Gaussian anamorphosis in the analysis step of the EnKF: a joint state variable / observation approach

Javier Amezcua*, Peter Jan van Leeuwen
Department of Meteorology, University of Reading

August 24, 2014

Abstract

The analysis step of the (ensemble) Kalman filter is optimal when (i) the distribution of the background is Gaussian, (ii) state variables and observations are related via a linear operator, and (iii) the observational error is of additive nature and has Gaussian distribution. When these conditions are largely violated, a pre-processing step known as Gaussian anamorphosis can be applied. The objective of this procedure is to obtain state variables and observations that better fulfil the Gaussianity conditions in some sense.

In this work we analyse GA from a joint perspective, paying attention to the effects of transformations in the joint state variable / observation space. First, we study transformations for state variables and observations that are independent from each other. Then, we introduce a targeted joint transformation which objective is to obtain joint Gaussianity in the transformed space. We focus primarily in the univariate case, and briefly comment on the multivariate one.

A key point of this paper is that, when (i)-(iii) are violated, using the analysis step of the EnKF will not recover the exact posterior density in spite of any transformations one may perform. These transformations, however, provide approximations of different quality to the Bayesian solution of the problem. Using an example in which the Bayesian posterior can be analytically computed, we assess the quality of the analysis distributions generated after applying the EnKF analysis step in conjunction with different GA options. The value of the targeted joint transformation is particularly clear for the case when the prior is Gaussian, the marginal density for the observations is close to Gaussian, and the likelihood is a Gaussian mixture.

* *Corresponding author address:* Javier Amezcua, University of Reading, PO Box 243 RG66BB UK
E-mail: j.AmezcuaEspinosa@reading.ac.uk

1 Introduction

It is often the case, when estimating a variable of interest, that one only counts with imperfect sources of information. For example, to determine the value of an atmospheric variable at a given location, one can rely on a short term forecast and observations of the variable, both of which contain errors. Data assimilation (DA) is the process of combining these sources of information in a way that is optimal in some predefined sense (see e.g. Cohn 1997).

This paper deals with a particular aspect of sequential DA methods. These methods have two steps. In the forecast step, the state estimator is evolved in time following some dynamical model, along with some measure of its uncertainty. Whenever an observation becomes available, the information from this observation is combined with that provided by the forecast (also called background) to produce a better estimate (denominated analysis). This is known as the analysis step.

In the present work we focus only on the latter step. Hence, we consider that when an observation arrives we have already got a background estimate (regardless of the way this was obtained). We consider both the background and the observations to contain random errors with some prescribed probability density functions (pdf's). Under such probabilistic framework, the aim of the analysis step is to obtain the posterior pdf of the variable of interest. In theory, this can be achieved through a direct application of Bayes theorem. Nonetheless, in practice this can result a difficult task since a complete representation of the distributions for the prior and the likelihood is required.

When dealing with full pdf's is not possible, one can work with summary statistics for both the background and the likelihood. For example, the analysis equations of the Kalman filter (KF: Kalman, 1960; Kalman and Bucy 1961) provide a method to update the first two moments (mean and covariance) of the state variable from background to analysis. In large scale applications, such as numerical weather prediction (NWP) or oceanography, the background statistics are usually obtained from samples (ensemble Kalman filter (EnKF) Evensen, 2006).

Under special conditions: (i) Gaussianity in the prior, (ii) linearity of the observation operator and (iii) Gaussianity in the additive observational error density, the solution given by the analysis step of the KF/EnKF provides the sufficient statistics of the Bayesian solution (the sampling nature of the EnKF obviously introduces statistical error in this case). If these conditions are not fulfilled, the application of the (En)KF analysis equations is suboptimal, but it can still be useful.

In some cases, however, the deviation from these conditions can be quite important. This happens, for example, when the prior is multimodal or when it does not have the statistical support (domain) of the Gaussian distribution. The latter is the case of positive definite variables such as precipitation (see e.g. Lien et al, 2013), and bounded quantities such as relative humidity. Large deviations from Gaussianity in the prior are not uncommon in many fields, for example in physical-biological models (Bertino et al, 2003; Simon and Bertino, 2009; Beal et al, 2010; Doron et al, 2011). Non-Gaussian pdf's can also result from the deformation of an original Gaussian pdf during the forecast step when the model is strongly nonlinear (Miller et al, 1994; Miller et al ,1999). Problems can also arise if the likelihood presents extreme non-Gaussian features.

In these cases, either of two options can be taken. One can select an analysis step based on methods that do not require Gaussianity, e.g. the rank histogram filter (Anderson, 2010). On the other hand, one can still apply the (En)KF analysis step, in conjunction with a procedure known as Gaussian anamorphosis (GA). This involves transforming the state variable and observations $\{\mathbf{x}, \mathbf{y}\}$ into new variables $\{\tilde{\mathbf{x}}, \tilde{\mathbf{y}}\}$ which present Gaussian features (details will be given in section **??**). The (En)KF analysis equations are computed using the new variables, and the resulting analysis is mapped back into the original space using the inverse of the transformation. GA is a well-known technique in the geostatistics community (see e.g. Wackernagel 2003), and it was introduced to the DA community in a seminal paper by Bertino et al. (2003). Since then, it has

been explored in different works (e.g. Zhou et al, 2011; Brankart et al, 2012; Simon and Bertino, 2012).

A possible drawback of anamorphosis is that as a result of the transformations a (generally nonlinear) observation operator is introduced in the new space (Bocquet et al, 2010). Although one can apply the EnKF with nonlinear observation operators (see e.g. Hunt et al, 2007), it seems undesirable to solve one problem (non-Gaussianity) at the cost of creating another. A central idea in this work is that different transformations in the state variable and/or observations can achieve different objectives: marginal Gaussianity in the state variables, marginal Gaussianity in the observations, joint Gaussianity in the pair {state variable, observation}. As we will see, different transformations will bring different side effects. Is there an optimal strategy to follow when performing anamorphosis? If not, how do different transformations compare? These questions are at the heart of this paper.

This paper has 3 main objectives. The first is to study anamorphosis transformations using a joint statistical approach between state variables and observations. The second is to visualise the effect that different transformations have on the joint probability space in which the EnKF is used. The third is to introduce a targeted joint state-variable/observation transformation which maps the pair of an arbitrary prior probability and arbitrary likelihood into a joint Gaussian space. In order to assess the performance of the different transformations, we choose an example in which we are able to compute analytically the posterior pdf of the model variables for different given observations. It is against these exact posteriors that we compare the EnKF-generated analysis pdf's.

The rest of this paper is organized as follows: section 2 discusses the DA analysis step in more detail, the probabilistic formulation and the EnKF solution. Section 3 introduces and explains the concept of GA. Section 4 discusses the implementation of GA, studying the existing methods and introducing a new targeted joint state-variable/observation transformation. In section 5 we perform study cases of the methods discussed in section 4. Section 6 includes the conclusions and some discussion.

Some remarks on notation will be useful before starting. We will try to follow (sometimes loosely) the convention of Ide et al (1997) in what respects to sequential DA. Pdf's will be denoted as $p_{\boldsymbol{\xi}}(\boldsymbol{\xi})$, while cumulative density functions (cdf's) will be denoted as $P_{\boldsymbol{\xi}}(\boldsymbol{\xi})$. If we want to explicitly include the parameters when referring to any distribution, this will be done with a semicolon in the argument, e.g. $p_{\boldsymbol{\xi}}(\boldsymbol{\xi};\boldsymbol{\theta}_{\boldsymbol{\xi}})$. The symbol $\sim$ should be read 'distributed as'. We will use $E_{\mathbf{x}}$ to denote expected value, with the subindex indicating the pdf with respect to which this operation is computed. For example,

$$E_{\mathbf{x}}[\boldsymbol{\xi}] = \int_{-\infty}^{\infty} \boldsymbol{\xi} p_{\mathbf{x}}(\boldsymbol{\xi}) d\boldsymbol{\xi} \tag{1}$$

Similarly, $Cov_{\mathbf{x}}[\cdot]$ denotes covariance, with the same meaning for the subindex. The Gaussian distribution will appear frequently in this work. For the sake of brevity, if the random variable $\xi \in \mathbb{R}^1$ follows a Gaussian distribution with mean μ_ξ and variance σ_ξ^2, we will denote its probability density function (pdf) as:

$$p_\xi(\xi) = \phi\left(\xi;\mu_\xi,\sigma_\xi\right) \equiv \phi\left(\frac{\xi-\mu_\xi}{\sigma_\xi}\right) \equiv \frac{1}{\sqrt{2\pi}\sigma_\xi}\exp\left(-\frac{(\xi-\mu_\xi)^2}{2\sigma_\xi^2}\right) \tag{2}$$

and its cdf as:

$$P_\xi(\xi) = \Phi\left(\xi;\mu_\xi,\sigma_\xi\right) \equiv \Phi\left(\frac{\xi-\mu_\xi}{\sigma_\xi}\right) \equiv \frac{1}{\sqrt{2\pi}\sigma_\xi}\int_{-\infty}^{\xi}\exp\left(-\frac{(t-\mu_\xi)^2}{2\sigma_\xi^2}\right) dt \tag{3}$$

For some examples we will also use exponential random variables (rv's). The pdf for this distribution is:

$$p_x(x) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}}, x \geq 0 \tag{4}$$

where $\lambda > 0$ is a scale factor.

2 Analysis step: Bayesian and EnKF solutions

In this section we make use of transformations of rv's; basic concepts of this topic can be found in appendix A. Let $\mathbf{x} \in \mathbb{R}^{N_x}$ denote the vector of state variables, and consider it follows a prior distribution $p_{\mathbf{x}}(\mathbf{x})$. In the most general case, the observations $\mathbf{y} \in \mathbb{R}^{N_y}$ follow the relationship:

$$\mathbf{y} = \hat{h}(\mathbf{x}, \boldsymbol{\eta}) \tag{5}$$

where $\hat{h} : \mathbb{R}^{N_x \times N_y} \to \mathbb{R}^{N_y}$ is a nonlinear observation operator, and $\boldsymbol{\eta} \in \mathbb{R}^{N_y}$ represents the observational error, which follows a distribution $p_{\boldsymbol{\eta}}(\boldsymbol{\eta})$. Consider there exists an inverse $\boldsymbol{\eta} = \hat{h}^{-1}(\mathbf{y}; \mathbf{x})$ (as a function of $\mathbf{y}$), then the likelihood $p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x})$ –conditional pdf of $\mathbf{y}$ given $\mathbf{x}$– can be written as:

$$p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x}) = p_{\boldsymbol{\eta}}(\hat{h}^{-1}(\mathbf{y}; \mathbf{x})) \left| det \left[\frac{\partial}{\partial \mathbf{y}} \hat{h}^{-1}(\mathbf{y}; \mathbf{x}) \right] \right| \tag{6}$$

where $\left| det \left[\frac{\partial}{\partial \mathbf{y}} \hat{h}^{-1}(\mathbf{y}; \mathbf{x}) \right] \right|$ is the absolute value of the determinant of the Jacobian matrix of the transformation. The joint pdf of $\mathbf{x}$ and $\mathbf{y}$ is the product of the likelihood and the prior. The posterior distribution can be computed via Bayes' theorem as:

$$p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}|\mathbf{y}) = \frac{p_{\mathbf{x},\mathbf{y}}(\mathbf{x}, \mathbf{y})}{\int_{-\infty}^{\infty} p_{\mathbf{x},\mathbf{y}}(\mathbf{x}, \mathbf{y}) d\mathbf{x}} = \frac{p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x}) p_{\mathbf{x}}(\mathbf{x})}{p_{\mathbf{y}}(\mathbf{y})} \tag{7}$$

The denominator $p_{\mathbf{y}}(\mathbf{y})$ is the marginal pdf of the observations, and can often be treated as a normalisation factor of the posterior pdf, since it does not depend on $\mathbf{x}$.

This is the most general solution for the DA analysis step. Nonetheless, obtaining the posterior pdf is not an easy task in many occasions, since it requires full knowledge of the two densities involved in the product of the numerator of (**??**). Let us step back and discuss a considerably simpler case; we will build on more complicated cases later. Hence, let (**??**) become:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \boldsymbol{\eta} \tag{8}$$

where $\mathbf{H} \in \mathbb{R}^{N_y \times N_x}$ is a linear operator and the observational error is additive. Define $\boldsymbol{\mu}^b = E_{\mathbf{x}}[\mathbf{x}] \in \mathbb{R}^{N_x}$, $\mathbf{B} = Cov_{\mathbf{x}}[\mathbf{x}] \in \mathbb{R}^{N_x \times N_x}$ (denoted σ^{b^2} in the univariate case), $\mathbf{R} = Cov_{\mathbf{y}|\mathbf{x}}[\mathbf{y}|\mathbf{x}] \in \mathbb{R}^{N_y \times N_y}$ (denoted σ^{o2} in the univariate case), and $E_{\boldsymbol{\eta}}[\boldsymbol{\eta}] = 0$. One can get a minimum variance estimator for the analysis mean $\boldsymbol{\mu}^a \in \mathbb{R}^{N_x}$ as:

$$\boldsymbol{\mu}^a = \boldsymbol{\mu}^b + \mathbf{K}(\mathbf{y} - \mathbf{H}\boldsymbol{\mu}^b) \tag{9}$$

where $\mathbf{K} \in \mathbb{R}^{N_x \times N_y}$ is known as gain, and is computed as:

$$\mathbf{K} = \mathbf{B}\mathbf{H}^{\mathbf{T}}(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathbf{T}} + \mathbf{R})^{-1} \tag{10}$$

The covariance of the analysis is computed as:

$$\mathbf{A} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B} \tag{11}$$

Expressions (**??**) and (**??**) are the KF analysis equations (Kalman 1960; Kalman and Bucy 1961). If, besides having a linear observation operator and additive observational error, both $p_{\mathbf{x}}(\mathbf{x})$ and $p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x})$ are multivariate Gaussians, then using these equations is optimal. This means they produce the sufficient statistics of the full Bayesian posterior. Note that if the set $\{\mathbf{x}, \mathbf{y}\}$ has a joint multivariate Gaussian distribution, then the aforementioned conditions are automatically fulfilled.

