Author/s:
Bishop, CH

Title:
The GIGG-EnKF: Ensemble Kalman filtering for highly skewed non-negative uncertainty distributions

Date:
2016-04-01

Citation:
Bishop, C. H. (2016). The GIGG-EnKF: Ensemble Kalman filtering for highly skewed non-negative uncertainty distributions. Quarterly Journal of the Royal Meteorological Society, 142 (696), pp.1395-1412. https://doi.org/10.1002/qj.2742.

Persistent Link:
https://hdl.handle.net/11343/291086

The GIGG-EnKF: Ensemble Kalman Filtering for highly skewed non-negative uncertainty distributions.

CRAIG H. BISHOP
Naval Research Laboratory, Monterey, CA, USA

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/qj.2742

Corresponding author address: Craig H. Bishop, Naval Research Laboratory, Marine Meteorology Division, 7 Grace Hopper Ave, Stop 2, Building 702, Room 212, Monterey, CA 93943-5502. E-mail: bishop@nrlmry.navy.mil

Abstract: Observations and predictions of near-zero non-negative variables such as aerosol, water vapor, cloud, precipitation and plankton concentrations have uncertainty distributions that are skewed and better approximated by gamma and inverse-gamma probability distribution functions (pdfs) than Gaussian pdfs. Current Ensemble Kalman Filters (EnKFs) yield suboptimal state estimates for these variables. Here, we introduce a variation on the EnKF that accurately solves Bayes' theorem in univariate cases where the prior forecasts and error prone observations given truth come in (gamma, inverse-gamma) or (inverse-gamma, gamma) or (Gaussian, Gaussian) distribution pairs. Since its multivariate extension is similar to an EnKF, we refer to it as the GIGG-EnKF or GIGG where GIGG stands for Gamma, Inverse-Gamma and Gaussian. The GIGG-EnKF enables near-zero semi-positive definite variables with highly skewed uncertainty distributions to be assimilated without the need for observation bias inducing lognormal or Gaussian anamorphosis non-linear transformations. In the special case that all observations are treated as Gaussian, the GIGG-EnKF gives identical results to the original EnKF. A multi-grid-point and multi-variable idealized system was used to compare and contrast the data assimilation performance of the GIGG with that of both the perturbed observation and deterministic forms of the EnKF. This test system featured variables and observation types whose uncertainty distributions approximate Gaussian, gamma and inverse-gamma distributions. The normalized analysis error variance of the GIGG ensemble mean was found to be significantly smaller than that of the EnKFs. The higher moments of the analyzed ensemble distributions were tested by subjecting the

ensemble members to non-linear "forecast" mappings. The normalized mean square error of the mean of the corresponding GIGG forecast ensemble was found to be less than a 3rd of that obtained from either form of the original EnKF.

Key Words: Non-Gaussian, Ensemble Kalman Filter, Lognormal, Gamma, Inverse-Gamma, Skewed, Data Assimilation, State Estimation.

1. Introduction

The accurate probabilistic estimation of the state of some system from prior information and recent observations is fundamental to a wide range of endeavors in science, engineering and finance. In meteorology and oceanography, the process of combining prior information from a model forecast with recent observations is called data assimilation (Kalnay, 2004). Ensemble data assimilation attempts to represent the distribution of truth after the assimilation of recent observations in terms of a K member ensemble of state estimates.

From a theoretical perspective, current operational data assimilation techniques (Rabier et al., 2000, Xu et al., 2005, Bonavita et al., 2012, Clayton et al., 2013, Wang et al., 2013, Kuhl et al, 2013) are designed to be accurate in the case where the distribution of truth given prior information is a multi-variate Gaussian distribution and the distribution of observations given a particular truth is also Gaussian and independent of the value of the truth. The degree of inaccuracy in these systems increases as the non-

Gaussianity of these systems increases. Near zero positive definite variables whose distributions are such that their standard deviations are of the same order of magnitude as their means are inevitably skewed and non-Gaussian. Such variables include high impact weather and climate variables such as those pertaining to aerosols (O'Neill et al., 2000, Saide et al., 2012) rainfall (Errico et al., 2001, Simpson, 1972, Husak et al., 2007), water-vapor-mixing-ratio (Kliewer et al., 2015), cloud-water/ice concentrations (Willis, 1984, Vivekanandan et al., 2004), phytoplankton (Pelc et al., 2012) and sea-ice (Wadhams et al., 1987, Lange and Eicken, 1991).

Some of these distributions are better approximated by gamma, inverse-gamma or lognormal probability density functions (pdfs) than Gaussian pdfs. Figure 1 compares and contrasts Gaussian, gamma, inverse-gamma and lognormal pdfs.

When the mean of a gamma, inverse-gamma or lognormal pdf is the same order of magnitude as its standard deviation, its shape is very different to the corresponding Gaussian pdf (blue and green curves). While the Gaussian distribution is symmetric and permits negative values, with gamma, inverse-gamma and lognormal pdfs, only non-negative quantities are assigned non-zero probability densities and the distributions are highly skewed.

In contrast, when the standard deviations are much smaller than the means (red curves on Figure 1), gamma, inverse-gamma and lognormal pdfs all take on shapes very similar to the corresponding Gaussian distributions. Note that while both inverse-gamma and lognormal pdfs always assign a probability density of zero to the value zero, the

gamma pdf allows large values of probability density to be assigned to the value zero. A secondary difference is the fact that for the same relative variance, the inverse gamma and lognormal pdfs are more highly skewed than the gamma pdf. Though qualitatively very similar to the lognormal pdf, the inverse-gamma pdf generally assigns less probability density to values near zero than the lognormal pdf.

The gamma and lognormal pdf have been widely used to characterize prior climatologies of rainfall (Simpson, 1972, Cho et al., 2004, Husak et al., 2007), rain and cloud parameterizations (Clark, 1974, Bougeault, 1982, Willis, 1984, Vivekanandan et al., 2004). Campbell (1995, Figure 4) shows that the lognormal pdf gives a reasonable fit of observed distributions of chlorophyll fluorescence. O'Neill et al. (2000) used lognormal distributions to approximate observed aerosol optical depths. Errico et al. (2001) used the lognormal distribution to characterize the distribution of forecast errors due to uncertainty in convective parameterizations. Lien et al's (2013) Figure 1a suggests that a gamma pdf would closely match the climatological pdf of rainfall in their simplified global atmospheric model. However, in their data assimilation experiments, Lien et al. (2013) assimilate a Gaussian anamorphosis transformation of precipitation observations in a data assimilation scheme designed for Gaussian uncertainties. Kliewer et al. (2015) used the lognormal pdf to characterize the prior distribution of water vapor mixing ratio in a data assimilation scheme designed for Gaussian variables. Similarly, Saide et al. (2015) assumed that both the background uncertainty associated with aerosol

number concentration and the uncertainty associated with an observation of cloud droplet number concentration could both be described by lognormal distributions.

If one could find an observing instrument for which the pdf of error-prone observed values y^o given a single true value y was given by the lognormal likelihood pdf $L\left(y^o \mid y\right) = \frac{1}{y^o \sigma\sqrt{2\pi}} \exp\left[-\frac{\left[\ln\left(y^o\right) - \ln\left(y\right)\right]}{2\sigma^2} \right]$ then it would be trivial to include this observation type in a data assimilation scheme designed to assimilate variables having Gaussian uncertainty distributions because, in this case, the pdf of $\ln\left(y^o\right)$ is Gaussian with a variance of σ^2 and a mean that is unbiased in the sense that the average value of $\ln\left(y^o\right)$ over all possible observations given the true value $\ln\left(y\right)$ is equal to $\ln\left(y\right)$; i.e., $\left\langle \ln\left(y^o\right) \right\rangle = \ln\left(y\right)$. Hence, in this case, one could just take the logarithm of the observation and assimilate it into a data assimilation scheme designed for Gaussian variables. The simplicity of this approach may be part of the reason that it was employed by both Kliewer et al. (2015) and Saide et al. (2015).

However, in order for this simple approach to yield unbiased analyses, $\ln\left(y^o\right)$ must be an unbiased estimate of $\ln\left(y\right)$. But if indeed $\left\langle \ln\left(y^o\right) \right\rangle = \ln\left(y\right)$ then, by the standard formula for the mean of a lognormal distribution, $\left\langle y^o \right\rangle = \exp\left[\ln\left(y\right) + \frac{\sigma^2}{2} \right] = y \exp\left(\frac{\sigma^2}{2} \right)$ and hence the raw observation y^o would be a

biased estimate of the true value y and that bias would depend on the measurement error variance. Conversely, if y^o was an unbiased estimate of y then $\ln\left(y^o\right)$ would be a biased estimate of $\ln\left(y\right)$ and the approach of using a data assimilation scheme designed for Gaussian variables to assimilate the logarithm of raw observations would lead to *systematically biased state estimates*.

It can be shown that non-linear transformations of unbiased observations result in transformed observations that are biased. This bias depends on the observation error of representation, which depends on the strength of local gradients and sub-grid scale eddies, which are both extremely difficult to accurately predict. Consequently, the problem of removing the flow dependent observation bias induced by taking non-linear transformations of observations (such as a logarithmic or Gaussian anamorphosis transformation) is non-trivial.

In order to more accurately estimate the state of variables whose prior and observation likelihood pdfs can be approximated by the curves given in figure 1 while avoiding the flow-dependent-observation-bias problems incurred by assimilating non-linear functions of the observations, we introduce a new ensemble based approach (the GIGG-EnKF) that is not only accurate for symmetric Gaussian like distributions but also for gamma and inverse-gamma distributions. In addition to being perfectly suited to the assimilation of variables whose uncertainty is best described by gamma and/or inverse gamma pdfs, given the similarity of gamma and inverse-gamma pdfs to log normal pdfs,

the GIGG-EnKF could be used on variables whose uncertainty has previously been modelled as lognormal. These variables include water-vapor-mixing ratio (Kliewer et al., 2015), aerosol number concentration and droplet number concentration (Saide et al., 2015). They also include the previously mentioned variables whose prior climatologies approximately follow gamma, lognormal or inverse gamma pdfs such as rainfall, chlorophyll fluorescence, and ice concentrations.

The multi-variate GIGG-EnKF algorithm may be viewed as a particular form of Anderson's (2003) generalized two-stage ensemble filter. In the first stage, the prior ensemble forecast of an observation is updated by assimilating a single observation using an appropriate choice of one of three distinct equation sets: the GIG equation set, which is designed for a prior ensemble drawn from a Gamma pdf and an Inverse-Gamma observation likelihood; an IGG equation set, which is designed for an Inverse Gamma prior and a Gamma likelihood; and a G or Gaussian equation set designed for Gaussian forecast and observation uncertainty. In the second stage, linear regression is used to estimate the changes in the state estimate of the extended model-plus-observed-variables state vector due to the assimilation of the observation. The procedure is then repeated for the next observation, and so on, until all observations are assimilated.

It can be shown (e.g. Bishop et al., 2015) that all-at-once assimilation is identical to serial assimilation for, say, a perturbed observations EnKF with no covariance localization. Hence, one could choose to group together all the observations likely to satisfy Gaussian assumptions and then assimilate them all-at-once or in batches using

ones preferred flavor of perturbed observations ensemble data assimilation scheme: EnKF, 4D-variational or otherwise. As detailed in Section 5, in this scenario, only the non-Gaussian observations would be serially assimilated using the GIGG-EnKF.

Other approaches for dealing with non-Gaussianity include variable transformation approaches (e.g. Zhou et al., 2011 and Amezcua and van-Leeuwen, 2014), particle filters (e.g. van Leeuwen and Ades, 2013) and higher order EnKFs (e.g. Hodyss, 2012 and Hodyss and Campbell, 2013). The performance differences between these methods, the aforementioned lognormal approaches and the GIGG no doubt depend on the extent to which the particular test system considered satisfies the assumptions of the various methods. A detailed exposition of these performance differences is beyond the scope of this paper.

The high profile of EnKFs in state estimation across a wide variety of fields is evident in the number of papers and citations associated with them found in scholarly databases. In meteorology, they are widely used in research (e.g. Anderson et al., 2009) and also in operations (e.g. Houtekamer et al., 2005 and Whitaker et al., 2008). This paper shows how the GIGG variation allows the EnKF to accurately accommodate a much broader range of uncertainty distributions with few coding changes and little additional computational expense.

Section 2 describes how GIGG produces an analysis ensemble in the special case where the prior distribution can be described by a gamma pdf and the likelihood pdf of observations given truth is given by an inverse-gamma pdf; we refer to this particular

case as the GIG case. Section 3 does the same for the case where the prior is an inverse-gamma pdf and the observation likelihood is a gamma pdf; we refer to this case as the IGG case. Section 4 measures the similarity of the pdfs randomly sampled by the GIGG-EnKF with the true posterior pdfs over a range of parameter values. Section 5 shows how Anderson's (2003) ensemble regression approach can be used to arrive at a multivariate perturbed observations GIGG-EnKF that can be summarized in a concise algorithm. Section 6 describes the idealized multi-grid point, multi-variable system used to compare and contrast the performance of the GIGG filter with that of the perturbed observations and deterministic form of the EnKF (see Tippett et al. 2003 for a description of these two types of EnKFs). Results of the experiments with the idealized model are discussed and interpreted in Section 7. Section 8 summarizes the paper.

2. Univariate GIG (gamma prior and inverse-gamma observation likelihood)

2.1 Analytic solution

Let y_{ij}^f denote the ith member of a prior ensemble forecast of the jth observation y_j^o of the true state y_j. We suppose that the continuous prior pdf of true values y_j that y_{ij}^f randomly samples is given by a gamma (γ) pdf of the form

$$\rho_{prior}\left(y_j\right)=\frac{1}{\Gamma\left[\left(P_j^r\right)^{-1}\right]}\frac{1}{y_j}\left(\left(P_j^r\right)^{-1}\frac{y_j}{\left\langle y_j^f\right\rangle}\right)^{\left(P_j^r\right)^{-1}}\exp\left(-\left(P_j^r\right)^{-1}\frac{y_j}{\left\langle y_j^f\right\rangle}\right)$$
$$=\frac{1}{\Gamma\left[\left(P_j^r\right)^{-1}\right]}\left(\frac{\left(P_j^r\right)^{-1}}{\left\langle y_j^f\right\rangle}\right)^{\left(P_j^r\right)^{-1}}\left\{\left(y_j\right)^{\left[\left(P_j^r\right)^{-1}-1\right]}\exp\left(-\left(P_j^r\right)^{-1}\frac{y_j}{\left\langle y_j^f\right\rangle}\right)\right\} \quad (1)$$

where $\left\langle y_j^f\right\rangle$ is the mean of the prior distribution and $P_j^r=\frac{\mathrm{var}\left(y_j\right)}{\left\langle y_j^f\right\rangle^2}$ is the type 1 relative error variance of the prior distribution. Note that if a K member ensemble forecast is available that randomly samples this prior distribution then one can make the approximations $\left\langle y_j^f\right\rangle\approx\frac{1}{K}\sum_{i=1}^{K}y_{ij}^f=\overline{y_j^f}$ and $P_j^r\approx\frac{\frac{1}{K-1}\sum_{i=1}^{K}\left(y_{ij}^f-\overline{y_j^f}\right)^2}{\left(\overline{y_j^f}\right)^2}$. Examples of gamma pdfs are given in Figure 1.

To keep track of the notation used in this paper, refer to Table 1.

Note that in order to make the connection between gamma pdfs and the error variances used in EnKFs clearer, in (1) we have deliberately departed from the usual practice of describing gamma pdfs in terms of shape and scale parameters k and θ. The mean and relative variance are related to these parameters via the equations $k=\left(P_j^r\right)^{-1}$ and $\theta=\left\langle y_j^f\right\rangle P_j^r$.

