

# *DADA: data assimilation for the detection and attribution of weather and climate-related events*

Article

Accepted Version

Hannart, A., Carrassi, A., Bocquet, M., Ghil, M., Naveau, P., Pulido, M., Ruiz, J. and Tandeo, P. (2016) DADA: data assimilation for the detection and attribution of weather and climate-related events. *Climatic Change*, 136 (2). pp. 155-174. ISSN 0165-0009 Available at <http://centaur.reading.ac.uk/90280/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

Publisher: Springer

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in

the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

## **CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

1 **DADA: Data Assimilation for the Detection and**  
2 **Attribution of Weather and Climate-related Events**

3 **A. Hannart · A. Carrassi · M. Bocquet ·**  
4 **M. Ghil · P. Naveau · M. Pulido · J.**  
5 **Ruiz · P. Tandeo**

6  
7 Received: date / Accepted: date

8 **Abstract** We describe a new approach that allows for systematic causal at-  
9 tribution of weather and climate-related events, in near-real time. The method  
10 is designed so as to facilitate its implementation at meteorological centers by

---

A. Hannart  
IFAECI, CNRS-CONICET-UBA  
Pab. II, piso 2, Ciudad Universitaria  
1428 Buenos Aires, Argentina  
Tel.: +5411-4787-2693  
Fax: +5411-4788-3572  
E-mail: alexis.hannart@cima.fcen.uba.ar

A. Carrassi  
Mohn-Sverdrup Center, Nansen Environmental and Remote Sensing Center, Bergen, Norway

M. Bocquet  
CEREA, joint laboratory École des Ponts ParisTech and EDF R&D, Université Paris-Est,  
Champs-sur-Marne, France

M. Ghil  
Ecole Normale Supérieure, Paris, France  
University of California, Los Angeles, USA

P. Naveau  
LSCE, CNRS, Gif-sur-Yvette, France

M. Pulido  
Dept. of Physics, Universidad Nacional del Nordeste, Corrientes, Argentina

J. Ruiz  
IFAECI, CNRS/CONICET/UBA, Buenos Aires, Argentina

P. Tandeo  
Télécom Bretagne, Brest, France

relying on data and methods that are routinely available when numerically forecasting the weather. We thus show that causal attribution can be obtained as a by-product of *data assimilation* procedures run on a daily basis to update numerical weather prediction (NWP) models with new atmospheric observations; hence, the proposed methodology can take advantage of the powerful computational and observational capacity of weather forecasting centers. We explain the theoretical rationale of this approach and sketch the most prominent features of a “data assimilation–based detection and attribution” (DADA) procedure. The proposal is illustrated in the context of the classical three-variable Lorenz model with additional forcing. The paper concludes by raising several theoretical and practical questions that need to be addressed to make the proposal operational within NWP centers.

**Keywords** Event attribution · Data assimilation · Causality theory · Modified Lorenz model

## 1 Background and motivation

Providing causal assessments about episodes of extreme weather or unusual climate conditions is an important topic in the climate sciences: it arises from the multiple needs for public dissemination, litigation in a legal context, adaptation to climate change or simply improvement of the science associated with these events (Stott et al., 2013). The approach widely used so far was introduced one decade ago by M.R. Allen and colleagues (Allen, 2003; Stone and Allen, 2005); it originates from best practices in epidemiology (Greenland and Rothman, 1998) and is referred to as probabilistic event attribution (PEA).

In the PEA approach, one evaluates the extent to which a given external climate forcing — such as solar irradiation, greenhouse gas (GHG) emissions, ozone or aerosol concentrations — has changed the probability of occurrence of an event of interest. For this purpose, one thus needs to compute two probabilities: (i) the probability of occurrence of the event in an ensemble of model simulations representing the observed climatic conditions, which simulates the actual occurrence probability in the real world, referred to as *factual*; and (ii) the probability of occurrence of the event in a second ensemble of model simulations, representing this time the alternative world that might have occurred had the forcing of interest been absent, referred to as *counterfactual*.

Denoting by  $p_1$  and  $p_0$  the probabilities of the event occurring in the factual world and in the counterfactual world respectively, the so-called fraction of attributable risk (FAR) is then defined:

$$\text{FAR} = 1 - \frac{p_0}{p_1} \quad (1)$$

The FAR has long been interpreted as the fraction of the change in likelihood of an event which is attributable to the external forcing. Over the past decade, most causal claims have been following from the FAR and its uncertainty, resulting in statements such as “*It is very likely that over half the risk of*

European summer temperature anomalies exceeding a threshold of  $1.6^\circ\text{C}$  is attributable to human influence.” (Stott et al., 2004). Hannart et al. (2015) have recently shown that, under realistic assumptions, the FAR may also be interpreted as the so-called *probability of necessary causation* (PN) associated — in a complete and self-consistent theory of causality (Pearl, 2000) — with the causal link between the forcing and the event. The FAR thus corresponds to only one of the two facets of causality in such a theory, while the *probability of sufficient causation* (PS) is its second facet. In this setting,

$$\text{PN} = 1 - \frac{p_0}{p_1}, \quad (2a)$$

$$\text{PS} = 1 - \frac{1 - p_1}{1 - p_0}, \quad (2b)$$

$$\text{PNS} = p_1 - p_0, \quad (2c)$$

where PNS is the *probability of necessary and sufficient causation*. Pearl (2000) provides rigorous definitions of these three concepts, as well as a detailed discussion of their meanings and implications. It can be seen from Eqs. (2) that causal attribution requires to evaluate the two probabilities,  $p_0$  and  $p_1$ , which is therefore the central methodological question of PEA.

So far, most case studies have used large ensembles of climate model simulations in order to estimate  $p_1$  and  $p_0$  based on a variety of methods. However, this general approach has a very high computational cost and is difficult to implement in a timely and systematic way. As recognized by Stott et al. (2013), this remains an open problem: “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 studies)”. Several research initiatives are presently addressing this real-time attribution challenge. For instance, the **weather@home** system (Massey et al., 2014) in the context of the World Weather Attribution initiative (<http://www.climatecentral.org/wwa>), the system proposed by Christidis et al. (2013), or the Weather Risk Attribution Forecast system (<http://www.csag.uct.ac.za/~daithi/forecast/>) aim at meeting those requirements within the conventional ensemble-based approach.

The purpose of this article is to introduce a new methodological approach that addresses the latter overarching operational challenge. Our proposal relies on a class of powerful statistical methods for interfacing high-dimensional models with large observational datasets. This class of methods originates from the field of weather forecasting and is referred to as *data assimilation* (DA) (Bengtsson et al., 1981; Ghil and Malanotte-Rizzoli, 1991; Talagrand, 1997).

Section 2 explains the rationale of the approach proposed herein, presents a brief overview of DA, and outlines the most prominent technical features of a “data assimilation–based detection and attribution” (DADA) approach.

95 Section 3 illustrates the proposal by implementing it on a version of the clas-  
96 sical Lorenz convection model (Lorenz, 1963, L63 hereafter) subject to an  
97 additional constant force. Finally, in Section 4, we discuss the main strengths  
98 and limitations of the DADA approach, and highlight several theoretical and  
99 practical research questions that need to be addressed to make it potentially  
100 operational within weather forecasting centers in a near future.

## 101 2 Methodology

### 102 2.1 General rationale

103 In an operational context, a significant difficulty of PEA is that events of inter-  
104 est are usually rare, i.e. they occur in regions of the climate system’s attractor  
105 that are reached quite rarely. It may hence require a very large ensemble of  
106 simulations for the numerical model representing the climate system to reach  
107 the relevant region of the attractor. This requirement is particularly relevant  
108 if the event is defined narrowly, based on multiple features that might involve  
109 some combination of the atmospheric circulation, of the climate system’s ther-  
110 modynamic state, and of the impacts associated with the event. Simulating a  
111 sufficiently large number of occurrences of such an event for a robust evalua-  
112 tion of  $p_1$  and  $p_0$  may then be computationally very costly, and a brute force  
113 approach based on an unconstrained ensemble may become unaffordable in an  
114 operational context.

115 The first general idea underlying the DADA proposal is that the latter com-  
116 putational burden may be greatly reduced by constraining the model to ex-  
117 plore only the relevant region of its state space where the event under scrutiny  
118 is defined to occur. Such a selective exploration of a high-dimensional state  
119 space is not new. The constrained simulation of very rare events using com-  
120 plex dynamical models has been studied extensively (e.g., Harris and Kahn  
121 (1951); Del Moral and Garnier (2005)) and is referred to as Rare Event Sam-  
122 pling (RES). RES methods are based on importance sampling and probabilis-  
123 tic large-deviation theory (Bucklew, 2004), and they are commonly used in  
124 several areas — such as queueing, reliability, telecommunication (Heidelberg,  
125 1995) — but their adaptation to a climate context has only recently started  
126 (Wouters and Bouchet, 2015).

127 The second general idea of the DADA proposal is to take a shortcut along  
128 this path: DA methods present the key advantage of being already operational  
129 in weather forecasting centers to routinely update an atmospheric model with  
130 new observations in order to initialize the forecast, and we argue that they  
131 can be used simultaneously to solve the class of problems addressed by RES  
132 methods. Carrassi et al. (2008, and references therein) have already used a  
133 similarly selective exploration of a reduced number of phase space dimensions  
134 in the context of DA methods designed to control chaotic dynamics.

135 For the purposes of PEA, we show that, by assimilating the observed trajec-  
136 tory of an event into a model, one can obtain as a by-product the probability

density function (PDF) associated with this trajectory. PEA is then obtained by assimilating the observations of the event twice, first in the factual setting of the model and second in its counterfactual setting, and then by computing the FAR as the ratio of the two PDF values thus obtained.

Heuristically speaking, if an observed event is incompatible with the counterfactual world but compatible with the factual one — according to the standard approach of defining the existence of a causal link (Pearl, 2000; Hannart et al., 2015) — then assimilation will act as a *crucial experiment*, since the event’s observed trajectory will be easy to assimilate in the factual setting and difficult to assimilate in the counterfactual one, merely because the counterfactual setting physically precludes the existence of such a trajectory.

In Subsection 2.2, we formulate this general rationale in probabilistic terms and discuss the relevance of the approach. We then show in Subsection 2.3 that, given a similar set of hypotheses as those that underly the majority of operational DA methods, it is possible to quantify the extent to which an observed trajectory is compatible with the model physics — either factual or counterfactual — or not. This quantification in an operational context is at the core of the DADA approach and it greatly facilitates real-time PEA.

