Estimating model evidence using data assimilation

Application to the attribution of climate-related events

Marc Bocquet[1], Alberto Carrassi[2], Alexis Hannart[3] & Michael Ghil[4,5]
and the DADA Team

(1) CEREA, joint laboratory École des Ponts ParisTech and EDF R&D, Université Paris-Est, Champs-sur-Marne, France.
(2) Nansen Environmental and Remote Sensing Center, Bergen,Norway
(3) IFAECI, CNRS-CONICET-UBA, Buenos Aires, Argentina.
(4) École Normale Supérieure, Paris, France.
(5) University of California, Los Angeles, USA

Workshop on Stochastic Weather Generators
17-20 May, 2016, Vannes, France

Model evidence – Definition & Interpretation

▶ Model evidence $p(\mathbf{y}_{K:}|\,\mathcal{M}) = p(\mathbf{y}_{K:}) = p(\mathbf{y}_K, \mathbf{y}_{K-1}, \ldots, \mathbf{y}_{-\infty})$
$\Rightarrow$ Likelihood of *(the infinite from the far past)* observation sequence

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

- It is the marginal likelihood of the data.
- $p(\mathbf{x})$ plays the role of the *prior*.
- It depends on the underlying dynamics (the model).
- If the dynamics is ergodic $\Longrightarrow$ $p(\mathbf{x})$ is the invariant distribution on its attractor.
- $p(\mathbf{y}_{K:})$ reflects this invariant distribution at a level of accuracy dependent on the observation model.
- We define $p(\mathbf{y}_{K:})$ as the *climatological model evidence*.

Model evidence – Its applications

- The marginal likelihood can be used as a general metric for model selection and comparison.
- Comparing the skill of several candidate models (or model settings, or boundary conditions) in representing a given observed phenomenon.
- Calibrating the state-space model's parameters based on the observed data by maximizing the marginal likelihood.
- Quantifying the evidence supporting several (potentially conflicting) theoretical hypothesis associated to the physics of a phenomena.
- Evidencing the existence (or non-existence) of a causal relationship between an external forcing and an observed response.

Statistical/climatological model evidence

- The climatological model evidence embeds significant global information about the system of interest.
- The accuracy of its estimate is limited by the computational difficulty implied by the large dimension of typical relevant problems.
- Its climatological character does not permit to adapt it to the present condition of the system.
- The model evidence has been used in various contexts, including inverse problems.
- A typical example is the estimation of the error statistics in the source term inversion, or discriminate sources (*i.e.* nuclear release - Winiarek *et al.*, 2011 *Atmos. Env.*)

From statistical/climatological to conditional/contextual evidence

- We propose using data assimilation (DA) to
  1. compute model evidence,
  2. and/or narrow it to the present-time context.

Proposal: (*leap of faith*) use instead the conditional/contextual model evidence

$$p(\mathbf{y}_{K:}) \longrightarrow p(\mathbf{y}_{K:1}|\mathbf{y}_{0:})$$

- $\mathbf{y}_{K:1}$ is used for evidence diagnostic.
- $\mathbf{y}_{0:}$ is used to specify the context.

From statistical/climatological to conditional/contextual evidence

- It narrows the climate perspective to the context at t_0:

$$p(\mathbf{y}_{K:1}|\mathbf{y}_{0:}) = \int \mathrm{d}\mathbf{x}_0 \, p(\mathbf{y}_{K:1}|\mathbf{x}_0) p(\mathbf{x}_0|\mathbf{y}_{0:}) \tag{2}$$

  - The conditional pdf $p(\mathbf{x}_0|\mathbf{y}_{0:})$ plays the role of the *prior* density and substitutes $p(\mathbf{x})$ of the climatological model evidence.
  - $p(\mathbf{y}_{K:1}|\mathbf{y}_{0:})$ is much easier to compute than $p(\mathbf{y}_{K:})$ (but yet complicated in many relevant situations)
  - $p(\mathbf{x}_0|\mathbf{y}_{0:})$ is the natural *(albeit approximate)* outcome of the DA machinery.
  - We will show that $p(\mathbf{y}_{K:1}|\mathbf{x}_0)$ can also be estimated via DA.