Now suppose that the pdf $L\left(y_j^o \mid y_j\right)$ of observed concentrations y_j^o given a true y_j is described by an inverse gamma pdf so that for a fixed truth y_j

$$L\left(y_j^o \mid y_j\right) = [y_j\left(\tilde{R}_j^r\right)^{-1}]^{\left[\left(\tilde{R}_j^r\right)^{-1}+1\right]} \frac{\left(y_j^o\right)^{-\left(\tilde{R}_j^r\right)^{-1}-2} \exp\left(-\left(\tilde{R}_j^r\right)^{-1} \frac{y_j}{y_j^o}\right)}{\Gamma\left(\left(\tilde{R}_j^r\right)^{-1}+1\right)}$$
$$= \frac{\left(\tilde{R}_j^r\right)^{-\left[\left(\tilde{R}_j^r\right)^{-1}+1\right]} \left(y_j^o\right)^{-\left(\tilde{R}_j^r\right)^{-1}-2}}{\Gamma\left(\left(\tilde{R}_j^r\right)^{-1}+1\right)} y_j^{\left[\left(\tilde{R}_j^r\right)^{-1}+1\right]} \exp\left(-\left(\tilde{R}_j^r\right)^{-1} \frac{y_j}{y_j^o}\right) \quad (2)$$

where $\tilde{R}_j^r = \frac{\mathrm{var}\left(y_j^o - y_j\right)}{\left(y_j\right)^2 + \mathrm{var}\left(y_j^o - y_j\right)} = \frac{\mathrm{var}\left(\varepsilon^o\right)}{\left(y_j\right)^2 + \mathrm{var}\left(\varepsilon^o\right)}$ is the type 2 relative observation error variance and ε^o is an observation error. Note that, by definition, $0 \le \tilde{R}_j^r \le 1$. It can be shown that the mean of the distribution of observations y_j^o given a fixed y_j is equal to y_j. Thus, (2) ensures that the mean observation error is zero and hence that the observation is unbiased.

Inverse gamma pdfs are often expressed in terms of the parameters α and β where $\alpha = \left(\tilde{R}_j^r\right)^{-1} + 1$ and $\beta = y_j\left(\tilde{R}_j^r\right)^{-1}$. See Fig. 1 for examples of inverse gamma pdfs.

By Bayes' theorem, the posterior pdf of y_j given y_j^o is

$$\rho_{\mathrm{post}}(y_j \mid y_j^o) = \frac{L(y_j^o \mid y_j)\rho_{\mathrm{prior}}(y_j)}{\int_0^\infty L(y_j^o \mid y_j)\rho_{\mathrm{prior}}(y_j)dy_j} \quad (3)$$

In Appendix A, we show that with (1) and (2), this posterior is given by the gamma pdf

$$\rho_{\text{post}}(y_j \mid y_j^o) = \frac{1}{\Gamma\left(\left(\Pi_j^r\right)^{-1}\right)}\left(\frac{\left(\Pi_j^r\right)^{-1}}{\left\langle y_j^a \right\rangle}\right)^{\left(\Pi_j^r\right)^{-1}} \left\{ \left(y_j\right)^{\left(\left(\Pi_j^r\right)^{-1}-1\right)} \exp\left[-\left(\Pi_j^r\right)^{-1}\frac{y_j}{\left\langle y_j^a \right\rangle}\right]\right\} \qquad (4)$$

with type 1 relative error variance Π_j^r and mean $\left\langle y_j^a \right\rangle$ respectively given by

$$\begin{aligned} \Pi_j^r &= \left[\left(\tilde{R}_j^r\right)^{-1} + \left(\tilde{P}_j^r\right)^{-1}\right]^{-1} = \tilde{P}_j^r - \tilde{P}_j^r\left(\tilde{P}_j^r + \tilde{R}_j^r\right)^{-1}\tilde{P}_j^r, \text{ and} \\ \frac{1}{\left\langle y_j^a \right\rangle} &= \frac{1}{\left\langle y_j^f \right\rangle} + \frac{\tilde{P}_j^r}{\tilde{R}_j^r + \tilde{P}_j^r}\left[\frac{1}{y_j^o} - \left(\tilde{R}_j^r + 1\right)\frac{1}{\left\langle y_j^f \right\rangle}\right]. \end{aligned} \qquad (5)$$

In (5) $\tilde{P}_j^r = \dfrac{\text{var}\left(y_j\right)}{\left\langle y_j^f \right\rangle^2 + \text{var}\left(y_j\right)} = \left[\left(P_j^r\right)^{-1} + 1\right]^{-1}$ is the type 2 relative forecast error variance.

Note the high degree of similarity between the equation for the posterior type 1 relative error covariance and that for the analysis error covariance of a Kalman filter (see, for example, Kalnay, 2004). Note also that in the equation for the mean of the posterior distribution, a counterpart of the Kalman gain matrix appears. However, this gain matrix is applied to a difference of the inverse of the observation and an inverse of the prior mean and hence, in this respect, the update equation for the posterior mean is radically different to the Kalman update. This difference is particularly noticeable in the limit of infinite observation error variance where $\tilde{R}_j^r \to 1$ and (5) becomes

$$\begin{aligned} \left(\Pi_j^r\right)^{-1} &= \left(\tilde{P}_j^r\right)^{-1} + 1 = \left(P_j^r\right)^{-1} + 2 \\ \frac{1}{\left\langle y_j^a \right\rangle} &= \frac{1}{\left\langle y_j^f \right\rangle}\left(\frac{1-\tilde{P}_j^r}{1+\tilde{P}_j^r}\right) + \frac{1}{y_j^o}\left(\frac{\tilde{P}_j^r}{1+\tilde{P}_j^r}\right) \end{aligned}$$

Thus, both the posterior mean and Type 1 relative error variance are changed by observations even when the observation error variance is infinite. This is a direct consequence of the fact that an inverse-gamma observation likelihood always assigns zero probability density to the possibility of the truth being equal to zero when the observation is non-zero – even when the observation error variance is infinite; i.e. there is structure and hence information in the observation likelihood even when its positive tail has become so long that its variance is infinite. Similarly, the posterior mean is different from the observed value even when the forecast error variance tends to infinity $\tilde{P}_j^r -> 1$. This is because as $\tilde{P}_j^r -> 1$ the probability density of the prior truth being equal to zero tends to infinity at the same time that the variance tends to infinity. This pronounced effect on the shape of the prior causes it to influence the shape of the posterior even though the prior variance is infinite. To summarize: unlike Gaussian pdfs, gamma and inverse-gamma pdfs do not become flat as their variances become infinite; hence, infinite variance gamma and inverse-gamma pdfs are not uninformative. Note however, that these infinite variance cases are of purely academic interest because the variance of the prior is bounded by the climatological variance of the variable being considered and observational instruments are unlikely to be deployed unless the variance of their likelihood pdfs is less than the climatological variance.

Also note that even though the gamma pdf can assign finite *probability-density* to the value zero, it is incapable of describing a variable that has a finite *probability* of being

equal to zero. To see this, recall that in order to obtain the probability of a variable lying between two distinct values, one integrates the variable's pdf between these two values. Hence, to obtain the probability of zero occurring one must integrate the pdf from zero to some infinitesimally small positive value. For gamma pdfs that assign finite probability densities to zero, this integral will be equal to zero in the limit of a vanishingly small interval about zero. Hence, despite the fact that gamma pdfs can assign non-zero probability densities to zero, they do not provide a good model of the distribution of a variable that has a finite probability of being equal to zero. Nevertheless, the gamma pdf could be adapted for such variables (*e.g.* rain-rate) by using a pdf that is a linear sum of a Dirac delta function at zero and a gamma pdf.

2.2 GIGG analysis ensemble generation for GIG case

To generate an analysis ensemble whose density is consistent with (4) from a forecast ensemble that is drawn from a gamma distribution, we determine the posterior sample mean $\overline{y_j^a}$ using

$$\frac{1}{\overline{y_j^a}} = \frac{1}{\overline{y_j^f}} + \frac{\tilde{P}_j^r}{\tilde{R}_j^r + \tilde{P}_j^r}\left[\frac{1}{y_j^o} - \left(\tilde{R}_j^r + 1\right)\frac{1}{\overline{y_j^f}}\right] \qquad (6)$$

where the overbar indicates the ensemble mean. The update of the ensemble perturbations is a little less straight-forward.

In the perturbed observations approach for generating a posterior ensemble (Burgers et al., 1998), one uses a stochastic form of the Kalman equation for updating the

mean. It is made stochastic by replacing the prior mean by an individual member of the prior ensemble and by replacing the observation by a perturbed observation. The perturbed observation is created by simply adding a random draw from a Gaussian distribution whose mean is equal to the actual observation and whose variance is equal to the observation error variance. The result is a stochastic equation that produces an ensemble whose density would satisfy Bayes' theorem in the special case of a Gaussian prior, a Gaussian observation-likelihood and a very large ensemble size.

Here, we give a corresponding stochastic equation for the case of a gamma prior and an inverse-gamma observation-likelihood. Using (i) trial and error together with the constraint that the type 1 relative variance of the posterior ensemble must satisfy the first line of eq. (5), and (ii) the fact that the density function derived from the posterior ensemble should closely approximate the analytical posterior pdf, the stochastic equation

$$\frac{y_{ji}^{a}-\overline{y_{j}^{a}}}{\overline{y_{j}^{a}}}=\frac{\left(y_{ji}^{f}-\overline{y_{j}^{f}}\right)}{\sqrt{\left(\overline{y_{j}^{f}}\right)^{2}+\mathrm{var}\left(y_{ji}^{f}\right)}} +\tilde{P}_{j}^{r}\left(\tilde{P}_{j}^{r}+\tilde{R}_{j}^{r}\right)^{-1}\left[\frac{\left(y_{ji}^{gig}-\left\langle y_{j}^{gig}\right\rangle\right)}{\sqrt{\left\langle y_{j}^{gig}\right\rangle^{2}-2\,\mathrm{var}\left(y_{ji}^{gig}\right)}}-\frac{\left(y_{ji}^{f}-\overline{y_{j}^{f}}\right)}{\sqrt{\left(\overline{y_{j}^{f}}\right)^{2}+\mathrm{var}\left(y_{ji}^{f}\right)}}\right] \quad (7)$$

was developed. In (7), y_{ji}^{gig} is the counterpart of the perturbed observations used in EnKFs but it is not a Gaussian variable. Experimentation showed that in order for (7) to produce an ensemble with the correct posterior density, y_{ji}^{gig} had to be a random sample

from the pdf that is directly proportional to the function of y_j obtained when y_j^o is fixed at its observed value in the observation-likelihood pdf given by (2). Appendix B shows that this pdf is the gamma pdf with type 1 relative variance R_j^{gig} and mean $\left\langle y_j^{gig} \right\rangle$ given by $\left(R_j^{gig}\right)^{-1} = \left[\left(\tilde{R}_j^r\right)^{-1} + 2\right] = \left[\left(R_j^r\right)^{-1} + 3\right]$ and $\left\langle y_j^{gig} \right\rangle = \dfrac{\left[\left(\tilde{R}_j^r\right)^{-1} + 2\right]}{\left(\tilde{R}_j^r\right)^{-1}} y_j^o = \dfrac{\left(R_j^{gig}\right)^{-1}}{\left(R_j^{gig}\right)^{-1} - 2} y_j^o$, (8)

respectively. Since $\mathrm{var}\left(y_{ji}^L\right) = R_j^{gig} \left\langle y_j^{gig} \right\rangle^2$ and $\left(R_j^{gig}\right)^{-1} = \left[\left(\tilde{R}_j^r\right)^{-1} + 2\right]$, it follows that

$$\tilde{R}_j^r = \frac{1}{\left(R_j^{gig}\right)^{-1} - 2} = \frac{R_j^{gig}}{1 - 2R_j^{gig}} = \frac{\left\langle y_j^{gig} \right\rangle^2 R_j^{gig}}{\left\langle y_j^{gig} \right\rangle^2 - 2R_j^{gig} \left\langle y_j^{gig} \right\rangle^2} = \frac{\mathrm{var}\left(y_{ji}^{gig}\right)}{\left\langle y_j^{gig} \right\rangle^2 - 2\,\mathrm{var}\left(y_{ji}^{gig}\right)} \qquad (9)$$

and hence

$$\left\langle \left[\frac{\left(y_{ji}^{gig} - \left\langle y_j^{gig} \right\rangle\right)}{\sqrt{\left\langle y_j^{gig} \right\rangle^2 - 2\,\mathrm{var}\left(y_{ji}^{gig}\right)}} \right]^2 \right\rangle = \tilde{R}_j^r . \qquad (10)$$

As shown in Appendix C, equation (10) ensures that the analysis perturbations generated by (7) are consistent with the theoretically determined posterior relative error variance given by the first line of (5). Specifically, Appendix C shows that

$$\left\langle \left(\frac{y_{ji}^a - \overline{y_j^a}}{\overline{y_j^a}} \right)^2 \right\rangle = \tilde{P}_j^r - \tilde{P}_j^r \left(\tilde{P}_j^r + \tilde{R}_j^r\right)^{-1} \tilde{P}_j^r = \Pi_j^r . \qquad (11)$$

Although, we have only been able to analytically prove that (6) and (7) give the correct 1st and 2nd moments of the posterior distribution, in Section 4 we show that, for large

ensembles, (6) and (7) yield ensembles whose associated density is very close to the true posterior pdf .

We conclude this section by making the somewhat obvious observation that individual ensemble members can be derived from (7) using the fact that

$$y_{ji}^{a} = \left[\frac{y_{ji}^{a} - \overline{y_{j}^{a}}}{\overline{y_{j}^{a}}} \right] \overline{y_{j}^{a}} + \overline{y_{j}^{a}} \,. \tag{12}$$

3. Univariate IGG (inverse-gamma prior and gamma observation likelihood)

3.1 Analytic solution

Now consider the case where the prior distribution of true values y_j of the observed variable is given by an inverse-γ pdf of the form

$$\rho_{prior}\left(y_j\right) = \frac{\left[\left(\tilde{P}_j^r\right)^{-1}\left\langle y_j^f \right\rangle\right]^{\left[\left(\tilde{P}_j^r\right)^{-1}+1\right]}}{\Gamma\left(\left[\left(\tilde{P}_j^r\right)^{-1}+1\right]\right)} y_j^{-\left[\left(\tilde{P}_j^r\right)^{-1}+2\right]} \exp\left(-\left(\tilde{P}_j^r\right)^{-1}\frac{\left\langle y_j^f \right\rangle}{y_j}\right) \tag{13}$$

The scale parameter β and shape parameter α typically associated with inverse-gamma functions are related to the prior type 2 relative error variance $\tilde{P}_j^r$ and prior mean $\left\langle y_j^f \right\rangle$ via the equations

$$\alpha = \left[\left(\tilde{P}_j^r\right)^{-1}+1\right] \text{ and } \beta = (\alpha-1)\left\langle y_j^f \right\rangle = \left(\tilde{P}_j^r\right)^{-1}\left\langle y_j^f \right\rangle \tag{14}$$

Now suppose that the observation likelihood pdf $L\left(y_j^o \mid y_j\right)$ is given by the gamma pdf

$$L\left(y_j^o \mid y_j\right)=\left(\frac{1}{\Gamma\left[\left(R_j^r\right)^{-1}\right]}\left(R_j^r\right)^{-\left(R_j^r\right)^{-1}}\left(y_j^o\right)^{\left(\left(R_j^r\right)^{-1}-1\right)}\right) y_j^{-\left(R_j^r\right)^{-1}} \exp\left(-\left(R_j^r\right)^{-1}\frac{y_j^o}{y_j}\right) \quad (15)$$

where $R_j^r=\dfrac{\mathrm{var}\left(y_o^j\right)}{\left(y_j\right)^2}=\dfrac{\mathrm{var}\left(\varepsilon^o\right)}{\left(y_j\right)^2}$ is the type 1 relative observation error variance. In this case the shape parameter k and scale parameter θ normally used to describe gamma pdfs are related to y_j and R_j^r through the relations $k=\left(R_j^r\right)^{-1}$ and $\theta=y_j R_j^r$. Note that for a fixed true value y_j of the observed variable, the mean of the pdf of observations given by (15) satisfies $\left\langle y_j^o\right\rangle=y_j$ and hence equation (15) describes the stochastic errors of an observing instrument that is unbiased.