## 2.2 Probabilistic description of the method

Let  $\mathbf{y}_t$  denote the  $d$ -dimensional vector of observations at discrete times  $\{t = 0, 1, \dots, T\}$ . Here,  $\mathbf{y} = \{\mathbf{y}_t : 0 \leq t \leq T\}$  corresponds, for instance, to the full set of all available meteorological observations over a time interval covering the event of interest, no matter the diversity and source of the data; typically, the latter include ground station networks, satellite measurements, ship data, and so on, cf. Bengtsson et al. (1981, Preface, Fig. 1) or Ghil and Malanotte-Rizzoli (1991, Fig. 1). In the present probabilistic context of PEA, the observed trajectory  $\mathbf{y}$  is viewed as a realization of a random variable denoted  $\mathbf{Y} = \{\mathbf{Y}_t : 0 \leq t \leq T\}$ , i.e. there exists an  $\omega \in \Omega$  such that  $\mathbf{Y}(\omega) = \mathbf{y}$  — where  $\Omega$  denotes the sample space of all possible outcomes and encompasses observational error, as well as internal variability.

In event attribution studies, it is recognized that defining the *occurrence* of an event, i.e. selecting a subset  $\mathcal{F} \subset \Omega$ , depends on a rather arbitrary choice. Yet this choice has been shown to greatly affect causal conclusions (Hannart et al., 2015). For instance, a generic and fairly loose event definition is arguably prone to yield a low level of evidence with respect to both necessary and sufficient causality while, on the other hand, a tighter and more specific event definition is prone to yield a stringent level for necessary causality but a reduced one for sufficient causality.

Indeed, it is quite intuitive that many different factors should usually be *necessary* to trigger the occurrence of a highly specific event and conversely, that no single factor will ever hold as a *sufficient* explanation thereof. For the class of *unusual* events at stake in PEA, where both  $p_0$  and  $p_1$  are very small, we arguably lean towards specific definitions that inherently result in

few sufficient causal factors or none. This conclusion immediately follows from Eq. (1b), which yields  $PS \simeq 0$  when both  $p_0$  and  $p_1$  are very small.

Usually, an event occurrence is defined in PEA based on an *ad hoc* scalar index  $\phi(\mathbf{Y})$  exceeding a threshold  $u$ , i.e.  $p_i = P(\phi(\mathbf{Y}) \geq u)$ ; from now on, we associate  $i = 0$  with the counterfactual and  $i = 1$  with the factual world. While this definition may be already quite restrictive for  $u$  large, it is defensible to restrict the event definition even further. Such a strategy may reduce an already negligible PS but it also may increase PN by a greater amount; one thus expects to gain more than is lost in this trade-off. In particular, this will be the case if additional features, not accounted for in  $\phi(\mathbf{Y})$ , can be identified that will allow one to further discriminate between the two worlds.

Following this strategy, a central element of our proposal is to use the tightest possible event occurrence definition, i.e. the trajectory  $\mathbf{y}$  exactly as it was observed, namely the singleton event  $\{\mathbf{Y} = \mathbf{y}\}$ . This singleton event has probability zero in both worlds, i.e.  $p_1 = p_0 = 0$ . Indeed, the full sequence of observations  $\mathbf{y}$ , exactly as it occurred, is unique. Quoting the Greek philosopher Heraclitus “*You cannot step into the same river twice, for other waters are continually flowing in*”: the exact same sequence  $\mathbf{y}$  never occurred before and will never occur again. Our proposed singleton event definition may thus arguably match with the suggestion of Trenberth et al. (2015) that “*a different framing is desirable which asks why extremes unfold the way they do*” in so far as it focuses on the event exactly as it happened and is thereby able to spot the detailed physical features of the event that made it “*unfold the way it did*”. However, by contrast with Trenberth et al. (2015), our proposed singleton event definition is not conditional on the circulation: the observed vector  $\mathbf{y}$  may perfectly include circulation-related observations.

One may find surprising that a causal analysis of such a zero probability event is possible. However, in the context of the aforementioned causal theory, such a causal analysis is definitely possible and meaningful. Indeed, the fact that  $p_1$  and  $p_0$  are null does not imply that the associated probability of necessary causation PN is null. Generally speaking, the ratio of two quantities that tends to zero may well converge to a finite quantity (e.g. the derivative of a differentiable function). Likewise, here the singleton set  $\{\mathbf{Y} = \mathbf{y}\}$  may be viewed as the limit of the sphere of radius  $r$  centered in  $\mathbf{y}$  when the radius  $r$  tends to zero, i.e.  $\{\mathbf{Y} = \mathbf{y}\} = \lim_{r \rightarrow 0} \{\|\mathbf{Y} - \mathbf{y}\| \leq r\}$ . It is clear that when  $r \rightarrow 0$ , then  $p_0 \rightarrow 0$  and  $p_1 \rightarrow 0$ . It is also straightforward to show that the limit of  $PN = 1 - p_0/p_1$  is then finite. More specifically, we have:

$$PN = 1 - \frac{f_0(\mathbf{y})}{f_1(\mathbf{y})} \quad (3)$$

where we denote  $f_i$  the PDF of  $\mathbf{Y}$  in world  $i$ . By contrast, the quantity  $1 - (1 - p_1)/(1 - p_0)$  converges to zero when  $p_0$  and  $p_1$  tends to zero, thus the probability of sufficient causation PS associated with the singleton event  $\{\mathbf{Y} = \mathbf{y}\}$  is always zero. Our DADA proposal thus intentionally sacrifices the evidence of sufficiency, in the hope of maximizing the evidence of necessity.



Our betting on the singleton set is thus justifiable already based on the above theoretical considerations. This choice, moreover, is also motivated by having a highly simplifying implication from a practical standpoint. Evaluating the PDF of  $\mathbf{Y}$  at a single point  $\mathbf{Y} = \mathbf{y}$  is indeed, under many circumstances, considerably easier than evaluating the probability  $P(\phi(\mathbf{Y}) \geq u)$  required in the conventional approach. Appendix A gives a concrete illustration of this situation, and Figure 1 shows the details of the latter evaluation for a scalar AR(1) process (panel *a*, as well as its associated accuracy (panels *b* and *c*), and the computational cost as the sample size  $n$  varies (panel *d*); the latter cost is much larger than the one of applying the DADA approach consisting in evaluating the PDF at a single point. This simple example confirms the large computational discrepancy between the two approaches. The reason for the discrepancy is quite simple: evaluating the conventional probability requires integrating a PDF over a predefined domain, instead of a one-off evaluation at a single point. Because both the domain of integration and the PDF may have potentially complex shapes, one cannot expect, in general, that the requisite integral be amenable to analytical treatment. Hence numerical integration is the default option: no matter how efficient an integration scheme one applies, it will require evaluating the PDF at many points and is thus as many times more costly computationally than just evaluating  $f(\mathbf{y})$  at a single point.

In order to obtain the PDF of  $\mathbf{Y}$ , the class of dynamic, statistical models referred to as *Hidden Markov Models* (HMMs; e.g. Ihler et al. (2007)) is relevant in the context of PEA. Indeed, the dynamics of a climate event can usually be represented by using a numerical climate model. Denoting  $\mathbf{X}_t$  the  $N$ -dimensional state vector at time  $t$  of the numerical model, we can assume:

$$\mathbf{X}_{t+1} = \mathbf{M}(\mathbf{X}_t, \mathbf{F}_t) + \mathbf{v}_t, \quad (4a)$$

$$\mathbf{Y}_t = \mathbf{H}(\mathbf{X}_t) + \mathbf{w}_t \quad (4b)$$

where Equation (4a) describes the dynamics of the state vector, with  $\mathbf{M}$  the numerical model operator,  $\mathbf{v}_t$  a stochastic term representing modeling error, and  $\mathbf{F}_t$  a prescribed forcing. Equation (4b) maps the state vector  $\mathbf{X}_t$  to our observations  $\mathbf{Y}_t$  at any time  $t$ , where  $\mathbf{H}$  is the so-called observation or forward operator and  $\mathbf{w}_t$  is a stochastic term representing observational error. The problem of interest here is thus to derive the likelihoods  $f_0(\mathbf{y})$  and  $f_1(\mathbf{y})$  of the observation  $\mathbf{y}$  when using the counterfactual and factual forcings, by using the HMM setting of Equation (4).

DA can be viewed as a class of inference methods designed for the above HMM setting. While inferring the unknown state vector trajectory  $\mathbf{X}$  given the observed trajectory  $\mathbf{y}$  is the main focus of DA, the likelihood  $f(\mathbf{y})$  can also be obtained as a side product thereof, as we will immediately clarify below. Therefore, with DA able to derive the two likelihoods  $f_0(\mathbf{y})$  and  $f_1(\mathbf{y})$ , and the latter two being the keys to causal attribution in our approach, one should be capable of moving towards near-real-time, systematic causal attribution of weather- and climate-related events.

## 267 2.3 Brief overview of data assimilation

268 DA was initially developed in the context of numerical weather forecasting,  
269 in order to initialize the model's state variables  $\mathbf{X}$  based on observations  $\mathbf{y}$   
270 that are incomplete, diverse, unevenly distributed in space and time and are  
271 contaminated by measurement error (Bengtsson et al., 1981; Talagrand, 1997).  
272 Over the past decades, those methods have grown out of their original applica-  
273 tion field to reach a wide variety of topics in geophysics such as oceanography  
274 (Ghil and Malanotte-Rizzoli, 1991), atmospheric chemistry, geomagnetism, hy-  
275 drology, and space physics, among many other areas (Robert et al., 2006;  
276 Cosme et al., 2010; Kondrashov et al., 2011; Bocquet, 2012; Martin et al.,  
277 2014).

278 DA is already playing an increasing role in the climate sciences, having be-  
279 ing applied, for instance, to initialize a climate model for seasonal or decadal  
280 prediction (Balmaseda et al., 2009), to constrain a climate model's parameters  
281 (Kondrashov et al., 2008; Ruiz et al., 2013), to infer carbon cycle fluxes from  
282 atmospheric concentrations (Chevallier, 2013), or to reconstruct paleoclimatic  
283 fields out of sparse and indirect observations (Bhend et al., 2012; Roques et  
284 al., 2014). In the context of D&A, Lee et al. (2008) actually tested a DA-like  
285 approach to include the effects of the various forcings over the last millennium,  
286 in addition to other paleoclimate proxy data, in combined climate reconstruc-  
287 tion and detection analysis. The present work thus follows a general trend in  
288 climate studies.