It is important to mention that the marginal pdf of the observations is $p_{\mathbf{y}}(\mathbf{y}; \boldsymbol{\mu}^y, \boldsymbol{\Sigma}^y)$, again a multivariate Gaussian, this time with mean and covariance:

$$\begin{aligned} \boldsymbol{\mu}^y &= \mathbf{H}\boldsymbol{\mu}^b \\ \boldsymbol{\Sigma}^y &= \mathbf{H}\mathbf{B}\mathbf{H}^\mathbf{T} + \mathbf{R} \end{aligned} \tag{12}$$

Now, let us partially relax the assumptions on the likelihood. For nonlinear observation operators and additive error, the observation equation is:

$$\mathbf{y} = h(\mathbf{x}) + \boldsymbol{\eta} \tag{13}$$

It should be clear that in this case (**??**) simplifies to $p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x}) = p_{\boldsymbol{\eta}}(\mathbf{y} - h(\mathbf{x}))$. A first order (linear) approximation to the KF analysis equations, known as extended KF (EKF, see e.g. Jazwinski, 1970) can be written as:

$$\begin{aligned} \boldsymbol{\mu}^a &= \boldsymbol{\mu}^b + \mathcal{K}(\mathbf{y} - h(\boldsymbol{\mu}^b)) \\ \mathbf{A} &= (\mathbf{I} - \mathcal{K}\mathcal{H})\mathbf{B} \\ \mathcal{K} &= \mathbf{B}\mathcal{H}^\mathbf{T}(\mathcal{H}\mathbf{B}\mathcal{H}^\mathbf{T} + \mathbf{R})^{-1} \end{aligned} \tag{14}$$

where $\mathcal{H} \in \mathbb{R}^{N_y \times N_x}$ is the tangent linear operator of h, i.e. the Jacobian matrix $\mathcal{H} = \left.\frac{\partial h}{\partial \mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}^b}$.

Formulating the KF analysis equations for a general observation operator as indicated in (**??**) is much more complicated (and further away from optimal conditions), since it would require the linearisation of $\hat{h}(\mathbf{x}, \boldsymbol{\eta})$ with respect to $\boldsymbol{\eta}$ to express:

$$\mathbf{y} = h(\mathbf{x}, \boldsymbol{\eta})\Big|_{\boldsymbol{\eta}=\mathbf{0}} + \left.\frac{\partial h}{\partial \boldsymbol{\eta}}\right|_{\boldsymbol{\eta}=\mathbf{0}} \boldsymbol{\eta} + O(\boldsymbol{\eta}^2) \tag{15}$$

This approximation of course will only be accurate for small $\boldsymbol{\eta}$.

To end this section, it is useful to discuss the analysis step of the ensemble KF (EnKF; see e.g. Evensen 2006). This is a Monte Carlo implementation of the KF, and uses sample statistics. Let us denote the background ensemble as $\mathbf{X}^b = \left[\mathbf{x}_1^b, \ldots, \mathbf{x}_M^b\right]$, where $\mathbf{X}^b \in \mathbb{R}^{N_x \times M}$. The sample mean is:

$$\bar{\mathbf{x}}^b = \frac{1}{M}\sum_{m=1}^{M} \mathbf{x}_m^b \tag{16}$$

An ensemble of perturbations around the mean can be defined as: $\mathbf{X}'^b = \mathbf{X}^b - \bar{\mathbf{x}}^b\mathbf{1}^T$, where $\mathbf{1} \in \mathbb{R}^M$. Then, the sample covariance is:

$$\mathbf{P}^b = \frac{1}{M-1}\mathbf{X}'^b\mathbf{X}'^{bT} \tag{17}$$

The KF analysis equations update both mean and covariance, but in the analysis step of the EnKF it is necessary to update each one of the M ensemble members. This can be done deterministically (ensemble square root filters: Tippett et al, 2003), or stochastically (perturbed-observations EnKF; Burgers et al., 1998). In this work we focus on the stochastic formulation (henceforth EnKF will refer to perturbed-observations EnKF), where each ensemble member is updated as:

$$\mathbf{x}_m^a = \mathbf{x}_m^b + \mathbf{K}(\mathbf{y}_m - \mathbf{H}\mathbf{x}_m^b) \tag{18}$$

where $\mathbf{K}$ is defined as before but using the sample covariances, and the perturbed observations $\{\mathbf{y}_m\}$ are samples from $p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x})$. In particular, if the error is additive they relate to the actual observations by $\mathbf{y}_m = \mathbf{y} + \boldsymbol{\eta}_m$, where $\boldsymbol{\eta}_m$ is a particular realization of the observational error. By

construction, the KF analysis equation for the mean is fulfilled if the perturbed observations are generated such that $\bar{\mathbf{y}} = \mathbf{y}$, where $\bar{\mathbf{y}}$ is the sample mean. The KF analysis equation for the covariance is fulfilled in an statistical sense.

In the case of nonlinear observation operator, (**??**) would be written as

$$\mathbf{x}_m^a = \mathbf{x}_m^b + \mathcal{K}(\mathbf{y}_m - h(\mathbf{x}_m^b)) \tag{19}$$

In the EKF analysis equation, the computation of $\mathcal{K}$ involves calculating $\mathcal{H}$. In the analysis step of the EnKF one can avoid computing this Jacobian by using the ensemble (Hunt et al 2007). First, one maps $\mathbf{X}^b \in \mathbb{R}^{N_x \times M}$ into observational space using the nonlinear observation operator to get a new ensemble $\mathbf{Y}^b \in \mathbb{R}^{N_y \times M}$. Explicitly:

$$\mathbf{Y}^b = [\mathbf{y}_1^b = h(\mathbf{x}_1^b); \mathbf{y}_2^b = h(\mathbf{x}_2^b); \cdots; \mathbf{y}_M^b = h(\mathbf{x}_M^b)] \tag{20}$$

Then a sample mean $\bar{\mathbf{y}}^{\mathbf{b}}$ can be computed, as well as an ensemble of perturbations around this mean $\mathbf{Y}'^b = \mathbf{Y}^b - \bar{\mathbf{y}}^b \mathbf{1}^T$. Finally, $\mathcal{K}$ is computed as:

$$\mathcal{K} = \mathbf{X}'^b \mathbf{Y}'^{b^T} (\mathbf{Y}'^b \mathbf{Y}'^{b^T} + (M-1)\mathbf{R})^{-1} \tag{21}$$

For the EnKF analysis step, the quality of the sample estimators does depend on the ensemble size M, and this size should be related to the number of unstable modes in the model. It is not within the objectives of this paper to consider the effect of ensemble size, since what we want to evaluate is the *exact* solution produced by the analysis step of the EnKF when computed in different spaces, and how does it compare to the actual Bayesian posterior. For this reason, we will consider effectively infinite ensemble size ($M = 10^6$ in all our experiments) such that $\bar{\mathbf{x}}^b \to \boldsymbol{\mu}^b$ and $\mathbf{P}^b \to \mathbf{B}$.

3 Anamorphosis

In section **??** we stated 3 conditions that ensured optimality in the application of the (En)KF analysis step. For the moment, let us assume that conditions (ii) and (iii) are fulfilled, and focus on non-Gaussian priors. Two cases –for $x \in \mathbb{R}^1$– that result challenging for the application of the EnKF analysis step are illustrated in figure **??**. In both cases the likelihood has been kept Gaussian and centered at the (directly observed) state variable.

In the left panel, the prior (blue line) is bimodal, a mixture of two Gaussians centered in $x = -2$ and $x = 2$ with equal variance $\sigma^2 = \frac{1}{4}$. The prior mean is $x = 0$, corresponding to a region where $p_x(x)$ is close to zero. By assimilating an observation (red line) at $y = x = \frac{1}{3}$, the EnKF incorrectly constructs a unimodal analysis pdf (green line) which does not resemble at all the Bayesian posterior (magenta line). In fact, the analysis pdf is centered in a region where the posterior pdf is close to zero.

In the right panel, the prior is an exponential distribution with $\lambda = 1$. This is a positive-definite variable, and the Bayesian posterior (corresponding to an observation at $y = x = \frac{2}{3}$) correctly captures this information, since $p_{x|y}(x|y) = 0 \; \forall x < 0$. The analysis pdf given by the KF, however, yields non-zero probabilities for negative values of x. In reality, physical observations of a non-negative variable will not be negative. An additive error with Gaussian distribution cannot be used in practice: either a truncated nonsymmetric distribution is likely to be used, or negative values will be mapped to zero. Moreover, in these cases the nature of the observational error tends not be additive, but multiplicative for instance. Nonetheless, for the purpose of illustration, we allow the existence of negative observations.

To avoid the mentioned problems, one can transform the state variable before applying the EnKF. The ultimate goal of this procedure is to map $\mathbf{x}$, a variable with an arbitrary multivariate pdf $p_{\mathbf{x}}(\mathbf{x})$, into a new state variable $\tilde{\mathbf{x}}$ with multivariate Gaussian pdf $p_{\tilde{\mathbf{x}}}(\tilde{\mathbf{x}})$. Then, the KF analysis equations can be applied to the transformed variables, and these updated values can be mapped back into the original space. This mapping process is known as Gaussian anamorphosis.

In the univariate case ($x \in \mathbb{R}^1$), applying GA is conceptually not complicated (aside from the implementation aspects). One could use analytic functions such as logarithm or Box-Cox transformations, but these are not guaranteed to improve the distribution in general (Simon and Bertino, 2009). A better solution is to make use of the integral probability transform theorem (IPT) and solve for the new variable as (for details see appendix A):

$$\tilde{x} = g(x) = P_{\tilde{x}}^{-1}(P_x(x)) \tag{22}$$

The moments of the target Gaussian variable $\tilde{x}$ are set to be those of the original ensemble (see section 4.4 Bertino et al, 2003). Of course, in the implementation one can transform x into a standard Gaussian rv $N(0, 1)$, and then translate and scale the values correspondingly to get $\tilde{x}$.

The actual prior $p_x(x)$, and consequently the cdf $P_x(x)$, are rarely known perfectly. Hence, to apply the IPT, the first step is to empirically estimate $P_x(x)$. This can be done using the ensemble. Then, a set of percentiles of this empirical cdf are mapped to the same percentiles of the cdf of a target normal distribution. A piecewise linear transformation can be used to get the intermediate values, and special care has to be taken when dealing with the tails (Simon and Bertino, 2009; Beal et al, 2010). The quality in the estimation of $P_x(x)$ clearly depends on the size of the ensemble. In order to increase the sample size, one can make use of values of the variable at different times and consider a stationary climatological pdf. Refinements to this idea include time-evolving anamorphosis functions. Simon and Bertino (2009), for example, construct the GA function for the state variables from a window of three months centered on the datum in a 3-D ecosystem model.

The multivariate case is considerably more difficult. Strictly speaking, it requires a joint multivariate transformation. A multivariate version of the IPT exists (Genest and Rivest, 2001),

but its application is not simple. Besides, checking for the joint Gaussianity of a multivariate spatial law is quite computationally demanding (Bertino et al 2003). For this reason, implementation of GA in large models is often done univariately, i.e. a different function is applied for each one of the components in the state variable vector:

$$\tilde{\mathbf{x}} = g(\mathbf{x}); \qquad \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \vdots \\ \tilde{x}_N \end{bmatrix} = \begin{bmatrix} g_1(x_1) \\ g_2(x_2) \\ \vdots \\ g_N(x_N) \end{bmatrix} \tag{23}$$

For field variables, one can either consider them to have homogeneous distributions, or one can apply local anamorphosis functions at different gridpoints (Doron et al 2011; Zhou et al, 2011). Another option for the multivariate case is to rotate the space to get uncorrelated variables by performing principal components analysis (PCA). It is not straightforward, however, that the updated variables will follow the same PCA, since the transformations are nonlinear (see the discussion in Bocquet et al, 2010). Moreover, residual correlations may remain (Pires and Perdigao, 2007). A more complicated approach involving copulas has been suggested by Scholzel and Freidrichs (2008).