In Appendix D, we show that the posterior pdf given an inverse-gamma prior and gamma observation likelihood is itself an inverse gamma pdf of the form

$$\rho_{\text{post}}\left(y_j \mid y_j^o\right)=\frac{\left[\left(\tilde{\Pi}_j^r\right)^{-1}\left\langle y_j^a\right\rangle\right]^{\left[\left(\tilde{\Pi}_j^r\right)^{-1}+1\right]}}{\Gamma\left(\left[\left(\tilde{\Pi}_j^r\right)^{-1}+1\right]\right)}\left(y_j\right)^{-\left(\left(\tilde{\Pi}_j^r\right)^{-1}+2\right)}\exp\left[-\left(\tilde{\Pi}_j^r\right)^{-1}\frac{\left\langle y_j^a\right\rangle}{y_j}\right] \quad (16)$$

where the posterior type 2 relative error variance $\tilde{\Pi}_j^r$ is given by

$$\tilde{\Pi}_j^r=\tilde{P}_j^r-\tilde{P}_j^r\left(\tilde{P}_j^r+R_j^r\right)^{-1}\tilde{P}_j^r \quad (17)$$

and where

$$\left\langle y_j^a\right\rangle=\left\langle y_j^f\right\rangle+\left(\frac{\tilde{P}_j^r}{\tilde{P}_j^r+R_j^r}\right)\left(y_j^o-\left\langle y_j^f\right\rangle\right). \quad (18)$$

Note that simply replacing the prior type 2 forecast error variance $\tilde{P}_j^r$ by the forecast error variance P_j and the type 1 relative observation error variance R_j^r by the observation error variance R_j in (17) and (18) would give the equations for the posterior error variance and mean for the case where the prior and observation likelihood distributions were both Gaussian. Hence, in contrast to the GIG case where the equations for the posterior mean is very different to that in the Gaussian case, in the IGG case, the equation for the posterior mean is isomorphic to the Gaussian case. Furthermore, unlike in the GIG case, as the observation error variance goes to infinity, the posterior mean becomes identical to the prior mean.

The similarity with the Gaussian case is further revealed by noting the average of the observation error variance over the entire prior is given by

$$\left\langle \mathrm{var}\left(\varepsilon_j^o\right)\right\rangle = \left\langle R_j^r \left(y_j\right)^2 \right\rangle = R_j^r \left\langle \left(y_j\right)^2 \right\rangle = R_j \qquad (19)$$

where R_j is the average of $\mathrm{var}\left(\varepsilon_j^o\right)$ over all possible values of the truth. Since the variance of the prior is given by $P_j = \tilde{P}_j^r \left\langle \left(y_j\right)^2 \right\rangle$, we can multiply the numerator and the denominator of the coefficient of $\left(y_j^o - \left\langle y_j^f \right\rangle\right)$ in (18) by $\left\langle \left(y_j\right)^2 \right\rangle$ to obtain

$$\left\langle y_j^a \right\rangle = \left\langle y_j^f \right\rangle + \frac{P_j}{P_j + R_j}\left(y_j^o - \left\langle y_j^f \right\rangle\right) \qquad (20)$$

Thus, for the analysis of a single observed variable, the IGG posterior mean is identical to the Gaussian posterior mean.

In deriving (20), we showed that $\tilde{P}_j^r\left(\tilde{P}_j^r + R_j^r\right)^{-1} = P_j^r\left(P_j^r + R_j^r\right)^{-1}$. Hence, right multiplying (17) by $\left(y_j^f\right)^2$ and then taking the expectation over the prior pdf gives

$$\tilde{\Pi}_j^r\left\langle\left(y_j^f\right)^2\right\rangle = P_j - P_j\left(P_j + R_j\right)^{-1}P_j. \qquad (21)$$

Since the right hand side of this equation is the formula for the posterior variance in the Gaussian case, (21) quantifies the relationship between the posterior Gaussian variance and the type 2 IGG posterior relative variance $\tilde{\Pi}_j^r$. The IGG posterior variance is obtained using the definition of the type 2 variance. Specifically,

$$\mathrm{var}\left(y_j^a\right) = \tilde{\Pi}_j^r\left(\left\langle y_j^a\right\rangle^2 + \mathrm{var}\left(y_j^a\right)\right) \Rightarrow \mathrm{var}\left(y_j^a\right) = \frac{\tilde{\Pi}_j^r}{\left(1-\tilde{\Pi}_j^r\right)}\left\langle y_j^a\right\rangle^2. \qquad (22)$$

Hence, as in the GIG case, the posterior IGG variance is proportional to the square of the analysis mean which, in turn is proportional to the specific value of the observation. In contrast, the Gaussian posterior variance is independent of the value of the observation. Thus, if a 2nd observation were to be assimilated using the posterior ensemble from the assimilation of the 1st observation as the prior for the assimilation of the 2nd observation, the IGG posterior mean would not be equal to the Gaussian posterior mean.

3.2 GIGG analysis ensemble generation for IGG case

From (20), the posterior ensemble mean $\overline{y_j^a}$ is obtained from the prior ensemble sample mean $\overline{y_j^f}$ using

$$\overline{y_j^a} = \overline{y_j^f} + \frac{\tilde{P}_j^r}{\tilde{P}_j^r + R_j^r}\left(y_j^o - \overline{y_j^f}\right). \tag{23}$$

Using (i) trial and error trial and error together with the constraint that the type 2 relative variance of the posterior ensemble must satisfy eq. (17), and (ii) the fact that the density function derived from the posterior ensemble should closely approximate the analytical posterior pdf, it was determined that good results could be obtained from,

$$\begin{aligned}\tilde{y}_{ji}^{a\,\prime} &= \frac{y_{ji}^a - \overline{y_j^a}}{\sqrt{\left(\overline{y_j^a}\right)^2 + \mathrm{var}\left(y_{ji}^a\right)}} = \frac{\left(y_{ji}^f - \overline{y_j^f}\right)}{\sqrt{\left(\overline{y_j^f}\right)^2 + \mathrm{var}\left(y_{ji}^f\right)}} \\ &+ \tilde{P}_j^r\left(\tilde{P}_j^r + R_j^r\right)^{-1}\left[\frac{\left(y_{ji}^{igg} - \left\langle y_j^{igg}\right\rangle\right)}{\sqrt{\left\langle y_j^{igg}\right\rangle^2 - \mathrm{var}\left(y_{ji}^{igg}\right)}} - \frac{\left(y_{ji}^f - \overline{y_j^f}\right)}{\sqrt{\left(\overline{y_j^f}\right)^2 + \mathrm{var}\left(y_{ji}^f\right)}}\right]\end{aligned} \tag{24}$$

where y_{ji}^{igg} is the i^{th} random sample from an inverse-gamma pdf whose characteristics are discussed below and where $\tilde{y}_{ji}^{a\,\prime}$ is the i^{th} analysis perturbation about the mean normalized by $\sqrt{\left(\overline{y_j^a}\right)^2 + \mathrm{var}\left(y_{ji}^a\right)}$. The tick symbol $'$ is to help mark the variable as a perturbation about the mean. For the previously considered GIG case, we let the perturbed observations be random draws from a pdf that was directly proportional to the probability density of obtaining the observed value y_j^o as a likelihood function of the true value y_j. This approach did not produce good results in the IGG case. Instead, accurate results

were obtained when we let y_{ji}^{igg} be a random sample from an inverse-gamma distribution with variance $\tilde{R}_j^{igg}$ and mean $\left\langle y_j^{igg} \right\rangle$ given by

$$\left(\tilde{R}_j^{igg}\right)^{-1} = \left[\left(R_j^r\right)^{-1} + 2\right] \text{ and } \left\langle y_j^{igg} \right\rangle = y_j^o, \qquad (25)$$

respectively. Note that (25) and the fact that $\left(R_j^r\right)^{-1} = k$ imply that the shape and scale parameters of the inverse-gamma function used to generate the perturbed observations are given by $\alpha_{igg} = k + 3$ and $\beta_{igg} = y_j^o\left(\alpha_{igg} - 1\right)$.

To understand the $\sqrt{\left\langle y_j^{igg} \right\rangle^2 - \text{var}\left(y_{ji}^{igg}\right)}$ term in (24) note that since, by definition, $\left(\tilde{R}_j^{igg}\right)^{-1} = \left[\left\langle y_j^{igg} \right\rangle^2 + \text{var}\left(y_{ji}^{igg}\right)\right] \Big/ \text{var}\left(y_{ji}^{igg}\right)$, rearranging the first equation in (25)_gives

$$R_j^r = \frac{1}{\left(\tilde{R}_j^{igg}\right)^{-1} - 2} = \frac{\text{var}\left(y_{ji}^{igg}\right)}{\left[\left\langle y_j^{igg} \right\rangle^2 + \text{var}\left(y_{ji}^{igg}\right)\right] - 2\,\text{var}\left(y_{ji}^{igg}\right)} = \frac{\text{var}\left(y_{ji}^{igg}\right)}{\left\langle y_j^{igg} \right\rangle^2 - \text{var}\left(y_{ji}^{igg}\right)} \qquad (26)$$

and hence

$$\left\langle \left[\frac{\left(y_{ji}^{igg} - \left\langle y_j^{igg} \right\rangle\right)}{\sqrt{\left\langle y_j^{igg} \right\rangle^2 - \text{var}\left(y_{ji}^{igg}\right)}} \right]^2 \right\rangle = R_j^r. \qquad (27)$$

Following an argument similar to that given in Appendix C, it is readily shown that property (27) ensures that the analysis perturbations generated by (23)-(25) are consistent with the true posterior variance given by (17) because

$$\mathrm{var}\left(\tilde{y}_{ji}^{a}{}'\right)=\left\langle\left(\frac{y_{ji}^{a}-\overline{y_{j}^{a}}}{\sqrt{\left(\overline{y_{j}^{a}}\right)^{2}+\mathrm{var}\left(y_{ji}^{a}\right)}}\right)^{2}\right\rangle=\tilde{P}_{j}^{r}-\tilde{P}_{j}^{r}\left(\tilde{P}_{j}^{r}+R_{j}^{r}\right)^{-1}\tilde{P}_{j}^{r}=\tilde{\Pi}_{j}^{r}. \qquad (28)$$

Although, we have only been able to analytically prove that (23)-(25) give the correct 1st and 2nd moments of the posterior distribution, in Section 4 we show that, for large ensembles, (23) and (24) yield ensembles whose associated pdf is very close to all of the moments of the true posterior pdf.

To obtain the posterior GIGG perturbations corresponding to the normalized perturbations given by (24), note that since $\tilde{y}_{ji}^{a}{}'=\dfrac{y_{ji}^{a}-\overline{y_{j}^{a}}}{\sqrt{\left(\overline{y_{j}^{a}}\right)^{2}+\mathrm{var}\left(y_{ji}^{a}\right)}}$, it follows that

$$\mathrm{var}\left(\tilde{y}_{ji}^{a}{}'\right)=\frac{\mathrm{var}\left(y_{ji}^{a}\right)}{\left(\overline{y_{j}^{a}}\right)^{2}+\mathrm{var}\left(y_{ji}^{a}\right)}\Rightarrow\mathrm{var}\left(y_{ji}^{a}\right)=\mathrm{var}\left(\tilde{y}_{ji}^{a}{}'\right)\left[\left(\overline{y_{j}^{a}}\right)^{2}+\mathrm{var}\left(y_{ji}^{a}\right)\right]$$

$$\Rightarrow\mathrm{var}\left(y_{ji}^{a}\right)=\left(\overline{y_{j}^{a}}\right)^{2}\frac{\mathrm{var}\left(\tilde{y}_{ji}^{a}{}'\right)}{1-\mathrm{var}\left(\tilde{y}_{ji}^{a}{}'\right)}\approx\left(\overline{y_{j}^{a}}\right)^{2}\frac{\frac{1}{K}\sum_{i=1}^{K}\left(\tilde{y}_{ji}^{a}{}'\right)^{2}}{1-\frac{1}{K}\sum_{i=1}^{K}\left(\tilde{y}_{ji}^{a}{}'\right)^{2}}. \qquad (29)$$

With this value of $\mathrm{var}\left(y_{ji}^{a}\right)$, the individual posterior ensemble members y_{ji}^{a} are obtained from the value of $\overline{y_{j}^{a}}$ given by (23) and the value of $\tilde{y}_{ji}^{a}{}'$ from (24) using

$$y_{ji}^{a}=\overline{y_{j}^{a}}+\tilde{y}_{ji}^{a}{}'\sqrt{\left(\overline{y_{j}^{a}}\right)^{2}+\mathrm{var}\left(y_{ji}^{a}\right)}. \qquad (30)$$

4. Test of univariate theory for GIGG

The above algebraic analyses only prove that the GIGG-EnKF returns a random sample ensemble with the same mean and variance as the true posterior distributions obtained from Bayes' theorem. To check the accuracy of the higher moments of the pdf that the GIGG samples, here we compare posterior pdfs empirically generated from large K=10^7 ensembles using GIGG with the corresponding true posterior pdfs given by Bayes' theorem. This very large ensemble size is used to minimize the sampling error in ensemble derived pdfs. For consistency, in the GIG (IGG) case, we randomly draw the prior K-member ensemble from a gamma (inverse-gamma) distribution and the pdf of observations given a truth is always assumed to be inverse-gamma (gamma).

To begin, consider the case where the prior distribution has a mean of unity and type 1 relative forecast error variance equal to 1 ($P_j^r = 1$) or equivalently, type 2 relative forecast error variance is equal to ½ $\left(\tilde{P}_j^r = 0.5\right)$. We compute the posterior pdfs for observed values of either 0.5 or 3 with Type 1 relative observation error variance of $R_j^r = 1/4$ (corresponding to Type 2 relative error variance $\tilde{R}_j^r = 1/5$).

To obtain the empirical posterior pdf from the 10^7 member ensemble, we divide the real line from y=0 to y=10 into 500 contiguous segments or bins of length 0.02 and (i) count how many posterior ensemble members fall into each bin, then (ii) divide this count by the total number of ensemble members $\left(K = 10^7\right)$ to obtain the frequency with which posterior ensemble members fell into each bin, and finally (iii) divided these frequencies by the segment lengths (0.02) to obtain the GIG and IGG ensemble estimates

of the posterior probability densities within each bin. The resulting GIGG ensemble based posterior pdf functions $\rho_{\text{ensemble}}\left(y_{ji} \mid y_j^o\right), i = 1, 2, ..., 500$ were then compared with the corresponding analytical values of the true posterior pdfs $\rho\left(y_{ji} \mid y_j^o\right)$ (obtained from (4) or (16) as appropriate). By plotting the ensemble based posterior pdfs as thick grey lines and then plotting the analytical pdfs as thin black lines, one can visually examine the extent to which the ensemble based pdf function $\rho_{\text{ensemble}}\left(y_{ji} \mid y_j^o\right)$ had the same shape as the true pdf $\rho\left(y_{ji} \mid y_j^o\right)$. To quantify the differences, we computed both the maximum difference (maxd) and the root mean square difference (rmsd) between the ensemble-based and analytical pdfs and then normalized these values by dividing them by the maximum value of the true posterior pdf. In equation form, these normalized measures are given by

$$\text{rmsd} = \frac{\sqrt{\frac{1}{500}\sum_{i=1}^{500}\left[\rho_{\text{ensemble}}\left(y_{ji} \mid y_j^o\right) - \rho\left(y_{ji} \mid y_j^o\right)\right]^2}}{\rho\left(y_j^{\text{mode}} \mid y_j^o\right)}, \text{ and}$$
$$\text{maxd} = \frac{\max\left|\rho_{\text{ensemble}}\left(y_{ji} \mid y_j^o\right) - \rho\left(y_{ji} \mid y_j^o\right)\right|}{\rho\left(y_j^{\text{mode}} \mid y_j^o\right)}. \qquad (31)$$

Figure 2 gives the results of these comparisons. The very close correspondence between the pdfs given by the GIG and IGG ensemble-based pdfs and the true posterior pdfs is illustrated by the fact that the lines depicting the analytical posterior pdf always lie within the boundaries of the thick grey lines marking the ensemble-based pdf. This

impression of very close correspondence is supported by the relatively small values of the of the maximum and root mean square differences (rmsd's) between the two curves given on each panel of Figure 2.