Conditional/contextual evidence: A useful iterative formula

- Contextual model evidence can be computed iteratively:

$$p(\mathbf{y}_{K:1}|\mathbf{y}_{0:}) = \prod_{k=1}^{K} p(\mathbf{y}_k|\mathbf{y}_{k-1:})$$

  - The contextual model evidence is the product of the individual contextual model evidences!
  - The individual context evidence $p(\mathbf{y}_k|\mathbf{y}_{k-1:})$ is often a tractable output of DA schemes.
  - We will use this decomposition in the sequel to compute model evidence using DA:

$$p(\mathbf{y}_k|\mathbf{y}_{k-1:}) = \int \mathrm{d}\mathbf{x}_k \, p(\mathbf{y}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{y}_{k-1:}) \tag{3}$$

  $p(\mathbf{y}_k|\mathbf{x}_k)$ is the observation likelihood and $p(\mathbf{x}_k|\mathbf{y}_{k-1:})$ is the forecast state pdf at t_k.

How to compute contextual model evidence
1. Importance sampling/Monte Carlo

- Use a DA method up to t_0.
- This provides (an approximation of) $p(\mathbf{x}_0|\mathbf{y}_{0:})$.
- Compute the model evidence via *Importance Sampling/Monte Carlo*:
  1. sample N members from $p(\mathbf{x}_0|\mathbf{y}_{0:})$ and forecast from t_0 to t_T
  2. weight each member trajectory $\mathbf{x}_{T:0}^{(k)}$ according to their likelihood $p(\mathbf{y}_{T:1}|\mathbf{x}_{T:0}^{(k)})$, so as to obtain

$$p(\mathbf{y}_{T:1}|\mathbf{y}_{0:}) \approx \sum_{k=1}^{N} p(\mathbf{y}_{K:1}|\mathbf{x}_0^{(k)}) \frac{1}{N}$$

- Importance sampling (IS): use an ensemble-DA method up to $t_0 \Longrightarrow$ the assumption $N \ll M$ can be used in steps (1) and (2)
- Evaluation of model evidence using a particle filter proposed by Reich and Cotter, 2015.

2. Filtering: Linear/Gaussian case – Kalman filter

- Use a DA method up to t_0.
- This provides (an approximation of) $p(\mathbf{x}_0|\mathbf{y}_{0:})$.
- In the Gaussian linear case (Kalman filter), the likelihood and forecast pdfs are Gaussian
- (It possible to show that) the contextual model evidence reads

$$p(\mathbf{y}_{K:1}|\mathbf{y}_{0:}) = \prod_{k=1}^{K} \frac{\exp\left(-\frac{1}{2}\left\|\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k^{\mathrm{f}}\right\|^2_{\mathbf{R}_k+\mathbf{H}_k\mathbf{P}_k^{\mathrm{f}}\mathbf{H}_k^{\mathrm{T}}}\right)}{\sqrt{(2\pi)^d\left|\mathbf{R}_k + \mathbf{H}_k\mathbf{P}_k^{\mathrm{f}}\mathbf{H}_k^{\mathrm{T}}\right|}} \tag{4}$$

$\mathbf{R}_k$, $\mathbf{P}_k^{\mathrm{f}}$: observation and forecast error covariances – $\mathbf{H}_k$: observation operator.
Notation remark: $\|\mathbf{x}\|^2_{\mathbf{A}} = \mathbf{x}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{x}$.

2. Filtering: nonlinear/Gaussian case – ensemble Kalman filter

Using an ensemble Kalman filter (**EnKF**), the Kalman filter formula for model evidence now becomes

$$p(\mathbf{y}_{K:1}|\mathbf{y}_{0:}) \simeq \prod_{k=1}^{K} \frac{\exp\left(-\frac{1}{2}\left\|\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k^{\mathrm{f}}\right\|^2_{\mathbf{R}_k+\mathbf{Y}_k\mathbf{Y}_k^{\mathrm{T}}}\right)}{\sqrt{(2\pi)^d\left|\mathbf{R}_k+\mathbf{Y}_k\mathbf{Y}_k^{\mathrm{T}}\right|}}$$

- $\mathbf{Y}_k = \mathbf{H}_k\mathbf{X}_k$ are the normalized observation anomalies, and $\mathbf{X}_k$ the forecast anomalies.
- The inversion of $\mathbf{R}_k + \mathbf{Y}_k\mathbf{Y}_k^{\mathrm{T}}$ can be done in ensemble subspace.
- The estimation is exact when the models are linear, the initial condition and observation errors are Gaussian and when the ensemble perturbation span the full range of uncertainty.