Up to this moment we have only considered transformations of the prior, but the observations can be transformed as part of a more general GA process, i.e.:

$$\begin{aligned} \tilde{\mathbf{x}} &= g_{model}(\mathbf{x}) \\ \tilde{\mathbf{y}} &= g_{obs}(\mathbf{y}) \end{aligned} \tag{24}$$

In the transformed space, $\tilde{\mathbf{y}}$ and $\tilde{\mathbf{x}}$ are related by the observation operator

$$\tilde{h} = g_{obs} \circ h \circ g_{model}^{-1} \tag{25}$$

where $\circ$ denotes function composition. In this space, for each one of the transformed ensemble members the EnKF analysis value can be obtained as (Bertino et al, 2003):

$$\tilde{\mathbf{x}}_m^a = \tilde{\mathbf{x}}_m^b + \tilde{\mathcal{K}}(\tilde{\mathbf{y}}_m - \tilde{h}(\tilde{\mathbf{x}}_m^b)) \tag{26}$$

To compute $\tilde{\mathbf{y}}_m$, Simon and Bertino (2012) propose to perturb the observations in the *original* space by sampling from $p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x})$, and then map each of the perturbed observations individually $\tilde{\mathbf{y}}_m = g_{obs}(\mathbf{y}_m)$. In this work we use said approach. The perturbed variables have associated covariance matrices $\tilde{\mathbf{B}}$ and $\tilde{\mathbf{R}}$, which can be computed directly from the the ensembles $\tilde{\mathbf{X}}^b$ and $\tilde{\mathbf{Y}}$. These covariance matrices are used for the computation of $\tilde{\mathcal{K}}$. If $\tilde{h}$ is nonlinear, then one uses the same procedure described at the end of section **??** for the computation of $\tilde{\mathcal{K}}$.

A crucial issue in GA is the choice of the transformations $g_{model}(\cdot)$ and $g_{obs}(\cdot)$, and the effect these choices will have in the observation operator in transformed space. In the next section we study different choices for these maps.

4 Choosing anamorphosis functions

We now discuss different ways to transform $\{\mathbf{x}, \mathbf{y}\}$ into new variables $\{\tilde{\mathbf{x}}, \tilde{\mathbf{y}}\}$, paying notice to the effects these transformations cause in the *joint* characteristics of state and observations. Is there a transformation that produces a Gaussian posterior $p_{\tilde{x}|\tilde{y}}(\tilde{\mathbf{x}}|\tilde{\mathbf{y}})$ in the transformed space? The search for this ideal case leads this section.

For the moment we focus on the univariate case ($x, y, \tilde{x}, \tilde{y} \in \mathbb{R}^1$). We start with a generalisation of (**??**) and consider joint bivariate forward transformations of the form:

$$\begin{aligned} \tilde{x} &= g_1(x, y) \\ \tilde{y} &= g_2(x, y) \end{aligned} \tag{27}$$

with the respective backward transformations:

$$\begin{aligned} x &= q_1(\tilde{x}, \tilde{y}) \\ y &= q_2(\tilde{x}, \tilde{y}) \end{aligned} \tag{28}$$

Then, if the joint pdf of $\{x, y\}$ in the original space is $p_{x,y}(x, y) = p_{y|x}(y|x)p_x(x)$, the joint pdf in the transformed space is (see appendix A for details):

$$p_{\tilde{x},\tilde{y}}(\tilde{x}, \tilde{y}) = p_{y|x}(q_2(\tilde{x}, \tilde{y})|q_1(\tilde{x}, \tilde{y}))p_x(q_1(\tilde{x}, \tilde{y})) \left| \frac{\partial q_1}{\partial \tilde{x}} \frac{\partial q_2}{\partial \tilde{y}} - \frac{\partial q_1}{\partial \tilde{y}} \frac{\partial q_2}{\partial \tilde{x}} \right| \tag{29}$$

We will now study different choices for (**??**). Throughout the rest of this section we will use the following example to visualise the effects of these choices in the joint state-variable/observation space. The prior pdf, likelihood, and observation equation are (refer to equation (**??**) for notation on Gaussian rv's):

$$\begin{aligned} p_x(x) &= \frac{1}{2}\phi\left(\frac{x+2}{1/2}\right) + \frac{1}{2}\phi\left(\frac{x-2}{1/2}\right) \\ p_\eta(\eta) &= \frac{4}{5}\phi\left(\eta + \frac{1}{4}\right) + \frac{1}{5}\phi\left(\frac{\eta-1}{1/2}\right) \\ y &= h(x, \eta) = x + \eta \end{aligned} \tag{30}$$

Both pdfs are Gaussian mixtures (GMs) with expected value equal to 0; $p_x(x)$ is symmetric while $p_\eta(\eta)$ is not. One can think of this choice for $p_x(x)$ to be plausible, but this type of distribution is rarely used for observational error. It could be seen as the result of the interaction of a simpler likelihood with a nonlinear observation operator. In any case, using GMs is convenient since they allow tractability of the analytical Bayesian posteriors (see appendix B for details), something very useful for illustration and evaluation purposes. Also, GMs can be used to approximate any smooth pdf.

The application of different anamorphosis functions for this example is illustrated in figure **??**. This figure has 5 panels, one for each transformation. In every panel we show the *joint* bivariate distribution of the state variables and observations (contour plot), the *marginal* distribution of the state variable (horizontal plot) and the *marginal* distribution of the observations (vertical plot). Also, we consider individual *given* observations (shown as color lines on top of the bivariate plot), and the effects of the transformations in these observations.

4.1 Independent transformations

The simplest case is to make the transformations for state variables and observations independent. This means $\tilde{x} = g_1(x, y) = g_1(x)$ and $\tilde{y} = g_2(x, y) = g_2(y)$. Then, (**??**) simplifies to:

$$p_{\tilde{x},\tilde{y}}(\tilde{x}, \tilde{y}) = p_{y|x}(g_2^{-1}(\tilde{y})|g_1^{-1}(\tilde{x}))p_x(g_1^{-1}(\tilde{x})) \left| \frac{\partial g_1^{-1}}{\partial \tilde{x}} \frac{\partial g_2^{-1}}{\partial \tilde{y}} \right| \tag{31}$$

Next, we list some choices for independent univariate transformations.
(a) Working in the original space.

In this trivial case, both transformations are the identity:

$$\begin{array}{ll} \tilde{x} = x; & g_1(\cdot) = 1 \\ \tilde{y} = y; & g_2(\cdot) = 1 \end{array} \tag{32}$$

This amounts to just applying the EnKF in the original space, but it serves as benchmark for comparison. As we see in panel (a) of figure **??**, for our example both $p_x(x)$ and $p_y(y)$ are bimodal, with $p_y(y)$ showing asymmetry, a consequence of the asymmetric likelihood. In the joint state-variables/observations space, this translates into two sell-separated areas of high probability. This is indeed a scenario that ensures non-optimality for the use of the EnKF.
(b) Transforming only x.

In this case only the state variable is transformed into a Gaussian rv. Hence the transformations are:

$$\begin{array}{ll} \tilde{x} = \Phi_{\tilde{x}}^{-1}(P_x(x)) & g_1(\cdot) = g_{x\to\tilde{x}}(\cdot) = \Phi_{\tilde{x}}^{-1}(P_x(\cdot)) \\ \tilde{y} = y & g_2(\cdot) = 1 \end{array} \tag{33}$$

where $\Phi_{\tilde{x}}(\cdot)$ explicitly indicates that the cdf in transformed space is that of a Gaussian rv (see notation defined in section **??**). As we can see in panel (b) of figure **??**, this transformation achieves a *marginal* Gaussian $p_{\tilde{x}}(\tilde{x})$, but does nothing either on $p_{\tilde{y}}(\tilde{y})$ or in the individual observation values. The joint pdf $p_{\tilde{x},\tilde{y}}(\tilde{x},\tilde{y})$ does not show the isolated peaks as before, but instead it has elongated features, a consequence of populating regions of the space variable around 0 which were previously unpopulated.
(c) Transforming both x and y with the same function.

This transformation is only possible when the domains of x and y are the same, as in our example with $h = 1$. This option cannot always be applied; it would be incorrect e.g. if $x \in \mathbb{R}$ and $y \in \mathbb{R}^+$. For the sake of completeness we include it in our discussion. In this case the maps would be:

$$\begin{array}{ll} \tilde{x} = \Phi_{\tilde{x}}^{-1}(P_x(x)) & g_1(\cdot) = g_{x\to\tilde{x}}(\cdot) = \Phi_{\tilde{x}}^{-1}(P_x(\cdot)) \\ \tilde{y} = \Phi_{\tilde{x}}^{-1}(P_x(y)) & g_2(\cdot) = g_{x\to\tilde{x}}(\cdot) = \Phi_{\tilde{x}}^{-1}(P_x(\cdot)) \end{array} \tag{34}$$

The application of this transformation in y does not guarantee anything characteristics for $p_{\tilde{y}}(\tilde{y})$. Panel (c) of figure **??** illustrates the effect of this transformation. While the state variable is indeed transformed into a Gaussian, we obtain a non-Gaussian and very peaked distribution for $p_{\tilde{y}}(\tilde{y})$, which translates in a very narrow bivariate pdf with respect to $\tilde{y}$. The individual observations are transformed, as depicted by the color lines.
(d) Transforming x and y marginally.

With the previous methods one achieved marginal Gaussianity in $\tilde{x}$, but not on $\tilde{y}$. One can apply the IPT to y and obtain marginal Gaussianity in $\tilde{y}$. The maps would then be:

$$\begin{array}{ll} \tilde{x} = \Phi_{\tilde{x}}^{-1}(P_x(x)) & g_1(\cdot) = g_{x\to\tilde{x}}(\cdot) = \Phi_{\tilde{x}}^{-1}(P_x(\cdot)) \\ \tilde{y} = \Phi_{\tilde{y}}^{-1}(P_y(y)) & g_2(\cdot) = g_{y\to\tilde{y}}(\cdot) = \Phi_{\tilde{y}}^{-1}(P_y(\cdot)) \end{array} \tag{35}$$

This transformation involves knowing the marginal distribution of the observations, or at least constructing an estimation. This is the approach used in Simon and Bertino (2009, 2012). In these works, the authors estimate a marginal climatological pdf for observations using values from an extended time period. Panel (d) of figure **??** shows the effects of this transformation. As we can see, both marginal pdf's are Gaussian. The individual observations are transformed (as in panel (c)). The joint pdf, however, looks very different from a bivariate Gaussian; recall that whereas bivariate Gaussianity implies marginal Gaussians, the opposite is not true (e.g. Casella and Berger, 2002).

4.2 Joint state-variable/observation transformations

In section **??**, the objectives of the proposed transformations became progressively become more ambitious. The last case achieves marginal Gaussianity in both $\tilde{x}$ and $\tilde{y}$. Still, with independent transformations we are not able to guarantee any particular characteristics for the relationship between state variables and observations in the transformed space. We now introduce a joint state-variable/observation transformation which has precisely this objective: to transform the pair $\{x, y\}$ (with arbitrary joint pdf) into the pair $\{\tilde{x}, \tilde{y}\}$ (with joint Gaussian pdf). Consequently as a by-product, the marginal and conditional pdfs in this space will also be Gaussian. Our algorithm can be divided in 3 steps. These are listed next and also depicted in figure **??**.

(i) The first step corresponds to the upper row of figure **??**. In this step we pre-design a transformed space (right panel) which is joint Gaussian and that shares statistical characteristics with the original space (left panel). In the transformed space we set the prior as $p_{\tilde{x}}(\tilde{x}) = \phi(\tilde{x}; \tilde{\mu}^b, \tilde{\sigma}^b)$, and the likelihood as $p_{\tilde{y}|\tilde{x}}(\tilde{y}|\tilde{x}) = \phi(\tilde{y}; \tilde{H}\tilde{x}, \tilde{\sigma}^o)$. The moments of $p_{\tilde{x}}$ are estimated by the sample moments of the ensemble in the original space, i.e. $\{\tilde{\mu}^b = \mu^b, \tilde{\sigma}^b = \sigma^b\}$, and the observational error is prescribed (or deduced) from the original likelihood, i.e. $\tilde{\sigma}^o = \sigma^o$. $\tilde{H}$ is a linear observation operator (in our example we choose the identity).

(ii) The second step corresponds to the middle row of figure **??**. We map both x and $\tilde{x}$ into $w \sim U[0, 1]$, i.e. a r.v. with uniform distribution in the interval $[0, 1]$. This is done by simply applying the IPT to both variables:

$$\begin{aligned} w &= P_x(x) \\ w &= P_{\tilde{x}}(\tilde{x}) = \Phi_{\tilde{x}}(\tilde{x}) \end{aligned} \tag{36}$$

The previous procedure is just the application of (**??**) with an extra intermediate step. Let us focus on the spaces $\{w, y\}$ and $\{w, \tilde{y}\}$. Since the marginal pdf of w is simply $p_w(w) = 1\ I_{[0,1]}(w)$ –where $I_{[0,1]}(\cdot)$ is the indicator function–, the joint pdf $p_{w,y}(w, y)$ coincides with the conditional pdf $p_{y|w}(y|w)$, i.e. $p_{w,y}(w, y) = p_{y|w}(y|w)$. The same applies to the $\{w, \tilde{y}\}$ case, i.e. $p_{w,\tilde{y}}(w, \tilde{y}) = p_{\tilde{y}|w}(\tilde{y}|w)$.