A key question is whether the GIGG would perform similarly well with other ratios of forecast to observation error variance. Currently, we have no analytic theory for the accuracy of the 3rd and higher moments of the GIGG pdfs and hence can only assess their accuracy via numerical experimentation. For the GIG case, we consider the range of relative variances P_j^r and $\tilde{R}_j^r$ values given by

$$\tilde{R}_j^r = 2^{-m} \text{ and } P_j^r = 2^{1-n}, m = 1,2,..,7 \text{ and } n = 1,2,...,7 \qquad (32)$$

For the observation distribution, this gives a range of type 2 (type 1) relative error variance ranging from 0.5 (1) to 1/128 (1/127). For the forecast distribution, it gives type 1 relative error variances ranging from 1, where the variance is the same size as the mean to 1/64 where the variance is a 64th of the mean. Figure panels 3a and 3c contours the "maxd" values as defined by (31) as a function of the exponent indices m (abscissa) and n (ordinate) as defined in equation (32). The worst case occurs for the $y_j^o = 3$ case (Figure 3c) when $P_j^r = R_j^r = 1$ (recall that $\tilde{R}_j^r = 0.5 => R_j^r = 1$) where a relative maximum error of 2.5% occurs. Figure 4a compares the analytical and ensemble pdfs for this "worst case". It shows that this error is still too small to separate the thick grey line depicting $\rho_{\text{ensemble}}\left(y_{ji} \mid y_j^o\right)$ from the thin black line depicting the true posterior pdf $\rho\left(y_{ji} \mid y_j^o\right)$.

For the IGG case, we consider the range of Type 1 relative variances P_j^r and R_j^r given by

$$R_j^r = 2^{1-m} \text{ and } P_j^r = 2^{1-n}, m = 1,2,..,7 \text{ and } n = 1,2,...,7 \quad (33)$$

For both forecast and observation distributions this gives relative error variances ranging from 1 to 1/64. Figure panels 3b and 3d contours the "maxd" values as defined by (31) as a function of the exponent indices m (abscissa) and n (ordinate) as defined in equation (33). (Note that the top right corner of 3d is white because a floating-point overflow error prevented the analytical solution $\rho\left(y_{ji} \mid y_j^o\right)$ from being computed for this pair of parameters and hence no difference could be computed for this point. Also, note that this area of parameter space is not particularly interesting because, in this area, the relative error variances are so small that the pdfs are well described by Gaussian pdfs).

The worst case occurs when both the forecast and observation relative error variances are the largest. Specifically, it occurs when both the type 1 relative observation and forecast error variances are equal to 1 for the $y_j^o = 3$ case. In this case, the relative maximum error is just over 10%.

Figure 4b compares the analytical and ensemble pdfs for this "worst IGG case". It shows that in this case, the error is just large enough to separate the thick grey curve depicting $\rho_{\text{ensemble}}\left(y_{ji} \mid y_j^o\right)$ from the thin black curve depicting the true posterior pdf $\rho\left(y_{ji} \mid y_j^o\right)$. The second worst IGG case occurs when the relative observation error

variance is reduced to 0.5. Figure 4c shows that in this case, the thin black curve $\rho\left(y_{ji} \mid y_j^o\right)$ again fails to separate completely separate from the thick grey curve $\rho_{\text{ensemble}}\left(y_{ji} \mid y_j^o\right)$.

For these worst case scenarios, we also varied the value of the observation to see if the error was sensitive to the observed value. These experiments showed that the error was a very weakly increasing function of the observed value and hence consistent with the hypothesis that Figure 4 gives a qualitatively correct measure of GIGG accuracy for all possible observed values.

In conclusion, the univariate GIGG-EnKF has been shown to generate a posterior ensemble whose pdf is extremely close to the true pdf for a wide range of forecast and observation uncertainty pdfs provided that these distributions come in the conjugate pairs associated with the GIG and IGG cases. Apart from the worst case for the IGG filter shown in figure 4b, by eye, it is difficult to distinguish the GIGG posterior pdfs from the true posterior pdf.

5. Extension to multi-variate high-dimensional systems and the Gaussian case

Anderson (2003, Section 2c) describes a generalized two stage multi-variate ensemble filter. The first stage involves using the best available method to find a posterior ensemble of state estimates of an observed variable given an observation of the variable and a prior ensemble state estimate of the variable. The second stage involves propagating this information to the rest of the extended model-plus-observed-variables

state vector using regression. The multi-variate GIGG-EnKF follows this overall format exactly, but for the sake of clarity we reiterate some of its details and its relationship to other EnKFs.

The ideal multi-variate system for the EnKF-GIGG is one in which the forecast and observation uncertainty associated with some observations matches the GIG case, other observations match the IGG case and yet other observations match the Gaussian assumptions of the well-established perturbed observation univariate EnKF single observation update equation given by

$$y_{ji}^{a} = y_{ji}^{f} + \frac{P_{jj}^{oo}}{P_{jj}^{oo} + R_j}\left(y_{ji}^{o} - y_{ji}^{f}\right), \qquad (34)$$

where the perturbed observation $y_{ji}^{o} = y_{j}^{o} + \varepsilon_{ji}^{o}$ and ε_{ji}^{o} is a random normal variable with mean zero and *non-relative* observation error variance R_j; P_{jj}^{oo} is the *non-relative* forecast error variance. The subscript j indicates the j^{th} observation, the subscript i indicates the i^{th} ensemble member. The superscript "oo" indicates that the covariance referred to by P is between two observed variables. (Later we will use the superscript *"mo"* indicate covariances between a model variable and an observed variable).

To see how the GIGG-EnKF can be extended to multivariate systems, it is helpful to recall the multivariate form of the EnKF. Consider the serial assimilation of p observations $y_j^o, j = 1, 2, ..., p$, where $x_{\mu i}^{f}, \mu = 1, 2, .., n;\ i = 1, 2, ..., K$ denotes the i^{th} member of a K-member ensemble forecast of the μ^{th} model variable and

$y_{ji}^f, j=1,2,..,p;\ i=1,2,...,K$ denotes the i^{th} member of the ensemble forecast of the j^{th} observation. Note that $y_{ji}^f = h_j\left(\mathbf{x}_i^f\right)$ where h_j is the operator that maps the full prior model state vector $\mathbf{x}_i^f$ of the ith ensemble member to the jth observation. Note that with these definitions $P_{jj}^{oo} = \text{var}\left(y_j^f\right) \approx \frac{\sum_{i=1}^{K}\left(y_{ji}^f - \overline{y_j^f}\right)^2}{K-1}$, similarly, the covariance of the error in the forecast of the kth and the jth observation is given by $P_{kj}^{oo} = \text{covar}\left(y_k^f, y_j^k\right) \approx \frac{\sum_{i=1}^{K}\left(y_{ki}^f - \overline{y_k^f}\right)\left(y_{ji}^f - \overline{y_j^f}\right)}{K-1}$ while the covariance of the error in the forecast of the μth model variable with the jth observation is given by $P_{\mu j}^{mo} = \text{covar}\left(x_\mu^f, y_j^k\right) \approx \frac{\sum_{i=1}^{K}\left(x_{\mu i}^f - \overline{x_\mu^f}\right)\left(y_{ji}^f - \overline{y_j^f}\right)}{K-1}$ where the superscript "*mo*" is to indicate that the covariance is between a model variable and an observed variable. With these definitions, the update of the state estimate of the kth observation from the assimilation of the j^{th} observation is given by

$$y_{ki}^a = y_{ki}^f + \frac{P_{kj}^{oo}}{P_{kj}^{oo} + R_j}\left(y_{ji}^o - y_{ji}^f\right), \text{ for } i=1,2,...,K \qquad (35)$$

Since (34) implies that

$$\frac{1}{P_{jj}^{oo} + R_j}\left(y_{ji}^o - y_{ji}^f\right) = \frac{1}{P_{jj}^{oo}}\left(y_{ji}^a - y_{ji}^f\right) = \frac{1}{\text{var}\left(y_j^f\right)}\left(y_{ji}^a - y_{ji}^f\right) \qquad (36)$$

it follows that for a single ensemble member, (35) can be rewritten in the form

$$y_{ki}^{a} = y_{ki}^{f} + \frac{P_{kj}^{oo}}{\mathrm{var}\left(y_{j}^{f}\right)}\left(y_{ji}^{a} - y_{ji}^{f}\right) = y_{ki}^{f} + \frac{\mathrm{covar}\left(y_{k}^{f}, y_{j}^{f}\right)}{\mathrm{var}\left(y_{j}^{f}\right)}\left(y_{ji}^{a} - y_{ji}^{f}\right) \qquad (37)$$

Similarly, since the EnKF update of the μ^{th} model state variable of the i^{th} ensemble member by the j^{th} observation is given by

$$\begin{aligned} x_{\mu i}^{a} &= x_{\mu i}^{f} + \frac{P_{\mu j}^{mo}}{P_{jj}^{oo} + R_{j}}\left(y_{ji}^{o} - y_{ji}^{f}\right) \\ &= x_{\mu i}^{f} + \frac{P_{\mu j}^{mo}}{\mathrm{var}\left(y_{j}^{f}\right)}\left(y_{ji}^{a} - y_{ji}^{f}\right) \text{ (where eq 36 has been used)} \\ &= x_{\mu i}^{f} + \frac{\mathrm{covar}\left(x_{\mu}^{f}, y_{j}^{f}\right)}{\mathrm{var}\left(y_{j}^{f}\right)}\left(y_{ji}^{a} - y_{ji}^{f}\right) \end{aligned} \qquad (38)$$

Equations (37) and (38)[1] serve to demonstrate that provided one has tools to accurately estimate the posterior ensemble analysis $y_{ji}^{a}, i = 1, 2, .., K$ for a single observation then one can turn this into a posterior ensemble analysis of the state of every model variable and every observation whose prior covaries with it. If the y_{ji}^{a} used in (37) and (38) are obtained using the EnKF equation (34), then (37) and (38) deliver the EnKF update. However, if the observation-likelihood and prior uncertainty pdfs do not satisfy the Gaussian assumptions of the EnKF then the y_{ji}^{a} provided by the EnKF will grossly misrepresent the true posterior distribution and (37) and (38) will propagate this mis-representation to the extended model-observation state vector. In contrast, if the non-Gaussian observation-likelihood and prior uncertainty pdfs approximately satisfy the

[1] Note that (37) and (38) express the same idea as eq's (3.6)-(3.8) of Anderson and Collins (2007)

assumptions of the GIG (IGG) equations then using the GIG (IGG) equations to obtain then y_{ji}^{a} provides an accurate representation of the true posterior distribution and (37) and (38) propagate this accurate information to the extended model-observation state vector. Hence, given knowledge of the nature of the observation-likelihood and prior pdfs, one ought to be able to improve on the performance of the EnKF by allowing the GIG or IGG equation sets to generate y_{ji}^{a} for those observations where the GIG or IGG pdf assumptions are more accurate than the Gaussian pdf assumptions. The GIGG-EnKF algorithm implements this idea and it is concisely summarized by the following pseudo-code which describes how to update an extended model-observation state vector comprising *n* model state variables and *p* observations:

for $j = 1:p$

Step 1: Decide whether forecast and observation uncertainty associated with y_j^o is best approximated by GIG, IGG or Gaussian assumptions.

Step 2: if (GIG) then use (6), (7) and (12) to obtain $y_{ji}^a,\ i = 1,2,...,K$;
else if (IGG) then use (23), (24) and (30) to obtain $y_{ji}^a,\ i = 1,2,...,K$;
else if (Gaussian) then use (34) to obtain $y_{ji}^a,\ i = 1,2,...,K$;

Step 3: Find corresponding analysis ensemble for observations and model variables

$$y_{ki}^a = y_{ki}^f + \frac{\text{covar}\left(y_k^f, y_j^f\right)}{\text{var}\left(y_j^f\right)}\left(y_{ji}^a - y_{ji}^f\right), \text{ for } k = 1,2,...,p;\ i = 1,2,...,K$$

$$x_{\mu i}^a = x_{\mu i}^f + \frac{\text{covar}\left(x_\mu^f, y_j^f\right)}{\text{var}\left(y_j^f\right)}\left(y_{ji}^a - y_{ji}^f\right), \text{ for } \mu = 1,2,...,n;\ i = 1,2,...,K$$

Step 4: Let the analysis ensemble be the prior ensemble for the next observation

$$y_{ki}^f = y_{ki}^a, \text{ for } k = 1,2,...,p;\ i = 1,2,...,K \qquad (39)$$

$$x_{\mu i}^f = x_{\mu i}^a, \text{ for } \mu = 1,2,...,n;\ i = 1,2,...,K$$

end

Note that steps 1 and 2 amount to the 1st stage of Anderson's (2003) generalized two stage ensemble filter, while step 3 is Anderson's 2nd stage linear regression step. While accurate with linear dynamics and Gaussian uncertainty pdfs, this linear regression step is likely to be imperfect when the dynamics are non-linear and uncertainty pdfs are non-Gaussian. Future research will consider the impact of replacing this linear regression step with some sort of higher order regression step. Also as in Anderson (2003), step 4 of the GIGG-EnKF turns the posterior multivariate analysis ensemble obtained after the

assimilation of a single observation into the prior ensemble for the assimilation of the next observation. Thus, the perturbed observations form of the generalized two-stage ensemble filter described in Anderson (2003) is almost identical to the GIGG-EnKF algorithm. The only required code changes is to include step 1 and step 2 in stage 1 of this filter. Anderson and collaborators (Anderson et al., 2007) have made their generalized two-stage ensemble filter freely available to the research community as part of the Data Assimilation Research Testbed (DART).

If accurate, the linear regression used in both Anderson's (2003) generalized two-stage ensemble filter solves Bayes' theorem when all uncertainty pdfs can be described in terms of Gaussians

If one chose to assimilate all the observations as Gaussian variables (39) would yield the serial observation processing form of the perturbed observation EnKF described in Burgers et al. (1998) – or equivalently, the single ensemble, single observation processing form of the EnKF described in Houtekamer and Mitchell (2001). A difference between the GIGG-EnKF formulation and these EnKFs is that these EnKFs use the update equation in the form given by (35) and the first line of (38) whereas the GIGG-EnKF uses the forms given by (37) and the last line of (38). Nevertheless, the code changes that would be required to turn these EnKFs into a GIGG-ENKF seem fairly minor.

Note that it would be straight forward to apply a localization function to the covariance terms in (39) to suppress spurious ensemble correlations.

With all observations as Gaussian variables and no covariance localization, the analyses produced by (39) would be independent of the order in which observations were assimilated (e.g. Bishop et al., 2015). A quantification of the extent to which such "independence of the order" is maintained when some of the observations are assimilated with either the GIG or IGG equations is beyond the scope of this paper and is left for future work.

The issue of how one should decide whether to assimilate an observation using the GIG, IGG or Gaussian update equations is worthy of discussion. We will consider two cases: one in which the observation likelihoods are clearly and distinguishably gamma, inverse-gamma or Gaussian and one in which either gamma or inverse-gamma could adequately model the observation likelihood pdf.

Suppose that one could separate the observations into groups with clearly distinguished inverse-gamma, gamma and Gaussian observation likelihood distributions. Choosing to assimilate each group of these observations using the GIG, IGG and Gaussian equations, respectively, regardless of the prior pdf of the state estimate of the observation would have appealing consequences. First, this approach would deliver the true posterior distribution when the prior pdfs were gamma, inverse-gamma and Gaussian pdfs, respectively; second, when the prior distributions were of some other form, it would nevertheless transmit qualitatively correct information about the nature of the observational uncertainty to the posterior ensemble.

For those observation types for which both inverse-gamma pdfs and gamma pdfs adequately describe an inaccurately known observation uncertainty, the choice could be based on whether the prior ensemble state estimate of the observation best approximated a gamma or inverse-gamma pdf. A quantitative approach for doing this would be to fit both a gamma pdf and an inverse-gamma pdf to the ensemble distribution and then choose the pdf that assigns a higher likelihood to the ensemble distribution that actually occurred. For such a test, a consistent method to fit the gamma and inverse-gamma pdfs to the prior ensemble would be the maximum likelihood method (e.g. Choi and Wette, 1969).

A less accurate but likely more computationally efficient pdf fitting approach could be based on the fact that while the square of the skewness of the gamma distribution is four times as large as its type 1 relative error variance, the square of the skewness of the inverse-gamma distribution is more than sixteen times its type 1 relative error variance. Hence, a method based on using the sample skewness of the ensemble to quickly predict the results of a test based on the more rigorous maximum likelihood method might also be useful. Detailed analyses of such methods for optimizing the choice between GIG, IGG and Gaussian update equations are beyond the scope of this paper and left for future work.