3. Smoothing

- Use a DA method up to t_0.
- This provides (an approximation of) $p(\mathbf{x}_0|\mathbf{y}_{0:})$.
- Use a **smoothing DA method** to estimate $p(\mathbf{y}_{T:1}|\mathbf{y}_{0:})$
- Compute the evidence with a **saddle-point approximation** (Laplace method):

$$\begin{aligned} p(\mathbf{y}_{K:1}|\mathbf{y}_{0:}) &= \int \mathrm{d}\mathbf{x}_0 \, p(\mathbf{y}_{T:1}|\mathbf{x}_0)p(\mathbf{x}_0|\mathbf{y}_{0:}) \\ &\propto \int \mathrm{d}\mathbf{x}_0 \, \exp\left(-\frac{1}{2}\sum_{k=1}^{K} ||\mathbf{y}_k - H_k M_{k\leftarrow 0}(\mathbf{x}_0)||^2_{\mathbf{R}_k} + \ln p(\mathbf{x}_0|\mathbf{y}_{0:})\right) \\ &= \int \mathrm{d}\mathbf{x}_0 \, \exp\left(-\mathcal{L}(\mathbf{y}_{K:1}, \mathbf{x}_0)\right) \\ &\simeq \frac{\exp\left(-\mathcal{L}(\mathbf{y}_{K:1}, \mathbf{x}_0^*)\right)}{\sqrt{|\nabla^2_{\mathbf{x}_0}\mathcal{L}(\mathbf{x}_0^*)|}} \end{aligned} \tag{5}$$

- How to obtain $\mathbf{x}_0^*$ characterizes the problem: 4DVar, En-smoother...

3. Smoothing – En4DVar

Ensemble-4DVar (**En4DVar**) (see nomenclature by *Lorenc, 2013*)

- The state is parameterized in the ensemble space: $\mathbf{x}_0 = \overline{\mathbf{x}}_0 + \mathbf{X}_0\mathbf{w}$
- The evidence is marginalized over the coefficient vector $\mathbf{w}$:

$$p(\mathbf{y}_{K:1}|\mathbf{y}_{0:}) = \int \mathrm{d}\mathbf{w}\, p(\mathbf{y}_{K:1}|\mathbf{w})p(\mathbf{w}|\mathbf{y}_{0:}) \tag{6}$$

- Using the Laplace approximation on the analysis (the minimum $\mathbf{w}^\star$) the evidence reads:

$$p(\mathbf{y}_{K:1}|\mathbf{y}_{0:}) \simeq \frac{\exp\left(-\frac{1}{2}\sum_{k=1}^{K}\left\|\mathbf{y}_k - H_k\mathcal{M}_{k:0}(\mathbf{x}_0^\star)\right\|^2_{\mathbf{R}_k} - \frac{1}{2}\left\|\mathbf{w}^\star\right\|^2\right)}{\sqrt{(2\pi)^{Kd}\prod_{k=1}^{K}|\mathbf{R}_k|\left|\mathbf{I}_N + \sum_{k=1}^{K}\left(\mathbf{Y}_k^\star\right)^{\mathrm{T}}\mathbf{R}_k^{-1}\mathbf{Y}_k^\star\right|}} \tag{7}$$

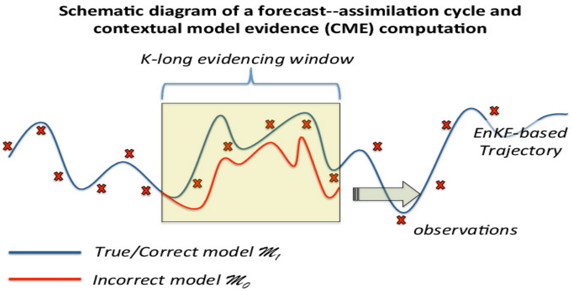
3 . Smoothing – IEnKS

(Quasi-Static) Iterative Ensemble Kalman Smoother (**IEnKS**)