For our example, these distributions are illustrated the second row of figure **??**, and they are shown in better detail in figure **??**. In the left panel of this figure we depict $p_{w,y}(w, y)$, the center panel depicts $p_{w,\tilde{y}}(w, \tilde{y})$ and the right panel is the difference between the two (we can do this subtraction because y and $\tilde{y}$ have the same support $(-\infty, \infty)$). In the left panel, we can see the effect of having a GM as prior, as we can see two well separated regions in the joint pdf, with a division at $w = 0.5$. We can also notice the effect of the non-symmetric likelihood: the distance between the contours in the upper part of the colored strip is less than the distance between those in the lower part. These effects are not present in the center panel. In fact, we need a way to convert the left panel into the center panel; this is the purpose of the next step.

(iii) The last step is depicted in the bottom row of figure **??**. For the last step we design a transformation from y to $\tilde{y}$ such that the given $p_{w,y}(w, y)$ becomes the prescribed $p_{w,\tilde{y}}(w, \tilde{y})$. This is equivalent –as we have explained before– to transforming $p_{y|w}(y|w)$ into $p_{\tilde{y}|w}(\tilde{y}|w)$. Hence, for each and every value of w, we can state the following equation of cumulative likelihoods:

$$P_{\tilde{y}|w}(\tilde{y}|w) = P_{y|w}(y|w) \tag{37}$$

Although it is not always possible to obtain explicitly, the solution of this equation is of the form $\tilde{y} = \tilde{y}(w, y)$. Solving this equation for each and every value of w completes the construction of the map from $\{x, y\}$ into $\{\tilde{x}, \tilde{y}\}$. To summarize, the transformation we just devised is formed by the forward and backward maps:

$$\begin{aligned} \tilde{x} &= g_{x\to\tilde{x}}(x) \qquad & x &= g^{-1}_{x\to\tilde{x}}(\tilde{x}) \\ \tilde{y} &= g_{biv}(x, y) \qquad & y &= q_{biv}(\tilde{x}, \tilde{y}) \end{aligned} \tag{38}$$

In figure **??** we show the form of these maps in our study case: forward transformations in the top row and backward transformations in the bottom row. The top left panel shows the simple IPT-based transformation from x to $\tilde{x}$. For the region $-1 < x < 1$ the graph looks almost horizontal, but it is not. This consequence comes from the fact that in this region $p_x(x)$ is close to zero, while this same region contains the largest probability mass for $p_{\tilde{x}}(\tilde{x})$, so the slope of the map in this region is extremely small. The top right figure shows the joint transformation $\tilde{y} = \tilde{y}(x, y)$, which is the solution of (**??**) in terms of y and w, but with the values of w replaced by the corresponding x for the plot. The bottom left panel shows the transformation from $\tilde{x}$ to x. Again, it is a simple IPT-based implementation. This time we observe an almost vertical behaviour of the graph near $\tilde{x} = 0$, for the reasons stated before. On the other hand $y = y(u, v)$ is the solution of (**??**) in terms of y and w, with the values of w replaced by the corresponding values of $\tilde{x}$ for the plot. The plot would suggest a discontinuity around $\tilde{x} = 0$, but this is not the case, there is only a sharp change not captured at the resolution of the graph. This behaviour is associated to the characteristics around $x = 0$ previously described.

The method we just described can be considered a special instance of the multivariate Rosenblatt transform (1952). Furthermore, the statistical characteristics of the joint state-variable/observation space $\{\tilde{x}, \tilde{y}\}$ constructed with this method fulfil conditions (i)-(iii). One could -at first sight- consider this to be an optimal transformation. Things are not that simple, however, and the complication comes from the mapping of the *given* individual observations. The issue is that a fixed value y_0 in $\{x, y\}$ is not fixed anymore in $\{\tilde{x}, \tilde{y}\}$, it becomes a function $\tilde{y}_0 = \tilde{y}_0(x, y_0)$ (actually a function of x since y_0 is a fixed value). This can easily be seen in panel (e) of figure **??**. By construction the obtained joint distribution is bivariate Gaussian (and consequently the marginals are Gaussian as well), but *fixed* observations are no longer horizontal lines, instead their values depend on $\tilde{x}$. This leads to a conceptual complication: in the $\{\tilde{x}, \tilde{y}\}$ space we are not finding a posterior in the proper sense (or estimating its first two moments, since we are using the EnKF). In this space, the posterior $p_{\tilde{x}|\tilde{y}}(\tilde{x}|\tilde{y})$ is the pdf along a horizontal line of fixed $\tilde{y}$. But we do not have fixed $\tilde{y}$'s, instead we have functions. Does this mean we are actually estimating a probability of the form $p_{\tilde{x}|\tilde{y}(\tilde{x})}(\tilde{x}|\tilde{y}(\tilde{x}))$ instead of $p_{\tilde{x}|\tilde{y}}(\tilde{x}|\tilde{y})$? This may not be as big as a problem if we update individually each ensemble member (as we do), it would be more problematic if we were using a deterministic square root filter, updating mean and covariance, and constructing the ensemble members after that. Fortunately, by using perturbed observations we sample directly from the likelihood. This avoid bias, as indicated in Simon and Bertino (2009). Finally, it is important to mention that if we wanted the Bayesian solution in the transformed space we would need to compute the corresponding normalisation factor, in this case $p(\tilde{y}(\tilde{x}))$, which cannot be considered a constant with respect to $\tilde{x}$. Fortunately we do not require this factor after mapping the sample back to the original space.

The EnKF analysis equation in the transformed space is simply (**??**), and it is linear. For $\tilde{\sigma}^{b2}$ we use the sample covariance in transformed space. The observational error variance $\tilde{\sigma}^{o2}$ is prescribed based on the characteristics of the likelihood in the original space. One could compute the empirical observational covariance in the transformed space, but one could not be averaging over straight lines, but functions of $\tilde{x}$ (see the previous discussion). Hence, this could lead to an overestimation or underestimation of the actual observational covariance.

4.3 Transformations in the multivariate case

Performing GA in the multivariate case ($\mathbf{x}, \tilde{\mathbf{x}} \in \mathbb{R}^{N_x}$ and $\mathbf{y}, \tilde{\mathbf{y}} \in \mathbb{R}^{N_y}$) is considerably more difficult. As mentioned in section **??**, the simplest way is to do independent transformations for each one of the state variables (see equation (**??**)) and observations. If there are variables that are neither transformed nor observed, they are still affected by the transformations via the corresponding covariances. For illustration, consider a two variable system in which the first

variable is indirectly observed, i.e. $\mathbf{x} = [x_1\ x_2]^T$ and $\mathbf{y} = [y_1] = [h(x_1, \eta_1)]$. Even if the unobserved x_2 is not transformed, the update from background to analysis of this variable is different in the original space than in one in which a GA of the form $x_1 \rightarrow \tilde{x}_1,\ y_1 \rightarrow \tilde{y}_1$ is performed. We can see this if we develop explicitly (**??**) for this variable in the two cases:

$$\begin{aligned} x_2^a - x_2^b &= \frac{Cov(x_2^b, y_1^b)}{Var(y_1^b) + Var(y_1)}(y_1 - y_1^b) \\ x_2^a - x_2^b &= \frac{Cov(x_2^b, \tilde{y}_1^b)}{Var(\tilde{y}_1^b) + Var(\tilde{y}_1)}(\tilde{y}_1 - \tilde{y}_1^b) \end{aligned} \tag{39}$$

where $y_1^b = h(x_1^b)$ and $\tilde{y}_1^b = \tilde{h}(\tilde{x}_1^b)$. The anamorphosis functions should guarantee that $Var(y_1^b) \approx Var(\tilde{y}_1^b)$ and that $Var(y_1) \approx Var(\tilde{y}_1)$. Hence, the crucial part is the way in which the anamorphoisis function changes the covariance between the observed and the unobserved variable. This is, the change from $Cov(x_2^b, y_1^b)$ to $Cov(x_2^b, \tilde{y}_1^b)$.

Now, let us think again about targeted joint transformations. Following our rationale in section **??**, the ultimate goal in this case would be to go from the space $\{\mathbf{x}, \mathbf{y}\}$ with a general nonlinear operator h to a space $\{\tilde{\mathbf{x}}, \tilde{\mathbf{y}}\}$ with a joint multivariate distribution and a linear observation operator $\tilde{\mathbf{H}}$. This may not be possible in general, depending on the precise behaviour of h.

For the moment, we can propose a modest solution. Let us consider that there is a set of L variables that are observed as: $[y_1 = h_1(x_1); \cdots ; y_L = h_L(x_L)]$. Then, we can perform the proposed joint transformations for each pair $\{x_l, y_l\}$. The effect of these transformation into other variables will still be communicated through covariance, just as in (**??**). To explore the full problem, one could start with a simple system such as the one described in this subsection (2 variables, one observed, one not). Can we replace a joint trivariate transformation by a sequence of two joint bivariate ones? This is one of the ideas we are exploring at the moment.

5 Experiments

In this section we study the analysis pdf's that result of performing the EnKF analysis step in combination with the transformations described earlier. For the sake of brevity, we will take the following short notation when describing the 5 spaces in which the EnKF analysis step is applied. Its application in $\{x, y\}$ is denoted as $K\{x, y\}$, in $\{\tilde{x} = g_{x\to\tilde{x}}(x), y\}$ as $K\{\tilde{x}, y\}$, in $\{\tilde{x} = g_{x\to\tilde{x}}(x), \tilde{y} = g_{x\to\tilde{x}}(y)\}$ as $K\{\tilde{x}, \tilde{y}_*\}$, in $\{\tilde{x} = g_{x\to\tilde{x}}(x), \tilde{y} = g_{y\to\tilde{y}}(y)\}$ as $K\{\tilde{x}, \tilde{y}\}$, and finally in $\{\tilde{x} = g_{x\to u}(x), \tilde{y} = g_{biv}(x, y)\}$ as $K\{\tilde{x}, \tilde{y}_{biv}\}$.

Figure **??** shows the results of assimilating an observation at $y_0 = -\frac{1}{3}$ in the system (**??**). The true Bayesian posterior (black line) is bimodal, with a considerably taller peak in the negative values. The analysis pdf produced by $K\{x, y\}$ (blue line) does not resemble this at all, instead it generates a pdf centred close to zero with and a hint of bimodality. Note that a Gaussian analysis pdf is not produced (as it was the case in the left panel of figure **??**) because in the current experiment the likelihood is not Gaussian and the perturbed observations were produced using the correct likelihood.

For the other 5 cases the resulting empirical posteriors are indeed bimodal. $K\{\tilde{x}, \tilde{y}_*\}$ (magenta line) produces almost symmetric peaks. $K\{\tilde{x}, \tilde{y}\}$ (green line) gives more probability to the wrong mode. $K\{\tilde{x}, y\}$ (red line) and the bivariate transformation $K\{\tilde{x}, \tilde{y}_{biv}\}$ (cyan line) gives higher probability to the left peak, resembling the actual Bayesian posterior.

5.1 An objective assessment of the quality of the EnKF-generated analysis

The previous discussion was rather qualitative. We now use the Kullback-Leibler divergence (D_{KL}, see e.g. Cover and Thomas, 2001) to quantitatively compare the EnKF-generated analysis pdfs with respect to the Bayesian posteriors. For two continuous pdfs $p(x)$ and $q(x)$, this quantity is defined as:

$$D_{KL}(p, q) = \int_{-\infty}^{\infty} \ln\left(\frac{p(x)}{q(x)}\right) p(x)dx \tag{40}$$

The definition for D_{KL} can be interpreted as the expected value of the logarithmic difference between the probabilities $p(\cdot)$ and $q(\cdot)$, evaluated over $p(\cdot)$. D_{KL} quantifies the information gain from q to p (Bocquet et al, 2010). Note that $0 \leq D_{KL}(p, q) < \infty \ \forall \ \{p, q\}$, and $D_{KL}(p, q) = 0$ if and only if the two densities are equal almost everywhere. Roughly speaking, the larger the value of this quantity the more different the two distributions are. In our case, $p(x)$ is the exact Bayesian posterior, whereas $q(x)$ is the EnKF-generated analysis pdf. We compute the D_{KL} numerically after dividing the data in J_{bins} bins, and using the following expression:

$$D_{KL}(p, q) = \sum_{j=1}^{J_{bins}} \ln\left(\frac{p_j}{q_j}\right) p_j \tag{41}$$

We will consider an experimental setting which stems from a generalisation of the system (**??**). The prior and (additive) observational error pdf's are:

$$p_x(x) = \alpha_{x1}\ \phi\left(\frac{x - \mu_{x1}}{\sigma_{x1}}\right) + \alpha_{x2}\ \phi\left(\frac{x - \mu_{x2}}{\sigma_{x2}}\right)$$
$$p_\eta(\eta) = \alpha_{\eta1}\ \phi\left(\frac{\eta - \mu_{\eta1}}{\sigma_{\eta1}}\right) + \alpha_{\eta2}\ \phi\left(\frac{\eta - \mu_{\eta2}}{\sigma_{\eta2}}\right)$$

and the observation operator is the identity. We will choose three combinations of parameters:

(a) The prior is a GM and the likelihood is Gaussian (GM-G).

$$\begin{array}{lll} \alpha_x = \{1/2, 1/2\} & \mu_x = \{-2, 2\} & \sigma_x = \{1/2, 1/2\} \\ \alpha_\eta = \{1, 0\} & \mu_\eta = \{0, \cdot\} & \sigma_\eta = \{1, \cdot\} \end{array} \tag{42}$$

(b) Both the prior and the likelihood are GMs (GM-GM).

$$\begin{array}{lll} \alpha_x = \{1/2, 1/2\} & \mu_x = \{-2, 2\} & \sigma_x = \{1/2, 1/2\} \\ \alpha_\eta = \{4/5, 1/5\} & \mu_\eta = \{-1/4, 1\} & \sigma_\eta = \{1, 1/2\} \end{array} \tag{43}$$

(c) The prior is Gaussian and the likelihood is a GM (G-GM).

$$\begin{array}{lll} \alpha_x = \{1, 0\} & \mu_x = \{0, \cdot\} & \sigma_x = \{1, \cdot\} \\ \alpha_\eta = \{4/5, 1/5\} & \mu_\eta = \{0, 1\} & \sigma_\eta = \{2/3, 1/4\} \end{array} \tag{44}$$

Furthermore, recall that we are assessing the quality of the empirical distributions that approximate $p_{x|y}(x|y)$, which depend on a *given* observation. The Bayesian posterior and the analysis pdfs generated by the EnKF analysis step –which were plotted in figure **??**– were based on a *single* observation. For the current experiment, however, we reconstruct the distributions for a range of 21 different observation values. For each one of the 3 scenarios (a-c), for each one of the 5 transformations, and each one of the 21 given observations we compute D_{KL}. We plot this information in the left panels of figure **??**, a different colour for each different transformation. In this figure, the top row corresponds to the GM-G case, the centre row to the GM-GM case and the bottom row corresponds to the G-GM case. Of the 21 observation values we select 3, which we identify with the violet, black and orange vertical dotted lines in these panels. In the right panels we take those observations and plot the Bayesian posteriors associated to them, each Bayesian posterior is identified with the corresponding colour. The quality of the EnKF-generated analysis distributions in the 5 different spaces will depend on the shape of the real Bayesian posterior they are trying to emulate. This is discussed in detail for each scenario.

