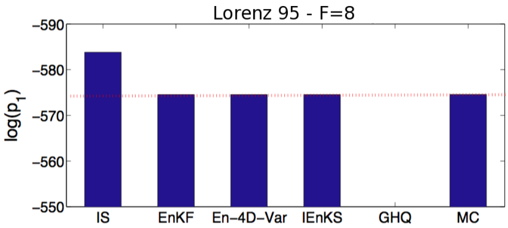
$$p(\mathbf{y}_{K:1}|\mathbf{y}_{0:}) = \int \mathrm{d}\mathbf{x}_0 \; p(\mathbf{y}_{K:1}|\mathbf{x}_0) p(\mathbf{x}_0|\mathbf{y}_{0:}) \tag{4}$$

with

- the likelihood of the observations
- the posterior density (state estimation DA product)

Estimating the CME using DA methods

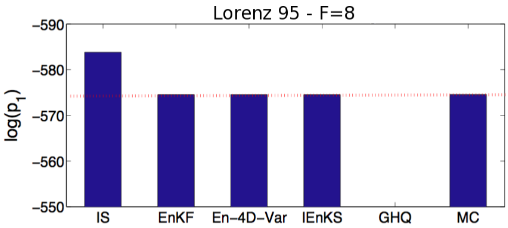
- ensemble Kalman filter
- 4D ensemble methods (En-4D-Var/IEnKS)



Carrassi et al. (2017)

Estimating the CME using DA methods

- ensemble Kalman filter
- 4D ensemble methods (En-4D-Var/IEnKS)



Carrassi et al. (2017)

Conclusions

- Accurate estimation of the CME using DA
- Accuracy related to DA method's sophistication
- Yet, not proportional

$\Rightarrow$ We use the EnKF formulation

CME formulation

The CME's EnKF formulation [$\mathbf{y}_{0:}$ is dropped for clarity]

$$p(\mathbf{y}_{K:1}) \approx \prod_{k=1}^{K} (2\pi)^{-\frac{d}{2}} |\boldsymbol{\Sigma}_k|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}[\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}^{\mathrm{f}}_k)]^{\mathrm{T}} \boldsymbol{\Sigma}_k^{-1} [\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}^{\mathrm{f}}_k)]\right\} \quad (5)$$

Hannart et al. (2016) ; *Carrassi et al. (2017)*

with $\boldsymbol{\Sigma}_k = \mathbf{H}_k \mathbf{P}^{\mathrm{f}}_k \mathbf{H}^{\mathrm{T}}_k + \mathbf{R}_k$ where

$\mathbf{P}^{\mathrm{f}}_k$: prior error covariance matrix at time k,
$\mathbf{R}_k$: observation error covariance matrix,

$\mathcal{H}_k$: observation operator at time k,
$\mathbf{H}_k$: its linearization.

CME formulation

The CME's EnKF formulation [$\mathbf{y}_{0:}$ is dropped for clarity]

$$p(\mathbf{y}_{K:1}) \approx \prod_{k=1}^{K} (2\pi)^{-\frac{d}{2}} |\boldsymbol{\Sigma}_k|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} [\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k^{\mathrm{f}})]^{\mathrm{T}} \boldsymbol{\Sigma}_k^{-1} [\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k^{\mathrm{f}})] \right\} \quad (5)$$

Hannart et al. (2016) ; *Carrassi et al. (2017)*

with $\boldsymbol{\Sigma}_k = \mathbf{H}_k \mathbf{P}_k^{\mathrm{f}} \mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k$ where

$\mathbf{P}_k^{\mathrm{f}}$: prior error covariance matrix at time k,
$\mathbf{R}_k$: observation error covariance matrix,

$\mathcal{H}_k$: observation operator at time k,
$\mathbf{H}_k$: its linearization.