289 Methodologically speaking, DA methods are traditionally grouped into two  
290 categories: sequential and variational (Ide et al., 1997, and references therein).  
291 Here, we concentrate on the sequential approach, but the two approaches are  
292 complementary and the choice of method depends on the specifics of the prob-  
293 lem at hand (Ghil and Malanotte-Rizzoli, 1991; Ide et al., 1997; Talagrand,  
294 1997). In the sequential approach (Ghil et al., 1981), the state estimate and  
295 a suitable estimate of the associated error covariance matrix are propagated  
296 in time until new observations become available and are used to update the  
297 state estimate. In practice, the evolution of the system of interest is retrieved  
298 — like in earlier, typically much smaller-dimensional applications (Kalman,  
299 1960; Jazwinski, 1970; Gelb, 1974) — through a sequence of prediction and  
300 analysis steps.

301 Abundant literature is available on DA and on Kalman-type filters. Kalman  
302 (1960) first presented the solution in discrete time for the case in which both  
303 the dynamic evolution operator  $M$  in Eq. (4a) and the observation operator  $H$   
304 in Eq. (4b) are linear, and the errors are Gaussian. Under these assumptions,  
305 the state-estimation problem for the system given by Eqs. (4a, 4b) has an  
306 exact solution given by the sequential Kalman filter (KF) equations (Appendix  
307 B). Further, the likelihood function  $f(\mathbf{y})$ , which is of primary importance for  
308 DADA, also has an exact expression under the above linearity and Gaussianity  
309 assumptions (Tandeo et al., 2014). Following the usual notations of DA, which

are detailed in Appendix B, the expression of the likelihood is given by:

$$f(\mathbf{y}) = \prod_{t=0}^T (2\pi)^{-\frac{d}{2}} |\boldsymbol{\Sigma}_t|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{y}_t - \mathbf{H}\mathbf{x}_t^f)' \boldsymbol{\Sigma}_t^{-1} (\mathbf{y}_t - \mathbf{H}\mathbf{x}_t^f) \right\} \quad (5)$$

with  $\boldsymbol{\Sigma}_t = \mathbf{H}\mathbf{P}_t^f\mathbf{H}' + \mathbf{R}$ . The proof of Eq. (5) is provided in Appendix C, and  $f(\mathbf{y})$  is typically computed by taking the logarithm of this equation to turn the product on the right-hand side into a sum.

The main interest of Eq. (5) is that, once the observations  $\mathbf{y}_t$  have been assimilated on the interval  $0 \leq t \leq T$ , the necessary ingredients  $\mathbf{x}_t^f$  and  $\mathbf{P}_t^f$  in Eq. (5) are available from the KF equations (Appendix B) and thus calculating  $f(\mathbf{y})$  is both straightforward and computationally inexpensive. The fundamental connections between this calculation, the HMM context, and Bayes theorem are further clarified in Appendix C.

Many difficulties arise in applying the simple ideas outlined here to geophysical models, which are typically nonlinear, have non-Gaussian errors and are huge in size (Ghil and Malanotte-Rizzoli, 1991). Most of these difficulties have been addressed by improving both sequential and variational methods in several ingenious ways (Bocquet et al., 2010; Kondrashov et al., 2011).

In particular, the Ensemble Kalman Filter (EnKF; Evensen, 2003)—in which the uncertainty propagation is evaluated by using a finite-size ensemble of trajectories—is now operational in numerical weather and oceanic prediction centers worldwide; see e.g. Houtekamer et al. (2005); Sakov et al. (2013). The EnKF is a convenient approximate solution to the filtering problem in a nonlinear, large-dimensional context. We simply note here that it can also be applied to obtain an approximation of the likelihood  $f(\mathbf{y})$  by substituting the approximate sequence  $\{(\hat{\mathbf{x}}_t^f, \hat{\mathbf{P}}_t^f) : t = 0, \dots, T\}$  that the EnKF produces into Eq. (5). This strategy is illustrated immediately below in the context of the L63 convection model subject to an additional constant force.

### 3 Implementation within the modified L63 model

#### 3.1 The modified model and its two worlds

A simple modification (Palmer, 1999) of the L63 model (Lorenz, 1963) has been extensively used for the purpose of illustrating methodological developments in both DA and PEA (e.g. Carrassi and Vannitsem, 2010; Stone and Allen, 2005). In the nonlinear, coupled system of three ordinary differential equations (ODEs) for  $x, y$  and  $z$  below,

$$\frac{dx}{dt} = \sigma(y - x) + \lambda_i \cos \theta_i, \quad \frac{dy}{dt} = \rho x - y - xz + \lambda_i \sin \theta_i, \quad \frac{dz}{dt} = xy - \beta z \quad (6)$$

the time-constant forcing terms in the  $x$ - and  $y$ -equation represent, in fact, an addition to the forcing hidden in the original L63 model. The latter forcing

is revealed by a well-known linear change of variables, in which  $x$  and  $y$  are left unchanged and  $z \rightarrow z + \rho + \sigma$  (Lorenz, 1963). In the new variables, the model of Eq. (6) will take the canonical form of a forced-dissipative system (Ghil and Childress, 1987, Sec. 5.4), with an extra forcing term  $-\beta(\rho + \sigma)$  in the  $z$ -equation, just like the original L63 model.

Here  $\lambda_i$  is the intensity of the additional forcing and  $\theta_i$  is its direction in world  $i = 0, 1$ : i.e.,  $\lambda_0 = 0$  represents a counterfactual world with no additional forcing, while  $\lambda_1 \neq 0$ . We take the parameters  $(\sigma, \rho, \beta)$  to equal their usual values  $(10, 28, 8/3)$  that yield the well-known chaotic behavior, and the (nondimensional) time unit  $t$  is interpreted as equaling days.

The ODE system given by (6) is discretized by using  $\Delta t = 0.01$  and  $t$  refers hereafter to the number of time increments  $\Delta t$ . This system is then turned into a HMM as described in Equation (4) by adding an error term  $\mathbf{v}_t$  assumed to be Gaussian and centered with covariance  $\mathbf{Q} = \sigma_Q^2 \mathbf{I}$ , where  $\mathbf{I}$  is the  $3 \times 3$  identity matrix. Furthermore, we assume that all three coordinates  $(x, y, z)$  of the state vector are observed, i.e. that  $\mathbf{H} = \mathbf{I}$ , and that the measurement error term  $\mathbf{w}_t$  is also Gaussian and centered, with covariance  $\mathbf{R} = \sigma_R^2 \mathbf{I}$ .

The HMM defined above is stationary, i.e. the PDF of the observed vector  $\mathbf{y}_t$  depends neither on  $t$  nor on the initial condition after a sufficiently long time  $t$  (Appendix D). In the factual world, the shape of the PDF is affected by the parameters  $(\lambda_1, \theta_1)$  of the forcing. In both worlds, the PDFs can be estimated, for instance, by using kernel density estimation applied to ensembles of simulations obtained for either forcing. In Figs. 2a,b, we plot the projections of both PDFs onto the plane associated with the greatest variance in the factual PDF. The difference between the two PDFs is shown in Fig. 2c; it emphasizes the existence of an area of the state space (represented in white), which is more likely to be reached in the factual world than in the counterfactual one.

Next, we define an event to occur for the sequence  $\{\mathbf{y}_t : t = 0, \dots, T\}$  if the scalar product  $\hat{\phi}'\mathbf{y}_t$  between the unit vector  $\hat{\phi}$  in the direction  $\phi$  and  $\mathbf{y}_t$ , i.e. the projection of  $\mathbf{y}_t$  onto the direction  $\phi$ , exceeds  $u$  for some  $0 \leq t \leq T$ , where  $\phi$  is a specified direction and  $u$  is a threshold chosen based on  $\phi$  so that  $p_1 = 0.01$ . Figure 2d shows a selection of sequences from both worlds in which an event did occur, where  $\phi$  was chosen to be the leading direction in the projection plane.

For this choice of  $\phi$ , the trajectories associated with event occurrence happen to all lie in the area of the state space which is more likely to be reached in the factual world than in the counterfactual one. Accordingly, the probability of the event in the former is found to be higher than in the latter, i.e.  $p_1 > p_0$ , and the occurrence of an event  $\{\max_{\{0 \leq t \leq T\}} \hat{\phi}'\mathbf{y}_t \geq u\}$  is thereby informative from a causal perspective.

Figure 2d also shows that the trajectories associated with the event in the two worlds — counterfactual (green) and factual (red) — appear to have slightly distinct features: the red trajectories are shifted towards higher values in the second direction, of highest-but-one variance. Such distinctions might help discriminate further between the two worlds in the DADA framework —

the circumstances under which such further discrimination is helpful will be discussed in Section 4.

### 3.2 DADA for the modified L63 model

The DADA procedure is illustrated in Fig. 3. We plot in panel (a) a trajectory of the state vector  $\mathbf{x}_t$  simulated under factual conditions, i.e. in the presence of the additional forcing (black solid line), along with the observations  $\{\mathbf{y}_t : 0 \leq t \leq T\}$  (gray dots), with  $T = 400$ . The EnKF is used to assimilate these observations into a factual model ( $i = 1$ ) that thus matches the true model  $M = M_1 = M(\lambda_1, \theta_1)$  used for the simulation: a reconstructed trajectory is obtained from the corresponding analyses  $\mathbf{x}_t^a$  (red solid line in panel (a)), cf. Eqs. (8), and the likelihoods  $f_1(\mathbf{y}_t)$  (red solid line in panel (c)) are obtained by application of Eq. (5), respectively.

Next, the assimilation is repeated in the counterfactual model ( $i = 0$ , i.e.  $\lambda = 0$ ) to obtain a second analysis of the trajectory, from the same observations; see green solid line in panel (a), for  $T = 400$ . The corresponding likelihoods  $f_0(\mathbf{y}_t)$  are shown in panel (c) as a green solid line. Comparing the trajectories of the two analyses in Fig. 3a shows that, even though the counterfactual analysis (green line) uses the same data as the factual analysis (red line), the former lies closer to the true trajectory (black line).

The local discrepancies between the trajectories estimated in the two worlds appear to be rather small at first glance, cf. panel (a), and so are the instantaneous differences between the associated factors on the right-hand side of Eq. (5); the latter are shown as gray rectangles in panel (c) of the figure. Still, the evidence in favor of the factual world accumulates as the time  $t$  over which the two trajectories differ, albeit by a small amount, lengthens. This cumulative difference in evidence,  $\log f_0(\mathbf{y}_t) - \log f_1(\mathbf{y}_t)$ , is reflected by a growing gap between the two curves, red and green, in panel (c), and by an associated high mean growth over time of the probability PN of necessary causation, cf. the black solid line in panel (d).