(a) Let us start with the GM-G case in the top row. Because of the settings (a bimodal prior symmetric with respect to $x = 0$ and a Gaussian likelihood), we expect $D_{KL}(y_0 = \xi)$ equal to $D_{KL}(y_0 = -\xi)$. This is indeed what we get for the 5 methods. For all the values of the observation, it is clear that $K\{x, y\}$ is the worst method. In particular, its highest D_{KL} value is for $y_0 = 0$ (vertical black line), since this y_0 gives rise to a bimodal posterior (black curve in right panel), a definite challenge for the EnKF applied in the original space. For the other two observational values ($y_0 = -1.8$ violet line, and $y_0 = 1.8$ orange line) the posteriors are close to Gaussians, and hence the D_{KL} values are lower. The next largest D_{KL} corresponds to $K\{\tilde{x}, y_*\}$. Again, the worst performance is for $y_0 = 0$, but there is a consistent gap for all observational values between this and the other methods. The D_{KL} values for the other 3 methods are very close. The performance of both $K\{\tilde{x}, y\}$ and $K\{\tilde{x}, \tilde{y}_{biv}\}$ is almost indistinguishable for all values of observations. In the interval $-0.75 \leq y_0 \leq 0.75$ $K\{\tilde{x}, \tilde{y}\}$ is outperformed by $K\{\tilde{x}, y\}$ and $K\{\tilde{x}, \tilde{y}_{biv}\}$, but outside this interval it is the best method overall.

(b) In the centre row we have the GM-GM case. Not surprisingly, $K\{x, y\}$ presents the worst performance, followed by $K\{\tilde{x}, \tilde{y}_*\}$, and both perform worst for observations that produce bimodal posteriors. This is again the case of $y_0 = 0$ (black vertical line) which produces a nonsymmetric bimodal posterior (right panel). The performance of $K\{\tilde{x}, y\}$ and $K\{\tilde{x}, \tilde{y}_{biv}\}$ is again very close to each other. It is interesting that for $-2.4 < y_0 < 0.6$ both $K\{\tilde{x}, y\}$ and $K\{\tilde{x}, \tilde{y}_{biv}\}$ have the best performance, whereas for $y_0 < -2.4$ and $y_0 > 0.6$ the best method is $K\{\tilde{x}, \tilde{y}\}$. We could identify that this particular method tends to have trouble with bimodal asymmetric posteriors.

(c) Finally we have the G-GM case in the bottom row. For this scenario, 3 transformations are exactly the same: $K\{x, y\}$, $K\{\tilde{x}, y\}$ & $K\{\tilde{x}, \tilde{y}_*\}$. The reason for this is that the

transformation $g_{x\to\tilde{x}}$ is the identity (the prior is already a Gaussian). This is why in the left bottom panel of the figure one line has the three corresponding colours (blue, red and magenta); we will refer to this line simply as $K\{x, y\}$. One can immediately notice that $K\{x, y\}$ and $K\{\tilde{x}, \tilde{y}\}$ have a very similar performance for most observations. The explanation is that with the settings of this experiment the marginal $p_y(y)$ is very close to being Gaussian, so that the transformation $g_{y\to\tilde{y}}$ is very close to the identity again. Still this method outperforms $K\{x, y\}$ for $y_0 < -0.4$. Note that the best performance for both methods is for large negative observations; as we can see for $y_0 = -1.8$ (vertical violet line) the Bayesian posterior is close to a Gaussian (right panel). This is not the case, however, for the posteriors produced by $y_0 = 0$ (black vertical line) and $y_0 = 1.8$ (orange vertical line). As it can be seen in the right panel of this row, both Bayesian posteriors are bimodal and asymmetric (black and orange lines). In this example we can truly appreciate the value of both bivariate transformations; for most part of the observation values they outperform the other methods, and the difference is especially significant for the challenging cases mentioned above.

6 Summary and discussion

The analysis step of the EnKF is optimal when the following 3 conditions are met: (i) the distribution of the prior is Gaussian, (ii) the observation operator that relates state variables and observations is linear, and (iii) the observational error is additive and follows a Gaussian distribution. The analysis step of the EnKF is often applied in spite of the violation of these conditions and still yields useful results. There are cases, however, when the departure from said conditions is too considerable. In these cases, a technique known as Gaussian anamorphosis is applied to convert these distributions into Gaussians before performing the analysis step.

The ultimate goal of GA would be to convert the set $\{\mathbf{x}, \mathbf{y}\}$ of arbitrary joint distribution into the set $\{\tilde{\mathbf{x}}, \tilde{\mathbf{y}}\}$ with a joint Gaussian distribution. This is not an easy objective at all, and a proper multivariate GA transformation is not straightforward to devise. For this reason, GA is often applied in a univariate manner. Thus, we have mostly restricted ourselves to the univariate case $x, y, \tilde{x}, \tilde{y} \in \mathbb{R}^1$. For this case, we have analysed GA transformations starting from the following classification: independent, i.e. transformations of the form $\tilde{x} = g_1(x), \tilde{y} = g_2(y)$, and joint state-variable/observation, i.e. transformations of the form $\tilde{x} = g_1(x, y), \tilde{y} = g_2(x, y)$.

For independent transformations (section **??**) we have studied some options: (a) an identity transformation –i.e. working in the original space– (denoted $K\{x, y\}$), (b) transforming only the state variable (denoted $K\{\tilde{x}, y\}$), (c) transforming both state variables and observations using the same function –applicable only when h is the identity– (denoted $K\{\tilde{x}, \tilde{y}_*\}$) and (d) transforming state variables and observations to obtain marginal Gaussianity for both (denoted $K\{\tilde{x}, \tilde{y}\}$).

One of the contributions of this work is the introduction of a targeted joint state-variables/observation transformation (section **??**) of the form $\tilde{x} = g_{x\to\tilde{x}}(x), \tilde{y} = g_{biv}(x, y)$, which is briefly outlined next. Having original distributions $p_x(x)$ and $p_{y|x}(y|x)$, we devise target Gaussian distributions $p_{\tilde{x}}(\tilde{x})$ and $p_{\tilde{y}|\tilde{x}}(\tilde{y}|\tilde{x})$ with prescribed parameters. Both $p_x(x)$ and $p_{\tilde{x}}(\tilde{x})$ are mapped into an auxiliary variable w $U[0, 1]$. Finally, an equality of cumulative likelihoods $P_{\tilde{y}|w}(\tilde{y}|w) = P_{y|w}(y|w)$ is solved for all w and this completes the transformation $\{x, y\} \to \{\tilde{x}, \tilde{y}\}$.

To test these transformations, we have selected a case in which the Bayesian posterior can be obtained analytically, in particular a directly observed GM prior – GM likelihood model with three settings: GM prior with Gaussian likelihood, GM prior with GM likelihood, and Gaussian prior with GM likelihood. We have compared the posterior pdf to the pdf's generated after applying the EnKF analysis step in conjunction with the different transformations. This resemblance has been evaluated using the Kullback-Leibler divergence (D_{KL}) for different given observations (figure **??**). To further understand the behaviour of the D_{KL} curves, we have plotted the Bayesian posteriors for 3 selected observational values (right panel of the same figure).

The truth is that, despite the application of any of the different transformations, the analysis step of the EnKF cannot exactly reconstruct the Bayesian posterior when conditions (i)-(iii) are not met in the original space. Still, one can get approximate solutions, and it is clear that some are better than others.

In all cases, $K\{x, y\}$ has the worst performance, highlighting the fact that severe deviations from Gaussianity in both the prior and likelihood can handicap the performance of the EnKF analysis step. The next method in increasing order of performance is $K\{\tilde{x}, \tilde{y}_*\}$. It seems that, at least for the situation we studied, applying the same transformation for both state variables and observations is not an appropriate strategy. In the first two cases (GM-G, GM-GM), 3 methods have very similar performance: $K\{\tilde{x}, y\}$, $K\{\tilde{x}, \tilde{y}\}$ and $K\{\tilde{x}, \tilde{y}_{biv}\}$. What are the sources of error for each one of these two methods, i.e. in what sense is the application of the EnKF analysis step not exact? For $K\{\tilde{x}, y\}$ and $K\{\tilde{x}, \tilde{y}\}$ the answer is the appearance of a nonlinear observation operator; for $K\{\tilde{x}, \tilde{y}_{biv}\}$ is the fact that the given observations are no longer fixed values but instead functions of the state variable (recall the coloured lines in panel (e) of figure **??**). In these two cases we studied these errors seem to lead to the same performance.

The real advantage of the bivariate transformation $K\{\tilde{x}, \tilde{y}_{biv}\}$ is appreciated in the G-GM case. In this case the transformation $g_{x\to\tilde{x}}$ is just the identity (since x is already Gaussian), and also $g_{y\to\tilde{y}}$ is very close to the identity since $p_y(y)$ turns to be close to Gaussian. In this scenario $K\{\tilde{x}, \tilde{y}_{biv}\}$ clearly outperforms the other methods when trying to reconstruct non-symmetric non-Gaussian posterior densities.

This work has two main limitations. First of all, we have considered the infinite ensemble size scenario –in fact all our experiments were done with $M = 10^6$–, which allows us to perfectly know and simulate the distributions $p_x(x)$ and $p_{y|x}(y|x)$. This is of course not the case in real applications. In general, an estimation of $p_x(x)$ has to be constructed empirically using the ensemble. The likelihood can often be considered to be prescribed (Bertino et al, 2003), but in occasions it is also necessary to construct an empirical estimation. For our bivariate method, the estimation of $p_x(x)$ can be done with the ensemble, but in general one does require a good knowledge of the likelihood. When ensemble sizes are small and the knowledge of $p_{y|x}(y|x)$ is not too precise, it is perhaps better to rely on a marginal transformation for both x and y (section **??**, method (d)). This is because one can increase the sample size by including state variables and observations for an extended time period and consider either stationary marginal distributions, or slowly-evolving ones (Simon and Bertino, 2009).

The second limitation is that we have restricted ourselves to the univariate case, and just briefly mentioned some ideas for the multivariate one (section **??**). GA implementations in large models is often done univariately (e.g. Simon and Bertino, 2012). To consider several variables at once would require multivariate anamorphosis. This is indeed a challenging and ongoing area of research (Scholzel and Friedrichs, 2008), and we hope that our insights on the univariate case may give guidance in the multivariate one. A further exploration of joint state variables – observations GA transformations for multivariate cases is part of our ongoing work.

A final comment must be stated. Our entire analysis has been restricted to the analysis step of the EnKF, and we have ignored any effects that the cycling of the forecast and analysis steps may bring. Therefore, the impacts of GA in forecast capabilities has not been assessed. In this sense, any benefit from the suggested transformations has not been proven. We are working towards satisfactorily answering these questions in the future.

7 Acknowledgments

The authors would like to acknowledge fruitful discussions with members of the Data Assimilation Research Center in Reading, in particular the members of the particle filter group. The extensive comments and reviews of three referees helped improve the presentation and content of the work considerably. The support of NERC (project number NE/I0125484/1) is kindly acknowledged.

Apprendix A: Transformations of random variables

Let x be a univariate random variable (rv) with pdf $p_x(x)$. Let g be a monotonic transformation and define $\tilde{x} = g(x)$. The distribution of $\tilde{x}$, denoted as $p_{\tilde{x}}(\tilde{x})$, is given by (see e.g. Casella and Berger, 2002):

$$p_{\tilde{x}}(\tilde{x}) = p_x(g^{-1}(\tilde{x})) \left| \frac{d}{d\tilde{x}} g^{-1}(\tilde{x}) \right| \tag{45}$$

where g^{-1} is the inverse of g. This inverse exists and is unique due to the monotonicity of g; if this condition is not met then one has to divide the sample space $\mathcal{X}$ into subsets $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_K$ in which g is monotonic, perform the transformation in each set, and then add.