The objective of this study

Problem in high dimension:

Ensemble DA methods suffer from sampling errors in high dimension and are usually used with localization

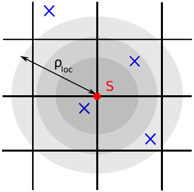
$\Rightarrow$ Crucial to consider how to deal with localization in the CME formulation

Domain localization

- Seperate analysis: DA performed for each model gridpoint $s \in \Gamma$
- Box car: Only the neighboring obs. are used in the analysis i.e. with $\mathbf{y}_{|s}$, $\mathbf{H}_{|s}$, $\mathbf{R}_{|s}$restricted to a disk around s of radius ρ_{loc}
- Tapering: a (diagonal) localization matrix $\mathbf{L}$ applied such that

$$\widetilde{\mathbf{R}}_{|s}^{-1} = \mathbf{L} \circ \mathbf{R}_{|s}^{-1} = (\mathbf{R}_{|s}^{-1})_{i,j} \cdot (\mathbf{L})_{i,j} \tag{6}$$

$(\mathbf{L})_{i,i}$ is equal to 1 if $i = s$ and decreases to 0 outside of the disk

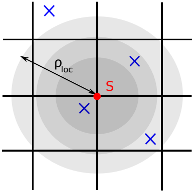


Domain localization

- Seperate analysis: DA performed for each model gridpoint $s \in \Gamma$
- Box car: Only the neighboring obs. are used in the analysis i.e. with $\mathbf{y}_{|s}$, $\mathbf{H}_{|s}$, $\mathbf{R}_{|s}$restricted to a disk around s of radius ρ_{loc}
- Tapering: a (diagonal) localization matrix $\mathbf{L}$ applied such that

$$\widetilde{\mathbf{R}}_{|s}^{-1} = \mathbf{L} \circ \mathbf{R}_{|s}^{-1} = (\mathbf{R}_{|s}^{-1})_{i,j} \cdot (\mathbf{L})_{i,j} \quad (6)$$

$(\mathbf{L})_{i,i}$ is equal to 1 if $i = s$ and decreases to 0 outside of the disk



$\Rightarrow$ Derive the CME for each gridpoint using $\mathbf{y}_{|s}$, $\mathbf{H}_{|s}$, $\widetilde{\mathbf{R}}_{|s}^{-1}$

DL-CME

At each gridpoint $s \in \Gamma$, it is possible to derive

$$p(\mathbf{y}_{K:1|s}) \approx \prod_{k=2}^{K} \int \mathrm{d}\mathbf{x}_k \, p(\mathbf{y}_{k|s}|\mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}|\mathbf{y}_{k-1:|s}) \int \mathrm{d}\mathbf{x}_0 \, p(\mathbf{y}_{1|s}|\mathbf{x}_0) p(\mathbf{x}_0|\mathbf{y}_{0:})$$

Local CME

$$p(\mathbf{y}_{K:1|s}) \approx \prod_{k=1}^{K} (2\pi)^{-\frac{\widetilde{d}}{2}} |\widetilde{\boldsymbol{\Sigma}}_k|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{y}_{k|s} - \mathbf{H}_{k|s}\mathbf{x}_k^{\mathrm{f}})^{\mathrm{T}} \widetilde{\boldsymbol{\Sigma}}_k^{-1} (\mathbf{y}_{k|s} - \mathbf{H}_{k|s}\mathbf{x}_k^{\mathrm{f}})\right\} \quad (7)$$

with $\widetilde{\boldsymbol{\Sigma}}_k = \mathbf{H}_{k|s}\mathbf{P}_k^{\mathrm{f}}\mathbf{H}_{k|s}^{\mathrm{T}} + \widetilde{\mathbf{R}}_{k|s}$ and $\widetilde{d}$ the size of $\mathbf{y}_{k|s}$.