In order to evaluate more systematically its performance and robustness compared to the conventional FAR approach, the DADA procedure was applied to a large sample of sequences  $\mathbf{y}_t$  of length  $T = 20$  simulated under diverse conditions. The sample explored all possible combinations of the triplet of parameters  $(\lambda_1, \sigma_Q, \sigma_R)$ , with ten equidistributed values each, for a total of  $10^3$  combinations; the ranges were  $0 \leq \lambda_1 \leq 40$ ,  $0.1 \leq \sigma_Q \leq 0.5$  and  $0.1 \leq \sigma_R \leq 1.0$ , respectively, with  $\theta_1 = -140^\circ$ . For each combination of  $(\lambda_1, \sigma_Q, \sigma_R)$ , ten directions  $\phi$  were randomly generated and  $u$  was defined based on  $\phi$  as in Sec. 3a above, so as to achieve  $p_1 \geq 0.01$ .

In order to estimate the corresponding conventional probabilities  $p_0$  and  $p_1$  of the associated event defined as  $\{\max_{\{0 \leq t \leq T\}} \phi' \mathbf{y}_t \geq u\}$ ,  $n = 50\,000$  sequences  $\mathbf{y}_t$  of length  $T = 20$  were simulated, by using a single sequence of length  $nT = 10^6$  and splitting it into  $n$  equal segments. Probabilities  $p_0$  and  $p_1$  were then directly estimated from empirical frequencies.

435 For each quintuplet of parameter values  $(\lambda_1, \sigma_Q, \sigma_R; \phi, u)$ , one hundred  
 436 sequences of observations  $\{\mathbf{y}_t : 0, \dots, T = 20\}$  were generated with a propor-  
 437 tion  $p_1/(p_1 + p_0)$  being simulated from the factual world and a proportion  
 438  $p_0/(p_1 + p_0)$  from the counterfactual one. All sequences were treated with the  
 439 DADA procedure — by applying DA to the synthetic observations according  
 440 to Eqs. (8a)–(8d) — and then Eq. (5) to obtain  $f_0(\mathbf{y})$  and  $f_1(\mathbf{y})$  from the re-  
 441 constructed trajectories. The a priori mean and covariance  $\mathbf{x}_0^f$  and  $\mathbf{P}_0^f$  required  
 442 as inputs to the DADA procedure were those associated with the PDF of the  
 443 attractor, given the forcing conditions ( $\lambda_1 \in [0, 40], \theta_1 = -140^\circ$ ) assumed for  
 444 each assimilation experiment. As a result, two probabilities PN of necessity  
 445 are finally obtained for each sequence  $\mathbf{y}_t$ ,  $\text{PN}_p = 1 - p_0/p_1$  for the conventional  
 446 approach and  $\text{PN}_f = 1 - f_0(\mathbf{y})/f_1(\mathbf{y})$  for the DADA approach.

447 We next wish to evaluate under various conditions how well the two prob-  
 448 abilities  $\text{PN}_p$  and  $\text{PN}_f$  perform with respect to discriminating between the  
 449 factual and counterfactual forcings. Consider a simple discrimination rule  
 450 whereby a trajectory  $\mathbf{y}_t$  is identified as factual for PN exceeding a given  
 451 threshold, and as counterfactual otherwise. The so-called receiver operating  
 452 characteristic (ROC) curve plots the rate of true positives as a function of the  
 453 rate of false positives obtained when varying the threshold in a binary classifi-  
 454 cation scheme from 0 to 1; it thus gives an overall visual representation of the  
 455 skill of our PN as a discriminative *score*.

456 The Gini (1921) index  $G$  was originally introduced as a measure of statisti-  
 457 cal dispersion intended to summarize the information contained in the Lorenz  
 458 (1905) curve that represents the income distribution of a nation’s residents;  $G$   
 459 may be viewed, though, more generally as a metric summarizing the dispersion  
 460 of any smooth curve that starts at the origin and ends at the point  $(1, 1)$  with  
 461 respect to the diagonal of the corresponding square. In particular, we use  $G$   
 462 here to summarize into a single scalar the ROC curve, which ranges from 0  
 463 for random discrimination to 1 for perfect discrimination.

464 Figure 4a shows ROC curves obtained over the entire sample of  $n = 50\,000$   
 465 sequences: they correspond to  $G = 0.35$  for the conventional method and to  
 466  $G = 0.82$  for the DADA method, i.e. the overall performance gap is more than  
 467 twofold. As expected, the performance of both methods is nil for  $\lambda_1 = 0$  and  
 468 it is very sensitive to the intensity of the forcing, cf. Fig. 4b.

469 Furthermore, the skill of the DADA method is boosted when decreasing the  
 470 level of model error, cf. Fig. 4c; this is an expected result, since DA becomes  
 471 more reliable when the model is more accurate, and when it is known to  
 472 be so. Ultimately, under perfect model conditions, i.e. as  $\sigma_Q \rightarrow 0$ , DADA  
 473 reaches perfect discriminative power, with  $G \rightarrow 1$ , no matter how small, but still  
 474 positive, the forcing is; see Fig. 4d. On the other hand, the level of observational  
 475 error  $\sigma_R$  appears to have but a limited effect on DADA performance for the  
 476 range of values considered, cf. Fig. 4e.

477 Finally, Fig. 4f shows that both methods perform better when the contrast  
 478 between  $p_0$  and  $p_1$  is strong, but the latter does not influence the gap between  
 479 the two methods, which remains nearly constant. This constant gap thus ap-  
 480 pears to quantify the additional power resulting from the extra discriminative

481 features that the PDF  $f(\mathbf{y})$  is able to capture on top of those associated with  
482 the probability  $P(\phi(\mathbf{y}) \geq u)$ .

## 483 4 Discussion and conclusions

484 Considerations rooted in the causality theory of Pearl (2000) have shown that  
485 the ratio between the factual likelihood  $f_1(\mathbf{y})$  and the counterfactual likeli-  
486 hood  $f_0(\mathbf{y})$  is relevant in studying causal attribution of weather- and climate-  
487 related events. In this paper, we described data assimilation (DA) methods  
488 and demonstrated that they are well suited for deriving  $f_0(\mathbf{y})$  and  $f_1(\mathbf{y})$   
489 from trajectories in the factual and the counterfactual worlds, respectively.  
490 Besides, these methods offer the key practical advantage of being already up-  
491 and-running in real time at meteorological centers.

492 Combining these two sets of considerations, theoretical and practical, opens  
493 a novel route towards real time, systematic causal attribution of weather- and  
494 climate-related events, thereby addressing a key challenge in the field of PEA  
495 at present (Stott et al., 2013).

### 496 4.1 Theoretical considerations

497 Implementing the DADA approach in the context of the L63 model in Sec-  
498 tion 3 allowed for a detailed step-by-step illustration of our methodological  
499 proposal. It also provided a basic test for an initial performance assessment,  
500 which showed an improved level of discriminating power with respect to the  
501 conventional approach outlined in Section 1. These results are promising, and  
502 their promise is easy to understand, given the fact that the DADA approach  
503 leverages the available information on the entire trajectory  $\mathbf{y}$ , as opposed to  
504 the single specific feature  $\phi(\mathbf{y}) \geq u$  in the conventional approach.

505 It is important, though, to stress that the term “performance” here should  
506 be considered with caution: improving discriminatory performance may or may  
507 not be a desirable outcome, depending on the causal question being asked.  
508 Hannart et al. (2015) and Otto et al. (2015) have shown that the causal ques-  
509 tion being formulated reflects the subjective interests of a particular class of  
510 end-users, and that the formulation itself may dramatically affect the answer.

511 For example, the question “*did anthropogenic CO<sub>2</sub> emissions cause the*  
512 *heatwave observed over Argentina during January 2014?*” has been tradition-  
513 ally treated by defining a “heatwave” in terms of a predefined temperature  
514 index reaching a predefined threshold, i.e., by a singular index exceeding a  
515 singular threshold. This class of questions matters for instance in the context  
516 of insurance disbursements, where a financial compensation may typically be  
517 triggered based on such an index exceedance. In this situation, the additional  
518 discriminatory power of DADA is meaningless because the DADA computa-  
519 tion does not address the question at stake: there is simply no alternative to  
520 computing the probabilities  $p_0$  and  $p_1$  of the index exceeding the threshold.

521 However, if the question is formulated instead as “*did anthropogenic CO<sub>2</sub>*  
 522 *emissions cause the atmospheric conditions observed over Argentina during*  
 523 *January 2014?*” — i.e., without specifying which feature of the observed se-  
 524 quence is most important — then improving discrimination makes perfect  
 525 sense and DADA becomes fully relevant. Furthermore, DADA is still fully rel-  
 526 evant even if the question is formulated more specifically as “*did anthropogenic*  
 527 *CO<sub>2</sub> emissions cause the damages generated in Argentina by the atmospheric*  
 528 *conditions of January 2014?*,” provided that a model relating atmospheric  
 529 observations to damages at every time step  $t$  along the trajectory of the phys-  
 530 ical model used in the assimilation is available and can be integrated into the  
 531 observation operator  $H$ .

532 On the other hand, the results of Section 3 should also be considered with  
 533 caution simply because the L63 testbed obviously differs in many respects from  
 534 the real situation envisioned for future applications, both in terms of model  
 535 dimension  $n$  and observation dimension  $d$ : in practice  $n$  will be very large and  
 536  $d \ll n$ , while here we took  $d = n = 3$ .

537 In particular, choosing a highly idealized, climatological prior distribution  
 538 on the initial condition  $\pi(\mathbf{x}_0)$  does not raise any difficulty under the tested  
 539 conditions nor does it influence significantly the outcome of the procedure  
 540 (not shown). The choice of  $\pi(\mathbf{x}_0)$ , however, may be an important problem in  
 541 practice, when  $d \ll n$ , and lead to potentially spurious results.

542 As a consequence, it may be both necessary and useful to further constrain  
 543 the so-called *background PDF*  $\pi(\mathbf{x}_0)$  by using the forecasts originating from  $\tau$   
 544 previous assimilation cycles, thus following the ideas of lagged-averaged fore-  
 545 casting (Hoffman and Kalnay, 1983; Dalcher et al., 1988). The evidence thus  
 546 obtained, though, will then also depend on previous observations over the “ini-  
 547 tialization” window  $[-\tau, \dots, -1]$  — i.e., it will no longer represent exclusively  
 548 the desired evidence  $f(\mathbf{y})$ . Besides, choosing  $\tau$  optimally to constrain the initial  
 549 background PDF in a satisfactory manner, while at the same time limiting the  
 550 latter unwanted dependence on previous observations, is a challenging question  
 551 that needs to be addressed.