One must be careful when transforming conditional probabilities. Consider $p_{y|x}(y|x)$. If one performs the transformation $\tilde{y} = g(y)$, it is clear by (**??**) that

$$p_{\tilde{y}|x}(\tilde{y}|x) = p_{y|x}(\tilde{g}^{-1}(\tilde{y})|x) \left| \frac{d}{d\tilde{y}} g^{-1}(\tilde{y}) \right|$$

On the other hand, if we still consider the same conditional probability $p_{y|x}(y|x)$ but with the transformation $\tilde{x} = g(x)$, the new conditional density is simply $p_{y|\tilde{x}}(y|\tilde{x}) = p_{y|x}(y|x = g^{-1}(\tilde{x}))$, since the transformation is performed on the variable upon which the pdf is conditioned.

The process becomes clearer if we consider the bivariate transformation for the pair $\{x, y\}$. Let this pair have a joint distribution $p_{x,y}(x, y) = p_{y|x}(y|x)p_x(x)$, and define the joint bivariate forward transformations:

$$\begin{aligned} \tilde{x} &= g_1(x, y) \\ \tilde{y} &= g_2(x, y) \end{aligned} \tag{46}$$

Consider these transformations to be invertible resulting in the following two backward transformations:

$$\begin{aligned} x &= q_1(\tilde{x}, \tilde{y}) \\ y &= q_2(\tilde{x}, \tilde{y}) \end{aligned} \tag{47}$$

Then, the transformed pair $\{\tilde{x}, \tilde{y}\}$ has the following joint distribution:

$$p_{\tilde{x},\tilde{y}}(\tilde{x}, \tilde{y}) = p_{x,y}(q_1(\tilde{x}, \tilde{y}), q_2(\tilde{x}, \tilde{y})) \, |det[J]| \tag{48}$$

where $|det(J)|$ is the absolute value of the determinant of the Jacobian matrix of the transformation, namely:

$$J = \begin{bmatrix} \frac{\partial}{\partial \tilde{x}} q_1(\tilde{x}, \tilde{y}) & \frac{\partial}{\partial \tilde{y}} q_1(\tilde{x}, \tilde{y}) \\ \frac{\partial}{\partial \tilde{x}} q_2(\tilde{x}, \tilde{y}) & \frac{\partial}{\partial \tilde{y}} q_2(\tilde{x}, \tilde{y}) \end{bmatrix} \tag{49}$$

In general, a joint multivariate transformation $\tilde{\mathbf{x}} = g(\mathbf{x})$, with $\mathbf{x} \in \mathbb{R}^{N_x}$, $\tilde{\mathbf{x}} \in \mathbb{R}^{N_x}$ and $g : \mathbb{R}^{N_x} \to \mathbb{R}^{N_x}$ will transform a joint pdf $p_x(\mathbf{x})$ into:

$$p_{\tilde{x}}(\tilde{\mathbf{x}}) = p_x(\mathbf{x} = g^{-1}(\tilde{\mathbf{x}})) \left| det \left[\frac{\partial}{\partial \tilde{\mathbf{x}}} g^{-1}(\tilde{\mathbf{x}}) \right] \right| \tag{50}$$

where $\frac{\partial}{\partial \tilde{\mathbf{x}}} g^{-1}(\tilde{\mathbf{x}}) \in \mathbb{R}^{N_x \times N_x}$ is the Jacobian matrix.

Another important concept to recall is the so-called integral probability theorem (IPT). If $P(x)$ is the cdf of x, then the variable $w = P(x)$ has uniform distribution in the interval $[0, 1]$. Multivariate extensions of this theorem exist, although the application is not straightforward as in the univariate case (Genest and Rivest, 2001). The IPT allows us to convert any rv into another; to transform $x \sim p_x(x)$ into $\tilde{x} \sim p_{\tilde{x}}(\tilde{x})$ one can write:

$$\tilde{x} = P_{\tilde{x}}^{-1}(P_x(x)) \tag{51}$$

One can check (**??**) by using (**??**) and defining $g(\cdot) = P_{\tilde{x}}^{-1}(P_x(\cdot))$. Then:

$$p_{\tilde{x}}(\tilde{x}) = p_x(g^{-1}(\tilde{x}))\frac{d}{d\tilde{x}}g^{-1}(\tilde{x}) = p_x(P_x^{-1}(P_{\tilde{x}}(\tilde{x})))\frac{d}{d\tilde{x}}(P_x^{-1}(P_{\tilde{x}}(\tilde{x}))) = \frac{\cancel{p_x(P_x^{-1}(P_{\tilde{x}}(\tilde{x})))}}{\cancel{p_x(P_x^{-1}(P_{\tilde{x}}(\tilde{x})))}}p_{\tilde{x}}(\tilde{x})$$

Appendix B: Exact Bayesian posteriors for GM priors and GM likelihoods

In this appendix, we analytically compute the marginal distribution for the observations $p_y(y)$ and the posterior distribution for the state variable $p_{x|y}(x|y)$ when both the prior probability and the likelihood have Gaussian mixture distributions. We limit the analysis to the univariate case $x \in \mathbb{R}^1$ with observation operator $h = 1$. Following the notation for Gaussian densities introduced in the text, we have:

$$p_x(x) = \sum_{j=1}^{J_b} \alpha_{bj} \phi\left(\frac{x - \mu_{bj}}{\sigma_{bj}}\right) \tag{52}$$

The first two moments of this distribution are:

$$\begin{aligned} \mu_b &= E[x] = \sum_{j=1}^{J_b} \alpha_{bj} \mu_{bj} \\ \sigma_b^2 &= Var[x] = \sum_{j=1}^{J_b} \alpha_{bj} ((\mu_{bj} - \mu_b)^2 + \sigma_{bj}^2) \end{aligned} \tag{53}$$

In a similar manner, the likelihood can be expressed as:

$$p_{y|x}(y|x) = \sum_{j=1}^{J_\eta} \alpha_{\eta j} \phi\left(\frac{y - (x + \mu_{\eta j})}{\sigma_{\eta j}}\right) \tag{54}$$

where the subscript η denotes the additive observational error in the observation equation $y = x + \eta$. The first two moments of the distribution are:

$$\begin{aligned} \mu_{y|x} &= E[y|x] = \sum_{j=1}^{J_b} \alpha_{\eta j} (x + \mu_{\eta j}) = x + \mu_\eta \\ \sigma_{y|x}^2 &= Var[y|x] = \sum_{j=1}^{J_b} \alpha_{\eta j} ((\mu_{\eta j} - \mu_\eta)^2 + \sigma_{\eta j}^2) \end{aligned} \tag{55}$$

where clearly $\mu_\eta = \sum_{j=1}^{J_b} \alpha_{\eta j} \mu_{\eta j}$. Notice that the variance $\sigma_{y|x}^2$ is independent of x.

The joint distribution of the state variables and observations is:

$$p_{x,y}(x, y) = \sum_{j=1}^{J_\eta} \sum_{j'=1}^{J_b} \alpha_{\eta j} \alpha_{bj'} \phi\left(\frac{y - (x - \mu_{\eta j})}{\sigma_{\eta j}}\right) \phi\left(\frac{x - \mu_{bj'}}{\sigma_{bj'}}\right) \tag{56}$$

Recalling that $p_y(y) = \int_{-\infty}^{\infty} p_{x,y}(x, y) dx$, the marginal distribution for the observations is:

$$p_y(y) = \sum_{j=1}^{J_\eta} \sum_{j'=1}^{J_b} \alpha_{\eta j} \alpha_{bj'} \phi\left(\frac{y - (\mu_{\eta j} + \mu_{bj'})}{\sqrt{\sigma_{\eta j}^2 + \sigma_{bj'}^2}}\right) \tag{57}$$

The first two moments of this distribution are:

$$\begin{aligned} \mu_y &= E[y] = \sum_{j=1}^{J_\eta} \sum_{j'=1}^{J_b} \alpha_{\eta j} \alpha_{bj'} (\mu_{\eta j} + \mu_{bj'}) \\ \sigma_y^2 &= Var[y] = \sum_{j=1}^{J_\eta} \sum_{j'=1}^{J_b} \alpha_{\eta j} \alpha_{bj'} ((\mu_{\eta j} + \mu_{bj'} - \mu_y)^2 + \sigma_{\eta j}^2 + \sigma_{bj'}^2) \end{aligned} \tag{58}$$

Using Bayes theorem, we can compute the posterior as:

$$p_{x|y}(x|y) = \frac{\sum_{j=1}^{J_\eta} \sum_{j'=1}^{J_b} w_{ajj'} \phi\left(\frac{x-\mu_{ajj'}}{\sigma_{ajj'}}\right)}{\sum_{j=1}^{J_\eta} \sum_{j'=1}^{J_b} w_{ajj'}} \tag{59}$$

where the subindex ẹnotes analysis. This is again a Gaussian mixture, in which the weights, means and variances of each one of the $J_b J_\eta$ Gaussian terms are:

$$\begin{aligned} w_{ajj'} &= \alpha_{\eta j}\alpha_{bj'}\phi\left(\frac{y-(\mu_{\eta j}+\mu_{bj'})}{\sqrt{\sigma_{\eta j}^2+\sigma_{bj'}^2}}\right) \\ \mu_{ajj'} &= \frac{\sigma_{\eta j}^2}{\sigma_{\eta j}^2+\sigma_{bj'}^2}\mu_{bj'} + \frac{\sigma_{bj'}^2}{\sigma_{\eta j}^2+\sigma_{bj'}^2}(y-\mu_{\eta j}) \\ \sigma_{ajj'}^2 &= \frac{\sigma_{\eta j}^2\sigma_{bj'}^2}{\sigma_{\eta j}^2+\sigma_{bj'}^2} \end{aligned} \tag{60}$$

Finally, the first two moments of this posterior distribution are:

$$\begin{aligned} \mu_{x|y} &= E\left[x|y\right] = \frac{\sum_{j=1}^{J_\eta} \sum_{j'=1}^{J_b} w_{ajj'}\mu_{ajj'}}{\sum_{j=1}^{J_\eta} \sum_{j'=1}^{J_b} w_{ajj'}} \\ \sigma_{x|y}^2 &= Var\left[x|y\right] = \frac{\sum_{j=1}^{J_\eta} \sum_{j'=1}^{J_b} w_{ajj'}\left((\mu_{ajj'}-\mu_a)^2+\sigma_{ajj'}^2)\right)}{\sum_{j=1}^{J_\eta} \sum_{j'=1}^{J_b} w_{ajj'}} \end{aligned}$$

In this paper we have used cases in which either the prior or the likelihood are simple Gaussians. These are obviously special cases of the aforementioned solution. Gaussian likelihood corresponds to $J_\eta = 1$, whereas Gaussian prior corresponds to $J_b = 1$.

References

J. L. Anderson, 2010. A non-Gaussian ensemble filter update for data assimilation. *Mon. Wea. Rev.*, **138**, 4186-4198.

D. Beal, P. Brasseur, J.-M. Brankart, Y. Ourmieres and J. Verron, 2010. Characterization of mixing errors in a coupled physical biochemical model of the North Atlantic: implications for nonlinear estimation using Gaussian anamorphosis. *Ocean Sci.*, **6**, 247-262.

L. Bertino, G. Evensen and H. Wackernagel, 2003. Sequential data assimilation techniques in oceanography. *Int. Stat. Rev.*, **71**, 223-241.

M. Bocquet, C. Pires and L. Wu, 2010. Beyond Gaussian statistical modelling in geophysical data assimilation. *Mon. Wea. Rev.*, **138**, 2997-3023.

J.-M. Brankart, C.-E. Testut, D. Beal, M. Doron, C. Fontana, M. Meinvielle, P. Brasseur and J. Verron, 2012. *Ocean Sci.*, **8**, 121-142.

G. Burgers, P. J. van Leeuwen and G. Evensen, 1998. Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719-1724.

G. Casella and R. Berger, 2002. Statistical inference. 2nd edition, *Duxbury advanced series.*

S. E. Cohn, 1997. An introduction to estimation theory. *J. Meteor. Soc. Japan*, **75**, 257-288.

T. M. Cover and A. J. Thomas, 2006. Elements of information theory. 2nd edition, *John Wiley & Sons.*

M. Doron, P. Brasseur and J.-M. Brankart. Stochastic estimation of biochemical parameters of a 3D ocean coupled phyisical-biological model: Twin experiments, 2011. *J. Marine Syst.*, **87**, 194-207.

G. Evensen, 2003. The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–367.

G. Evensen, 2006. Data assimilation: the ensemble Kalman filter. *Springer.*

C. Genest, L-P Rivest, 2001. On the multivariate probability integral transformation. *Stat. & Prob. Let.*, **53**, 391-399. M. Ghil, P. Malanotte-Rizzoli, 1991. Data assimilation in meteorology and oceanography. *Adv. Geophys.*, **33**, 141–266.