- The initial condition and the ensemble at t_0 are sequentially updated first by assimilating $\mathbf{y}_1$ in the first step, then $\mathbf{y}_2$ in the second step and so on until $\mathbf{y}_K$.
- The analysis at k, $\mathbf{x}_k^\star$, and the normalized anomaly matrix, $\mathbf{X}_k^\star$, are both defined at t_0, and used in the subsequent step.
- Each single contextual evidence $p(\mathbf{y}_k|\mathbf{y}_{k-1:})$ corresponds to the quasi-static analysis of the IEnKS (Bocquet and Sakov, 2014).
- The iterative formula for model evidence and Laplace approximation on the analysis of the IEnKS gives

$$p(\mathbf{y}_{K:1}|\mathbf{y}_{0:}) \simeq \prod_{k=1}^{K} \frac{\exp\left(-\frac{1}{2}\left\|\mathbf{y}_k - H_k\mathcal{M}_{k:0}(\mathbf{x}_k^\star)\right\|^2_{\mathbf{R}_k+\mathbf{Y}_k^\star\left(\mathbf{Y}_k^\star\right)^{\mathrm{T}}}\right)}{\sqrt{(2\pi)^d\left|\mathbf{R}_k+\mathbf{Y}_k^\star\left(\mathbf{Y}_k^\star\right)^{\mathrm{T}}\right|}} \tag{8}$$

Model evidence – Numerical experiments



Forced L63 Model

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \sigma(y-x) + \lambda_i \cos\theta \quad \frac{\mathrm{d}y}{\mathrm{d}t} = \rho x - y - xz + \lambda_i \sin\theta \quad \frac{\mathrm{d}z}{\mathrm{d}t} = xy - \beta z \tag{9}$$

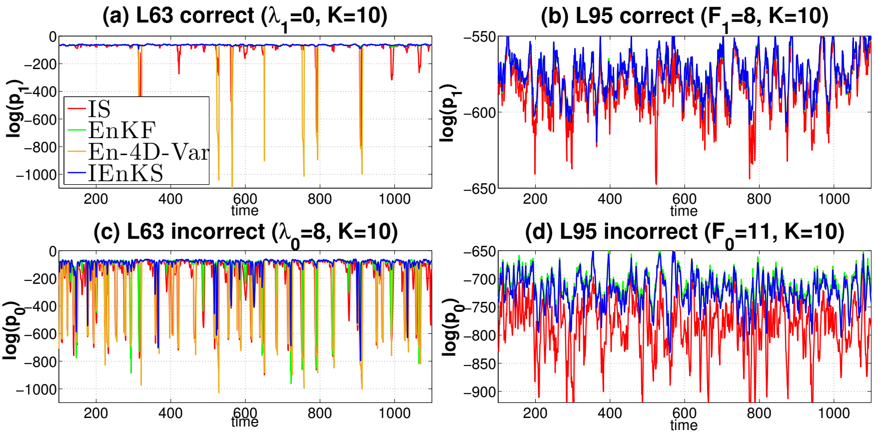
Correct $\lambda_1 = 0$; Incorrect $\lambda_0 \in \{-8:8\}$; $\theta = 140^\circ$.

L95 Model

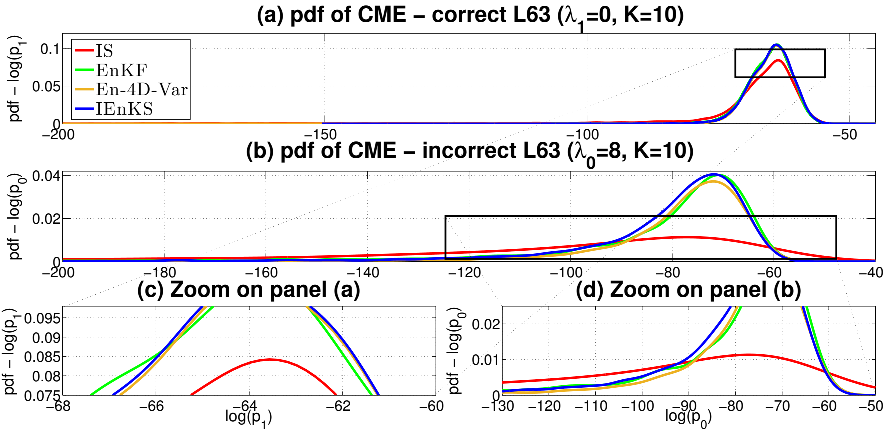
$$\frac{\mathrm{d}x_m}{\mathrm{d}t} = x_{m-1}\left(x_{m+1} - x_{m-2}\right) - x_m + F_i \quad m = 1, \dots, M \tag{10}$$

Correct $F_1 = 8$; Incorrect $F_0 \in \{5:11\}$.