552 More generally, the problem of evaluating the evidence  $f(\mathbf{y})$  is not new in  
 553 the HMM and DA literature; see, for instance, Baum et al. (1970); Hürzeler  
 554 and Künsch (2001); Pitt (2002) and Kantas et al. (2009). Various algorithms  
 555 are thus available to carry out this evaluation, depending on a number of key  
 556 assumptions — such as lack of Gaussianity or linearity — and on the inferential  
 557 setting chosen, e.g. particle filtering. These algorithms may provide accurate  
 558 and effective solutions to the above problem, as well as improved alternatives  
 559 to the Gaussian and linear approximation of Eq. (5), since the latter may  
 560 not be sufficiently accurate for successfully implementing the DADA approach  
 561 under realistic conditions.



## 562 4.2 Practical considerations

563 While we have shown here that the proposal of using DADA for event attribu-  
564 tions has intellectual merit, its main strength lies, in our view, in down-to-earth  
565 cost considerations. By design, the DADA approach allows one to piggyback  
566 at a low marginal cost on the large and powerful infrastructures already in  
567 place at several meteorological centers, in terms of both hardware and person-  
568 nel. These centers are capable of processing massive amounts of observational  
569 data with high-throughput pipelines on the world’s largest computational plat-  
570 forms, as opposed to requiring the design, set-up and maintenance of a new  
571 and large, PEA-specific infrastructure to collect observations and generate —  
572 under real time constraints — the many model simulations required by the  
573 conventional approach recalled in Section 1.

574 Taking a step back, it is useful to examine our proposal within the wider  
575 context of the emergence of so-called climate services. It is widely recog-  
576 nized that extending the scope of activity of meteorological centers from being  
577 “monoline” weather forecasting providers to becoming “multiline” climate ser-  
578 vices providers – encompassing, for instance, weather forecasting and weather  
579 event attribution as two service lines among several others – is a relevant  
580 strategic option (Hewitt et al., 2012). Such a strategy may foster the timely  
581 and cost-efficient emergence of the latter services by building upon techno-  
582 logical and infrastructure synergies with the former. For these reasons, our  
583 proposal is particularly relevant for, and could contribute to, the implementa-  
584 tion of the strategic option just outlined.

585 This being said, DADA can very well serve as a method for real time  
586 event attribution even for hypothetical climate services providers that focus  
587 uniquely or mainly on longer time scales, beyond a month, a season or a year.  
588 In such a context, DADA may allow for the assimilation of a broader range  
589 of observations, and in particular of ocean observations; it may, in fact, be  
590 important to include the latter in causal analysis when the event occurrence  
591 under scrutiny is defined over a sufficiently large time window.

592 Finally, it is important to remember that providing real-time attribution  
593 assessments is a major communication challenge, since different methods give  
594 different answers and different definitions of a specific event may also im-  
595 pact the outcome of an assessment — as mentioned above and as discussed  
596 recently by Trenberth et al. (2015). Various recent examples, such as the ongo-  
597 ing California drought have shown that divergences among experts may lead  
598 to confusion in the media and among stakeholders. In this respect, a detailed  
599 comparison of the DADA approach with other methods in realistic, real-time  
600 situations will be required before the method can be applied operationally.

601 **Acknowledgements** It is a pleasure to thank Fredi Otto and Dáithí Stone, who provided  
602 careful and constructive reviews of the original paper. This work has been supported by  
603 grant DADA from the Agence Nationale de la Recherche (ANR, France: AH and all co-  
604 authors) and by the Multi-University Research Initiative (MURI) N00014-12-1-0911 from  
605 the the U.S. Office of Naval Research (MG).

606 **Appendix A — Illustration of the computational benefit of the**  
 607 **DADA approach.** To illustrate the computational benefit, let  $\mathbf{Y}$  be for in-  
 608 stance a  $d$ -variate autoregressive process defined by  $\mathbf{Y}_{t+1} = \mathbf{A}\mathbf{Y}_t + \mathbf{w}_t$ , where  
 609  $\mathbf{w}_t$  is an i.i.d. noise having known PDF  $g(\cdot)$  and where  $\mathbf{A}$  has the usual prop-  
 610 erties that insure stationarity (Gardiner, 2004). We then have:

$$611 \quad f(\mathbf{y}) = \prod_{t=1}^T g(\mathbf{y}_t - \mathbf{A}\mathbf{y}_{t-1})\pi(\mathbf{y}_0), \quad (7a)$$

$$612 \quad P(\phi(\mathbf{Y}) \geq u) = \int_{\phi(\mathbf{y}) \geq u} \prod_{t=1}^T g(\mathbf{y}_t - \mathbf{A}\mathbf{y}_{t-1})\pi(\mathbf{y}_0) dy_{1,0} \dots dy_{d,0} \dots dy_{d,T}, \quad (7b)$$

614 with  $\pi(\cdot)$  the prior PDF on the initial state  $\mathbf{Y}_0$ . Equation (7a) shows that  $f(\mathbf{y})$   
 615 can be easily computed using a closed-form expression, while  $P(\phi(\mathbf{Y}) \geq u)$   
 616 in Eq. (7b) is an integral on  $d \times T + 1$  dimensions which must instead be  
 617 evaluated by using, for instance, a computationally quite costly Monte-Carlo  
 618 (MC) simulation.

619 **Appendix B — Data Assimilation.** The state-estimation problem for  
 620 the system given by Eqs. (4a, 4b) has an exact solution given by the following  
 621 sequential Kalman filter (KF) equations:

$$623 \quad \mathbf{x}_t^a = \mathbf{x}_t^f + \mathbf{K}(\mathbf{y}_t - \mathbf{H}\mathbf{x}_t^f), \quad (8a)$$

$$624 \quad \mathbf{P}_t^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}_t^f, \quad (8b)$$

$$625 \quad \mathbf{x}_{t+1}^f = \mathbf{M}\mathbf{x}_t^a, \quad (8c)$$

$$626 \quad \mathbf{P}_{t+1}^f = \mathbf{M}\mathbf{P}_t^a\mathbf{M}' + \mathbf{Q}. \quad (8d)$$

628 where  $'$  denotes the transpose operation. Here Eqs. (8a) and (8b) are referred  
 629 to as the analysis step and denoted by a superscript  $a$ , while the forecast step  
 630 is given by Eqs. (8c) and (8d), and is denoted by a superscript  $f$  (Ide et al.,  
 631 1997). The vector  $\mathbf{x}_t^a$  and the matrix  $\mathbf{P}_t^a$  are the mean and covariance of  $\mathbf{X}_t$   
 632 conditional on  $(\mathbf{Y}_1, \dots, \mathbf{Y}_t) = (\mathbf{y}_1, \dots, \mathbf{y}_t)$ ;  $\mathbf{K} = \mathbf{P}_t^f\mathbf{H}'(\mathbf{H}\mathbf{P}_t^f\mathbf{H}' + \mathbf{R})^{-1}$  is the so-  
 633 called Kalman gain matrix; while  $\mathbf{Q}$  and  $\mathbf{R}$  are the covariances associated with  
 634  $\mathbf{v}_t$  and  $\mathbf{w}_t$ , respectively. Following Wiener (1949), one distinguishes between  
 635 *filtering*, in which  $\mathbf{x}_t^a$  and  $\mathbf{P}_t^a$  are conditioned only on the previous and current  
 636 observations  $(\mathbf{y}_0, \dots, \mathbf{y}_t)$ , and *smoothing*, in which they are conditioned on the  
 637 entire sequence,  $0 \leq t \leq T$ . Furthermore, the sequential algorithm needs to  
 638 be initialized at time  $t = 0$  with  $\mathbf{x}_0^f$  and  $\mathbf{P}_0^f$ , which thus represent the a priori  
 639 mean and covariance of  $\mathbf{X}_0$ , respectively, and have to be prescribed by the user.

640 **Appendix C — Derivation of the model evidence.** In this appendix,  
 641 we outline the derivation of model evidence within a general Bayesian frame-  
 642 work, and we apply the latter to the narrower KF context to obtain Eq. (5).  
 643 Consider two consecutive cycles of a DA run, the first with state vector  $\mathbf{x}_t$  and  
 644 observation vector  $\mathbf{y}_t$  at instant  $t$  and the subsequent one with state vector

646  $\mathbf{x}_{t+1}$  and observation vector  $\mathbf{y}_{t+1}$  at instant  $t + 1$ . We plan to find a tractable  
 647 expression for the model evidence  $p(\mathbf{y}_t, \mathbf{y}_{t+1})$ .

648 The model evidence provided by the full sequence of observations  $\mathbf{y} =$   
 649  $(\mathbf{y}_0, \dots, \mathbf{y}_T)$  will be inferred by recursion, using the results of this two-observation  
 650 setting. In order to decouple the two cycles, one first has to spell out the  
 651 Bayesian inference  $p(\mathbf{y}_t, \mathbf{y}_{t+1}) = p(\mathbf{y}_t)p(\mathbf{y}_{t+1}|\mathbf{y}_t)$ . We look for a tractable ex-  
 652 pression for  $p(\mathbf{y}_{t+1}|\mathbf{y}_t)$  by further introducing the states  $\mathbf{x}_{t+1}$  and  $\mathbf{x}_t$  as inter-  
 653 mediate random variables:

$$654 \begin{aligned} p(\mathbf{y}_{t+1}|\mathbf{y}_t) &= \int_{\mathbf{x}_{t+1}} p(\mathbf{y}_{t+1}|\mathbf{y}_t, \mathbf{x}_{t+1})p(\mathbf{x}_{t+1}|\mathbf{y}_t) d\mathbf{x}_{t+1} \\ &= \int_{\mathbf{x}_{t+1}} p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}) \left\{ \int_{\mathbf{x}_t} p(\mathbf{x}_{t+1}|\mathbf{x}_t) p(\mathbf{x}_t|\mathbf{y}_t) d\mathbf{x}_t \right\} d\mathbf{x}_{t+1}, \end{aligned} \quad (9)$$

655 where  $p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1})$  is the likelihood of the observation vector  $\mathbf{y}_{t+1}$  conditional  
 656 on the state vector  $\mathbf{x}_{t+1}$  and it is known from Eq. (4b). The conditional PDF  
 657  $p(\mathbf{x}_t|\mathbf{y}_t)$  of  $\mathbf{x}_t$  on  $\mathbf{y}_t$  at instant  $t$  — which appears on the right-hand side of  
 658 the above equation — is referred to as the *analysis* PDF in the DA literature,  
 659 where it is denoted by a superscript  $a$  (Ide et al., 1997), and it constitutes  
 660 the main DA output. The integral  $\int_{\mathbf{x}_t} p(\mathbf{x}_{t+1}|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_t) d\mathbf{x}_t = p(\mathbf{x}_{t+1}|\mathbf{y}_t)$ ,  
 661 in which  $p(\mathbf{x}_{t+1}|\mathbf{x}_t)$  is known from the model dynamics given by Eq. (4),  
 662 propagates this analysis PDF further in time, to instant  $t + 1$ . Hence, the result  
 663 of this integration coincides with the forecast PDF, denoted by superscript  $f$   
 664 in the DA literature (Ide et al., 1997). It follows that this decomposition is  
 665 tractable using a DA scheme that is able to estimate the conditional and  
 666 forecast PDFs.