P. L. Houtekamer and H. I. Mitchell, 1998. Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811.

P. L. Houtekamer, H.L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, B. Hansen, 2005. Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Mon. Wea. Rev.*, **133**, 604–620.

B. R. Hunt, E. J. Kostelich and I. Szunyogh, 2007. Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter. *Phys. D*, **230**, 112-126.

K. Ide, P. Courtier, M. Ghil and A. C. Lorenc, 1997. Unified notation for data assimilation: operational, sequential and variational. *J. Meteo. Soc. Japan*, **75**, 181-189.

A. Jazwinski, 1970. Stochastic processes and filtering theory. *Academic Press.* New York, 376 pp.

R. E. Kalman, 1960. A new approach to linear filtering and prediction problems. *Trans. of the ASME, Jour. of Bas. Engin. D*, **82**, 35-45.

R. E. Kalman and R. Bucy, 1961. New results in linear filtering and prediction problems. *Trans. of the ASME, Jour. of Bas. Engin. D*, **83**, 95-108.

G. Y. Lien, E. Kalnay and T. Miyoshi, 2013. Effective assimilation of global precipitation: simulation experiments. *Tellus A*, **65**, 19915.

R. N. Miller, M. Ghil and F. Gauthiez, 1994. Advanced data assimilation in strongly nonlinear dynamical systems. *J. Atmos. Sci.*, **51**, 1037-1056.

R. N. Miller, E. F. Carter and S. T. Blue, 1999. Data assimilation into nonlinear stochastic models. *Tellus A*, **51A**, 167-194.

C. A. Pires and R. Perdigao, 2007. Non-Gaussianity and asymmetry of the winter monthly precipitation estimation from NAO. *Mon. Wea. Rev.*, **135**, 430-448.

M. Rosenblatt, 1952. Remarks on a Multivariate Transformation. *The Annals of Mathematical Statistics*, **23**, 3, 470-472.
D. Simon, 2006. Optimal State Estimation. *John Wiley & Sons*. New York. 552 pages.
E. Simon and L. Bertino, 2009. Application of the Gaussian anamorphosis to assimilation in a 3-D coupled physical-ecosystem model of the North Atlantica with the EnKF: A twin experiment. *Ocean Sci.*, **5**, 495-510.
E. Simon and L. Bertino, 2012. Gaussian anamorphosis extension of the DEnKF for combined state parameter estimation: Application to a 1D ocean ecosystem model. *Jour. of Mar. Sys.*, **89**, 1-18.
C. Scholzel and P. Friedrichs, 2008. Multivariate non-normally distributed random variables in climate research - introduction to the copula approach. *Nonlin. Proc. Geophys.*, **15**, 761-772. I. Szunyogh, E. J. Kostelich, G. Gyarmati, D. J. Patil, B. R. Hunt, E. Kalnay, E. Ott, J. A. Yorke, 2007. Assessing a local ensemble Kalman filter: Perfect model experiments with the national centers for environmental prediction global model. *Tellus A*, **57**, 528–545.
M. K. Tippett, J. L. Anderson, C. H. Bishop, T. M. Hamill, J. S. Whitaker, 2003. Ensemble square-root filters. *Mon. Wea. Rev.*, **131**, 1485–1490.
R. Von Mises, 1964. Mathematical theory of probability and statistics. *Academic Press*. New York.
H. Wackernagel, 2003. Multivariate Geostatistics. 3rd ed. *Springer*, 403 pp.
H. Zhou, J. J. Gomez-Hernandez, H.-J. Hendricks Franssen and L. Li, 2011. An approach to handling non-Gaussianity of parameters and state variables in ensemble Kalman filtering. *Advances in Water Resources*, **34**, 844-864.

Figures

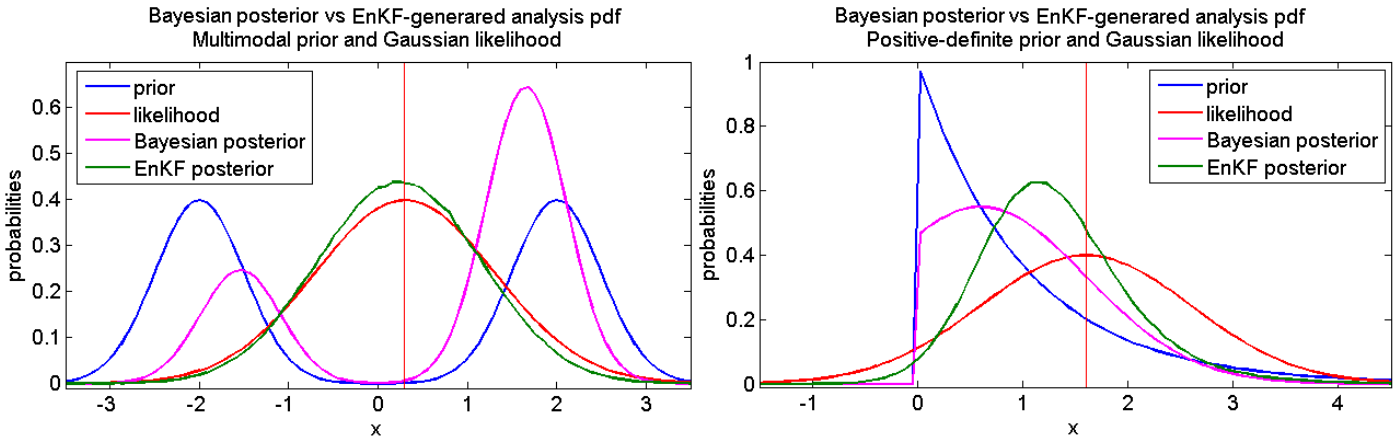

Figure 1: Comparison of the analysis pdfs obtained by a direct application of the EnKF analysis step (green line) with respect to the actual Bayesian posteriors (magenta line). The state variables have either a multimodal prior distribution (left), or they are positive-definite quantities (right). The EnKF analysis step is applied with $M = 10^6$.

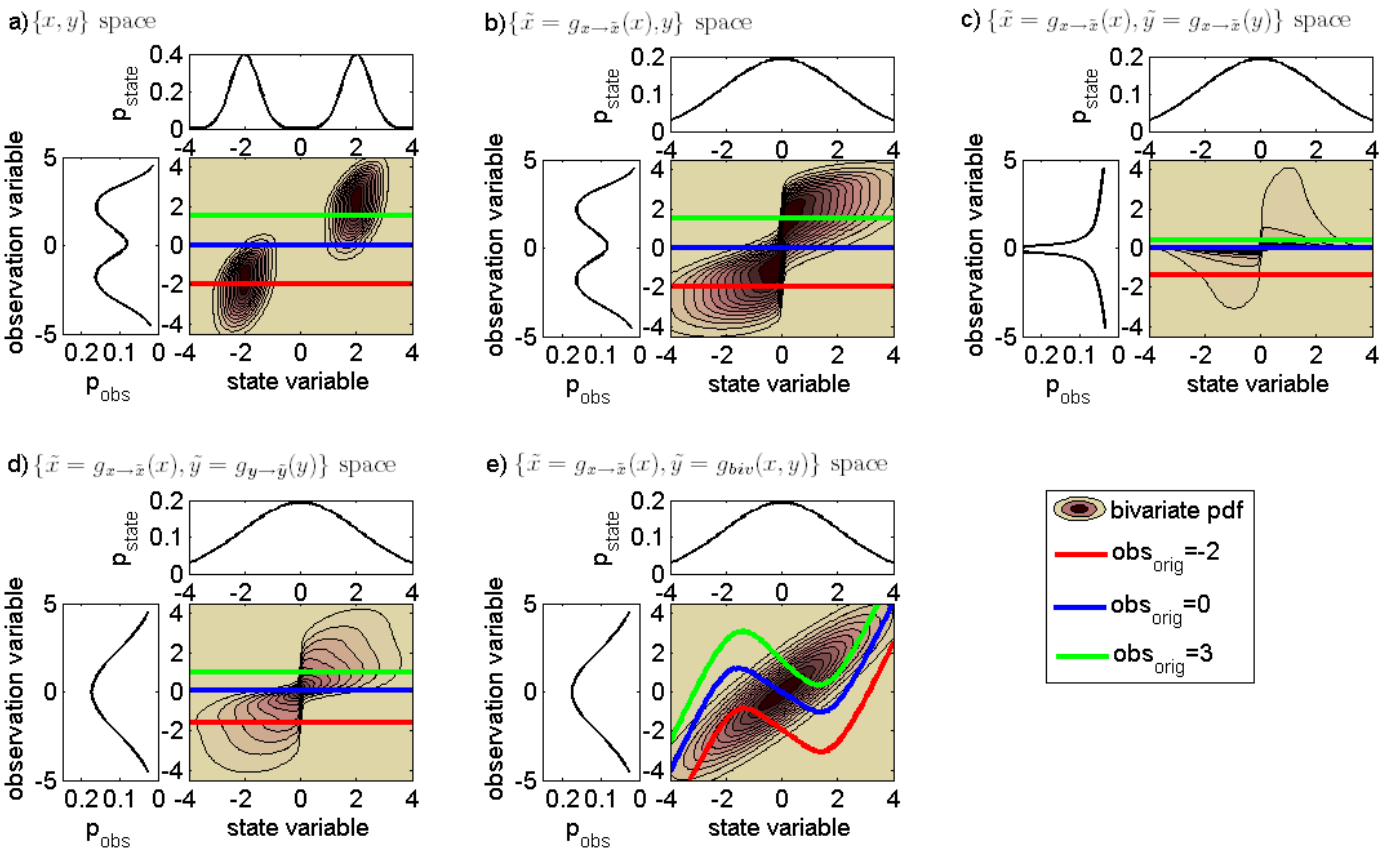

Figure 2: Bivariate distributions (contour plots) and marginal distributions (line plots) for state variables (horizontal) and observations (vertical) under 6 different transformations (panels (a)-(e)) described in the text. Individual *given* observations are identified with color lines in the contour plot, except for panel (f) where individual values of x are shown.

.

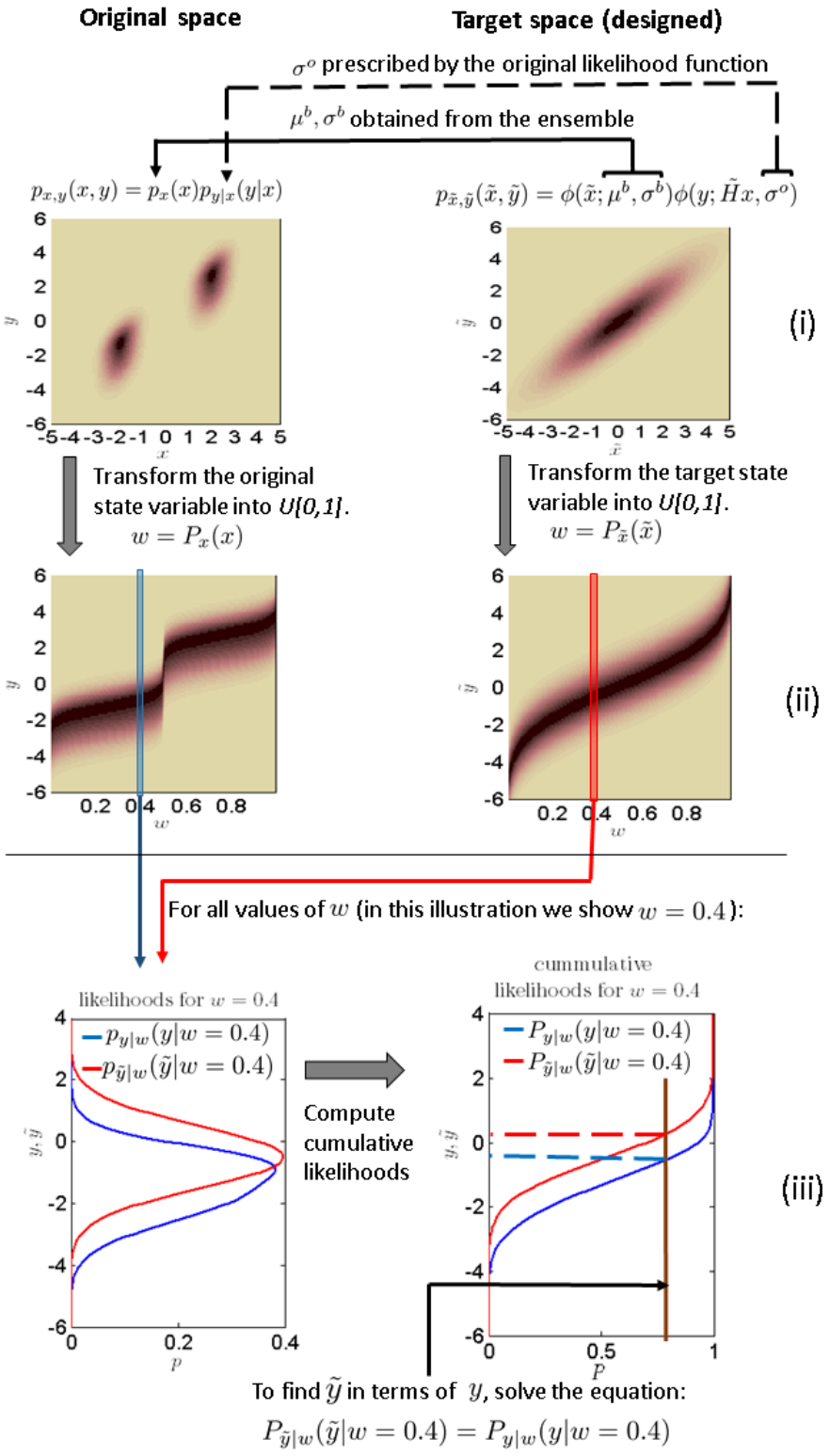

Figure 3: The process for the joint state-space/observations transformation described in section **??**. First (i), a target probability space is constructed using the statistical moments inferred or prescribed by the original variables. Second (ii), the state variables (both original and transformed) are mapped into a random variable w distributed $U[0,1]$. Finally (iii), for each w an equation of cumulative likelihoods is solved to find $\tilde{y}$ in terms of y.

