



The impact of nonlinearity in Lagrangian data assimilation

A. Apte¹ and C. K. R. T. Jones²

¹TIFR Centre for Applicable Mathematics, P. O. Bag 6503, Chikkabommasandra, Bangalore 560064, India

²Department of Mathematics, University of North Carolina, Chapel Hill NC, 27599 USA

Correspondence to: A. Apte (apte@math.tifrbng.res.in)

Received: 29 November 2012 – Revised: 17 April 2013 – Accepted: 19 April 2013 – Published: 23 May 2013

Abstract. The focus of this paper is on how two main manifestations of nonlinearity in low-dimensional systems – shear around a center fixed point (nonlinear center) and the differential divergence of trajectories passing by a saddle (nonlinear saddle) – strongly affect data assimilation. The impact is felt through their leading to non-Gaussian distribution functions. The major factors that control the strength of these effects is time between observations, and covariance of the prior relative to covariance of the observational noise. Both these factors – less frequent observations and larger prior covariance – allow the nonlinearity to take hold. To expose these nonlinear effects, we use the comparison between exact posterior distributions conditioned on observations and the ensemble Kalman filter (EnKF) approximation of these posteriors. We discuss the serious limitations of the EnKF in handling these effects.

1 Introduction

The assimilation of Lagrangian or pseudo-Lagrangian data from fluid flow, obtained by surface and subsurface drifters and floats in the ocean as well as by gliders and autonomous underwater vehicles (AUV), has now become an indispensable and widely used tool for studying the oceans and other natural water bodies (Chyba, 2009; The Argo Science Team, 2001). The set of mathematical techniques that use the data from these instruments for the purpose of state estimation are generally referred to as Lagrangian data assimilation (LaDA) (Ide et al., 2002; Kuznetsov et al., 2003). The emphasis in the recent past on studying various aspects of LaDA has been prompted mainly by the rapidly increasing availability and density of such data in a variety of situations. Another aspect of the LaDA problem that has received much less attention is its suitability for studying certain funda-

mental aspects of the data assimilation problem in general – namely, the effects of nonlinearity and the consequent non-Gaussianity of probability densities involved in assimilation problems.

The main emphasis of this paper will be precisely on these issues of nonlinearity using the Lagrangian data assimilation framework. We will use the Bayesian viewpoint of the data assimilation problem (Apte et al., 2007, 2008a) that is summarized later in Sect. 3. The central object of study in the Bayesian framework is the posterior probability distribution function (PDF) on the dynamic state of the system, i.e., the conditional PDF of the state conditioned on the observations. This posterior is a compilation of the available information, and approximating it as closely as possible is the main goal of data assimilation. We take the viewpoint that valuable insights on the interplay between the dynamics and statistics can be gained through comparisons between this posterior PDF and the results of other data assimilation methods that produce, in effect, different approximations of this posterior.

In particular, we focus on comparing the Markov chain Monte Carlo (MCMC) sampling of this posterior PDF with a commonly used data assimilation technique – namely, the ensemble Kalman filter (EnKF), (Evensen, 2009) which generates an ensemble of the states of the system, incorporating the observations sequentially using a Kalman-filter-based method, explained in detail later in Sect. 3. The ensemble produced by the EnKF method also samples, albeit approximately, the posterior PDF. In the limit of a large ensemble size, the ensemble of states resulting from the “update” step of the EnKF, or the various variants of it, is an exact representation of posterior PDF if the ensemble *prior* to the update is Gaussian. (Lei et al., 2010) If the EnKF is initialized with a Gaussian ensemble, the nonlinearity of the dynamical model may result in a significantly non-Gaussian prior ensemble that will update to a poor approximation of the posterior.

We argue that this discrepancy between the exact MCMC sampling of the posterior and a possibly poor approximation of it produced by EnKF can be fruitfully exploited to analyze and understand the effects of nonlinearity on data assimilation strategies. In particular, we will utilize this method to illustrate that the nonlinear shear and differential divergence effects, explained in detail below, can lead to the failure (or divergence) of the EnKF method precisely because of the non-Gaussianity of the prior distributions.

To understand the *nonlinear shear effect*, we will consider the situation occurring in a low-dimensional dynamical system wherein a fixed point is surrounded by periodic orbits. In the linear case (harmonic oscillator) the orbits rotate around the center fixed point with the same angular velocity. But this synchronicity is destroyed by nonlinearity, which will result in a differential speed of rotation around the center. The resulting prior distribution will then tend to spread around the center in an annular shape.

A similar story holds in the saddle case, although of course with different-shaped distributions emerging. In the case of a saddle fixed point of a linear system, nearby orbits diverge from the separatrix at an exponential rate that is the same for all the orbits. But different orbits near a saddle point of a nonlinear system diverge away from the separatrix at a different rate and this *differential divergence* gives rise to non-Gaussian distributions for this case. Since the case of saddle has been studied more extensively in previous works, (Kuznetsov et al., 2003; Salman et al., 2006) we put greater focus in this paper on the nonlinear shear phenomenon. Both these are discussed and illustrated in greater detail in Sect. 2.1.

These nonlinear effects may be mitigated in two ways: (1) if the frequency of observations is high enough, i.e., the time between observations small enough, then the shear effect will be less pronounced, or (2) if the covariance of the prior is small enough, then it will be dominated by the (Gaussian) observational likelihood. In both of these cases, the linear-based approximations implicit in any Kalman filter scheme will be effective. The point is, however, that working with a specific system, a priori control of the observational frequency and/or the prior covariance may be lacking. In the former case, the issue is that the specifications for the needed observational frequency will depend on a (potentially unknown) period in the system. As for the prior covariance, in sequential data assimilation it will have been built up through repeated applications of the dynamics and assimilation steps, and there is no guarantee that this will have any particular size.

In general, the role of data assimilation is not only to give information about the observed variables but also about the unobserved ones. A novel phenomena that our studies illustrate is about the effects of nonlinearity on unobserved variables. In particular, we will see cases when the marginal posterior given by EnKF for the observed variables and *some of the unobserved variables* agrees well with the exact marginal posterior for them, but the marginal posterior for *some of the*

other unobserved variables disagrees. This is because the information contained in the nonlinearity about the latter type of the unobserved variables is not captured by EnKF.

The *saddle effect* (Ide et al., 2002; Kuznetsov et al., 2003; Salman et al., 2006) has been known to be a cause of the failure of EnKF and the effect of the observational time on filtering has also been studied. (Apte et al., 2008b; Nurujjaman et al., 2012) The four main contributions of the present work are a detailed study of the precise origin of the EnKF divergence, the new phenomenon of EnKF failure due to the shear around a center, the exact role played by the prior covariance in these phenomena, and the effect of the nonlinearity on unobserved variables.

A brief outline of the paper is as follows. The next section contains a description of the specific dynamical model we use, along with an illustration of the shear and the saddle effect. In Sect. 3 we review the Bayesian framework and the EnKF method. Section 4 is devoted to describing how the shear manifests itself in the failure of EnKF and generally its effects on data assimilation. The saddle effect is more briefly described in Sect. 5, followed by Sect. 6 containing a summary and a discussion of some of the directions for further studies.