667 Next, let us apply the general Bayesian inference (9) to the case in which all  
 668 the PDFs involved are Gaussian; this requires, in turn, that both the dynamics  
 669 and observation models  $M$  and  $H$  be linear, and that the input statistics all  
 670 be Gaussian. In this case, the Kalman filter allows for the exact computation  
 671 of the PDFs mentioned in Eq. (9), which turn out to be Gaussian.

672 In the following,  $\mathcal{N}(\bar{\mathbf{x}}, \mathbf{P})$  designates the Gaussian PDF of mean  $\bar{\mathbf{x}}$  and co-  
 673 variance matrix  $\mathbf{P}$ . In this context, the analysis PDF at instant  $t$  is  $\mathcal{N}(\mathbf{x}_t^a, \mathbf{P}_t^a)$ ,  
 674 where  $\mathbf{x}_t^a$  and  $\mathbf{P}_t^a$  are the analysis state and error covariance matrix at instant  
 675  $t$ . As a result of the linearity assumptions, the forecast PDF at instant  $t + 1$  is  
 676 given by a Gaussian distribution  $\mathcal{N}(\mathbf{x}_{t+1}^f, \mathbf{P}_{t+1}^f)$ , where  $\mathbf{x}_{t+1}^f$  and  $\mathbf{P}_{t+1}^f$  are the  
 677 forecast state and error covariance matrix at instant  $t + 1$ . Further, the integra-  
 678 tion on  $\mathbf{x}_{t+1}$  in Eq. (9) can readily be performed under these circumstances,  
 679 with the outcome that  $p(\mathbf{y}_{t+1}|\mathbf{y}_t)$  is distributed as  $\mathcal{N}(\mathbf{H}\mathbf{x}_{t+1}^f, \mathbf{R} + \mathbf{H}\mathbf{P}_{t+1}^f\mathbf{H}')$ .

680 The desired model evidence  $f(\mathbf{y})$  can then be computed by recursion on  
 681 successive time steps as:

$$682 \begin{aligned} f(\mathbf{y}) &= p(\mathbf{y}_0) \prod_{t=1}^T (2\pi)^{-\frac{d}{2}} |\boldsymbol{\Sigma}_t|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{y}_t - \mathbf{H}\mathbf{x}_t^f)' \boldsymbol{\Sigma}_t^{-1} (\mathbf{y}_t - \mathbf{H}\mathbf{x}_t^f) \right\}; \end{aligned} \quad (10)$$

683 here  $p(\mathbf{y}_0)$  represents the prior PDF of the initial state,  $\boldsymbol{\Sigma}_t = \mathbf{R} + \mathbf{H}\mathbf{P}_t^f\mathbf{H}'$ ,  
 684 and this expression coincides with Eq. (5) and can be evaluated with the help  
 685 of any DA method that yields the forecast states and forecast error covariance

686 matrices, such as the KF or the EnKF. Note that the traditional standard  
 687 Kalman smoother would give the same result as the KF, since they share the  
 688 same forecasts.

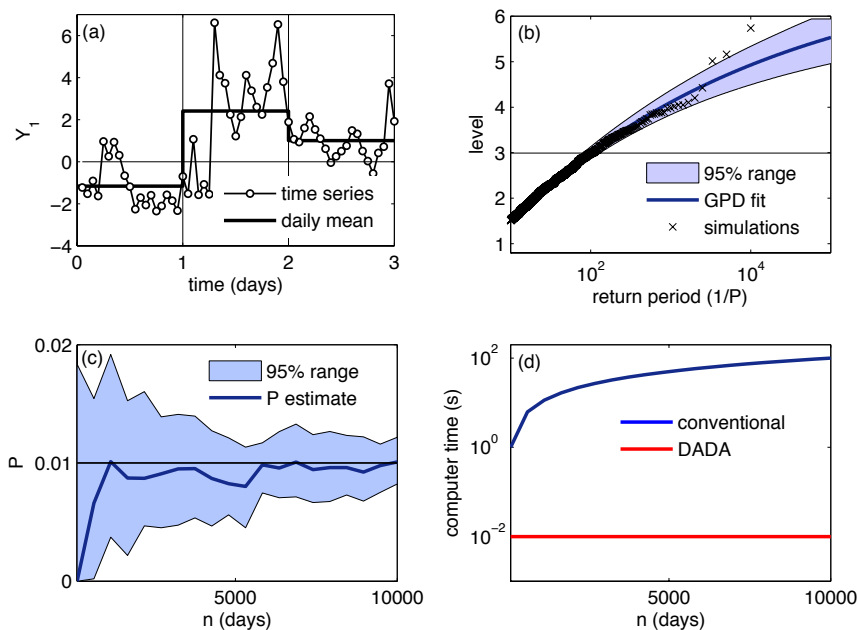
689 Finally, Eqs. (9) and (10) above show that the likelihood  $f(\mathbf{y})$  may be  
 690 obtained as a by-product of the inference on the state vector  $\mathbf{x}$ , which usually  
 691 is the main purpose in numerical weather prediction. This idea may actually  
 692 be highlighted in even greater generality by considering the equality:

$$693 \quad f(\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{x}|\mathbf{y})}. \quad (11)$$

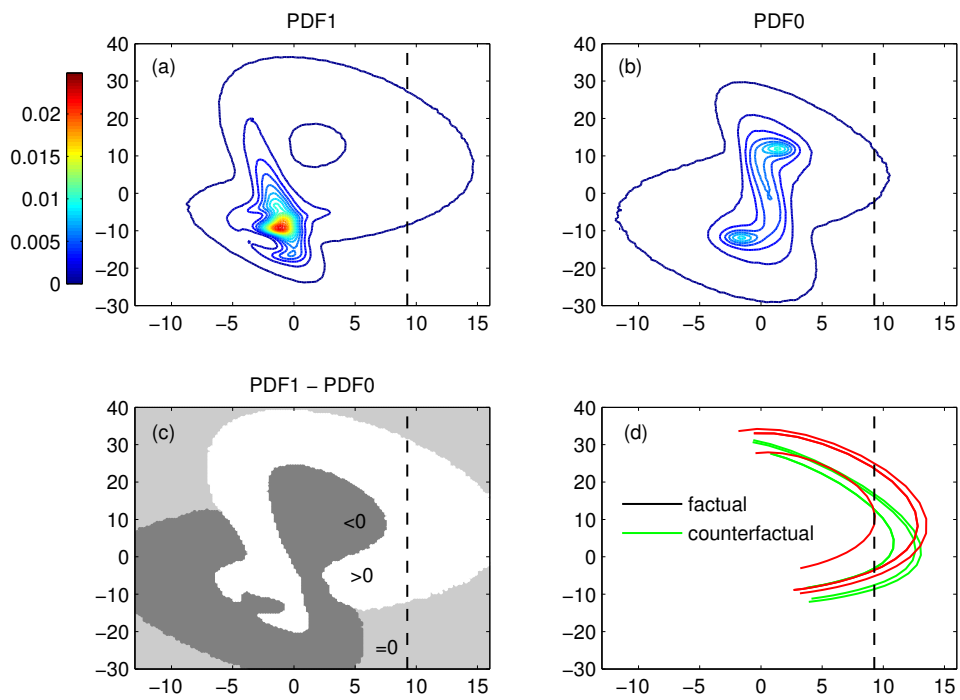
694 While Eq. (11) is a direct consequence of Bayes theorem, it also illustrates a  
 695 point that is arguably not so intuitive. The likelihood  $f(\mathbf{y})$  is obtained here as  
 696 the ratio of two quantities: a numerator  $p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$  that is a model premise  
 697 inherently postulated by Eqs. (4a) and (4b), and a denominator  $p(\mathbf{x}|\mathbf{y})$  that  
 698 may be viewed as the end result of the primary inference on  $\mathbf{x}$ . In other words,  
 699 estimating  $f(\mathbf{y})$  requires only a straightforward division, provided  $\mathbf{x}$  has been  
 700 previously inferred.

701 Equation (11) thus expresses with great clarity and simplicity a fundamen-  
 702 tal idea buttressing our proposal, as it provides a general theoretical justifica-  
 703 tion for the suggestion of deriving the likelihood from an inferential treatment  
 704 that focuses on  $\mathbf{x}$ . To put it succinctly, this equation basically says, “*He who*  
 705 *can do more can do less.*” In the context of DA, whose end purpose is to infer  
 706 the state vector  $\mathbf{x}$  out of an observation  $\mathbf{y}$  — i.e., the *more* part — it is possible  
 707 to obtain the likelihood as a by-product thereof — i.e., the *less* part — and  
 708 thus almost for free.