Because serial assimilation is known to be identical to all-at-once assimilation for *any* series of observations that are assimilated using the Gaussian uncertainty assumptions, one could choose to group together all the observations likely to satisfy the Gaussian

assumptions and then assimilate them all-at-once or in batches using ones preferred flavor of perturbed observations ensemble data assimilation: EnKF, 4D-variational or otherwise. This "in-batches" or "all-at-once" Gaussian assimilation step could be performed either before, in-between, or after the serial assimilation of the non-Gaussian observation types with the GIGG-EnKF. Experimentation or analytical progress beyond the scope of this paper will be required to determine the best approach.

6. Test of multi-variate GIGG-EnKF

a. Experimental set-up

The multi-variate GIGG filter was tested with a system capable of generating random multi-variate prior distributions on a 96 grid point periodic domain that feature a high degree of inter-variable correlations and three variables on each grid point. The three variables are (i) the zonal wind *u,* whose distribution is very close to Gaussian (ii) the square of the zonal wind u^2 whose distribution is approximately gamma when the prior mean of *u* is near zero, and (iii) a pseudo-dust concentration variable given by

$$dust = \frac{1}{100}\left(20 + u^2\right)^2 \quad (10^{-6} g / m^3) . \qquad (40)$$

At the risk of offending atmospheric dust and aerosol modelers, the 4th order polynomial on the right hand-side is labelled "*dust*" - primarily for simplicity of labeling but also because atmospheric dust uptake is sometimes modelled as being proportional to the 4th power of the near surface wind. The appeal of (40) for this study is that it always assigns

zero probability density to zero concentrations (like an inverse-gamma pdf) and that it generates a pdf with a very high degree of skewness when the prior mean of the u distribution is equal to zero. (Recall, that inverse-gamma distributions have a higher skewness than gamma distributions for the same relative error variance.) As far as the pdf of error prone observations given truth is concerned, we assume that these pdfs are given by Gaussian, inverse-gamma and gamma distributions for the u, u^2 and *dust* variables, respectively.

The *dust* and u^2 variables are simply observable non-linear functions of u and lack the usual independence of the model variables of a Numerical Weather Prediction (NWP) model. Nevertheless, labeling them as model variables provides a simple framework for understanding aspects of GIGG-EnKF performance in the presence of prior inter-variable correlations between variables with approximately, Gaussian, gamma and inverse-gamma priors.

To create random prior mean quasi-Gaussian zonal wind fields, a distribution was sampled with mean zero and covariance matrix depicted in Fig 5a. Note that the standard deviation associated with this covariance matrix is 10.5 ms^{-1}. The thick green line on the top panel of Fig. 6 gives an example of such a randomly obtained ensemble mean $\bar{\mathbf{u}}$. To obtain quasi-Gaussian perturbations around this mean, we randomly sampled an inverse gamma distribution with shifted mean $\bar{\mathbf{u}}+10^3$ and a very small type 1 relative error

variance[2] of 10^{-6}. The spatial covariance of the distribution used for the ensemble perturbations is shown in Fig 5b. Note that because the variance of the prior wind perturbations is given by the product of the type 1 relative variance and the mean of the distribution squared, the variance of wind perturbations is given by $\left(\bar{\mathbf{u}}+10^3\right)^2 10^{-6} \approx 1$. Consequently, Fig 5b shows a slight variation in the magnitude of the variance about unity that is weakly proportional to $\bar{\mathbf{u}}$ which, in this specific case, is given by the green line on the top panel of Fig. 6. Once a K member ensemble has been obtained in this way, the fields are "de-shifted" by subtracting the original shift of 10^3 from each of them. These perturbations were then added to $\bar{\mathbf{u}}$ to obtain the prior ensemble of zonal wind. Each grey line in the top panel of Fig. 6 depicts one of a prior K=10,000 member generated in this way. As in section 4, this very large ensemble size is used to minimize the sampling error in ensemble derived pdfs. For each member of the prior ensemble of zonal wind, a corresponding prior ensemble member of u^2 is obtained by squaring each element of u. Similarly, a corresponding *dust* ensemble member is obtained by applying (40) to each element of u. The grey lines in the middle and bottom panels of Figure 6 give the resulting prior ensemble of u^2 and *dust*, respectively.

One of the K members is selected to be the "true state". Contemporaneous error prone observations of each of the variables are indicated by the red dots on Fig. 6. These

[2] Note that the u field could have been produced by simply sampling a Gaussian distribution but the approach given here permitted tests (not reported on here) of treating Gaussian variables like gamma variables with small relative error variances.

observations are generated from this true state in differing ways depending on the observation type: for the zonal wind *u*, an independent random normal perturbation with mean zero and variance $R(u)=1\ (m/s)^2$ is added to the truth at each of the observation sites; for u^2 we randomly sample an inverse gamma distribution with type 1 relative error variance $R^r(u^2)=0.1$ and mean equal to the true value of u^2. To create the observations of the dust variable, we randomly sample a gamma distribution with mean equal to the true value and type 1 relative error variance equal to $R^r(\text{dust})=0.1$. In Fig. 6, the thick cyan line gives the true state of the system. .

As previously described, in the multivariate GIGG filter, observations are consecutively assimilated one after the other. Here, largely because of the assumed distribution of observation errors individual u, u^2 and *dust* observations were assimilated using the equations for the Gaussian, GIG and IGG cases. For the cases described in this paper, the order of assimilation was GIG observation assimilation followed by IGG observation assimilation followed by Gaussian observation assimilation. Although no rigorous assessment was made of the sensitivity of the results to the chosen order of assimilation, the few checks that were made revealed little sensitivity. To remove it as a possible source of performance differences, in this study we will only use ensemble sizes large enough to make ensemble covariance localization of negligible value. An example of the results of the assimilation procedure is the posterior distribution of states given by the black lines on Fig. 6.

To compare GIGG filter performance with EnKF performance, we also made state estimates by assimilating the same observations using the same prior ensemble and both the deterministic and perturbed observations form of Bishop et al.'s (2001) Ensemble Transform Kalman Filter (ETKF) given in Posselt and Bishop (2012). Since no ensemble covariance localization is used in our experiments, the perturbed observations form of the ETKF will give identical results to other perturbed observation EnKFs. The deterministic ETKF, will give identical posterior ensemble covariances to other deterministic EnKFs such as those of Anderson (2001) and Whitaker and Hamill (2002). However, the posterior perturbations may differ by an orthonormal rotation (Tippett et al., 2003).

EnKFs require an estimate of the overall observation error variance not the relative observation error variances used in specifying gamma and inverse-gamma pdfs. However, the observation error variance is a function of the *unknown* true state when the observation uncertainty is described by gamma or inverse gamma distributions. At the same time, the gain matrix for EnKFs is obtained by finding the gain that minimizes the mean square of analysis errors over all possible prior forecast and observation errors. Hence, it is appropriate to define the observation error variances to be used in the ETKF with the equations $R\left(u^2\right)=\overline{\left(u^2\right)^2 R^r\left(u^2\right)}=\left(\overline{\left(u^2\right)^2}\right)R^r\left(u^2\right)$ and, similarly, $R\left(dust\right)=\left(\overline{dust^2}\right)R^r\left(dust\right)$ where the overbar indicates that the average over the prior ensemble of forecasts of the specific variable being observed. Using these definitions,

Fig. 7 gives the posterior distributions obtained when the observations from Fig. 6 are assimilated using the perturbed observations ETKF. The perturbed observations for the ETKF were created for all variables using random Gaussian distributions with mean equal to the observed value and variance equal to $R(u)$, $R(u^2)$ or $R(dust)$ as appropriate.

Comparison of Fig's 6 and 7 indicates that the posterior mean values (thick yellow lines) produced by the GIGG and ETKF filters are somewhat similar. To more precisely measure the difference in the accuracy of the GIGG and ETKF analysis means, we use

$$u^a_{mse} = \frac{1}{n}\sum_{i=1}^{96}\left(\overline{u_i^a} - u_i^t\right)^2,$$

$$\left[\left(u^2\right)^a\right]^{rel}_{mse} = \frac{1}{n}\sum_{i=1}^{96}\left[\frac{\left(\overline{\left(u_i^2\right)^a} - \left(u_i^2\right)^t\right)}{\frac{1}{2}\left\{\overline{\left(u_i^2\right)^a} + \left(u_i^2\right)^t\right\}}\right]^2 \text{ and } \left[dust^a\right]^{rel}_{mse} = \frac{1}{n}\sum_{i=1}^{96}\left[\frac{\left(\overline{(dust)_i^a} - (dust)_i^t\right)}{\frac{1}{2}\left\{\overline{(dust)_i^a} + (dust)_i^t\right\}}\right]^2 \quad (41)$$

where the subscript i indicates the grid point at which the error is computed and the superscript t indicates the true value. Thus, (41) uses the mean square error to measure the error of the mean of the zonal wind analyses. For the u^2 and *dust* variables (41) normalizes the errors by the average of the analysis mean and the true value and then takes the square of this normalized error. This was done because the errors for these variables tend to be proportional to the magnitude of the variable being analyzed.

Comparison of the coverage of the black analysis lines in Fig's 6 and 7 indicates that the range of the posterior ETKF ensemble is substantially different to that of the GIGG filter. This is particularly noticeable near grid-point 50 in the *dust* field where it is clear that the range of ETKF posterior ensemble values of dust is significantly greater than that for the GIGG posterior ensemble. These range differences reflect the dissimilarity of the 2nd, 3rd and higher moments of the GIGG and ETKF posterior distributions. Such dissimilarities can have a profound effect on the accuracy of estimates of variables that are non-linear functions of the analysis ensemble. Examples of such non-linear functions include the electrical power produced by a wind turbine or the state of the atmosphere at some future time. Hence, as a measure of the inaccuracy in the higher moments of the analysis ensembles, we create non-linear functions u_i^f ,$\left(u_i^2\right)^f$ and $\left(dust\right)_i^f$ at the ith grid point from the analysis members at the ith grid point using the following simple non-linear relationships,

$$u_i^f = \left(u_i^a\right)^2, \left(u_i^2\right)^f = \left[\left(u_i^2\right)^a\right]^2 \text{ and } \left(dust\right)_i^f = \left[\left(dust\right)_i^a\right]^4 \qquad (42)$$

At the risk of offending the designers of numerical weather prediction models, and in spite of the fact that (42) would rapidly achieve floating point overflow if cycled, we will hereafter refer to the mappings in (42) as "forecasts" that are valid at some "forecast time" and will refer to the mappings in (42) as the forecast model.

The true state at the forecast time is obtained by applying the non-linear relationships given in (42) to the true values at the analysis time – (a perfect model

assumption). Denote the true values at the forecast time by $\left(u_i^f\right)^t$, $\left[\left(u_i^2\right)^f\right]^t$ and $\left(dust_i^f\right)^t$ for the u, u^2 and $dust$ variables, respectively. We then measure the inaccuracy of the non-linear forecasts using

$$\left[u^f\right]_{mse}^{rel} = \frac{1}{n}\sum_{i=1}^{96}\left[\frac{\left(\overline{u_i^f} - \left(u_i^f\right)^t\right)}{\frac{1}{2}\left\{\overline{u_i^f} + \left(u_i^f\right)^t\right\}}\right]^2, \left[\left(u^2\right)^f\right]_{mse}^{rel} = \frac{1}{n}\sum_{i=1}^{96}\left[\frac{\left(\overline{\left(u_i^2\right)^f} - \left(u_i^2\right)^t\right)}{\frac{1}{2}\left\{\overline{\left(u_i^2\right)^f} + \left(u_i^2\right)^t\right\}}\right]^2, \text{ and}$$
$$\left[dust^f\right]_{mse}^{rel} = \frac{1}{n}\sum_{i=1}^{96}\left[\frac{\left(\overline{\left(dust\right)_i^f} - \left[\left(dust\right)_i^f\right]^t\right)}{\frac{1}{2}\left\{\overline{\left(dust\right)_i^f} + \left[\left(dust\right)_i^f\right]^t\right\}}\right]^2 \quad (43)$$

Unlike (41), (43) normalizes the mean square error of the forecast of zonal wind. This is because the non-linear function used to make the forecast gives the zonal wind forecast errors a tendency to be proportional to the magnitude of the zonal wind.

It is of interest to compare the performance of the GIGG and ETKF filters over many independent trials in our idealized system. To this end 7x512 completely independent data assimilation trials were performed by repeating the random number generation process that led to the data assimilation problems depicted in Figures 6 and 7 using differing independent random number seeds. The performance of each method (using the same forecasts and observations) was then measured by taking averages of the accuracy measures given by (41) and (43) over each subset of 512 trials and also over the 7x512=3584 trials. Since the 10,000 ensemble members depicted in Fig's 6 and 7 would

be difficult to achieve at today's operational forecasting centers, we performed each of the 3584 trials using just K=250 ensemble members. This ensemble size is large enough to make the value of ensemble covariance localization negligible in our idealized system while small enough to be achievable by some present day meteorological forecasting centers. (Ensemble covariance localization is beyond the scope of this paper but it can be implemented in the GIGG-EnKF in the same way that it is implemented in Anderson's (2003) generalized two-stage ensemble filter.)

If any of the methods produced analysis ensemble members with negative values for the positive definite variables u^2 and *dust*, these negative values were set to zero after the analysis was completed and before the analysis ensemble was used to initialize non-linear forecasts. All of the methods produced negative values of u^2 and *dust* in some of the trials for some of the ensemble members. Since such negative values would not occur for any of the methods in a univariate case, the negative values are attributable to the linear regression used to update unobserved values.

To quantify the statistical significance of performance differences, we used two distinct approaches. First, since each subset of 512 trials is entirely independent of the other subsets, the chance of one of the methods producing lower errors in all 7 trials purely by chance is the same as obtaining seven heads from seven flips of a fair coin; i.e. it is equal to $1/2^7 = 1/128$. Hence, if it is found that one of the methods has lower errors than the other method in all 7 subsets of trials, one may reject the null hypothesis with

[100-(100/128)]%≈99% confidence[3]. Second, when comparing the performance of two methods, we compute the variance σ_{err}^2 of the domain averaged error difference over all 3584 trials. By the central limit theorem, the mean of the error difference over all 3584 trials should be normally distributed with a standard deviation equal to $\sqrt{\sigma_{err}^2/3584}$. Hence, if the average difference over all 3584 trials is different from zero by more than 3 standard deviations, one may again reject the null hypothesis with greater than 99% confidence.

To get a sense of the size of the error reduction (or increase) due to using GIGG instead of either the perturbed observation or deterministic forms of the ETKF relative to the size of the error, we use a percentage error reduction (per) measure which, for the example of a zonal wind forecast, is given by

$$\left[u^f\right]_{per} = 100\left\{\frac{\left\langle\left[u^f\right]_{mse}^{rel}(ETKF)\right\rangle - \left\langle\left[u^f\right]_{mse}^{rel}(GIGG)\right\rangle}{\frac{1}{2}\left[\left\langle\left[u^f\right]_{mse}^{rel}(ETKF)\right\rangle + \left\langle\left[u^f\right]_{mse}^{rel}(GIGG)\right\rangle\right]}\right\} \qquad (44)$$

where $\left\langle\left[u^f\right]_{mse}^{rel}(ETKF)\right\rangle$ gives the average $\left[u^f\right]_{mse}^{rel}$ incurred by a version of the ETKF over all 3584 trials and $\left\langle\left[u^f\right]_{mse}^{rel}(GIGG)\right\rangle$ is the same thing for the GIGG filter. Note that the measure $\left[u^f\right]_{per}$ is positive when GIGG produces smaller average errors than

[3] In the sense that if this rejection rule was used and there was no difference between the two methods, it would produce a false rejection in less than 1 out of 100 cases.

ETKF. Note also that the possible values of $\left[u^f\right]_{per}$ range between +200 and -200 as $\left\langle \left[u^f\right]_{mse}^{rel}(GIGG)\right\rangle$ ranges from zero to infinity. Also note that in the case that $\left\langle \left[u^f\right]_{mse}^{rel}(GIGG)\right\rangle = \frac{1}{3}\left\langle \left[u^f\right]_{mse}^{rel}(ETKF)\right\rangle$ then $\left[u^f\right]_{per} = 100$. The percentage error reduction over the 3584 trials was computed for u, u^2 and *dust* at the analysis time by replacing $\left[u^f\right]_{mse}^{rel}$ in (44) by u_{mse}^a, $\left[\left(u^2\right)^a\right]_{mse}^{rel}$ and $\left[dust^a\right]_{mse}^{rel}$, respectively, and at the forecast time for u^2 and *dust* by replacing $\left[u^f\right]_{mse}^{rel}$ in (44) by $\left[\left(u^2\right)^f\right]_{mse}^{rel}$ and $\left[dust^f\right]_{mse}^{rel}$, respectively.