2 Shear and saddle effects in linear shallow water Lagrangian dynamics

The specific dynamical system under consideration is the same as that used in Apte et al. (2008b), i.e., the following two Fourier modes of the velocity $(u, v)(x, y, t)$ and the height $h(x, y, t)$ of linear shallow water (LSW) equations with M Lagrangian floats.

$$\begin{aligned} u(x, y, t) &= -2\pi \sin(2\pi x) \cos(2\pi y) u_0 + \\ &\quad \cos(2\pi y) u_1(t), \\ v(x, y, t) &= 2\pi \cos(2\pi x) \sin(2\pi y) u_0 + \\ &\quad \cos(2\pi y) v_1(t), \\ h(x, y, t) &= \sin(2\pi x) \sin(2\pi y) u_0 + \\ &\quad \sin(2\pi y) h_1(t). \end{aligned} \quad (1)$$

The dynamics for the velocity field Fourier modes (u_0, u_1, v_1, h_1) and positions (x_i, y_i) , $i = 1, \dots, M$ of M Lagrangian drifters is given by the following set of ODE.

$$\begin{aligned} \dot{u}_0 &= 0, & \dot{u}_1 &= v_1, \\ \dot{v}_1 &= -u_1 - 2\pi h_1, & \dot{h}_1 &= 2\pi v_1, \\ \dot{x}_i(t) &= u(x_i(t), y_i(t), t), \\ \dot{y}_i(t) &= v(x_i(t), y_i(t), t), & i &= 1, \dots, M, \end{aligned} \quad (2)$$

where the functions u and v on the right-hand side are as given in Eq. (1). The LaDA problem is to infer the velocity field (the Fourier modes (u_0, u_1, v_1, h_1) in the above model) from the measurements of the positions (x_i, y_i) , $i =$

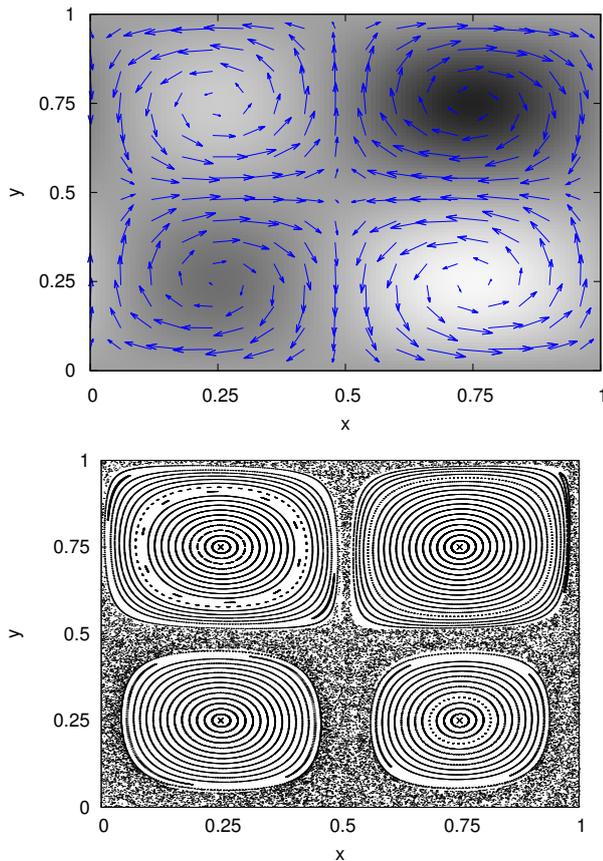


Fig. 1. Top panel shows the velocity field (arrows) and the height field (shaded) of the linear shallow water equations for typical initial conditions used in this paper. We will focus on Lagrangian dynamics near the saddle at $(x, y) = (0.5, 0)$ and the center at $(x, y) = (0.25, 0.25)$. The Poincaré plot at the bottom shows a number of different orbits – some quasi-periodic orbits in invariant circles, some periodic orbits, and some chaotic orbits near the two separatrices at $x = 0.5$ and $y = 0.5$. This plot shows that some parts of the phase space for the Lagrangian trajectories are chaotic, while some are regular.

$1, \dots, M$ of the M Lagrangian drifters. The following key features of this model make it suitable for studying the effects of nonlinearity in data assimilation.

1. As with other Lagrangian data assimilation problems, the dynamical model has a skew-product structure – the drifter dynamics depends on the velocity field but not the other way around – and only the drifter positions are observed.
2. For the LSW the velocity dynamics is linear, while all the nonlinearity is contained in the Lagrangian drifter dynamics, which is observed. Thus the observed component is highly nonlinear. This helps us in isolating the precise nonlinear effects.

3. Most importantly, the model shows the dynamical behavior of interest to our purposes – namely, presence of saddles and centers. Figure 1 shows typical velocity and height fields along with a representative Poincaré plot of the Lagrangian trajectories. Note that the velocity dynamics alone, i.e., the solutions of the first four of Eq. (2) are periodic in time with period $T_{\text{flow}} = 2\pi/\sqrt{1+4\pi^2} \approx 1$ and the Poincaré plot shows the positions of a number of drifters at times that are multiples of T_{flow} .

2.1 Illustrative examples

In order to understand the effects of shear and saddle on ensembles of initial conditions, we start with a simple ensemble of Lagrangian drifters along a straight line. Figure 2 shows the positions of these Lagrangian drifters after every 0.05 time units (which is small compared to the period of the velocity flow $T_{\text{flow}} \approx 1$). The blue dots show the positions of Lagrangian drifters after every 0.05 time step for trajectories in the time-independent velocity flow (i.e., with $u_0 = 1, (u_1, v_1, h_1) = (0, 0, 0)$), the red ones in a time-dependent flow (with initial conditions $u_0 = 1, (u_1, v_1, h_1) = (0.5, 0.34, 0.26)$), and the black ones in an ensemble of velocity flows (with initial conditions $u_0 = 1, h_1 = 0$ and (u_1, v_1) initial conditions chosen randomly to be uniform in the interval $(0, 1)$). (Note that here and in the rest of the paper, the dependence of u_1, v_1, h_1 on time will be understood and the plots show the values of these variables at the corresponding time relevant to that discussion.)

1. The figure clearly shows the strong shear effect – the trajectories near the center rotate much faster than those away from it, leading to strong distortion of the initial line of drifters. We will see later that this effect leads to initial Gaussian distributions being distorted into “annulus”-shaped non-Gaussian distributions (see, e.g., Fig. 4).
2. The shear effect is the basic feature of the time-independent flow, and the time-dependent flow modifies it only slightly. Thus, trajectories near the center contain information mostly about the first Fourier mode u_0 and much less about the other modes u_1, v_1, h_1 . This will again be revealed by the posterior distributions of these variables.
3. The middle panel shows dynamics near the separatrix at $x = 0.5$. We note that the origin of the saddle effect in data assimilation is not just the divergence near the saddle, which is a purely linear effect, but the “differential divergence” (as shown by the distortion of the line of initial conditions in the small inset at the top right). It is the latter of the two that leads to non-Gaussian distributions.

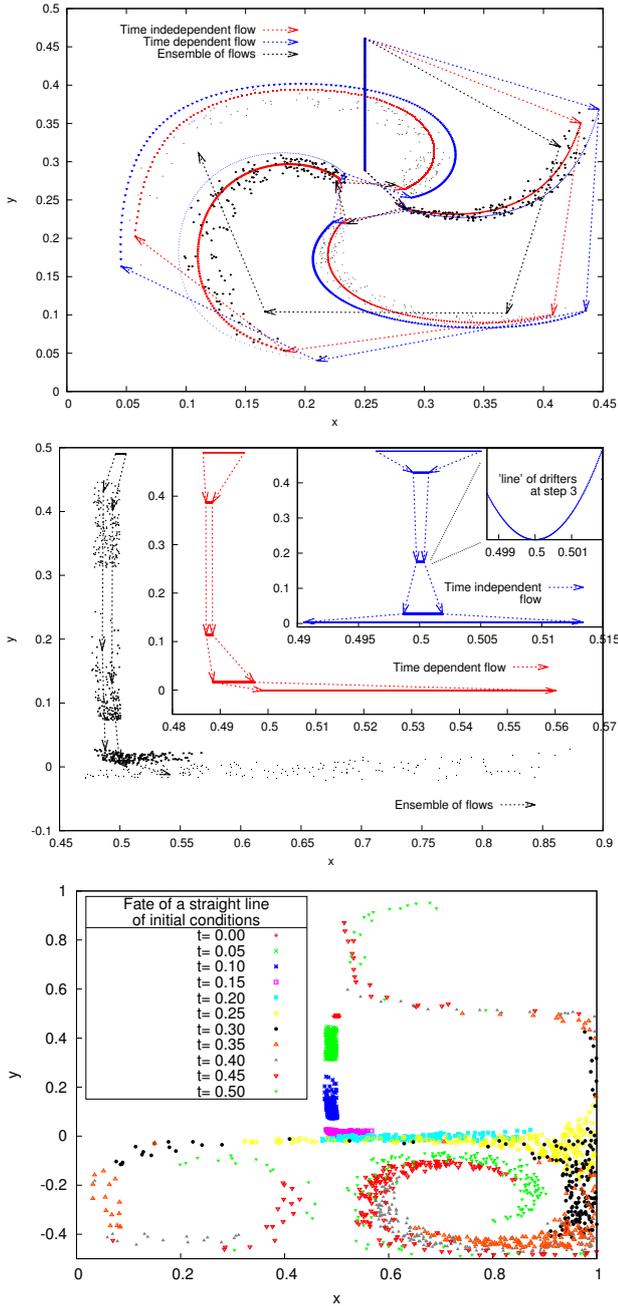


Fig. 2. The effect of shear and saddle on an ensemble of initial conditions of the Lagrangian drifter near a center at $(x, y) = (0.25, 0.25)$ (top panel) and near the saddle at $(x, y) = (0.5, 0.5)$ with separatrices $x = 0.5$ and $y = 0.5$ (middle panel). The divergence of trajectories starting near the saddle, within a time span of 0.50 in an ensemble of velocities, is shown in bottom panel.