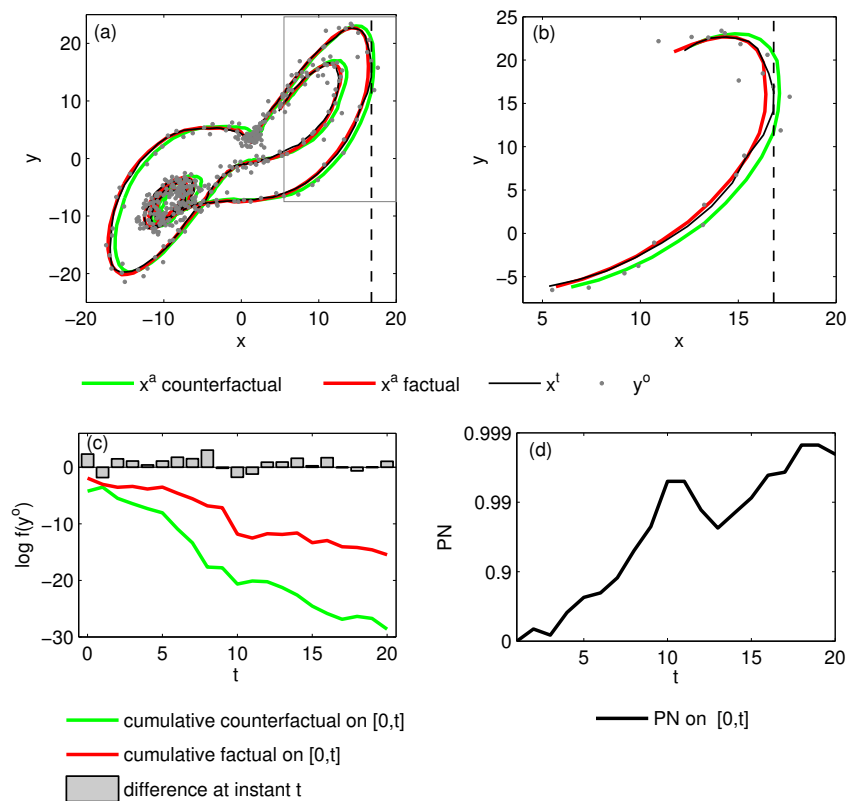
709 **Appendix D — PDF of the state vector.** We associate a label  $\omega \in \Omega$   
 710 with each realization of the random process  $\mathbf{v}_t$  that drive the model given by  
 711 Eq. (6). The PDF of the state vector  $\mathbf{x}_t$  can be obtained as the numerical  
 712 solution of the corresponding Fokker-Planck equation, and it is the mean over  
 713  $\Omega$  of the sample measures obtained for each realization  $\omega$  of the noise  $\mathbf{v}_t$   
 714 and (Chekroun et al., 2011, and references therein). Each sample measure is  
 715 supported on a random attractor that may have very fine structure and be  
 716 time-dependent (Chekroun et al., 2011, Figs. 1–3 and supplementary material),  
 717 but the PDF is supported smoothly, in the counterfactual world in which  
 718  $\lambda_0 = 0$ , on a “thickened” version of the fairly well-known strange attractor of  
 719 the original L63 model. The latter PDF represents its *attractor* in dynamic  
 720 system’s terminology.  
 721



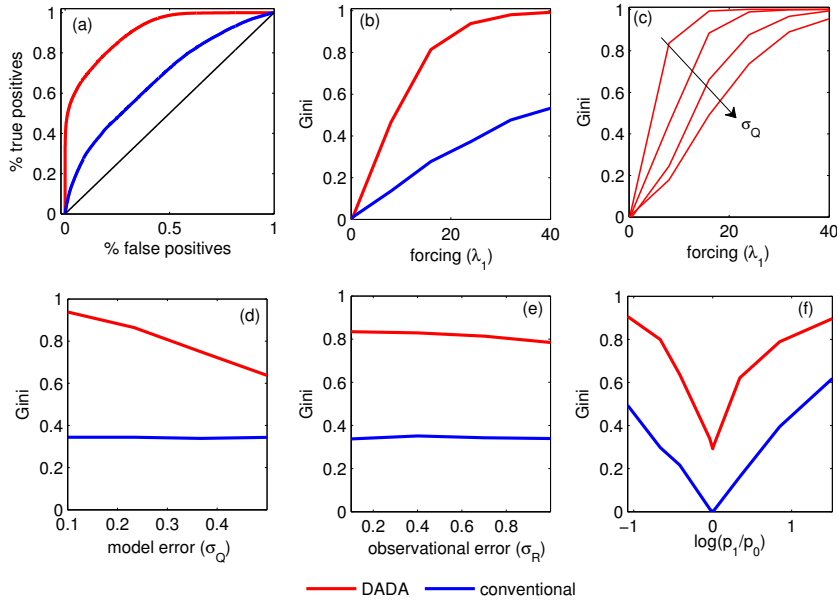
**Fig. 1** Illustration of the conventional PEA approach as applied to a univariate AR(1) process. (a) Observed time series (first component  $Y_1$ , dotted line) and daily average  $\phi(\mathbf{Y})$  (heavy solid line) over the three first days. (b) Threshold level (vertical axis) as a function of the return period (horizontal axis): simulated values (crosses); fit based on the Generalized Pareto distribution (GPD, heavy dark-blue line); uncertainty range at the 95% level (light blue area); and threshold value  $u = 3.1$  (light solid black line). (c) Estimated value of  $P = P(\phi(\mathbf{Y}) \geq u)$  (heavy dark-blue line) using a GPD fit as a function of the sample size  $n$  (horizontal axis); uncertainty range (light blue area); and true value  $P = 0.01$  (light solid black line). (d) Computational time on a desktop computer (seconds, vertical axis) as a function of sample size  $n$  (horizontal axis) required by the conventional method (dark blue line) and the DADA method (solid red line); the latter method is explained in Sections 2b and 3 below.



**Fig. 2** Two-dimensional (2-D) projections of the PDF of the modified L63 model; the projection is onto a plane defined by the two leading eigenvectors of the factual PDF shown in the first panel. (a) PDF of the factual attractor, with  $\lambda_1 = 20$  and  $\sigma_Q = 0.1$ ; and (b) PDF of the counterfactual attractor, with  $\lambda_0 = 0$ . (c) Difference between the factual and counterfactual PDFs. (d) Sample trajectories associated with an event occurrence originating from the factual (red solid lines) and counterfactual worlds (green solid lines); the vertical dashed line in all four panels indicates the threshold  $u$  with respect to the horizontal axis of largest variance in the factual PDF.



**Fig. 3** Sample trajectories from data assimilation (DA) in our modified L63 model. (a) True trajectory (black solid line) and the two trajectories reconstructed by DA in the factual ( $i = 1$ ) and counterfactual ( $i = 0$ ) worlds (red and green solid lines), respectively, over a long sequence,  $T = 400$ ; the values of  $\lambda_1$  and  $\theta_1$  here are the same as in Fig. 2, and the assimilated observations are shown as gray dots. (b) Same as panel (a) but zoomed over a short sequence,  $T = 20$ . (c) Logarithm of the cumulative evidences  $f_1(\mathbf{y})$  and  $f_0(\mathbf{y})$  (red and green lines, respectively) computed over the window  $[0, t \leq T]$ ; gray bars indicate the instantaneous differences between  $f_1(\mathbf{y}_t)$  and  $f_0(\mathbf{y}_t)$ . (d) PN computed over the window  $[0, t]$ .



**Fig. 4** Performance of the DADA and conventional methods (red vs. blue solid lines, respectively). (a) Receiver operating characteristic (ROC) curve: true positive rate as a function of false positive rate, when varying the cut-off level  $u$ , as obtained from the entire sample of  $n = 50\,000$  sequences; see text for details. (b) Gini index  $G$  as a function of forcing intensity  $\lambda_1$ . (c) Same as (b) for several values of  $\sigma_Q$  and for DADA only, with the black arrow indicating the direction of growing  $\sigma_Q$ . (d) Same as (b) but as a function of model error amplitude  $\sigma_Q$ . (e) Same as (b) but as a function of observational error amplitude  $\sigma_R$ . (f) Same as (b) as a function of the logarithmic contrast between the conventional probabilities  $\log p_1/p_0$ .



## References

- Allen M.R. (2003) Liability for climate change. *Nature*, 421:891–892.
- Arnold L. (1998) *Random Dynamical Systems*. Springer, 625 pp.
- Baum L.E., T. Petrie, G. Soules, N. Weiss (1970) A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*, 41(1):164–171.
- Balmaseda M.A., O.J. Alves, A. Arribas, T. Awaji, D.W. Behringer, N. Ferry, Y. Fujii, T. Lee, M. Rienecker, T. Rosati, D. Stammer (2009) Ocean initialization for seasonal forecasts, *Oceanography Special Issue*, 22(3).
- Bengtsson L., M. Ghil, E. Källén (Eds., 1981) *Dynamic Meteorology: Data Assimilation Methods*, Springer-Verlag, New York/Heidelberg/Berlin, 330 pp.
- Bhend J., J. Franke, D. Folini, M. Wild, S. Brönnimann (2012) An ensemble-based approach to climate reconstructions *Clim. Past*, 8:963–976.
- Bocquet M., C.A. Pires, L. Wu (2010) Beyond Gaussian statistical modeling in geophysical data assimilation. *Mon. Wea. Rev.*, 138:2997–3023.
- Bocquet M. (2012) Parameter-field estimation for atmospheric dispersion: application to the Chernobyl accident using 4D-Var. *Quart. J. Roy. Meteor. Soc.*, 138:664–681.
- Bucklew J.A. (2004) *Introduction to Rare Event Simulation*, Springer.
- Carrassi A, S. Vannitsem (2010) Model error and variational data assimilation: A deterministic formulation. *Mon. Wea. Rev.*, 138, 3369–3386.
- Carrassi A., M. Ghil, A. Trevisan, F. Uboldi (2008) Data assimilation as a nonlinear dynamical systems problem: Stability and convergence of the prediction-assimilation system. *Chaos: An Interdisciplinary Journal of Non-linear Science*, 18(2):023112.
- Chekroun M.D., E. Simonnet, M. Ghil, 2011: Stochastic climate dynamics: Random attractors and time-dependent invariant measures, *Physica D*, 240(21):1685–1700, doi :10.1016/j.physd.2011.06.005.
- Chevallier F. (2013) On the parallelization of atmospheric inversions of CO<sub>2</sub> surface fluxes within a variational framework. *Geosci. Model. Dev. Discuss.*, 6, 37–57.
- Christidis N., P.A. Stott, A. A. Scaife, A. Arribas, G. S. Jones, D. Copey, J. R. Knight, W. J. Tennant. (2013) A New HadGEM3-A-Based System for Attribution of Weather- and Climate-Related Extreme Events. *J. Clim.*, 26(9): 2756–2783.
- Cosme E., J.M. Brankart, J. Verron, P. Brasseur, M. Krysta (2006) Implementation of a reduced-rank, square-root smoother for ocean data assimilation. *Ocean Modelling*, 33, 87–100.
- Dalcher A., Kalnay E., Hoffman R.N. (1988) Medium-range lagged average forecasts. *Mon. Wea. Rev.*, 116, 402–416, doi: [http://dx.doi.org/10.1175/1520-0493\(1988\)116<0402:MRLAF;2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(1988)116<0402:MRLAF;2.0.CO;2).
- Del Moral P., J. Garnier, Genealogical particle analysis of rare events. *Ann. Appl. Probab.*, 15:4, 2496–2534.