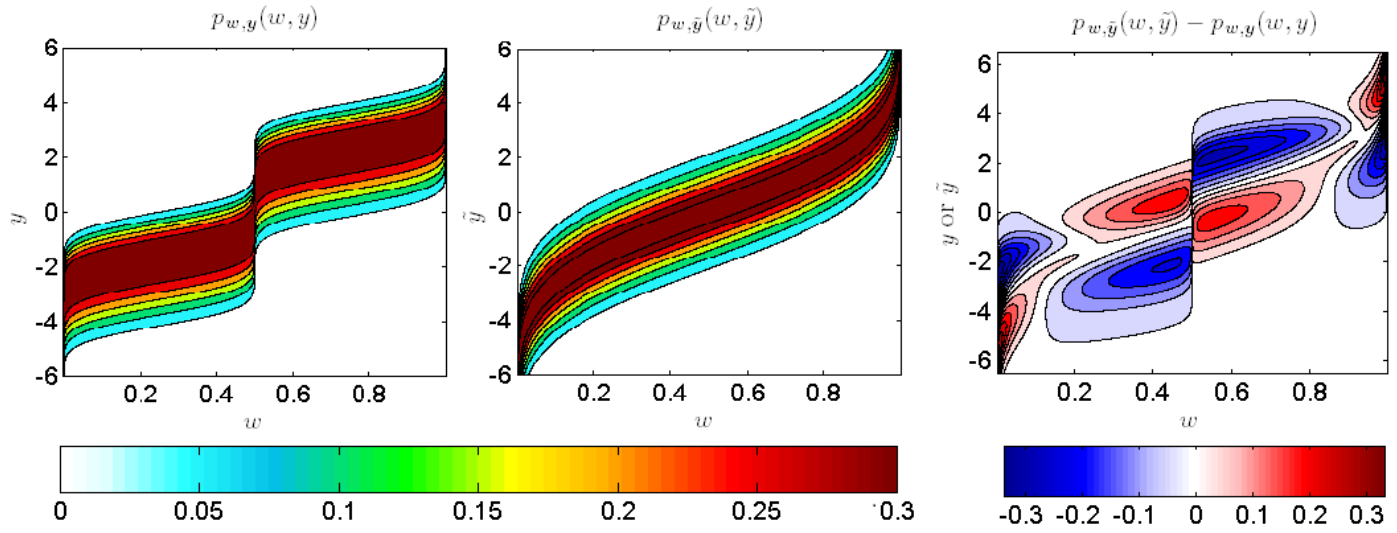

Figure 4: Joint pdfs for the spaces $\{w, y\}$ (left) and $\{w, \tilde{y}\}$ (center). The difference between the two densities is plotted in the right panel.

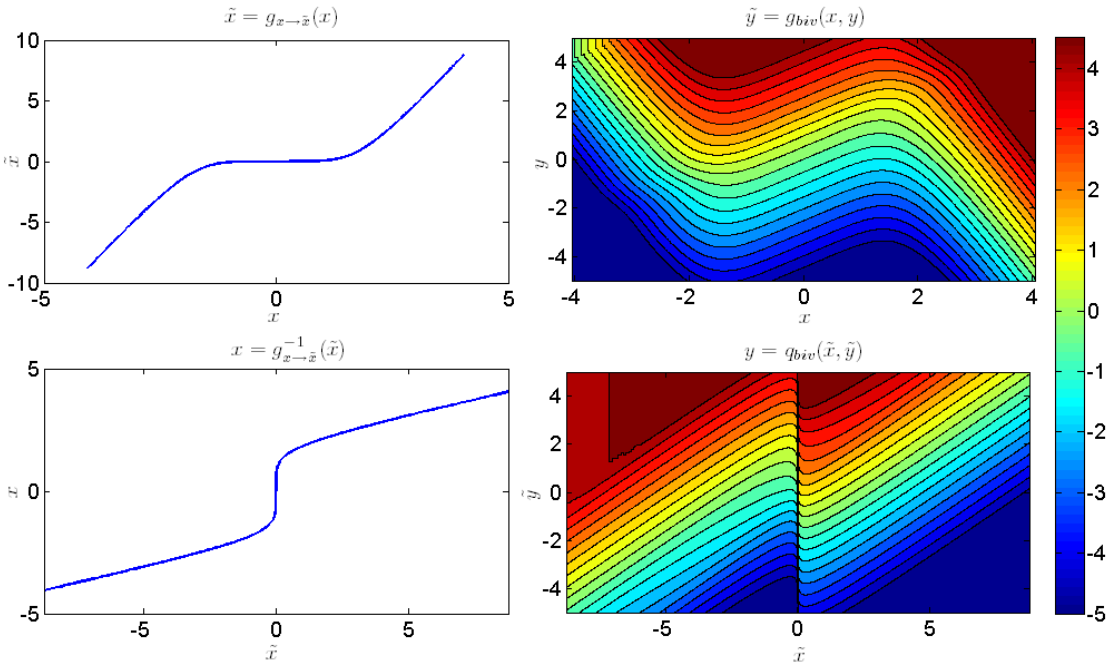

Figure 5: Joint bivariate transformations from the $\{x, y\}$ space (with GM marginals) to the $\{\tilde{x}, \tilde{y}\}$ space (with a joint bivariate Gaussian pdf). The first row show the forward transformations: the state variable is univariately transformed (left) whereas the observation is transformed in a joint bivariately manner (right). The backward transformations are presented in the bottom row.

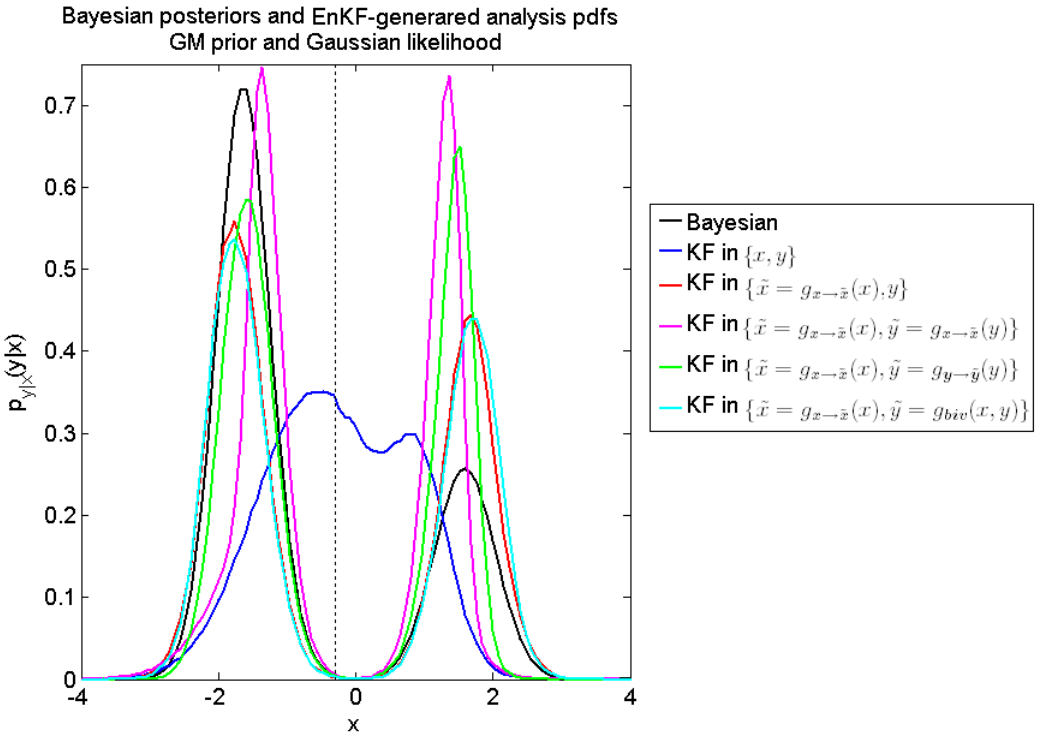

Figure 6: Comparison of the Bayesian posterior distribution (black line) with respect to the EnKF-generated analysis pdfs, with the EnKF analysis step applied in 5 different spaces (colour lines) for a given observation (dotted vertical line). Both the prior and likelihood in the original space are GMs.

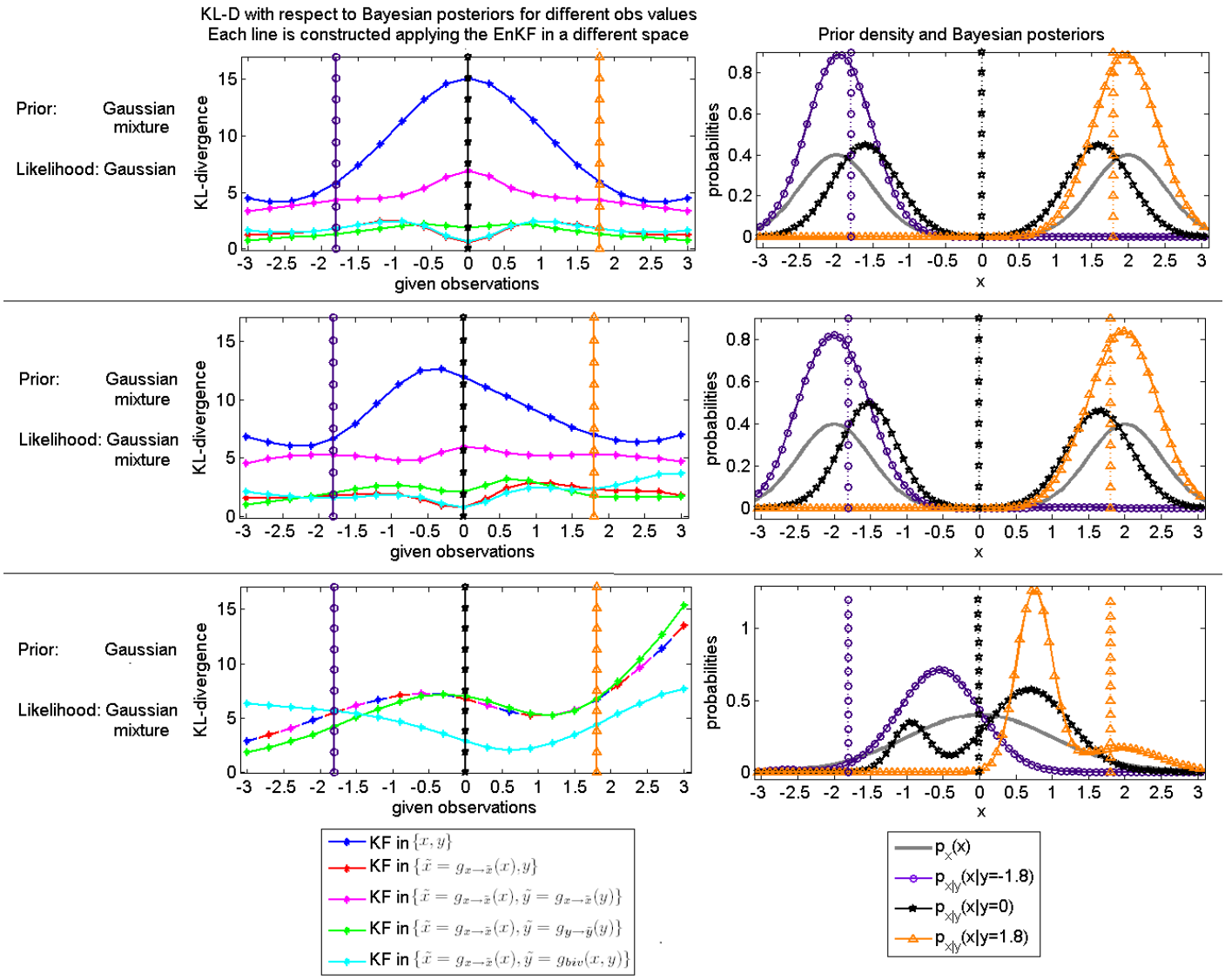

Figure 7: Assessing the quality of the EnKF-generated analysis pdf's for three cases: GM-G (top row), GM-GM (center row) and G-GM (bottom row). The left panels shows the D_{KL} for the EnKF-generated analysis pdf's with respect to the Bayesian posterior (colored lines) for different given observations (horizontal axis). In each case we choose 3 given observations (vertical lines with markers) and in the right panels we show the Bayesian posteriors associated with those 3 observations (colored lines with markers, the colors correspond to those of the vertical lines). The solid grey curve in these panels represents the prior for each case.