DL-CME

At each gridpoint $s \in \Gamma$, it is possible to derive

$$p(\mathbf{y}_{K:1|s}) \approx \prod_{k=2}^{K} \int \mathrm{d}\mathbf{x}_k \, p(\mathbf{y}_{k|s}|\mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}|\mathbf{y}_{k-1:|s}) \int \mathrm{d}\mathbf{x}_0 \, p(\mathbf{y}_{1|s}|\mathbf{x}_0) p(\mathbf{x}_0|\mathbf{y}_{0:})$$

Local CME

$$p(\mathbf{y}_{K:1|s}) \approx \prod_{k=1}^{K} (2\pi)^{-\frac{\widetilde{d}}{2}} |\widetilde{\mathbf{\Sigma}}_k|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{y}_{k|s} - \mathbf{H}_{k|s}\mathbf{x}_k^{\mathrm{f}})^{\mathrm{T}} \widetilde{\mathbf{\Sigma}}_k^{-1} (\mathbf{y}_{k|s} - \mathbf{H}_{k|s}\mathbf{x}_k^{\mathrm{f}})\right\} \quad (7)$$

with $\widetilde{\mathbf{\Sigma}}_k = \mathbf{H}_{k|s}\mathbf{P}_k^{\mathrm{f}}\mathbf{H}_{k|s}^{\mathrm{T}} + \widetilde{\mathbf{R}}_{k|s}$ and $\widetilde{d}$ the size of $\mathbf{y}_{k|s}$.

Euristic global estimator

Domain localized CME (DL-CME)

$$\widetilde{p}(\mathbf{y}_{K:1}) = \exp\left\{\sum_{s\in\Gamma} w(s) \ln\{p(\mathbf{y}_{|s})\}\right\}, \quad (8)$$

with $w(s)$, scalar weights inversely proportional to the localization radius.

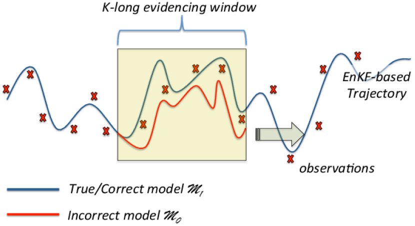
CME for model selection

Two models: $\mathcal{M}_0$ and $\mathcal{M}_1$
and their respective model evidences:

$$p_0(\mathbf{y}) = p(\mathbf{y}_{K:1}|\mathbf{y}_{0:}, \mathcal{M}_0) \text{ and } p_1(\mathbf{y}) = p(\mathbf{y}_{K:1}|\mathbf{y}_{0:}, \mathcal{M}_1)$$

Model selection indicator with global and domain localized CME:

- G-CME: $\Delta_p(\mathcal{M}_0, \mathcal{M}_1) = \ln\{p_1(\mathbf{y})\} - \ln\{p_0(\mathbf{y})\} > 0$, if $\mathcal{M}_1$ correct
- DL-CME: $\Delta_{\widetilde{p}}(\mathcal{M}_0, \mathcal{M}_1) = \ln\{\widetilde{p}_1(\mathbf{y})\} - \ln\{\widetilde{p}_0(\mathbf{y})\} > 0$, if $\mathcal{M}_1$ correct



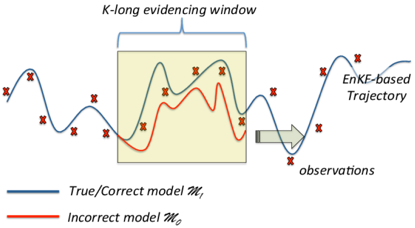
CME for model selection

Two models: $\mathcal{M}_0$ and $\mathcal{M}_1$
and their respective model evidences:

$$p_0(\mathbf{y}) = p(\mathbf{y}_{K:1}|\mathbf{y}_{0:}, \mathcal{M}_0) \text{ and } p_1(\mathbf{y}) = p(\mathbf{y}_{K:1}|\mathbf{y}_{0:}, \mathcal{M}_1)$$

Model selection indicator with global and domain localized CME:

- G-CME: $\Delta_p(\mathcal{M}_0, \mathcal{M}_1) = \ln\{p_1(\mathbf{y})\} - \ln\{p_0(\mathbf{y})\} > 0$, if $\mathcal{M}_1$ correct
- DL-CME: $\Delta_{\widetilde{p}}(\mathcal{M}_0, \mathcal{M}_1) = \ln\{\widetilde{p}_1(\mathbf{y})\} - \ln\{\widetilde{p}_0(\mathbf{y})\} > 0$, if $\mathcal{M}_1$ correct



The scope of the following experiments is to compare the G-CME's and the DL-CME's model selection abilities

L95 - Model selection problem

Lorenz-95 model

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + \mathrm{F}, \quad (9)$$

for $i = 1, ..., M = 40$ and F represents the external forcing.