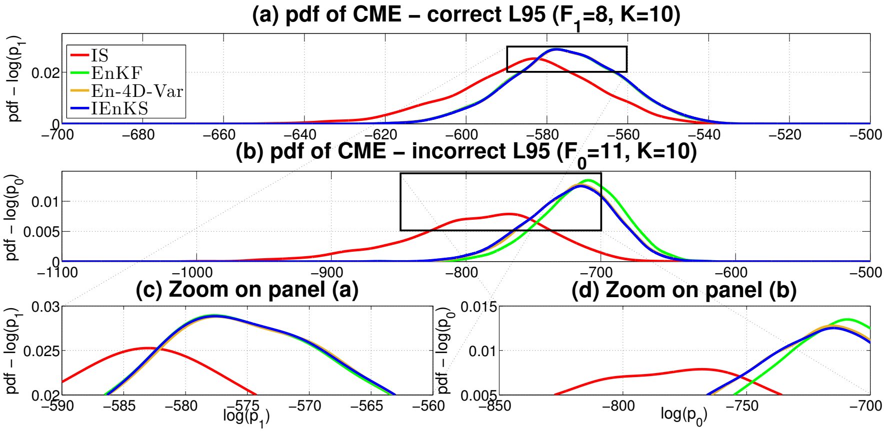
Model evidence time series



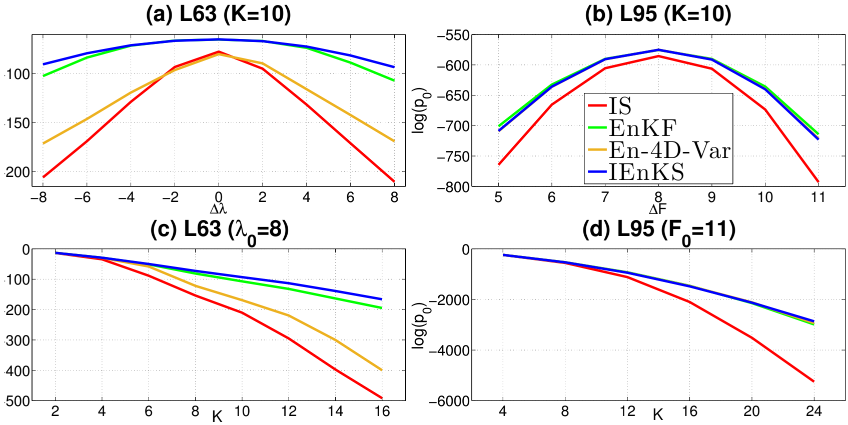
Pdf of model evidence – L63



Pdf of model evidence – L95



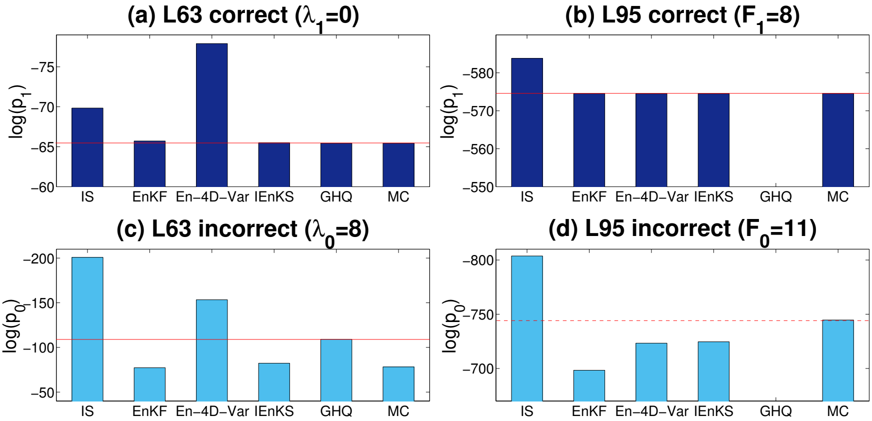
Model evidence *versus* forcing and length of window