7. Results

Figure 8 plots the percentage error reduction resulting from using the GIGG filter rather than the perturbed observations ETKF (Fig. 8a) or the deterministic ETKF (Fig. 8b). The percentage error reductions at the analysis time ranged from just 3.7% improvement for the *dust* variable (Fig. 8b) to 10.2% improvement for the u^2 variable (Fig. 8a). All of these improvements were statistically significant because (a) the GIGG filter produced smaller error in all 7 subsets of 512 trials for the comparison against the deterministic ETKF and also in the comparison against the perturbed observations ETKF, and (b) the lower 3σ bounds shown on Fig. 8 are all above zero.

Even though the GIGG-EnKF assimilates observations of zonal wind u in exactly the same way as the EnKF, the GIGG-EnKF analysis error variance for u is significantly smaller than the EnKF analysis error variance. This shows that the superior ability of the GIGG-EnKF to assimilate observations of u^2 and $dust$ translates into a superior analysis of the Gaussian variable u via the linear regression term that adjusts the u field after each u^2 and $dust$ observation is assimilated.

The percentage forecast error reduction due to using GIGG was much greater for the non-linear forecasts than for the analyses. It ranged from 122% for the *dust* variable (Fig. 8a) to 127% for the u^2 variable (Fig. 8b). These >100 % percentage error reductions imply that the normalized squared forecast error of the mean of the ensemble forecast made from the GIGG analysis ensemble is less than a 3rd of the corresponding normalized squared error from EnKFs.

As previously mentioned, the accuracy of the non-linear forecasts is sensitive to the accuracy of the 2nd and higher moments of the ensemble distribution of analyses. The large degree of superiority of the GIGG ensemble over both versions of ETKF ensembles after non-linear forecast mappings indicates that the GIGG's representation of these higher order moments is profoundly more accurate than that of either version of the ETKF.

To highlight the differences between the 2nd order and higher moments for the various filters, single observation experiments were performed in the case that the prior mean of zonal wind was set to zero at all grid points. In order to obtain smooth ensemble

pdfs, the ensemble size was reset to K=10,000 members for these experiments. Figure 9 compares and contrasts the posterior ensemble pdfs obtained from single observations of u^2. Comparison of Fig's 9a and 9b shows that (as was previously mentioned) the posterior ensemble variance from the GIGG filter depends on the value of the observation because the posterior ensemble standard deviation is proportional to the posterior ensemble mean. In contrast, Figures 9e and 9f show that the posterior variance for the deterministic ETKF is independent of the value of the observation. (The only reason that the posterior variance for the perturbed observations ETKF shown in Fig. 9c is different to that shown in 9d is because the perturbed observations ETKF produced physically impossible negative values of u^2 which were subsequently set to zero).

Comparison of the posterior skewness given on panels 9a, 9c and 9e shows that the posterior skewness given by the GIGG filter lies somewhere in between the posterior skewness given by the perturbed observations and deterministic ETKF. The fact that the prior and posterior skewness of the deterministic ETKF is essentially unchanged by data assimilation is consistent with the results of Lawson and Hansen (2004). Lawson and Hansen found that this inability of deterministic EnKFs to change the ensemble skewness degraded its performance relative to the perturbed observations form of the EnKF.

We speculate that it is these profound differences between the ETKF and GIGG's treatment of 2nd and 3rd posterior moments that cause the non-linear forecast superiority of GIGG over ETKF to be so much greater than its superiority at the analysis time.

8. **Summary**

The perturbed observations EnKF can only accurately sample the true posterior distribution given by Bayes' theorem when both forecast and observation uncertainty are described by Gaussian pdfs. The GIGG-EnKF algorithm (eq 39) retains the accuracy of the EnKF in the Gaussian case while lending it a high degree of accuracy in the following two non-Gaussian, highly skewed, non-negative cases:

(i) The ensemble forecast of an observation is a random sample from a gamma pdf *and* the pdf of observations given truth is an inverse-gamma pdf, and

(ii) The ensemble forecast of an observation is a random sample from an inverse-gamma pdf *and* the pdf of observations given the truth is a gamma pdf.

Consequently, the GIGG-EnKF is better suited than existing EnKFs in situations where the forecast and observation uncertainty are more accurately approximated by gamma or inverse-gamma distributions than by Gaussian distributions.

If the GIGG filter assimilates all observations with the Gaussian assumption, the posterior GIGG ensemble is identical to that of a perturbed observations EnKF. However, when (gamma, inverse-gamma) or (inverse-gamma, gamma) pdf pairs are assumed to best describe forecast and observation uncertainty, the posterior mean, standard deviation and skewness are all significantly different to that which would be obtained by the perturbed observations EnKF. For these variables, the GIGG posterior standard deviation is proportional to the GIGG posterior mean and hence depends on the value of the observation. In contrast, the EnKF posterior standard deviation is independent of the particular value of the variable that is observed. Posselt et al. (2014) identified this

inability of EnKFs to allow the posterior variance to change with the observed value as one of its fundamental limitations. The GIGG-EnKF usefully removes this limitation in situations where its assumptions are satisfied.

Skewness increases with the asymmetry of a pdf and may be viewed as a crude measure of non-Gaussianity. Posterior skewness was found to be largely independent of the observed value for all of the EnKFs. But it was very sensitive to the flavor of EnKF employed. As in Lawson and Hansen (2004), when the prior was skewed and hence qualitatively non-Gaussian, the deterministic EnKF (ETKF) returned a posterior distribution with almost the same skewness and same degree of non-Gaussianity. The GIGG posterior skewness was less than that of the deterministic EnKF and somewhat higher than that delivered by the perturbed observations EnKF.

A multi-grid point and multi-variable idealized system was used to compare and contrast the data assimilation performance of the GIGG filter with that of the EnKF. The GIGG delivered significant reductions in ensemble mean error at the analysis time.

To test the accuracy of the higher moments of an analysis ensemble a new or at least unsung method was introduced. It is based on the fact that the mean of the ensemble distribution obtained after applying a non-linear "forecast" mapping to each member of the ensemble is highly sensitive to the higher moments of the analysis ensemble. For this study, analysis ensemble variables were raised to either the 2^{nd} or 4^{th} power to create a "forecast" ensemble. The percentage error reduction of the forecast ensemble mean error due to the GIGG filter after these non-linear forecast mappings was found to be more

than an order of magnitude larger than the corresponding reduction associated with the analysis ensemble. Notably, the normalized squared forecast error of the mean of the ensemble forecast made from the GIGG analysis ensemble was less than a 3rd of the corresponding normalized squared forecast error from the EnKFs.

An alternative approach for the assimilation of non-Gaussian variables is to attempt to find a non-linear transformation of the observation that would make the observation errors in the transformed space be Gaussian. The appeal of this approach is that it would allow observations with non-Gaussian error statistics to be assimilated in a data assimilation scheme designed for Gaussian uncertainty distributions. However, as discussed in the Introduction, non-linear transformations of unbiased observations result in biased transformed observations. Furthermore, this bias depends on difficult to predict flow dependent variances of the observation error of representation.

The GIGG-EnKF enables near-zero semi-positive definite variables with highly skewed uncertainty distributions to be assimilated without observation bias inducing non-linear transformations.

Work underway will compare the effectiveness of the GIGG-EnKF against the EnKF in assimilating synthetic precipitation observations generated by a cloud model. The GIG equation set is particularly appropriate in this context. In addition, the GIGG-EnKF may be used to extend the univariate method described in Bishop and Satterfield (2013) and Bishop et al. (2013) for finding the distribution of true error variances given an imperfect ensemble variance to a multivariate method. This is because the pdf of ensemble

variances given a true error variance is well modelled as a gamma pdf while the prior climatological pdf of true flow dependent error variances is well-modelled by an inverse gamma pdf. Hence, the IGG flavor of the GIGG-EnKF may be of assistance in this endeavor.

Gamma and inverse-gamma pdfs are only bounded on one side. However, there are important atmospheric variables such as specific humidity that have known upper and lower bounds. Hopefully, the approach taken by the GIGG filter to accommodate variables that are bounded on one side can be extended to accommodate variables that are bounded on two sides.

Acknowledgement: This work was inspired, in part, by conversations with Derek Posselt and Daniel Hodyss on qualitative differences between the posterior distributions delivered by EnKFs and those of Monte-Carlo-Markov-Chains. The author gratefully acknowledges financial support from the Office of Naval Research Grant # N0001413WX00008.

Appendix A: **Evaluation of Bayes' theorem for GIG**

To evaluate equation (3) note that the dependence on y_j of the numerator of equation (3) is given by

$$
\begin{aligned}
& L(y_j^o \mid y_j)\rho_{\text{prior}}(y_j) \\
& \propto y_j^{\left[\left(\tilde{R}_j^r\right)^{-1}+1\right]}\left(y_j\right)^{\left(P_j^{r-1}-1\right)}\exp\left(-\left(P_j^r\right)^{-1}\frac{y_j}{\left\langle y_j\right\rangle}\right)\exp\left(-\left(\tilde{R}_j^r\right)^{-1}\frac{y_j}{y_j^o}\right) \\
& =\left(y_j\right)^{\left[\left(\tilde{R}_j^r\right)^{-1}+\left(P_j^r\right)^{-1}\right]}\exp\left[-\left(\frac{\left(P_j^r\right)^{-1}}{\left\langle y_j\right\rangle}+\frac{\left(\tilde{R}_j^r\right)^{-1}}{y_j^o}\right)y_j\right] \\
& =\left(y_j\right)^{\left[\left(\tilde{R}_j^r\right)^{-1}+\left(P_j^r\right)^{-1}\right]}\exp\left[-\left(\left(\tilde{R}_j^r\right)^{-1}+\left(P_j^r\right)^{-1}+1\right)\frac{\left(\frac{\left(P_j^r\right)^{-1}}{\left\langle y_j\right\rangle}+\frac{\left(\tilde{R}_j^r\right)^{-1}}{y_j^o}\right)}{\left(\left(\tilde{R}_j^r\right)^{-1}+\left(P_j^r\right)^{-1}+1\right)}y_j\right] \\
& =\left(y_j\right)^{\left(\left(\Pi_j^r\right)^{-1}-1\right)}\exp\left[-\left(\Pi_j^r\right)^{-1}\frac{y_j}{\left\langle y_j^a\right\rangle}\right]
\end{aligned}
\tag{A1}
$$

where

$$
\left(\Pi_j^r\right)^{-1}=\left(\tilde{R}_j^r\right)^{-1}+\left(P_j^r\right)^{-1}+1 \text{ and } \frac{1}{\left\langle y_j^a\right\rangle}=\frac{1}{\left(\left(\tilde{R}_j^r\right)^{-1}+\left(P_j^r\right)^{-1}+1\right)}\left(\frac{\left(P_j^r\right)^{-1}}{\left\langle y_j^f\right\rangle}+\frac{\left(\tilde{R}_j^r\right)^{-1}}{y_j^o}\right). \tag{A2}
$$

The y_j dependence of the last line of (A1) is isomorphic to that shown in equation (1). Hence, the posterior distribution must be a gamma pdf with a different set of parameters. Since it is only the numerator of equation (3) that imparts y_j dependence and since the role of the numerator of (3) is to ensure that the posterior pdf integrates to unity, it follows that the gamma pdf

$$\frac{1}{\Gamma\left(\left(\Pi_j^r\right)^{-1}\right)}\left(\frac{\left(\Pi_j^r\right)^{-1}}{\left\langle y_j^a\right\rangle}\right)^{\Pi_j^{r-1}}\left\{\left(y_j\right)^{\left(\left(\Pi_j^r\right)^{-1}-1\right)}\exp\left[-\left(\Pi_j^r\right)^{-1}\frac{y_j}{\left\langle y_j^a\right\rangle}\right]\right\}=\frac{L(y_j^o\mid y_j)\rho_{\text{prior}}(y_j)}{\int_0^\infty L(y_j^o\mid y_j)\rho_{\text{prior}}(y_j)dy_j} \quad \text{(A3)}$$

is the solution to Bayes' theorem where (A2) defines Π_j^r and $\left\langle y_j^a\right\rangle$.

It is of interest to try and manipulate these properties into forms that are more easily recognizable to data assimilation experts. First, let us define the type 2 relative forecast error variance $\tilde{P}_j^r=\dfrac{\text{var}\left(y_j\right)}{\left\langle y_j^f\right\rangle^2+\text{var}\left(y_j\right)}$ so that we can write

$$\left(\Pi_j^r\right)^{-1}=\left(\tilde{R}_j^r\right)^{-1}+\left(\tilde{P}_j^r\right)^{-1} \quad \text{(A4)}$$

This equation is like the information form of the analysis error covariance matrix. For the posterior mean, note that (A2) implies that

$$\frac{1}{\left\langle y_j^a\right\rangle}=\frac{1}{\left(\Pi_j^r\right)^{-1}}\left(\frac{\left(P_j^r\right)^{-1}}{\left\langle y_j^f\right\rangle}+\frac{\left(\tilde{R}_j^r\right)^{-1}}{y_j^o}\right)=\Pi_j^r\left(\frac{\left(\tilde{P}_j^r\right)^{-1}-1}{\left\langle y_j^f\right\rangle}+\frac{\left(\tilde{R}_j^r\right)^{-1}}{y_j^o}\right) \quad \text{(A5)}$$

Now because $\Pi_j^r\left(\left(\tilde{R}_j^r\right)^{-1}+\left(\tilde{P}_j^r\right)^{-1}\right)=1$, (A5) then implies that

$$\begin{aligned}\frac{1}{\left\langle y_j^a\right\rangle}-\frac{1}{\left\langle y_j^f\right\rangle}&=\Pi_j^r\left(\frac{\left(\tilde{P}_j^r\right)^{-1}-1}{\left\langle y_j^f\right\rangle}+\frac{\left(\tilde{R}_j^r\right)^{-1}}{y_j^o}\right)-\frac{1}{\left\langle y_j^f\right\rangle}\\ &=\Pi_j^r\left(\frac{\left(\tilde{P}_j^r\right)^{-1}-1}{\left\langle y_j^f\right\rangle}+\frac{\left(\tilde{R}_j^r\right)^{-1}}{y_j^o}\right)-\Pi_j^r\left(\left(\tilde{R}_j^r\right)^{-1}\frac{1}{\left\langle y_j^f\right\rangle}+\left(\tilde{P}_j^r\right)^{-1}\frac{1}{\left\langle y_j^f\right\rangle}\right) \quad \text{(A6)}\\ &=\Pi_j^r\left(\tilde{R}_j^r\right)^{-1}\left[\frac{1}{y_j^o}-\left(\tilde{R}_j^r+1\right)\frac{1}{\left\langle y_j^f\right\rangle}\right]\end{aligned}$$

and hence that

$$\begin{aligned}\frac{1}{\langle y_j^a \rangle} &= \frac{1}{\langle y_j^f \rangle} + \Pi_j^r \left(\tilde{R}_j^r\right)^{-1} \left[\frac{1}{y_j^o} - \left(\tilde{R}_j^r + 1\right)\frac{1}{\langle y_j^f \rangle}\right] \\ &= \frac{1}{\langle y_j^f \rangle} + \frac{1}{\left[\left(\tilde{R}_j^r\right)^{-1} + \left(\tilde{P}_j^r\right)^{-1}\right]}\left(\tilde{R}_j^r\right)^{-1}\left[\frac{1}{y_j^o} - \left(\tilde{R}_j^r + 1\right)\frac{1}{\langle y_j^f \rangle}\right] \\ &= \frac{1}{\langle y_j^f \rangle} + \frac{\tilde{P}_j^r \tilde{R}_j^r}{\tilde{R}_j^r + \tilde{P}_j^r}\left(\tilde{R}_j^r\right)^{-1}\left[\frac{1}{y_j^o} - \left(\tilde{R}_j^r + 1\right)\frac{1}{\langle y_j^f \rangle}\right] \\ &= \frac{1}{\langle y_j^f \rangle} + \frac{\tilde{P}_j^r}{\tilde{R}_j^r + \tilde{P}_j^r}\left[\frac{1}{y_j^o} - \left(\tilde{R}_j^r + 1\right)\frac{1}{\langle y_j^f \rangle}\right]\end{aligned} \tag{A7}$$