The models

- $\mathcal{M}_1$: $\mathrm{F} \equiv \mathrm{F}_1 = 8$
- $\mathcal{M}_0$: $\mathrm{F} \equiv \mathrm{F}_0$ varying

for $T = 10^5$ DA cycles

The observations

$\mathcal{M}_1$ traj. perturbed: $\epsilon \in \mathcal{N}(0, 1)$

Obs. error cov. matrix: $\mathbf{R} = \mathbf{I}_{40}$

Obs. grid: $\Delta_t = 0.05$ and $\mathbf{H}_k = \mathbf{I}_{40}$

DA setup

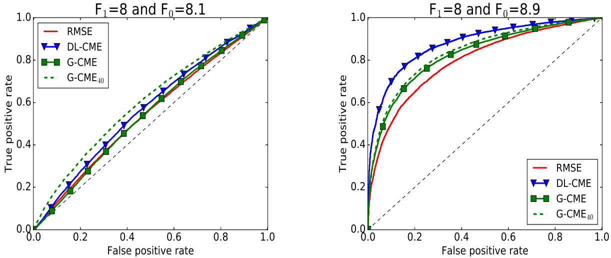
LETKF - 10 members

Localization radius: $\rho_{loc} = 5$ (tuned for $\mathcal{M}_0$)

Inflation: tuned for each model

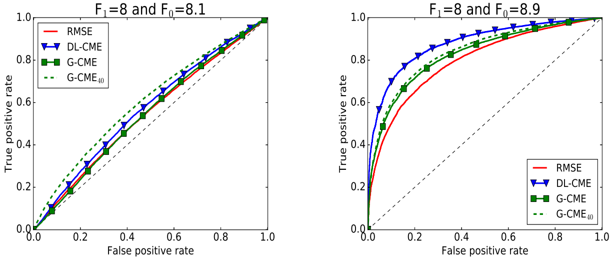
L95 - Sensitivity to the forcings

- ROC curves assess the quality of the selection indicators for various confidence thresholds, from a diagonal curve for random to 1 for perfect selection
- $\mathrm{F}_0 = 8.1$ and $\mathrm{F}_0 = 8.9$; $\rho_{loc} = 5$; $K = 1$



L95 - Sensitivity to the forcings

- ROC curves assess the quality of the selection indicators for various confidence thresholds, from a diagonal curve for random to 1 for perfect selection
- $F_0 = 8.1$ and $F_0 = 8.9$; $\rho_{loc} = 5$; $K = 1$



1- For $F_0 = 8.1$, all indicators close to random for the very close incorrect model

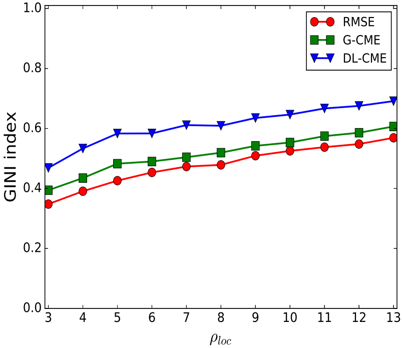
2- DL-CME still improves over the G-CME and the reference RMSE

3- The reference G-CME$_{40}$ remains the best indicator

4- For $F_0 = 8.9$, all indicators improve and the DL-CME outperforms G-CME$_{40}$

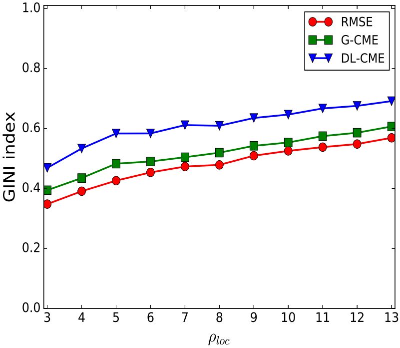
L95 - Sensitivity to localization

- GINI index quantifies a ROC curve performance, from 0 for random to 1 for perfect selection
- $F_0 = 8.5$; varying ρ_{loc} ; $K = 1$