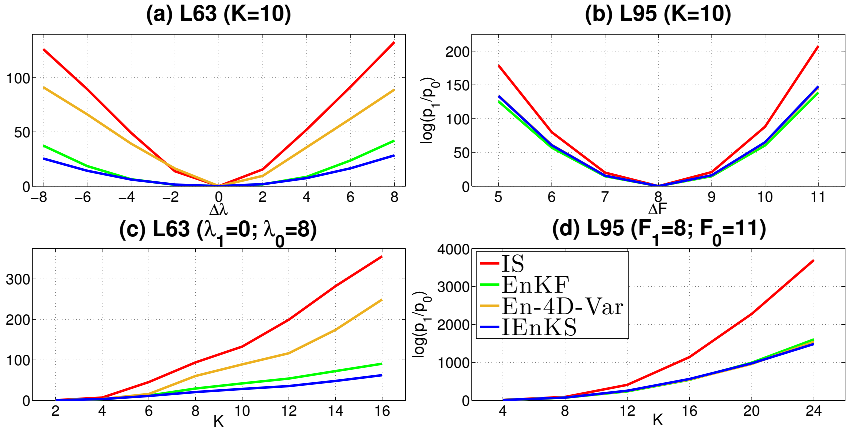
Computation of the integral: Which is right?

Comparison with:

1. Hermite-Gauss quadrature with 32 degrees - for L63
2. IS/Monte Carlo with 10^6 particles - for L63 and L95



Discriminating power *versus* forcing and length of window



The attribution problem

Detection and Attribution (D&A) : definition

- Formal definition (IPCC guidance paper, 2009):
  « *Attribution is the process of evaluating the relative contributions of multiple causal factors to a change or event with an assignment of confidence* »

⇒ Evidencing the causal influence of several factors

- Focus may be either on:
  - long term trends (e.g. global warming)
  - weather or climate-related events (e.g. heatwave)

Attribution – standard approach

- Define the event based on threshold exceedance of an ad-hoc climate index.
- Derive the likelihood of the event in two model ensembles:
  - factual world (e.g. HIST) => p1
  - counterfactual world (e.g. NAT) => p0
- Derive the « fraction of attributable risk » (FAR)

$$\mathrm{FAR} = 1 - \frac{p_0}{p_1}$$

« It is very likely (>90%) that CO_2 emissions have increased the frequency of occurrence of 2003-like heatwaves by a factor at least two.»

Limitations of the standard approach

- Limitations:
  - computationally costly procedure.
  - long delays in producing analysis after an event occurrence.

- The challenge is to design a system/procedure that allows for near real time event attribution.

  "The overarching challenge for the community is to move beyond research-mode case studies and to develop systems that can deliver regular, reliable and timely assessments in the aftermath of notable weather and climate-related events, typically in the weeks or months following (and not many years later as is the case with some research-mode)." Stott et al. (2013)

our philosophy: find ways to piggyback on meteorological centers' infrastructure, rather than build up a new one from scratch.

Event definition: Standard approach *versus* DADA

1. **Standard Approach** - Event occurrence based on *ad hoc* condition like a scalar index $\Phi(\mathbf{Y})$ exceeding a threshold $u \Rightarrow \mathbf{p}_i = \mathbf{P}(\Phi(\mathbf{Y}) \geq u)$
   - Hard to compute - Mathematically intractable in many practical cases
   - Simulation based
2. **DADA Proposal** - Use of the tightest possible occurrence definition $\Rightarrow$ $f_i(\mathbf{y}) = \{\omega \in \boldsymbol{\Omega} \quad | \quad ||\mathbf{Y}(\omega) - \mathbf{y}|| \leq h\} \quad h \to 0$ and $FAR = 1 - \frac{f_0(\mathbf{y})}{f_1(\mathbf{y})}$
   - In DADA we wish to evaluate $f_{0,1}(\mathbf{y})$ - The likelihood (model evidence) associated to the observation $\mathbf{y}$ of the event of interest in each two worlds (factual and counter-factual)
   - Easy to estimate! (sometimes ...)
   - The **model evidence** can be obtained as a **side-product of DA procedures** aimed at inferring the state-vector

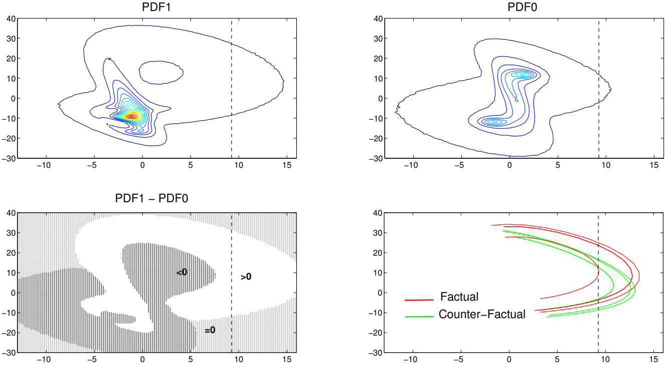
The D&A "grand challenge" associated to the design of an operational platform for near real time attribution of weather events may be solved by piggybacking on existing Data Assimilation meteorological routines.