- 766 Evensen G. (2003) The ensemble Kalman filter: theoretical formulation and  
767 practical implementation. *Ocean Dyn.* 53:343–367.
- 768 Gardiner C. (2004) *Handbook of Stochastic Methods for Physics, Chemistry*  
769 *and the Natural Sciences*. Publisher, pls.; no web tonite.
- 770 Gelb A. (Ed.) (1974) *Applied Optimal Estimation*. M.I.T. Press, Cambridge,  
771 MA, 374 pp.
- 772 Ghil M., S. Childress (1987) *Topics in Geophysical Fluid Dynamics: At-*  
773 *mospheric Dynamics, Dynamo Theory and Climate Dynamics*. Springer-  
774 Verlag, New York/Berlin, 485 pp.
- 775 Ghil M., P. Malanotte-Rizzoli (1991) Data assimilation in meteorology and  
776 oceanography, *Adv. Geophys.*, 33:141–266.
- 777 Ghil M., S. Cohn, J. Tavantzis, K. Bube, E. Isaacson (1981) Applications of  
778 estimation theory to numerical weather prediction. In: *Dynamic Meteorol-*  
779 *ogy: Data Assimilation Methods*, L. Bengtsson, M. Ghil, E. Källén (Eds.),  
780 Springer Verlag, pp. 139–224.
- 781 Gini C. (1921) Measurement of inequality of incomes. *Econ. J.* 31 (121):124–  
782 126. doi:10.2307/2223319.
- 783 Greenland S., K.J. Rothman (1998) Measures of effect and measures of associ-  
784 ation, Chapter 4 in Rothman, K. J., Greenland, S. (eds.), *Modern Epidemi-*  
785 *ology*, 2nd edn., Lippincott-Raven, Philadelphia, USA.
- 786 Hannart A., J. Pearl, F.E.L. Otto, P. Naveau, M. Ghil (2015). Counterfactual  
787 causality theory for the attribution of weather and climate-related events.  
788 *Bull. Am. Meteorol. Soc.*, in press.
- 789 Harris T. E., Kahn H. (1951). Estimation of particle transmission by random  
790 sampling. *Natl. Bur. Stand. Appl. Math. Ser.*, 12 27–30.
- 791 Heidelberg P. (1995). Fast simulation of rare events in queueing and reliability  
792 models. *ACM Trans. Modeling Computer Simulation*, 5 43–85.
- 793 Hegerl G.C., O. Hoegh-Guldberg, G. Casassa, M.P. Hoerling, R.S. Kovats, C.  
794 Parmesan, D.W. Pierce, P.A. Stott (2010): Good Practice Guidance Paper  
795 on Detection and Attribution Related to Anthropogenic Climate Change.  
796 In: *Meeting Report of the Intergovernmental Panel on Climate Change Ex-*  
797 *pert Meeting on Detection and Attribution of Anthropogenic Climate Change*  
798 [Stocker, T.F., C.B. Field, D. Qin, V. Barros, G.-K. Plattner, M. Tignor,  
799 P.M. Midgley, K.L. Ebi (eds.)]. IPCC Working Group I Technical Support  
800 Unit, University of Bern, Bern, Switzerland.
- 801 Hewitt C., S. Mason, D. Walland (2012) The Global Framework for Climate  
802 Services, *Nature Climate Change*, 2, 831–832.
- 803 Hoffman R.N., Kalnay, E. (1983) Lagged average forecasting, an alterna-  
804 tive to Monte Carlo forecasting. *Tellus*, 35A, 100–118, doi: 10.1111/j.1600-  
805 0870.1983.tb00189.x.
- 806 Houtekamer P.L., H.L. Mitchell, G. Pellerin, M. Buehner, M. Charron (2005)  
807 Atmospheric data assimilation with an ensemble Kalman filter: Results with  
808 real observations. *Mon. Wea. Rev.*, 133, 604–620.
- 809 Hume D. (1748) *An Enquiry Concerning Human Understanding*. Reprinted  
810 by Open Court Press (1958), LaSalle, IL, USA.

- 811 Hürzeler M., Künsch H.R. (2001) Approximation and maximising the likeli-  
812 hood for a general state-space model. In: *Sequential Monte Carlo Methods*  
813 *in Practice* [Doucet, A., De Freitas, J.F.G., Gordon N.J. (eds.)]. Springer-  
814 Verlag, New York, USA.
- 815 Ide K., P. Courtier, M. Ghil, A. Lorenc (1997) Unified notation for data as-  
816 similation: Operational, sequential and variational. *J. Meteor. Soc. Japan*,  
817 75:181–189.
- 818 Ihler A.T., S. Kirshner, M. Ghil, A.W. Robertson, P. Smyth (2007) Graphical  
819 models for statistical inference and data assimilation. *Physica D*, 230, 72–87,  
820 2007.
- 821 IPCC (2013) Summary for Policymakers. In: *Climate Change 2013: The Phys-  
822 ical Science Basis. Contribution of Working Group I to the Fifth Assessment*  
823 *Report of the Intergovernmental Panel on Climate Change* [Stocker, T.F., D.  
824 Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia,  
825 V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge,  
826 United Kingdom and New York, NY, USA.
- 827 Jazwinski A.H. (1970) *Stochastic and Filtering Theory*. Mathematics in Sci-  
828 ences and Engineering Series, Vol. 64. Academic Press, 376 pp.
- 829 Kalman R.E. (1960) A new approach to linear filtering and prediction prob-  
830 lems. *J. Basic Eng.*, 82D:33–45.
- 831 Kalnay E. (2002) *Atmospheric Modeling, Data Assimilation and Predictability*,  
832 Cambridge University Press, Cambridge, UK.
- 833 Kantas N., A. Doucet, S.S. Singh, J.M. Maciejowski (2009) An overview of se-  
834 quential Monte Carlo methods for parameter estimation. In: *General State-  
835 Space Models*, IFAC System Identification, no. M1.
- 836 Kondrashov D., C.J. Sun, M. Ghil (2008) Data assimilation for a coupled  
837 ocean-atmosphere model. Part II: Parameter estimation. *Mon. Wea. Rev.*,  
838 136, 50625076, doi: 10.1175/2008MWR2544.1.
- 839 Kondrashov D., Y. Shprits, M. Ghil (2011) Log-normal Kalman filter for as-  
840 simulating phase-space density data in the radiation belts. *Space Weather*,  
841 9, S11006, doi:10.1029/2011SW000726.
- 842 Le Dimet F.X., O. Talagrand (1986) Variational algorithms for analysis and  
843 assimilation of meteorological observations: Theoretical aspects. *Tellus*,  
844 38A:97–110.
- 845 Lee T.C.K., F.W. Zwiers, M. Tsao (2008) Evaluation of proxy-based millennial  
846 reconstruction methods. *Climate Dyn.*, 31, 263–281.
- 847 Lorenz E.N. (1963) Deterministic non-periodic flow. *J. Atmos. Sci.* 20:130–  
848 141.
- 849 Lorenz M.O. (1905) Methods of measuring the concentration of wealth.  
850 *Publications of the American Statistical Association*, 9 (70): 209219,  
851 doi:10.2307/2276207.
- 852 Martin M.J. et al. (2014) Status and future of data assimilation in operational  
853 oceanography. *J. of Oper. Ocean.*, in press.
- 854 Massey N., Jones R., Otto F.E.L., Aina T., Wilson S., Murphy J.M., Hassell  
855 D., Yamazaki Y.H., Allen M.R. (2014) *weather@home* — development and  
856 validation of a very large ensemble modelling system for probabilistic event

- 857 attribution. *Q. J. R. Meteorol. Soc.* doi: 10.1002/qj.2455
- 858 Otto F.E.L., Boyd, E., Jones, R.G., Cornforth, R.J., James, R., Parker, H.R.,  
859 Allen, M.R. (2015) Attribution of extreme weather events in Africa: a pre-  
860 liminary exploration of the science and policy implications. *Climatic Change*.
- 861 Palmer T.N. (1999) A non-linear dynamical perspective on climate prediction.  
862 *J. Clim.* 12:575–591.
- 863 Pearl J. (2000) *Causality: Models, Reasoning and Inference*, Cambridge Uni-  
864 versity Press, Cambridge, United Kingdom and New York, NY, USA.
- 865 Pitt M.K. (2002) Smooth particle filters for likelihood evaluation and maxim-  
866 sation. *Warwick Economic Research Papers*, No. 651.
- 867 Robert C., E. Blayo, J. Verron (2006) Comparison of reduced-order sequential,  
868 variational and hybrid data assimilation methods in the context of a Tropical  
869 Pacific ocean model. *Ocean Dynamics*, 56, 624–633.
- 870 Roques L., M.D. Chekroun, M. Cristofol, S. Soubeyrand, M. Ghil (2014) Pa-  
871 rameter estimation for energy balance models with memory. *Proc R. Soc.*  
872 *A*, 470, 20140349.
- 873 Ruiz J., M. Pulido, T. Miyoshi (2013) Estimating model parameters with  
874 ensemble-based data assimilation: A review. *JMSJ*, 91, 2, 79–99.
- 875 Sakov P., Counillon F., Bertino L., Lister K.A., Oke P.R., Korabely A. (2012)  
876 TOPAZ4: an ocean-sea ice data assimilation system for the North Atlantic  
877 and Arctic, *Ocean Sci.*, 8, 633–656, doi:10.5194/os-8-633-2012.
- 878 Stone D.A., M.R. Allen (2005) The end-to-end attribution problem: from emis-  
879 sions to impacts. *Clim. Change*, 71:303–318.
- 880 Stott P.A., et al. (2013) Attribution of weather and climate-related events,  
881 in *Climatic Science for Serving Society: Research, Modelling and Prediction*  
882 *Priorities*, G.R. Asrar and J. W. Hurrell (Eds.), Springer, in press.
- 883 Stott P.A., Stone D.A., Allen M.R. (2004) Human contribution to the Euro-  
884 pean heatwave of 2003. *Nature*, 432:610–614.
- 885 Talagrand O. (1997) Assimilation of observations, an introduction, *J. Meteor.*  
886 *Soc. Japan*, 75 (1B):191–209.
- 887 Tandeo P., Pulido M., Lott F. (2014), Offline parameter estimation using  
888 EnKF and maximum likelihood error covariance estimates: Application to  
889 a subgrid-scale orography parametrization. *Q. J. R. Meteorol. Soc.* doi:  
890 10.1002/qj.2357
- 891 Trenberth, K. E., Fasullo, J. T., Shepherd, T. G. (2015) Attribution of climate  
892 extreme events. *Nature Clim. Change* 5, 725–730.
- 893 Wiener N. (1949) *Extrapolation, Interpolation and Smoothing of Stationary*  
894 *Time Series, with Engineering Applications*. M.I.T. Press, Cambridge, MA,  
895 163 pp.
- 896 Wouters J. and F. Bouchet (2015) Rare event simulation of the chaotic Lorenz  
897 96 dynamical system. *Geophysical Research Abstracts, EGU General As-*  
898 *sembly 2015* Vol. 17, EGU2015-10421-1.