L95 - Sensitivity to localization

- GINI index quantifies a ROC curve performance, from 0 for random to 1 for perfect selection
- $F_0 = 8.5$; varying ρ_{loc} ; $K = 1$



1- The two CMEs have better selecting skills than the reference RMSE

2- The DL-CME shows a constant improvment over the G-CME

⇒ The DL-CME improvment doesn't seem sensitive to the tuning of ρ_{loc}

SPEEDY - Model selection problem

The SPEEDY model (*Molteni, 2003*)

A global atmospheric model resolving the large scale dynamic

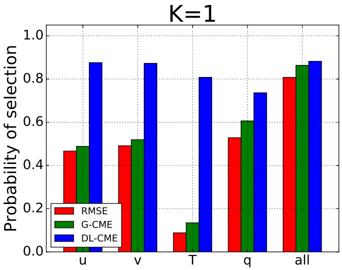
- Res.: $96 \times 48 \times 7 \sim \mathcal{O}(10^4)$
- Vor, Div, T, Q, $log(p_s)$
- Hydrostat., σ-coord, spectral-transf.
- Convect., condens., clouds, radiat.

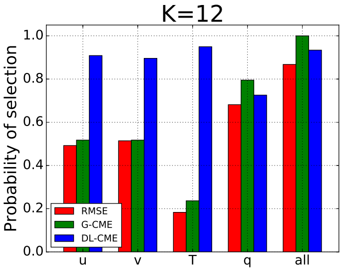
Twin experiment

- True trajectory: 5 month SPEEDY run (01/02-30/06/1983)
- 2 versions of the model: different convective relaxation time parameter
  - Correct parameter: $\tau_{cnv} = 6$hs
  - Incorrect parameter: $\tau_{cnv} = 5$hs50min
- Artificial observations on $[u, v, T, Q, p_s]$
  (Frequ.: 6h, Spat. distrib.: random on 1/2 x grid)
- DA: LETKF, 50 members *(Miyoshi, 2005, 2007)*

SPEEDY - Probability of selection

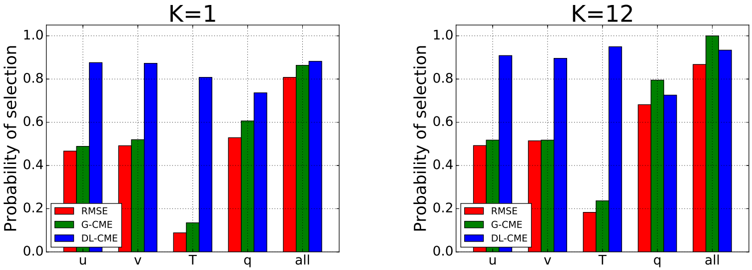
- Probabilities of selection: number of successfull selection
- DA using all obs. ; the CME computed for seperate var.
- $K = 1$ (6 hours) and $K = 12$ (3 days)





SPEEDY - Probability of selection

- Probabilities of selection: number of successfull selection
- DA using all obs. ; the CME computed for seperate var.
- $K = 1$ (6 hours) and $K = 12$ (3 days)



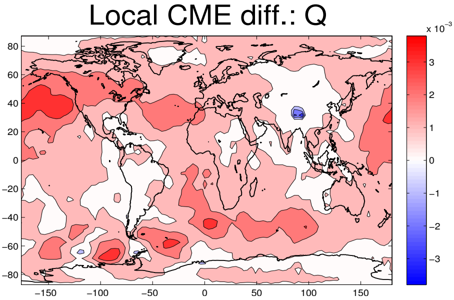
1- For (u,v,T), DL-CME has better selection skills (small impact of modified parameter)

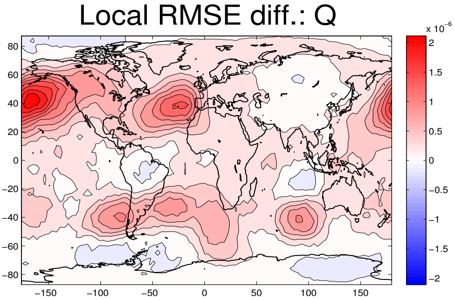
2- For Q, G-CME and DL-CME have closer selection skills