Extreme event attribution – DADA approach

- Forced L63 model with stochastic noise (Palmer, 1999, *J. Climate*):

$$\frac{dx}{dt} = \sigma(y-x) + \lambda_i \cos\theta_i + v_x \quad \frac{dy}{dt} = \rho x - y - xz + \lambda_i \sin\theta_i + v_y \quad \frac{dz}{dt} = xy - \beta z + v_z$$

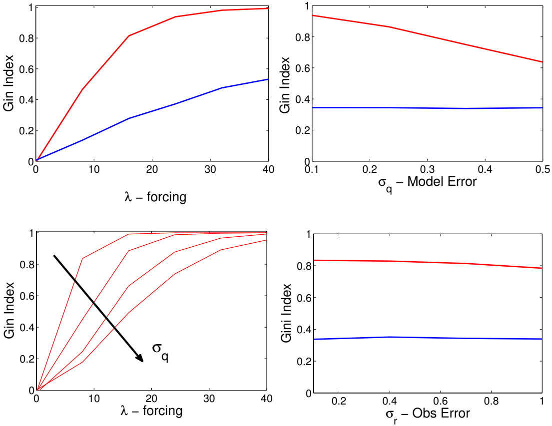
- $(v_x, v_y, v_z) \in \mathcal{N}(0, \sigma_Q \mathbf{I}_3)$
- Factual $\lambda_1 = 20$; Counter-factual $\lambda_0 = 0$



DADA application – Forced L63 Model

GINI index is a measure of the discriminating skill, ranging from 0 for no discrimination to 1 for perfect discrimination

DADA - STANDARD



Conclusion

- Data assimilation is used to compute model evidence
  1. in the contextualization $\Rightarrow$ (ensemble)-DA method used up to t_0,
  2. in the model evidence computation $\Rightarrow$ (ensemble)-DA method used to assimilate from t_1 to t_T.
- The accuracy of this DA-based estimate scales with the sophistication of the DA method (IS$\rightarrow$ EnKF $\rightarrow$ IEnKS).
- To be (further) assessed upon exact calculation of $p(\mathbf{y}_{T:0})$ (via quadrature or massive Monte Carlo).
- The model evidence is used as a new way to address the problem of attribution of climate-related event - DADA.
- The DADA approach has proven to be effective on low-order models and respond to the need of timely attribution evaluation.

Conclusion

Fundamental questions that requires attention:

1. The choice of the DA method is not trivial and (unsurprisingly) related to the degree of nonlinearity of the problem.
2. Model error is a central concern.
3. It is so also for standard methods of attribution.
4. Indeed it can mask the factor whose casual attribution is under scrutiny $\Rightarrow$ confuse discrimination between factual and counter-factual worlds.
5. A DA-based method for attribution has the potential to deal with issue taking advantage of the knowledge in model error treatment in DA schemes.

Key references

On the DADA approach

- Carrassi A., M. Bocquet, A. Hannart and M. Ghil, 2016: Estimating model evidence using Data assimilation. *Submitted to Q.J.Roy.Meteorol.Soc.*
- Hannart A., A. Carrassi, M. Bocquet, M. Ghil, P. Naveau, M. Pulido, J. Ruiz and P. Tandeo, 2016. DADA: Data Assimilation for the Detection and Attribution of Weather- and Climate-related Events. *Climatic Change*, IN PRESS

On the IEnKS

- Bocquet M. and P. Sakov. An iterative ensemble Kalman smoother. *Q. J. R. Meteorol. Soc.*, **140**:1521-1535, 2014.

On the use of model evidence

- 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.
- 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.
- Reich, S. and Cotter, C. Probabilistic Forecasting and Bayesian Data Assimilation. Cambridge University Press., chapter 9.1, 2015.