Appendix B: Perturbed observations for gamma prior and inverse gamma likelihood

Holding y_j^o constant in equation (2) gives

$$\begin{aligned}L\left(y_j^o \mid y_j\right) &\propto \left(y_j\right)^{\left(\left(\tilde{R}_j^r\right)^{-1}+1\right)} \exp\left(-\left(\tilde{R}_j^r\right)^{-1}\frac{y_j}{y_j^o}\right) \\ &= \left(y_j\right)^{\left\{\left[\left(\tilde{R}_j^r\right)^{-1}+2\right]-1\right\}} \exp\left(-\left[\left(\tilde{R}_j^r\right)^{-1}+2\right]\frac{\left(\tilde{R}_j^r\right)^{-1}}{\left[\left(\tilde{R}_j^r\right)^{-1}+2\right]}\frac{y_j}{y_j^o}\right) \\ &= \left(y_j\right)^{\left\{\left(R_j^{gig}\right)^{-1}-1\right\}} \exp\left(-\left(R_j^{gig}\right)^{-1}\frac{y_j}{\langle y_j^{gig} \rangle}\right),\end{aligned} \tag{B1}$$

where

$$\left(R_j^{gig}\right)^{-1} = \left[\left(\tilde{R}_j^r\right)^{-1}+2\right] \text{ and } \left\langle y_j^{gig} \right\rangle = \frac{\left[\left(\tilde{R}_j^r\right)^{-1}+2\right]}{\left(\tilde{R}_j^r\right)^{-1}} y_j^o \tag{B2}$$

Comparison of the form of (B1) with (1) makes it clear that the function of y_j obtained by holding y_j^o fixed in (2) is directly proportional to the gamma pdf given by

$$\rho_L\left(y_j\right)=\frac{1}{\Gamma\left[\left(R_j^{gig}\right)^{-1}\right]}\left(\frac{\left(R_j^{gig}\right)^{-1}}{\left\langle y_j^{gig}\right\rangle}\right)^{\left(R_L^r\right)^{-1}}\left(y_j\right)^{\left\{\left(R_j^{gig}\right)^{-1}-1\right\}}\exp\left(-\left(R_j^{gig}\right)^{-1}\frac{y_j}{\left\langle y_j^{gig}\right\rangle}\right) \tag{B3}$$

The equivalent gamma pdf scale parameter θ_L and shape parameter k_L are given by

$$\begin{aligned} k_L &= \left(R_j^{gig}\right)^{-1}=\left(\tilde{R}_j^r\right)^{-1}+2=\left(\alpha-1\right)+2=\alpha+1 \\ \theta_L &= \frac{\left\langle y_j^{gig}\right\rangle}{k_L}=\frac{\left[\left(\tilde{R}_j^r\right)^{-1}+2\right]}{\left(\tilde{R}_j^r\right)^{-1}}\frac{y_j^o}{\left[\left(\tilde{R}_j^r\right)^{-1}+2\right]}=\frac{y_j^o}{\left(\tilde{R}_j^r\right)^{-1}}=\frac{y_j^o}{\left(\alpha-1\right)} \end{aligned} \tag{B4}$$

Appendix C: Proof that GIG posterior perturbations have correct variance

To see that (10) ensures that the posterior perturbations given by the ensemble gig filter satisfy the first line of equation (5), note that from (7) and (10)

$$
\begin{aligned}
&\left\langle \left(\frac{y_{ji}^{a} - \overline{y_j^{a}}}{\overline{y_j^{a}}} \right)^2 \right\rangle \\
&= \left\langle \left\{ \frac{\left(y_{ji}^{f} - \overline{y_j^{f}} \right)}{\sqrt{\left(\overline{y_j^{f}} \right)^2 + \operatorname{var}\left(y_{ji}^{f} \right)}} \right. \right. \\
&\left. \left. + \tilde{P}_j^r \left(\tilde{P}_j^r + \tilde{R}_j^r \right)^{-1} \left[\frac{\left(y_{ji}^{gig} - \left\langle y_j^{gig} \right\rangle \right)}{\sqrt{\left\langle y_j^{gig} \right\rangle^2 - 2\operatorname{var}\left(y_{ji}^{gig} \right)}} - \frac{\left(y_{ji}^{f} - \overline{y_j^{f}} \right)}{\sqrt{\left(\overline{y_j^{f}} \right)^2 + \operatorname{var}\left(y_{ji}^{f} \right)}} \right] \right\}^2 \right\rangle \\
&= \left\langle \left(\frac{\left(y_{ji}^{f} - \overline{y_j^{f}} \right)}{\sqrt{\left(\overline{y_j^{f}} \right)^2 + \operatorname{var}\left(y_{ji}^{f} \right)}} \right)^2 \right\rangle - 2\tilde{P}_j^r \left(\tilde{P}_j^r + \tilde{R}_j^r \right)^{-1} \left\langle \left(\frac{\left(y_{ji}^{f} - \overline{y_j^{f}} \right)}{\sqrt{\left(\overline{y_j^{f}} \right)^2 + \operatorname{var}\left(y_{ji}^{f} \right)}} \right)^2 \right\rangle \\
&+ \tilde{P}_j^r \left(\tilde{P}_j^r + \tilde{R}_j^r \right)^{-1} \left\langle \left(\frac{\left(y_{ji}^{gig} - \left\langle y_j^{gig} \right\rangle \right)}{\sqrt{\left\langle y_j^{gig} \right\rangle^2 - 2\operatorname{var}\left(y_{ji}^{gig} \right)}} \right)^2 + \left(\frac{\left(y_{ji}^{f} - \overline{y_j^{f}} \right)}{\sqrt{\left(\overline{y_j^{f}} \right)^2 + \operatorname{var}\left(y_{ji}^{f} \right)}} \right)^2 \right\rangle \left(\tilde{P}_j^r + \tilde{R}_j^r \right)^{-1} \tilde{P}_j^r \\
&= \tilde{P}_j^r - 2\tilde{P}_j^r \left(\tilde{P}_j^r + \tilde{R}_j^r \right)^{-1} \tilde{P}_j^r + \tilde{P}_j^r \left(\tilde{P}_j^r + \tilde{R}_j^r \right)^{-1} \left(\tilde{P}_j^r + \tilde{R}_j^r \right) \left(\tilde{P}_j^r + \tilde{R}_j^r \right)^{-1} \tilde{P}_j^r \\
&= \tilde{P}_j^r - 2\tilde{P}_j^r \left(\tilde{P}_j^r + \tilde{R}_j^r \right)^{-1} \tilde{P}_j^r + \tilde{P}_j^r \left(\tilde{P}_j^r + \tilde{R}_j^r \right)^{-1} \tilde{P}_j^r \\
&= \tilde{P}_j^r - \tilde{P}_j^r \left(\tilde{P}_j^r + \tilde{R}_j^r \right)^{-1} \tilde{P}_j^r = \Pi_j^r
\end{aligned}
\tag{C1}
$$

as was required. Thus, the stochastic process given by (7) generates posterior perturbations with the required posterior relative error variance.

Appendix D: **Evaluation of Bayes' theorem for IGG**

To evaluate Bayes' theorem for the IGG case, note that

$$L(y_j^o \mid y_j)\rho_{\text{prior}}(y_j) \propto y_j^{-\left(R_j^r\right)^{-1}} y_j^{-\left[\left(\tilde{P}_j^r\right)^{-1}+2\right]} \exp\left(-\left(R_j^r\right)^{-1}\frac{y_j^o}{y_j}\right)\exp\left(\frac{-\left(\tilde{P}_j^r\right)^{-1}\left\langle y_j^f\right\rangle}{y_j}\right)$$
$$= \left(y_j\right)^{-\left(\left(\tilde{\Pi}_j^r\right)^{-1}+2\right)} \exp\left[-\left(\tilde{\Pi}_j^r\right)^{-1}\frac{\left\langle y_j^a\right\rangle}{y_j}\right] \tag{D1}$$

where

$$\left(\tilde{\Pi}_j^r\right)^{-1} = \left(R_j^r\right)^{-1} + \left(\tilde{P}_j^r\right)^{-1} \quad \text{and} \quad \left\langle y_j^a\right\rangle = \left(\tilde{\Pi}_j^r\right)\left[\left(R_j^r\right)^{-1} y_j^o + \left(\tilde{P}_j^r\right)^{-1}\left\langle y_j^f\right\rangle\right]. \tag{D2}$$

The y_j dependence of the last line of (D1) is isomorphic to that of the inverse gamma pdf given by (13). Thus, the normalization factor required to make the integral of this function is equal to (1) also has the same form as that in (13) and hence

$$\frac{\left[\left(\tilde{\Pi}_j^r\right)^{-1}\left\langle y_j^a\right\rangle\right]^{\left[\left(\tilde{\Pi}_j^r\right)^{-1}+1\right]}}{\Gamma\left(\left[\left(\tilde{\Pi}_j^r\right)^{-1}+1\right]\right)}\left(y_j\right)^{-\left(\left(\tilde{\Pi}_j^r\right)^{-1}+2\right)}\exp\left[-\left(\tilde{\Pi}_j^r\right)^{-1}\frac{\left\langle y_j^a\right\rangle}{y_j}\right] = \frac{L(y_j^o \mid y_j)\rho_{\text{prior}}(y_j)}{\int_0^\infty L(y_j^o \mid y_j)\rho_{\text{prior}}(y_j)dy_j} \tag{D3}$$

is the solution to Bayes' theorem for this inverse gamma prior, gamma likelihood (IGG) case where $\tilde{\Pi}_j^r$ and $\left\langle y_j^a\right\rangle$ are given by (D2), respectively.

It is of interest to try and manipulate these properties into forms that are more easily recognizable to atmospheric data assimilation researchers. First, note that $\left(\tilde{\Pi}_j^r\right)^{-1} = \left(R_j^r\right)^{-1} + \left(\tilde{P}_j^r\right)^{-1}$ implies that

$$\tilde{\Pi}_j^r = \frac{\tilde{P}_j^r R_j^r}{\tilde{P}_j^r + R_j^r} = \frac{\left(\tilde{P}_j^r + R_j^r\right)\tilde{P}_j^r}{\tilde{P}_j^r + R_j^r} - \frac{\left(\tilde{P}_j^r\right)^2}{\tilde{P}_j^r + R_j^r} = \tilde{P}_j^r - \tilde{P}_j^r\left(\tilde{P}_j^r + R_j^r\right)^{-1}\tilde{P}_j^r \tag{D4}$$

This equation is like the error form of the analysis error covariance matrix. For the posterior mean, note that (D2) implies that

$$\begin{aligned}\left\langle y_j^a \right\rangle &= \left(\tilde{\Pi}_j^r\right)\left[\left(R_j^r\right)^{-1} y_j^o + \left(\tilde{P}_j^r\right)^{-1}\left\langle y_j^f \right\rangle\right] = \frac{\tilde{P}_j^r R_j^r}{\tilde{P}_j^r + R_j^r}\left[\left(R_j^r\right)^{-1} y_j^o + \left(\tilde{P}_j^r\right)^{-1}\left\langle y_j^f \right\rangle\right] \\ &= \left\langle y_j^f \right\rangle - \frac{\tilde{P}_j^r + R_j^r}{\tilde{P}_j^r + R_j^r}\left\langle y_j^f \right\rangle + \frac{R_j^r}{\tilde{P}_j^r + R_j^r}\left\langle y_j^f \right\rangle + \frac{\tilde{P}_j^r}{\tilde{P}_j^r + R_j^r} y_j^o \\ &= \left\langle y_j^f \right\rangle + \frac{\tilde{P}_j^r}{\tilde{P}_j^r + R_j^r}\left(y_j^o - \left\langle y_j^f \right\rangle\right).\end{aligned} \tag{D5}$$

References

Amezcua J., van Leeuwen PJ.,2014. Gaussian anamorphosis in the analysis step of the EnKF: a joint state-variable/observation approach. *Tellus* A. ISSN 1600-0870.

Anderson JL., 2001. An Ensemble Adjustment Kalman Filter for Data Assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903.

Anderson JL., 2003. A Local Least Squares Framework for Ensemble Filtering. *Mon. Wea. Rev.*, **131**, 634–642.

doi: http://dx.doi.org/10.1175/1520-0493(2003)131<0634:ALLSFF>2.0.CO;2

Anderson JL, Collins N. 2007. Scalable Implementations of Ensemble Filter Algorithms for Data Assimilation. *J. Atmos. Oceanic Technol.*, **24**, 1452–1463.

doi: http://dx.doi.org/10.1175/JTECH2049.1

Anderson J, Hoar, T,Raeder, K., Liu, H., Collins, N., Torn, R. and A. Avellano, 2009. The Data Assimilation Research Testbed: A Community Facility. *Bull. Amer. Meteor. Soc.*, **90**, 1283–1296.

doi: http://dx.doi.org/10.1175/2009BAMS2618.1

Bishop CH, Etherton, BJ, Majumdar SJ, 2001. Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects. *Mon. Wea. Rev.* **129,** 420-436.

Bishop CH, Huang B, Wang, X, 2015. A non-variational consistent Hybrid ensemble filter. *Mon. Wea. Rev.* (In Press and early release version available on-line).

Bishop, CH, Satterfield EA. 2013. Hidden error variance theory 1: Exposition and analytic model. *Mon. Wea. Rev.*, **141,** 1454-1468.

Bishop, CH, Satterfield EA, Shanley KT. 2013. Hidden error variance theory 2: An instrument that reveals hidden error variance distributions from ensemble forecasts and observations. *Mon. Wea. Rev.*, **141,** 1469-1483.

Bonavita M, Isaksen L, Hólm E. 2012. On the use of EDA background error variances in the ECMWF 4D-Var. *Q.J.R. Meteorol. Soc.,* **138**: 1540–1559. doi: 10.1002/qj.1899

Bougeault P. 1982. Cloud-Ensemble Relations Based on the Gamma Probability Distribution for the Higher-Order Models of the Planetary Boundary Layer. *J. Atmos. Sci.*, **39**, 2691–2700.

doi: http://dx.doi.org/10.1175/1520-0469(1982)039<2691:CERBOT>2.0.CO;2

Burgers G., van Leeuwen PJ, Evensen G. 1998. Analysis Scheme in the Ensemble Kalman Filter. *Mon. Wea. Rev.*, **126**, 1719–1724.

Campbell, JW. 1995. The lognormal distribution as a model for bio-optical variability in the sea. *Journal of Geophysical Research Oceans* **100**, 13237–13254.

Cho HK, Bowman KP, North GR, 2004. A Comparison of Gamma and Lognormal Distributions for Characterizing Satellite Rain Rates from the Tropical Rainfall Measuring Mission. *J. Appl. Meteor.*, **43**, 1586–1597.

doi: http://dx.doi.org/10.1175/JAM2165.1

Choi SC, Wette R. 1969. Maximum Likelihood Estimation of the Parameters of the Gamma Distribution and Their Bias, *Technometrics*, 11(4) 683–690.

Clark TL. 1974. A Study in Cloud Phase Parameterization Using the Gamma Distribution. *J. Atmos. Sci.*, **31**, 142–155.

doi: http://dx.doi.org/10.1175/1520-0469(1974)031<0142:ASICPP>2.0.CO;2.

Clayton AM, Lorenc AC. Barker DM. 2013. Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *Q.J.R. Meteorol. Soc.*, **139**: 1445–1461. doi: 10.1002/qj.2054.

Errico RM, Stensrud DJ, Raeder KD. 2001. Estimation of the error distributions of precipitation produced by convective parametrization schemes. *Q.J.R. Meteorol. Soc.*, **127**: 2495–2512. doi: 10.1002/qj.49712757802.

Hodyss D. 2012. Accounting for Skewness in Ensemble Data Assimilation. *Mon. Wea. Rev.*, **140**, 2346–2358.

Hodyss D, Campbell WF, 2013. Square Root and Perturbed Observation Ensemble Generation Techniques in Kalman and Quadratic Ensemble Filtering Algorithms. *Mon. Wea. Rev.*, **141**, 2561–2573.

Houtekamer, PL, Mitchell HL. 2001. A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation. *Mon. Wea. Rev.*, **129**, 123–137.