4. The middle panel clearly shows that the trajectories near the saddle are strongly affected by the time-dependent flow since even small perturbations of the time-independent flow lead to drastically different dy-

namics near the saddle at $(0.5, 0.5)$. (This is seen from the fact that there is a large difference between the blue, red, and black lines, which are the trajectories in time-independent, time-dependent, and an ensemble of time-dependent flows, respectively.) Thus, the trajectories near the saddle contain information about precisely those modes u_1, v_1, h_1 of the velocity about which the trajectories near the center do not contain much information. Thus, a judicious mix of the trajectories near the centers and saddles will be required to get enough knowledge of the full velocity field. This has implications for the so-called “drifter placement strategies”, as discussed in Salman et al. (2008).

5. The divergence near the different saddles leads to drastically different dynamics of the drifters that are initially nearby, as shown in the panel at the bottom. This sensitivity to initial conditions poses problems for data assimilation methods in general, as discussed later. (For clarity, the positions at different times are shown in different colors. Note also that the velocity flow is different for each different drifter.)

We will see later in Sects. 4–5 the implications for data assimilation of all these basic facts about the dynamics of nonlinear systems. Before beginning that discussion, we will describe the Bayesian formulation of data assimilation and the ensemble Kalman filter in the following section.

3 Posterior distribution: exact sampling and ensemble Kalman filter approximation

We will use the Bayesian viewpoint outlined in detail in Apte et al. (2007, 2008a). Here we will only state a few key points for reference. We will consider deterministic models

$$\frac{dx}{dt} = f(x, t), \quad x(t_0) = x_0 \sim \zeta, \tag{3}$$

with solution $x(t) = \Phi(x_0; t_0, t)$ for n -dimensional state vector $x(t) \in \mathbb{R}^n$ with initial conditions drawn from a prior with probability density function $p_\zeta(x_0)$. The data, or the observations, are taken at discrete times and are subject to noise whose statistical characteristics, or the probability densities, are assumed to be known. Thus, the observation $y_k \in \mathbb{R}^m$ at time t_k is modeled as a random vector

$$y_k = h(x(t_k)) + \eta_k = h(\Phi(x_0; t_k)) + \eta_k,$$

where $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$.

Remark 1

For the specific dynamical model used in this paper – namely, the linear shallow water equations with M Lagrangian drifters – the state vector or the control vector is $x(t) \equiv$

$(u_0(t), u_1(t), v_1(t), h_1(t), x_1(t), y_1(t), \dots, x_M(t), y_M(t))$. Since the observations are only of the drifter location, the observation vector is $y_k = (x_1(t_k), y_1(t_k), \dots, x_M(t_k), y_M(t_k))$, resulting in an observation operator $h(x(t)) = [0 \ I]x(t)$, which is a projection on the drifter position components of the state $x(t)$. In this section we will only discuss the general setting, and hence in this section, instances of x will refer to the state vector of the model and those of y to the observations, consistent with generally accepted notation in data assimilation literature. (Ide et al., 1997) In the rest of the paper, (x, y) will refer to the positions of Lagrangian drifters.

We concatenate the observations at various times t_1, \dots, t_K to write the total observational random vector $y^T = (y_1^T, \dots, y_K^T)$ as

$$y = H(x_0) + \eta, \tag{4}$$

where

$$H(x_0)^T = (h(x(t_1)))^T, \dots, h(x(t_K))^T$$

and

$$\eta^T = (\eta_1^T, \dots, \eta_K^T).$$

We assume that we are given the probability density function $p_\eta : \mathbb{R}^{mK} \rightarrow \mathbb{R}$ of the random vector η . Thus, for a given set of the observations \hat{y} , which are a realization of the random vector y , the posterior probability $P_{\text{ex}}(x(t_0)|\hat{y}_1, \dots, \hat{y}_K)$ for the initial condition x_0 conditioned on observations up to time t_K is obtained from Bayes' theorem:

$$\begin{aligned} P_{\text{ex}}(x(t_0)|\hat{y}_1, \dots, \hat{y}_K) &:= p(x_0|\hat{y}) \\ &= \frac{p_\eta(\hat{y} - H(x_0))p_\zeta(x_0)}{p(\hat{y})} \\ &\propto p_\eta(\hat{y} - H(x_0))p_\zeta(x_0), \end{aligned} \tag{5}$$

where

$$p(y) = \int p_\eta(y - H(x_0))p_\zeta(x_0)dx_0$$

is a function of the observations y alone, and hence $p(\hat{y})$ is a constant for a given realization \hat{y} . In particular the constant of proportionality in Eq. (5) does not depend upon x_0 . This distribution can be pushed forward to obtain conditional distribution $P_{\text{ex}}(x(t)|\hat{y}_1, \dots, \hat{y}_K)$ for the state at any later time $t > 0$. Note that when $t < t_K$, this is a *smoothing distribution*, whereas $P_{\text{ex}}(x(t_K)|\hat{y}_1, \dots, \hat{y}_K)$ is the *filtering distribution*.

In this paper we have used Markov chain Monte Carlo (MCMC) methods in order to get samples from this posterior distribution $\pi(z) := P_{\text{ex}}(x(t_0)|\hat{y}_1, \dots, \hat{y}_K)$ for the initial condition $z := x_0$. The details of the method are detailed in Apte et al. (2008b). Here we only give a brief description of the Metropolis–Hastings algorithm we use. The Markov chain is generated by using a proposal

$$z^* = z_n + \alpha \Lambda \nabla_z \ln \pi(z_n) \delta_n + \sqrt{2\delta_n \Lambda} \omega_n,$$

where ω_n are iid standard Gaussian random variables. Thus, $z^* \sim \mathcal{N}(\mu(z_n), \Sigma)$ with $\mu(z) = z + \alpha \Lambda \nabla_z \ln \pi(z) \delta_n$ and $\Sigma = 2\delta_n \Lambda$. This has a distribution

$$q(z, z^*) \propto \exp \left[-\frac{1}{2} \|z^* - \mu(z)\|_\Sigma^2 \right],$$

where $\|z\|_\Sigma^2 = z^T \Sigma^{-1} z$. The standard Metropolis–Hastings criterion is used for accepting or rejecting the proposed state: if $\alpha = \min\{(\pi(z^*)q(z^*, z_n)/\pi(z_n)q(z_n, z^*)), 1\} > u_n$, then $z_{n+1} = z^*$; otherwise $z_{n+1} = z_n$, where $u_n \sim U(0, 1)$ are iid uniform random variables. The parameter $\alpha = 1$ corresponds to the Metropolis-adjusted Langevin algorithm, (Robert and Casella, 1999; Roberts and Rosenthal, 2001) while $\alpha = 0$ gives the usual random walk Metropolis–Hastings (Robert and Casella, 1999). The appropriate choice of the proposal covariance Λ , step-size δ_n , and their adaptation is discussed in detail in Ref. (Apte et al., 2008b, Sect. (5.2)).