3- For $K = 12$, static covariance hyp. may be ill-adapted for long evidence window

Evidence maps

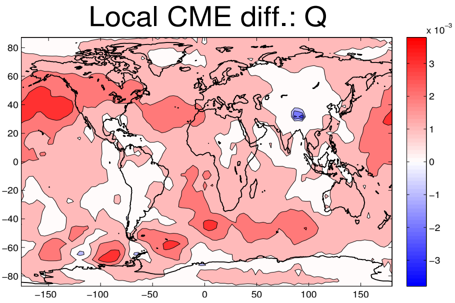
- Maps of differences for local CME and local RMSE averaged over 5 months

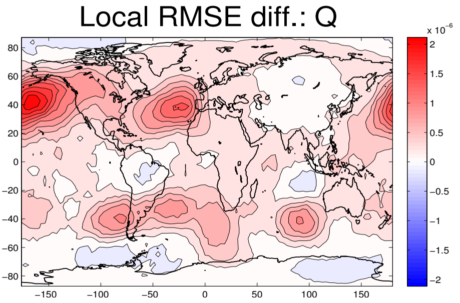




Evidence maps

- Maps of differences for local CME and local RMSE averaged over 5 months





1- The local CME map reveals different geographical information

2- This information could be used to understand the impact of the altered param.

Conclusions

- Model evidence is a useful statistic tool (*Winiarek et al., 2011 ; Elsheikh et al., 2014 ; Carson et al., 2016 ...*)
- *Carrassi et al. (2017)* proved a CME can be computed using DA
- We developed a new CME formulation taking into account localization for high dimensional applications
- We showed its skills as a model selection metric
- We exhibited the spatial diagnosing potential of local CME
- Applications of the CME:
  - Extreme event attribution (*Hannart et al., 2016*)
  - Parameter estimation (*Carrassi et al., 2017*)
  - Model selection (*Metref et al., 2017*)
  - Climate change attribution (Ongoing work)

References

- **Carrassi A., M. Bocquet, A. Hannart and M. Ghil:** *Estimating model evidence using data assimilation.* Q. J. R. Meteorol. Soc., 143: 866-880. **2017**
- **Carson J, Crucifix M., Preston S., and RD. Wilkinson:** *Bayesian model selection for the glacial-interglacial cycle.* ArXiv preprint arXiv:1511.03467. **2015**
- **Elsheikh A., I. Hoteit I and M. Wheeler:** *Efficient bayesian inference of subsurface flow models using nested sampling and sparse polynomial chaos surrogates.* Comput. Methods Appl. Mech. Engrg., 269: 515537. **2014**
- **Hannart A., A. Carrassi, M. Bocquet, M. Ghil, P. Naveau, M. Pulido, J. Ruiz and P. Tandeo:** *DADA: Data assimilation for the detection and attribution of weather- and climate-related events.* Clim. Change., 136: 155-174. **2016**
- **Metref S., J. Ruiz, A. Hannart, M. Bocquet, A. Carrassi and M. Ghil:** *Estimating model evidence using ensemble-based data assimilation with localization - The model selection problem.* Q. J. R. Meteorol. Soc. **In preparation**
- **Miyoshi T.:** *Ensemble Kalman filter experiments with a primitive-equation global model.* Ph.D. dissertation, University of Maryland. **2005**
- **Miyoshi T., S. Yamane and T. Enomoto:** . *Localizing the error covariance by physical distances within a Local Ensemble Transform Kalman Filter (LETKF).* SOLA, 3: 89-92. **2007**
- **Molteni F.:** *Atmospheric simulations using a GCM with simplified physical parameterizations. I: model climatology and variability in multi- decadal experiments.* Clim. Dyn., 20: 175-191. **2003**
- **Winiarek V., J. Vira, M. Bocquet, M Sofiev and O. Saunier:** *Towards the operational estimation of a radiological plume using data assimilation after a radiological accidental atmospheric release.* Atmos. Env., 45: 2944-2955. **2011**