Continuing the study begun in Apte et al. (2008b), we argue in this paper that this Bayesian posterior density $P_{\text{ex}}(x(t)|\hat{y}_1, \dots, \hat{y}_K)$ – henceforth referred to as *exact posterior* – can be used as a key tool in understanding the effects of nonlinearity on data assimilation problems. (Zhou et al., 2006) The method we use for this purpose is to compare this exact posterior with an approximation of it given by the *ensemble Kalman filter (EnKF)* algorithm (Evensen, 2009), which we also outline below for further reference in this paper, using the same notation as the above discussion.

EnKF begins with an ensemble $\{x_u^1(t_0), \dots, x_u^N(t_0)\}$ of N samples of the state at initial time, drawn from the prior density $p_\zeta(x_0)$. For each of the observational time instances t_1, \dots, t_K , the following steps are performed:

1. Evolve the “updated” ensemble $\{x_u^1(t_{i-1}), \dots, x_u^N(t_{i-1})\}$ from time t_{i-1} to t_i to get the “forecast” ensemble $\{x_f^1(t_i), \dots, x_f^N(t_i)\}$. I.e., $x_f^j(t_i) = \Phi(x_u^j(t_{i-1}); t_{i-1}, t_i)$ for each $j = 1, \dots, N$.
2. Calculate the ensemble mean $\bar{x}_f(t_i) = (\sum_{k=1}^N x_f^k(t_i))/N$ and ensemble covariance

$$\bar{P}_f(t_i) = \frac{1}{N-1} \left[\sum_{k=1}^N \delta_f^k (\delta_f^k)^T \right]$$

where $\delta_f^k = (x_f^k(t_i) - \bar{x}_f(t_i))$. In practice, for large system size n , typical EnKF algorithms avoid calculating the $n \times n$ covariance $\bar{P}_f(t_i)$, by using appropriate matrix identities involving the much smaller $n \times N$ matrix formed with δ_f^k as columns.

3. Find an updated ensemble at time t_i in such a fashion that the ensemble mean $\bar{x}_u(t_i)$ and covariance $\bar{P}_u(t_i)$ of this updated ensemble are the same as the mean and covariance given by the following equations, which are basically the update equations of the Kalman filter:

$$\bar{x}_u(t_i) = \bar{x}_f(t_i) + K (\hat{y}_i - H \bar{x}_f(t_i)) \tag{6}$$

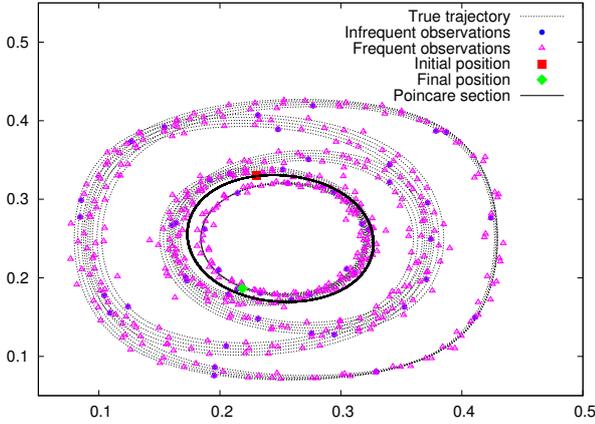


Fig. 3. The trajectory used to study the effect of shear on data assimilation

and

$$\bar{P}_u(t_i) = (I - KH)\bar{P}_f(t_i), \quad (7)$$

where

$$K = \bar{P}_f(t_i)H^T \left(H\bar{P}_f(t_i)H^T + R_i \right)^{-1} \quad (8)$$

and $H = \nabla h(\bar{x}_f(t_i))$. Thus, it is clear that this update process is nonunique since only the first two moments of this updated or “posterior” density are specified by the above two equations. This nonuniqueness gives rise to the many different versions of the EnKF available in the literature, such as the ensemble transform Kalman filter (ETKF) (Bishop et al., 2001) or the ensemble adjustment Kalman filter (EAKF) (Anderson, 2001). In our numerical experiments described below, we implement the perturbed observations EnKF (Burgers et al., 1998), (Evensen, 2009, pp. 41-44) with large enough ensemble size. We also choose an ensemble size between 10^4 and 10^5 in many of the cases shown below and this is chosen so that increasing the ensemble size does not affect the results. Also note that because of the large ensemble size, the results are expected to be independent of the choice of Kalman filter such as perturbed observation EnKf or EAKF or ETKF.

3.1 Comparison between EnKF and exact sampling

For any observation time t_i , we will consider the following four ensembles.

1. The ensemble of samples drawn from the exact prior $P_{\text{ex}}(x(t_i)|\hat{y}_1, \dots, \hat{y}_{i-1})$, i.e., distribution for state at time t_i conditioned on observations up to time t_{i-1} , denoted by $p_i^{\text{ex,pr}}$, and shown consistently with black in all the plots.

2. The ensemble drawn from the EnKF approximation of this prior $P_{\text{kf}}(x(t_i)|\hat{y}_1, \dots, \hat{y}_{i-1})$, denoted by $p_i^{\text{kf,pr}}$ and shown with magenta.
3. The ensemble drawn from the exact posterior $P_{\text{ex}}(x(t_i)|\hat{y}_1, \dots, \hat{y}_i)$, denoted by $p_i^{\text{ex,po}}$ and shown with red.
4. The ensemble drawn from the EnKF approximation of this posterior $P_{\text{kf}}(x(t_i)|\hat{y}_1, \dots, \hat{y}_i)$, denoted by $p_i^{\text{kf,po}}$ and shown with blue.

Note that the first two of these are obtained by pushing forward from time t_{i-1} to time t_i the ensemble of samples drawn from $P_{\text{ex}}(x(t_{i-1})|\hat{y}_1, \dots, \hat{y}_{i-1})$ and from $P_{\text{kf}}(x(t_{i-1})|\hat{y}_1, \dots, \hat{y}_{i-1})$, respectively. The EnKF is considered to “fail” when the $P_{\text{kf}}(x(t_i)|\hat{y}_1, \dots, \hat{y}_i)$ is a poor approximation of the $P_{\text{ex}}(x(t_i)|\hat{y}_1, \dots, \hat{y}_i)$. This is to be contrasted with what is commonly known as divergence of EnKF (or filtering/smoothing methods in general), which refers to comparing the results with a “true” state of the system that is known (Brett et al., 2011). We will later also comment on the approximation of the “true” state both by the exact and the EnKF posteriors.

We will see later that in many cases including those when EnKF fails, the latter two distributions, i.e., the posteriors, are approximately Gaussian. The main deciding factor for failure or success of EnKF will be seen to be the Gaussianity or the lack of it for the prior $P_{\text{kf}}(x(t_{i-1})|\hat{y}_1, \dots, \hat{y}_{i-1})$. This is the main reason for considering the above four distributions. Of course, if the dynamical system is linear, a Gaussian $P_{\text{kf}}(x(t_{i-1})|\hat{y}_1, \dots, \hat{y}_{i-1})$ will lead to Gaussian $P_{\text{kf}}(x(t_i)|\hat{y}_1, \dots, \hat{y}_{i-1})$, and thus EnKF will not fail, as has been known to be the case.

4 The shear effect near the center

In this section we will focus on comparison of exact Bayesian posterior and the EnKF posterior for a trajectory, shown in Fig. 3, that is near the center at $(x, y) = (0.25, 0.25)$. The true initial condition is $(u_0, u_1, v_1, h_1, x, y) = (1.0, 0.5, 0.8, 0.7, 0.23, 0.33)$. In all these experiments, the observational error covariance R is taken so that $\sqrt{R} = \text{diag}(0.005, 0.003)$. For this specific trajectory, we will discuss the following three different numerical experiments.

- In the first case, we choose a relatively large time between the observations $T_{\text{obs}}^L = t_i - t_{i-1} = 0.1$, and these observations are shown by filled circles in Fig. 3. The prior in this case is Gaussian with mean $(0.9, 0.2, 0.2, 0.2, 0.2, 0.3)$ and broad covariance Σ_{ζ}^L , where $\sqrt{\Sigma_{\zeta}^L} = \text{diag}(1.0, 0.7, 0.7, 0.7, 0.005, 0.005)$. Note that we have chosen the velocity covariance to be large but the drifter position covariance to be small.

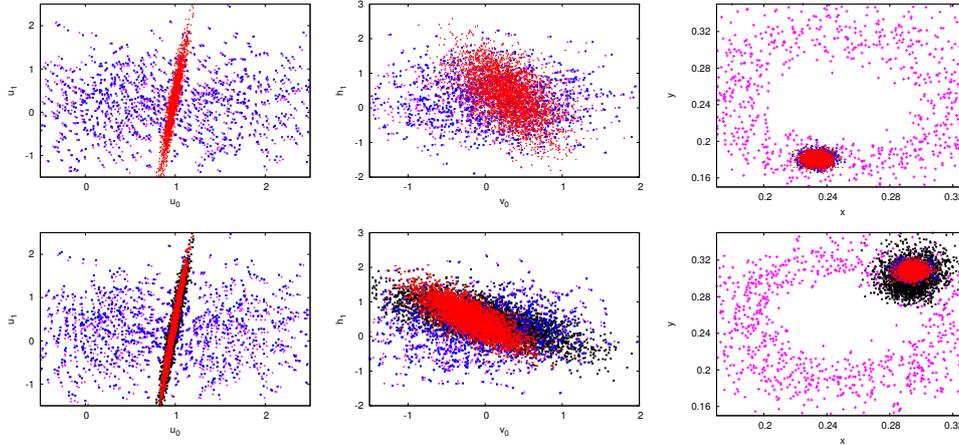


Fig. 4. The four distributions $p_i^{kf,pr}$ (magenta), $p_i^{ex,pr}$ (black), $p_i^{kf,po}$ (blue), and $p_i^{ex,po}$ (red) for the update step of the first observation $i = 1$ at $t = T_{obs}^L$ (top row), and at the second observation $i = 2$ at $t = 2T_{obs}^L$ (bottom row), with $T_{obs}^L = 0.1$. The scatter plots are for (u_0, u_1) (left column), (v_1, h_1) (middle column), and (x, y) (right column) variables. Note that the observations are of the (x, y) variables. The distribution at time $t = 0$ has a “broad” covariance of Σ_ζ^L with $\sqrt{\Sigma_\zeta^L} = \text{diag}(1.0, 0.7, 0.7, 0.7, 0.005, 0.005)$.

None of the remarks we make below about the posterior are affected by this choice since the drifter positions are observed with error covariance comparable to this prior covariance.

- In the second case, we choose the same time $T_{obs}^L = 0.1$ between observations as in the first case (and also the same realization of the observations), but choose a tight prior covariance Σ_ζ^S with $\sqrt{\Sigma_\zeta^S} = \text{diag}(0.1, 0.07, 0.07, 0.07, 0.005, 0.005)$.
- In the last case, we choose the broad prior covariance Σ_ζ^L as in the first case, but choose a smaller time between the observations: $T_{obs}^S = 0.01$. These frequent observations are shown by open triangles in Fig. 3.

We will see that in the first case, the EnKF fails to approximate the posterior, whereas in the latter two cases, it provides a pretty accurate representation of the posterior. We will see in Sect. 4.1 that this is precisely because of the shear near the center. The above choice of the specific trajectory is only representative of this phenomena, which is observed in all the trajectories that show qualitatively the same feature, namely the shear. Our studies so far do not lead to quantifying the effect of varying T_{obs} or the prior covariance Σ_ζ , but we are currently working on this aspect.

4.1 The failure of EnKF update steps due to shear effect

Figure 4 shows the exact and EnKF priors and the posteriors $p^{ex,pr}$ (black), $p^{kf,pr}$ (magenta), $p^{ex,po}$ (red), and $p^{kf,po}$ (blue) (see Sect. 3.1 for the notation) for the update step involving the first and second observation, i.e., at times $t = 0.1, 0.2$, for

the first of the three cases described above, i.e., broad prior with covariance Σ_ζ^L at the initial time and infrequent observations with $T_{obs}^L = 0.1$. Note that for the first observation, $p_1^{kf,pr} = p_1^{ex,pr}$ since this is simply the distribution at time $t = 0$ pushed forward up to time $t = T_{obs}^L$. Since the posterior after incorporating the first observation using EnKF is different from the exact posterior, i.e., $p_1^{ex,po} \neq p_1^{kf,po}$, the prior at other observation times is different for EnKF and exact sampling method, i.e., $p_i^{ex,pr} \neq p_i^{kf,pr}$ for $i > 1$.

It is clear from this figure that the main cause of the failure of EnKF in this case is the highly non-Gaussian nature of the prior distribution $p_1^{kf,pr}$. Clearly, a Gaussian approximation of this prior will have a mean that is near the center $(x, y) = (0.25, 0.25)$, which is a very low probability state. Of course, since the drifter position is observed with high accuracy, the EnKF posterior in the position variables is very close to the exact posterior. But the EnKF fails in the update of the *unobserved* velocity components – the EnKF posterior in the velocity variables is a poor approximation of the exact posterior. We also see that since the shear around the center is mainly affected by the time independent mode u_0 and not much by the time-dependent modes (u_1, v_1, h_1) , the update in u_0 is the worst for EnKF, whereas the update in other velocity components is reasonably good.

In the case of infrequent observations with $T_{obs}^L = 0.1$ but with a prior that is also broad in the drifter position components, e.g., $\sqrt{\Sigma_\zeta} = \text{diag}(1.0, 0.7, 0.7, 0.7, 0.5, 0.5)$, the above discussed distributions (not shown) are almost identical to those shown in Fig. 4, except those at the first observation time $t = T_{obs}^L$ (where the prior in position variables have a very large covariance). This is because the large prior in position components only affects $p_1^{ex,pr}$, which is very broad

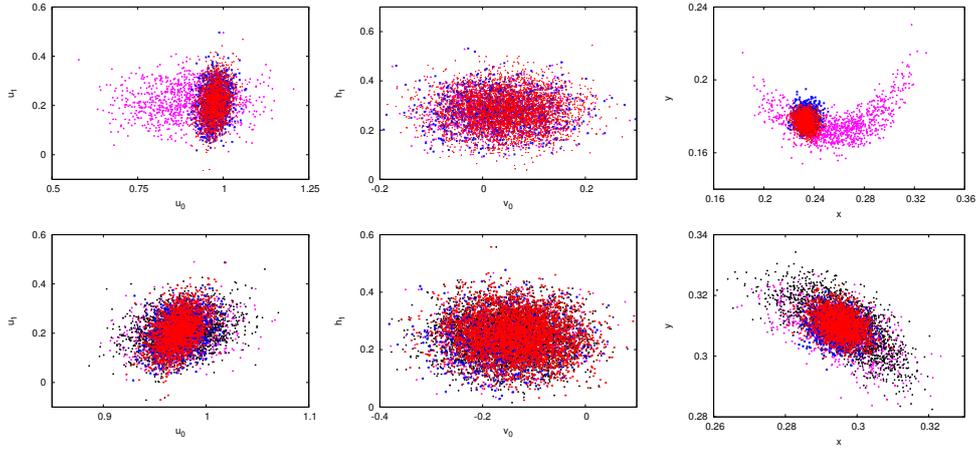


Fig. 5. Same as Fig. 4, except that the prior distribution at time $t=0$ has a “narrow” covariance with $\sqrt{\Sigma_\zeta^S} = \text{diag}(0.1, 0.07, 0.07, 0.07, 0.005, 0.005)$.

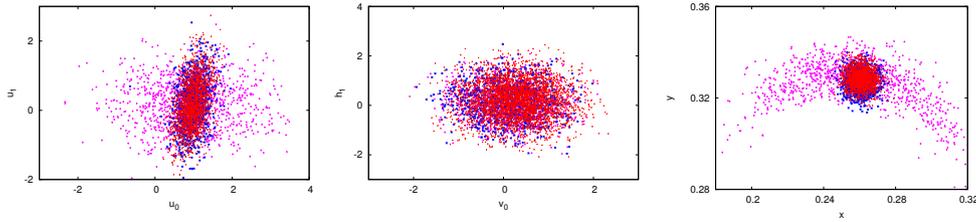


Fig. 6. Same as Fig. 4, except that $T_{\text{obs}}^S = 0.01$ and distributions at only the first observation time $t = T_{\text{obs}}^S = 0.01$ are shown.

in this case. But by assimilating the observation of the position that has a much smaller error covariance, the posterior $p_1^{\text{ex,po}}$ is seen to be identical to that shown in the above figure.

In contrast, Fig. 5 shows the prior and posterior distributions for the second of the three cases, i.e., still with infrequent observations with $T_{\text{obs}}^L = 0.1$ but using a prior that is narrow in velocity components. We see that even at the first observation time, the prior $p_1^{\text{kf,pr}}$ is not as severely non-Gaussian as in the previous case, leading to a good approximation of the posterior, i.e., $p_1^{\text{kf,po}} \approx p_1^{\text{ex,po}}$. Because of this the EnKF posterior gives a much better approximation of the exact one at later times, as seen in the second row.

Another way by which the effect of shear in producing the non-Gaussianity is suppressed is obviously by decreasing the time between the observations, as shown in Fig. 6. The prior chosen is the same as the first case presented above, but the time between the observations is $T_{\text{obs}}^L = 0.01$. The resemblance with Fig. 5 is quite striking. In fact, the posterior and prior distributions at the second observation time are not qualitatively different from those shown in Fig. 5 and hence not shown.

In order to quantify this effect, we will compare the so-called degree of freedom (DOF) for signal, denoted $d_s(t)$, which in the case of data assimilation has the form from

Zupanski et al. (2007),

$$d_s(t) = \text{tr} \left[\mathbf{I} - \Sigma(p(t))(\Sigma_\zeta)^{-1} \right]. \quad (9)$$

Here, Σ_ζ is the covariance of the prior distribution at time $t=0$, while $\Sigma(p)$ is the covariance of the distributions $p^{\text{ex,pr}}$, $p^{\text{ex,po}}$, $p^{\text{kf,pr}}$, or $p^{\text{kf,po}}$. The larger the value of d_s , the more information is gained through assimilation and it asymptotes to $\text{tr}(\mathbf{I})$. This is shown in Fig. 7 for both the broad prior (top row) and the narrow prior (bottom row). The left column shows the degree of freedom for the distributions on all the six variables of the model, while the right column shows it for the marginal distributions on only the velocity variables (u_0, u_1, v_1, h_1). We notice that in the case of a broad prior, the exact posterior has more information than the prior at all the times. In fact, in our numerical experiments, at time $t=5.0$, the value of d_s for the posterior approaches 6, which is the theoretical maximum, as shows in Table 1. On the other hand, the EnKF posterior is unable to gain such information. As indicated before, when the prior is narrow, the EnKF posterior has as much information as the exact posterior.

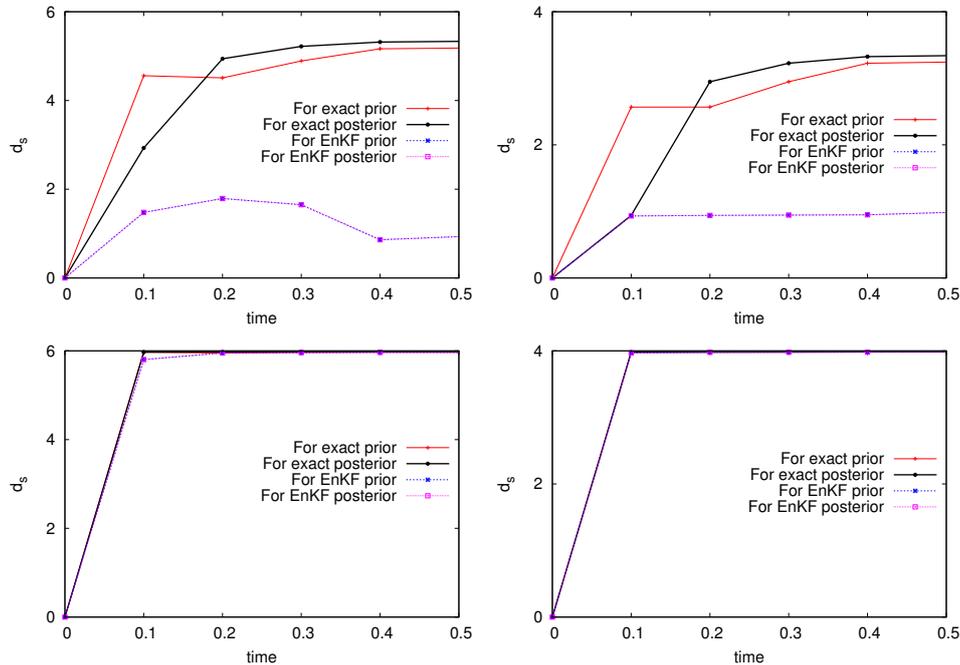


Fig. 7. The degree of freedom Eq. (9) for various distributions described in the text. Left column is for the full distributions on all the six variables of the model $(u_0, u_1, v_1, h_1, x, y)$, while the right column is for the velocity variables (u_0, u_1, v_1, h_1) . Top row is for the case of broad prior (see Fig. 4), bottom row for narrow prior (see Fig. 5).

Table 1. The degree of freedom Eq. (9) for various distributions described in the text.

Time	Broad prior				Narrow prior			
	$p^{ex,pr}$	$p^{ex,po}$	$p^{kf,pr}$	$p^{kf,po}$	$p^{ex,pr}$	$p^{ex,po}$	$p^{kf,pr}$	$p^{kf,po}$
0.0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.1	4.559	2.930	1.478	1.478	2.568	0.940	0.931	0.931
0.2	4.510	4.940	1.791	1.791	2.568	2.948	0.940	0.940
0.3	4.891	5.218	1.650	1.650	2.948	3.226	0.945	0.945
0.4	5.164	5.317	0.861	0.861	3.226	3.325	0.949	0.949
5.0	5.980	5.979	4.143	4.143	3.980	3.979	2.549	2.549

4.2 Effect of nonlinearity on assimilation over longer time periods

Figure 8 shows the posterior distributions at time $t = 5.0$, conditioned on 50 observations in the first two of the above three cases with $T_{obs}^L = 0.1$ (top and middle rows, with broad and narrow priors at $t = 0$, respectively) and conditioned on 500 observations in the last of the three cases with $T_{obs}^S = 0.01$ (bottom row). We see that in the last of the three cases, when the prior is broad and the time between the observations is short, the EnKF approximates the posterior very well. At the other extreme, when the prior is broad and the time between the observations is long, the EnKF gives a very poor approximation of the posterior. In the case when the prior is narrow but the time between the observations is long, the

EnKF does produce distributions whose covariance is comparable to that of the exact distributions but the mean is systematically biased compared to the exact distribution.

5 Saddle effect

The nonlinear divergence of trajectories, as illustrated in Fig. 2, poses a challenge for data assimilation methods in general. In this section we describe the comparison of the EnKF and the exact posteriors in order to understand in detail the origin of the saddle effect, using a chaotic trajectory, shown in Fig. 9, which is near the saddle point at $(x, y) = (0.5, 0)$. The true initial condition is $(u_0, u_1, v_1, h_1, x, y) =$

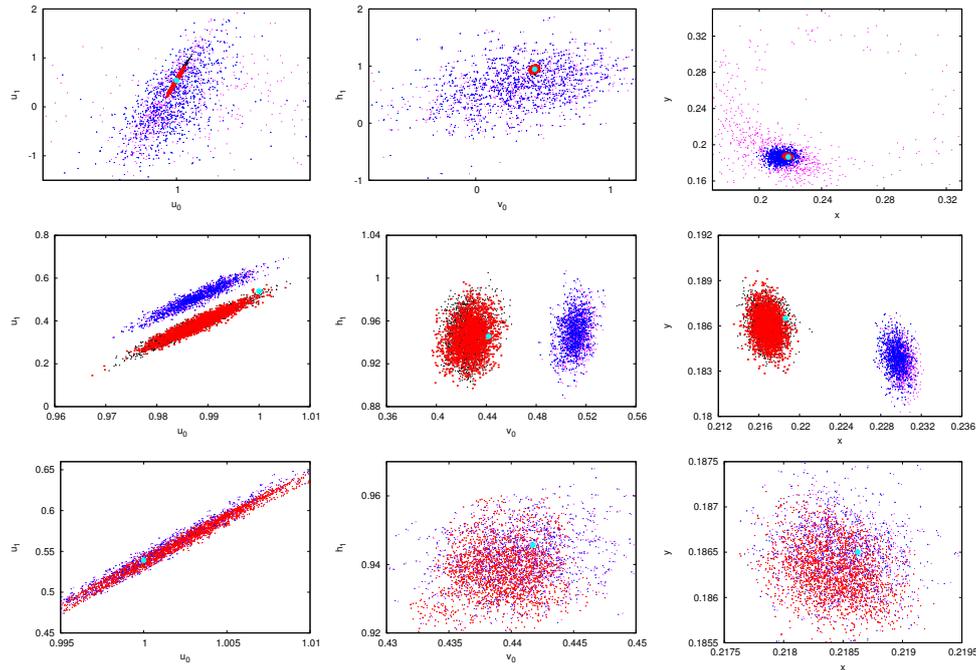


Fig. 8. The four posterior distributions $p_i^{kf,pr}$ (magenta), $p_i^{ex,pr}$ (black), $p_i^{kf,po}$ (blue), and $p_i^{ex,po}$ (red) for the first two of the cases described above, with $T_{obs}^L = 0.1$, for the update step of the last observation $i = 50$ at $t = 50T_{obs}^L = 0.5$: broad prior (top) and narrow prior (middle row). The bottom row is for the third case with $T_{obs}^S = 0.01$, but for the update of the 500th observation, thus at the same $t = 500T_{obs}^S = 0.5$. Note that the true values of the variables at this time are $(u_0, u_1, v_1, h_1, x, y) = (1.0, 0.539, 0.442, 0.946, 0.21862, 0.18650)$, shown by the cyan dot in the figures.

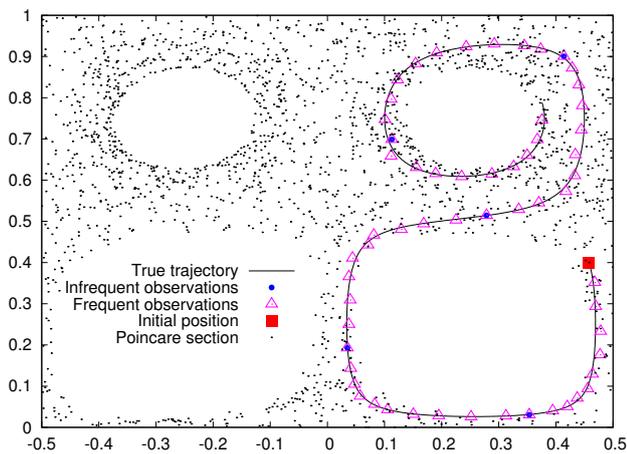


Fig. 9. The trajectory used to study the effect of saddles on data assimilation. Note that the Poincaré plot, shown by black dots, of this trajectory shows that it is a chaotic trajectory.

$(1.0, 0.5, 0.8, 0.7, 0.458, 0.4)$, and the observational error covariance R is fixed so that $\sqrt{R} = \text{diag}(0.005, 0.003)$.

As in the case of the center, Fig. 10 shows the posterior and prior distributions when assimilating frequent and infrequent observations with a broad and a narrow prior using the

EnKF as well as using exact sampling. The broad conclusions we can draw from these are the same as in the case of the center, except that the cause of these effects is different. In particular, the prior at each observational time is still non-Gaussian in the case when the time between the observations is large: $T_{obs}^L = 0.1$ and the prior has a broad prior covariance Σ_ξ^L with $\sqrt{\Sigma_\xi^L} = \text{diag}(1.0, 0.7, 0.7, 0.7, 0.005, 0.005)$ (first and second row of Fig. 10 for assimilation of first and second observations, respectively). In contrast with the data for the trajectory around the center, the observations in this case do contain information about the time-dependent modes (u_1, v_1, h_1) , as noted in Sect. 2.1. Thus, the exact posteriors in all the components are much more narrow compared with assimilation of trajectory near a center. Hence, we see here an example of how information about different aspects of velocity flow is obtained through different trajectories, either near the saddle or near the centers. Similar conclusions have been discussed and used to propose different drifter placement strategies in Salman et al. (2008).

Recall that in the case of trajectory near the center, we looked at the effect of assimilating observations over long time periods, on both the exact and the EnKF posteriors. In contrast, for chaotic trajectories such as the one shown in Fig. 9, it becomes exceedingly difficult to effectively sample the posterior distributions on initial conditions, conditioned

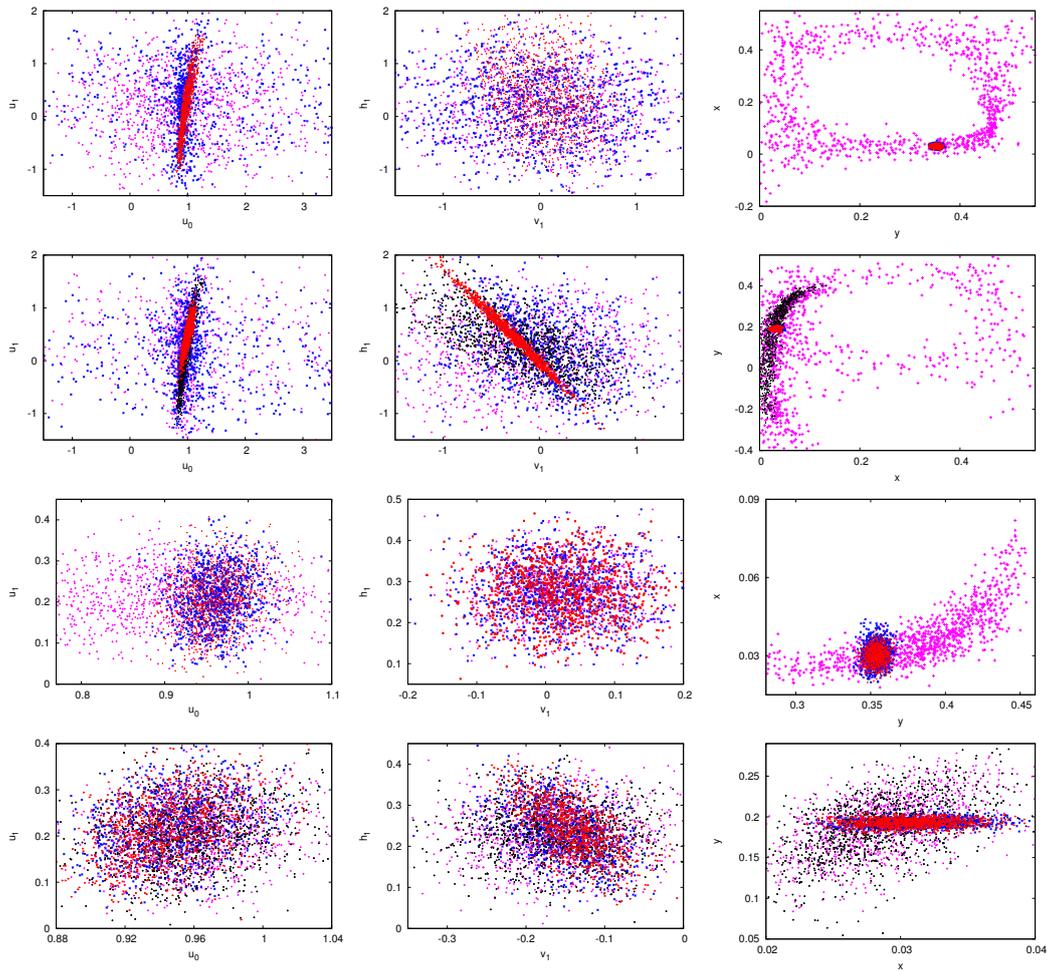


Fig. 10. Same as Figs. 4 (broad prior and infrequent observations, top row is update step at the first observation, second row at second observation) and 5 (narrow prior and infrequent observations, third row is update step at the first observation, bottom row at second observation), except that the observations from the trajectory near the saddle are assimilated.

on more than a few observations. This is clearly because of the sensitive dependence of such trajectories on initial conditions. It is known that assimilation of such trajectory data may lead to the divergence of EnKF – the so-called saddle effect (Kuznetsov et al., 2003) – in the sense that the EnKF estimates and the true trajectory diverge from each other. Our current work or other previous studies of the Lagrangian data assimilation problem have *not* yet given sufficient indications about the conditions under which the EnKF distribution converges to the exact distribution in the case of trajectories near the saddle, though there are previous results proving convergence of the EnKF distribution to the exact distribution when the dynamics is linear, but to a distribution that may not be the exact Bayesian distribution in the case of nonlinear dynamics. (Le Gland et al., 2011)

6 Discussion and future directions

This paper discusses the effects of nonlinearity on data assimilation. Adopting the Bayesian framework, we work with two different DA methods – namely, the exact Markov chain Monte Carlo sampling of the posterior distribution for the initial conditions of the model conditioned on observations over a given time period, and the ensemble Kalman filter (EnKF) approximation of this distribution. Since the EnKF update step is exact only when the prior distribution at each step is Gaussian, the comparison of the exact posterior and the EnKF posterior gives us information about which aspects of nonlinearity play a significant role in DA. Some of the main conclusions we draw from this comparison are summarized below.

The flow near an elliptic fixed point, or center, in a nonlinear dynamical systems generically has a shear or differential speed of rotation. Specifically, the initial conditions near the

center rotate faster than those farther away. The probability densities on the phase space evolving under such flow naturally develop non-Gaussianity.

The flow near a hyperbolic fixed point, or saddle, in a nonlinear system is qualitatively and quantitatively similar to the flow near a saddle in the linearized system. But in the case of a nonlinear system, the saturation of the diverging parts leads to the eventual development of characteristic non-Gaussian features of the densities.

The presence of such non-Gaussian densities leads to the failure of data assimilation techniques such as the ensemble Kalman filter (EnKF). We emphasize our viewpoint, also presented in (Apte et al., 2008b), that the failure or success of data assimilation scheme such as the EnKF is measured by comparing its outcome with the exact posterior distribution Eq. (5) given by Bayes' rule. This is to be contrasted with the so-called divergence of EnKF (or filter divergence in general), which refers to the growing difference between the filter estimate and the true state of the system.

The comparisons between the exact posterior and the EnKF posterior reveal that the shear effect is one of the main causes of EnKF failure. In fact, even in the case of failure of EnKF for chaotic trajectories near the saddle, it is the differential speed of divergence and ultimately the shear that causes the failure of the EnKF.

We note that the EnKF fails when the observations are infrequent and the prior at the initial time is broad. We show using examples that the effects of nonlinearity are less prominent either when the observations are frequent or when the prior is narrow. We also emphasize the effects of nonlinearity on the marginal posterior distributions of unobserved variables. In our numerical experiments, we have taken the observations to be relatively accurate, and because of which the marginal posterior distribution in the observed variables is very close to the observational likelihood, but the nonlinearity leads to a poor EnKF approximation of the marginal posterior of unobserved variables. Using numerical observations for the case of trajectories near the center, we surmise that this is precisely the cause of EnKF divergence.

Finally, We have hinted at the connection between the failure of the EnKF (mismatch between exact posterior and EnKF posterior) and the filter divergence (mismatch between filter estimate and the true state of the system) over long periods of time. But because of sensitivity to initial conditions, sampling the exact (smoothing) posterior conditioned on data over long time periods becomes prohibitively difficult for the MCMC method we have used for the studies in this paper.

Acknowledgements. The authors would like to thank the two referees for valuable comments that significantly improved the discussion, in particular, about quantifying the effects of nonlinearity. This work was supported by the Office of Naval Research, grant numbers N00014-11-1-0087 and N00014-12-1-0257.

Edited by: O. Talagrand

Reviewed by: two anonymous referees

References

- Anderson, J. L.: An Ensemble Adjustment Kalman Filter for Data Assimilation, *Mon. Weather Rev.*, 129, 2884, 2001.
- Apte, A., Hairer, M., Stuart, A., and Voss, J.: Sampling the posterior: an approach to non-Gaussian data assimilation, *Physica D*, 230, 50–64, 2007.
- Apte, A., Jones, C., Stuart, A., and Voss, J.: Data assimilation: Mathematical and statistical perspectives, *Int. J. Numer. Meth. Fluids*, 56, 1033–1046, 2008a.
- Apte, A., Jones, C., and Stuart, A. M.: A Bayesian approach to Lagrangian data assimilation, *Tellus A*, 60, 336–347, 2008b.
- Bishop, C., Etherton, B., and Majumdar, S.: Adaptive sampling with the ensemble transform Kalman filter. part I: theoretical aspects, *Mon. Weather Rev.*, 129, 420–436, 2001.
- Brett, C., Lam, K., Law, K., McCormick, D., Scott, M., and Stuart, A.: Stability of Filters for the Navier-Stokes Equation, 2011.
- Burgers, G., van Leeuwen, P. J., and Evensen, G.: Analysis scheme in the ensemble Kalman filter, *Mon. Weather Rev.*, 126, 1719–1724, 1998.
- Chyba, M.: *Ocean Eng.*, 36, 1, 1–132 (January 2009), Special Issue on “Autonomous Underwater Vehicles”, 2009.
- Evensen, G.: *Data Assimilation: the Ensemble Kalman Filter*, Springer, 2009.
- Ide, K., Courtier, P., Ghil, M., and Lorenc, A.: Unified notation for data assimilation: Operational, sequential and variational, *J. Meteor. Soc. Jpn.*, 75, 181–189, 1997.
- Ide, K., Kuznetsov, L., and Jones, C.: Lagrangian data assimilation for point-vortex system, *J. Turbulence*, 3, 053, , 2002.
- Kuznetsov, L., Ide, K., and Jones, C.: A Method for Assimilation of Lagrangian Data, *Mon. Weather Rev.*, 131, 2247, 2003.
- Le Gland, F., Monbet, V., and Tran, V.-D.: Large sample asymptotics for the ensemble Kalman filter, in: *Handbook on Nonlinear Filtering*, edited by: Crisan, D. and Rozovskii, B., 598–631, Oxford University Press, 2011.
- Lei, J., Bickel, P., and Snyder, C.: Comparison of ensemble Kalman filters under non-Gaussianity, *Mon. Weather Rev.*, 138, 1293–1306, 2010.
- Nurujjaman, M., Shivamurthy, S., Apte, A., Singla, T., and Parmananda, P.: Effect of discrete time observations on synchronization in Chua model and applications to data assimilation, *Chaos*, 22, 023125, 2012.
- Robert, C. and Casella, G.: *Monte Carlo statistical methods*, Springer, New York, 1999.
- Roberts, G. and Rosenthal, J.: Optimal scaling for various Metropolis-Hastings algorithms, *Statistical Science*, 16, 351–367, 2001.
- Salman, H., Kuznetsov, L., Jones, C., and Ide, K.: A Method for Assimilating Lagrangian Data into a Shallow-Water Equation Ocean Model, *Mon. Weather Rev.*, 134, 1081–1101, 2006.

Salman, H., Jones, C., and Ide, K.: Using flow geometry for drifter deployment in Lagrangian data assimilation, *Tellus A*, 60, 321–335, 2008.

The Argo Science Team: Argo: The Global Array of Profiling Floats, in: *Observing the Oceans in the 21st Century*, edited by Koblinsky, C. and Smith, N., pp. 248–258, GODAE Project Office and Bureau of Meteorology, Melbourne, 2001.

Zhou, Y., McLaughlin, D., and Entekhabi, D.: Assessing the performance of the ensemble Kalman filter for land surface data assimilation, *Mon. Weather Rev.*, 134, 2128–2142, 2006.

Zupanski, D., Hou, A. Y., Zhang, S. Q., Zupanski, M., Kummerow, D., and Cheung, S. H.: Applications of information theory in ensemble data assimilation, *Q. J. R. Meteorol. Soc.*, 1545, 1533–1545, 2007.