Houtekamer PL., Mitchell HL, Pellerin G, Buehner M, Charron M, Spacek L, Hansen B. 2005. Atmospheric Data Assimilation with an Ensemble Kalman Filter: Results with Real Observations. *Mon. Wea. Rev.*, **133**, 604–620.

Husak GJ, Michaelsen J, Funk C. 2007. Use of the gamma distribution to represent monthly rainfall in Africa for drought monitoring applications. *Int. J. Climatol.*, **27**: 935–944. doi: 10.1002/joc.1441.

Kalnay E, 2004. Atmospheric modelling, data assimilation and predictability. Cambridge University Press, pp 1-341.

Kliewer AJ, Fletcher SJ, Jones AS, Forsythe, JM. 2015. Comparison of Gaussian, logarithmic transform and mixed Gaussian–lognormal distribution based 1DVAR microwave temperature–water-vapour mixing ratio retrievals. *Q.J.R. Meteorol. Soc.*. doi: 10.1002/qj.2651

Kuhl DD, Rosmond TE, Bishop CH, McLay J, Baker NL, 2013. Comparison of Hybrid Ensemble/4DVar and 4DVar within the NAVDAS-AR Data Assimilation Framework. *Mon. Wea. Rev.*, **141**, 2740–2758.

Lange MA, Eicken H. 1991. The sea ice thickness distribution in the northwestern Weddell Sea, *J. Geophys. Res.*, **96**(C3), 4821–4837, doi:10.1029/90JC02441.

Lawson GW, Hansen JA, 2004: Implications of Stochastic and Deterministic Filters as Ensemble-Based Data Assimilation Methods in Varying Regimes of Error Growth. *Mon. Wea. Rev.*, **132**, 1966–1981.

doi: http://dx.doi.org/10.1175/1520-0493(2004)132<1966:IOSADF>2.0.CO;2

Lien G-Y, Kalnay E, Miyoshi T. 2013. Effective assimilation of global precipitation: simulation experiments. *Tellus A*, **65**. ISSN 1600-0870. doi:http://dx.doi.org/10.3402/tellusa.v65i0.19915.

O'Neill NT, Ignatov A, Holben BN, Eck TF. 2000. The lognormal distribution as a reference for reporting aerosol optical depth statistics: Empirical tests using multi-year, multi-site AERONET sunphotometer data. *Geophys. Res. Lett.*, **27**, 3333–3336, doi:10.1029/2000GL011581.

Pelc JS, Simon E, Bertino L, El Serafy G, Heemink AW. 2012. Application of model reduced 4DVAR to a 1D ecosystem model. *Ocean Modelling,* **57-58**, 43-58. doi:10.1016/j.ocemod.2012.09.003

Posselt DJ, Bishop CH. 2012. Nonlinear parameter estimation: Comparison of an Ensemble Kalman Smoother with a Markov chain Monte Carlo algorithm. *Mon. Wea. Rev.* **140**, 1957-1974.

Posselt DJ, Hodyss D, and Bishop CH, 2014: Errors in Ensemble Kalman Smoother Estimates of Cloud Microphysical Parameters. *Mon. Wea. Rev.*, **142**, 1631–1654.

doi: http://dx.doi.org/10.1175/MWR-D-13-00290.1

Rabier F, Järvinen H, Klinker E, Mahfouf, J-F, Simmons A. 2000. The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Q.J.R. Meteorol. Soc.*, **126**: 1143–1170.

Saide PE, Carmichael GR, Spak SN, Minnis P, Ayers JK. 2012. Improving aerosol distributions below clouds by assimilating satellite-retrieved cloud droplet number. *Proceedings of the National Academies of Sciences*, **109**, 11939–11943.

Simpson J. 1972. USE OF THE GAMMA DISTRIBUTION IN SINGLE-CLOUD RAINFALL ANALYSIS. *Mon. Wea. Rev.*, **100**, 309–312.

doi: http://dx.doi.org/10.1175/1520-0493(1972)100<0309:UOTGDI>2.3.CO;2.

Tippett MK, Anderson JL, Bishop CH, Hamill TM, Whitaker JS. 2003. Square root ensemble filters. *Mon Wea. Rev.* **131**, 1485-1490.

van Leeuwen PJ, Ades M. 2013. Efficient fully nonlinear data assimilation for geophysical fluid dynamics. *Mon. Wea. Rev.,* **137**, 4089–4114.

Vivekanandan J, Zhang G, Brandes E. 2004. Polarimetric Radar Estimators Based on a Constrained Gamma Drop Size Distribution Model. *J. Appl. Meteor.*, **43**, 217–230

doi: http://dx.doi.org/10.1175/1520-0450(2004)043<0217:PREBOA>2.0.CO;2

Wadhams P, Lange MA, Ackley SF. 1987. The ice thickness distribution across the Atlantic sector of the Antarctic Ocean in midwinter, *J. Geophys. Res.*, **92**(C13), 14535–14552, doi:10.1029/JC092iC13p14535.

Wang X, Parrish D, Kleist D, Whitaker J. 2013. GSI 3DVar-Based Ensemble–Variational Hybrid Data Assimilation for NCEP Global Forecast System: Single-Resolution Experiments. *Mon. Wea. Rev.*, **141**, 4098–4117.

Willis PT. 1984. Functional Fits to Some Observed Drop Size Distributions and Parameterization of Rain. *J. Atmos. Sci.*, **41**, 1648–1661.

doi: http://dx.doi.org/10.1175/1520-0469(1984)041<1648:FFTSOD>2.0.CO;2

Whitaker JS, Hamill TM. 2002. Ensemble Data Assimilation without Perturbed Observations. *Mon. Wea. Rev.*, **130**, 1913–1924.

Whitaker JS, Hamill TM, Wei X, Song Y, Toth Z. 2008. Ensemble Data Assimilation with the NCEP Global Forecast System. *Mon. Wea. Rev.*, **136**, 463–482.

Xu L, Rosmond T, Daley R. 2005. Development of NAVDAS-AR: formulation and initial tests of the linear problem. *Tellus A*, **57**: 546–559.

Zhou H, Gómez-Hernández JJ, Hendricks-Franssen, Li L. 2011. An approach to handling non-Gaussianity of parameters and state variables in ensemble Kalman filtering. *Advances in Water Resources* **34**, 844–864.

Table 1: Symbols in order of appearance

y_j^o	jth observation
y_j	true value of jth observation
$\varepsilon^o = y_j^o - y_j$	observation error
y_{ji}^f	ith random sample from the prior distribution of y_j
$\overline{y_j^f}$	mean of K random samples from the prior distribution of y_j
$\left\langle y_j^f \right\rangle$	true mean of prior distribution of y_j
$\mathrm{var}\left(y_j^f\right)$	true variance of prior distribution of y_j
$P_j^r = \dfrac{\mathrm{var}\left(y_j^f\right)}{\left\langle y_j^f \right\rangle^2}$	type 1 relative error variance of prior
$\tilde{P}_j^r = \dfrac{\mathrm{var}\left(y_j^f\right)}{\left\langle y_j^f \right\rangle^2 + \mathrm{var}\left(y_j^f\right)}$	type 2 relative error variance of prior
$P_j = \mathrm{var}\left(y_j^f\right)$	variance of prior
$R_j^r = \dfrac{\mathrm{var}\left(\varepsilon^o\right)}{\left(y_j\right)^2}$	type 1 relative observation error variance
$\tilde{R}_j^r = \dfrac{\mathrm{var}\left(\varepsilon^o\right)}{\left(y_j\right)^2 + \mathrm{var}\left(\varepsilon^o\right)}$	type 2 relative observation error variance
$R_j = \left\langle R_j^r \left(y_j^f\right)^2 \right\rangle$	average of observation error variance over prior
$\Pi_j^r = \dfrac{\mathrm{var}\left(y_j^a\right)}{\left\langle y_j^a \right\rangle^2}$	type 1 relative analysis error variance
$\tilde{\Pi}_j^r = \dfrac{\mathrm{var}\left(y_j^a\right)}{\left\langle y_j^a \right\rangle^2 + \mathrm{var}\left(y_j^a\right)}$	type 2 relative analysis error variance
y_{ji}^{gig} $\left(y_{ji}^{igg}\right)$	ith perturbed observation for ensemble gig (igg) filter
$\left\langle y_{ji}^{gig} \right\rangle$ $\left(\left\langle y_{ji}^{igg} \right\rangle\right)$	mean of perturbed observations for ensemble gig (igg) filter
R_j^{gig} $\left(R_j^{igg}\right)$	type 1 relative variance of perturbed observations for gig (igg)
$\tilde{R}_j^{gig}$ $\left(\tilde{R}_j^{igg}\right)$	type 2 relative variance of perturbed observations for gig (igg)

Table 1: Symbols in order of appearance (continued)

$\tilde{y}_{ji}^{a} = \frac{y_{ji}^{a} - \overline{y_j^a}}{\sqrt{\left(\overline{y_j^a}\right)^2 + \mathrm{var}\left(y_{ji}^{a}\right)}}$	normalized univariate analysis perturbation from igg filter
$P_{jj}^{oo} = P_j^o = \mathrm{var}\left(y_j^f\right)$	variance of prior
$P_{kj}^{oo} = \mathrm{covar}\left(y_k^f, y_j^f\right)$	covariance of kth observation with jth observation in the prior
$x_{\mu i}^{f}$	ith member of ensemble forecast of μth model variable
$P_{\mu j}^{mo} = \mathrm{covar}\left(x_\mu^f, y_j^f\right)$	covariance of μth model variable with jth observation in the prior

Figure Captions

Figure 1: All of the above pdfs pertain to distributions with standard deviations of unity. The means of the blue, green and red lines are 1,2, and 11, respectively. Going from top to bottom, panels refer to Gaussian, gamma, inverse-gamma and lognormal pdfs, respectively. Note that the gamma, inverse-gamma and lognormal pdfs are all similar to the Gaussian pdfs when the mean is large compared with the standard deviation (red lines).

Figure 2: Left two panels pertain to GIG filter, right two panels to IGG filter. In all panels, the mean of the prior is equal to 1 and is indicated by the dashed vertical line while the dashed curve gives the prior pdf. In the top (bottom) two panels, the observed value is equal to 0.5 (3) and is indicated by the dot-dash vertical line while the dot-dash curve gives the likelihood probability density of this observation occurring as a function of the true value of y_j. In all panels, the posterior mean is given by the vertical solid line while the true posterior pdf $\rho\left(y_{ji} \mid y_j^o\right)$ is given by the curved solid line. Note that in all cases the posterior pdf is proportional to the product of the dashed and dot-dashed curves.

The corresponding GIG and IGG ensemble based estimates of these pdfs $\rho_{\text{ensemble}}\left(y_{ji} \mid y_j^o\right)$ are given by the thicker grey line. The normalized root mean square distance (rmsd) and maximum difference (maxd) between the ensemble based and analytical posterior pdfs are written on each panel.

Figure 3: Normalized maximal distance between $\rho\left(y_{ji} \mid y_j^o\right)$ and $\rho_{\text{ensemble}}\left(y_{ji} \mid y_j^o\right)$ as a function of the type 1 relative variances P_j^r and R_j^r. Left two panels pertain to GIG filter, right two panels to IGG filter. Top two panels pertain to an observed value of 0.5, bottom two panels to an observed value of 3. The prior mean was equal to 1 in all cases. In all panels, the numbers on the abscissa and ordinate axes give the exponents *m* and *n* that define the observation and forecast relative error variance, respectively. Equation (32) {33} defines this relationship for the GIG {IGG} filter.

Figure 4: As in Figure 2, the difference between the curve given by the thin black curve $\rho\left(y_{ji} \mid y_j^o\right)$ and the thick grey line $\rho_{ensemble}\left(y_{ji} \mid y_j^o\right)$ indicates the error in the new ensemble filters presented in this paper. In (a) and (b) are shown the worst cases for the GIG and IGG filters, respectively. These worst cases correspond to highly uninformative prior and likelihood distributions for which the type prior and observation relative error

variance is given by unity. In (c) is shown, the second worst case for the IGG filter in which the type 1 relative observation error variance is equal to 0.5.

Figure 5: Covariance matrices of random vectors used to create prior distributions of the zonal wind. The prior mean is a random vector with mean zero and covariance matrix given by 5a. For the zonal wind u mean depicted by the thick green line in figure 6, the individual prior ensemble members of u depicted as colored lines in figure 6 were obtained by adding perturbations with mean zero and covariance matrix given by 5b to the zonal mean.

Figure 6: Example of system used to examine filter performance. Abscissa gives the index of the grid point on the periodic model domain. Ordinate gives the value of the model variable. Grey lines depict the 10000 members of the prior ensemble forecast. Thick green line gives the prior forecast ensemble mean. Cyan line depicts the truth. Red dots indicate the error prone observations of the truth. Black lines depict the posterior analysis ensemble obtained from the GIGG filter. Yellow line depicts the posterior analysis ensemble mean. (The yellow line is largely obscured by the truth (cyan line)). Top, middle and bottom panels refer to zonal wind u, u^2 and dust, respectively.

Figure 7: As in Figure 6, but for the perturbed observations ETKF.

Figure 8: Percentage reduction in average square error of analysis and forecast ensemble mean over the 3584 independent trials. In (a) and (b) is shown the reduction due to using GIGG instead of the perturbed observations ETKF and the deterministic ETKF, respectively. The red crosses give the mean percentage error reduction while the upper and lower "+" signs give the upper and lower 99% confidence intervals for each result.

Figure 9: Illustration of the stark differences between the higher order moments of the posterior ensemble distributions produced by the GIGG (panels a and b), perturbed observations ETKF (panels c and d) and deterministic ETKF (panels e and f). In all panels, the black dashed line gives the prior ensemble density, the vertical black dashed line gives the value of the prior ensemble mean (equal to 1 in all cases), the vertical dot-dashed line gives the value of the observation and the solid grey line is the pdf of the posterior ensemble, while the vertical solid grey line gives the posterior mean. The three left (right) panels pertain to an observation of 5 (0.5).

Figures

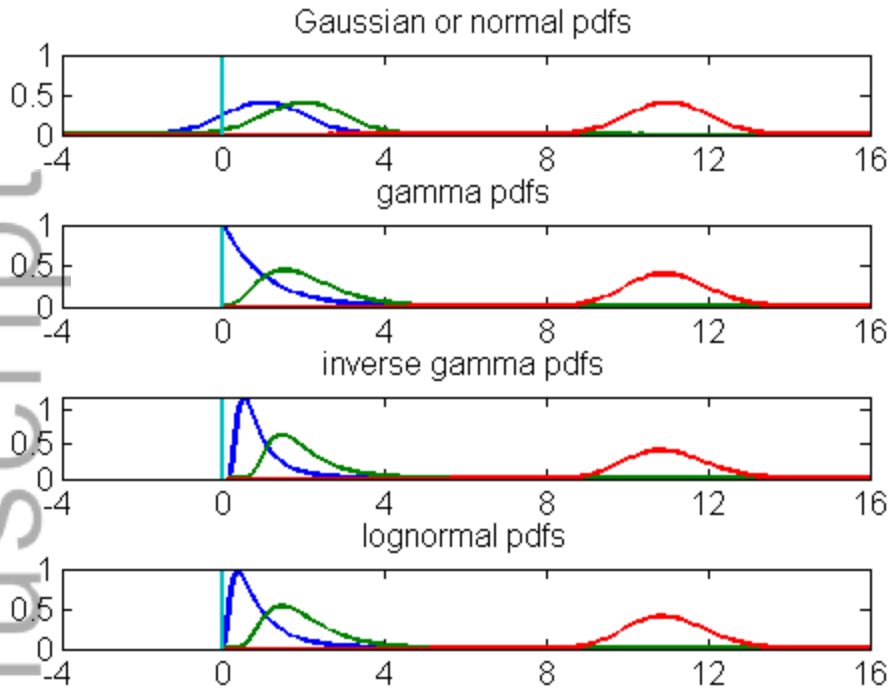

Figure 1: All of the above pdfs pertain to distributions with standard deviations of unity. The means of the blue, green and red lines are 1,2, and 11, respectively. Going from top to bottom, panels refer to Gaussian, gamma, inverse-gamma and lognormal pdfs, respectively. Note that the gamma,inverse-gamma and lognormal pdfs are all similar to the Gaussian pdfs when the mean is large compared with the standard deviation (red lines).

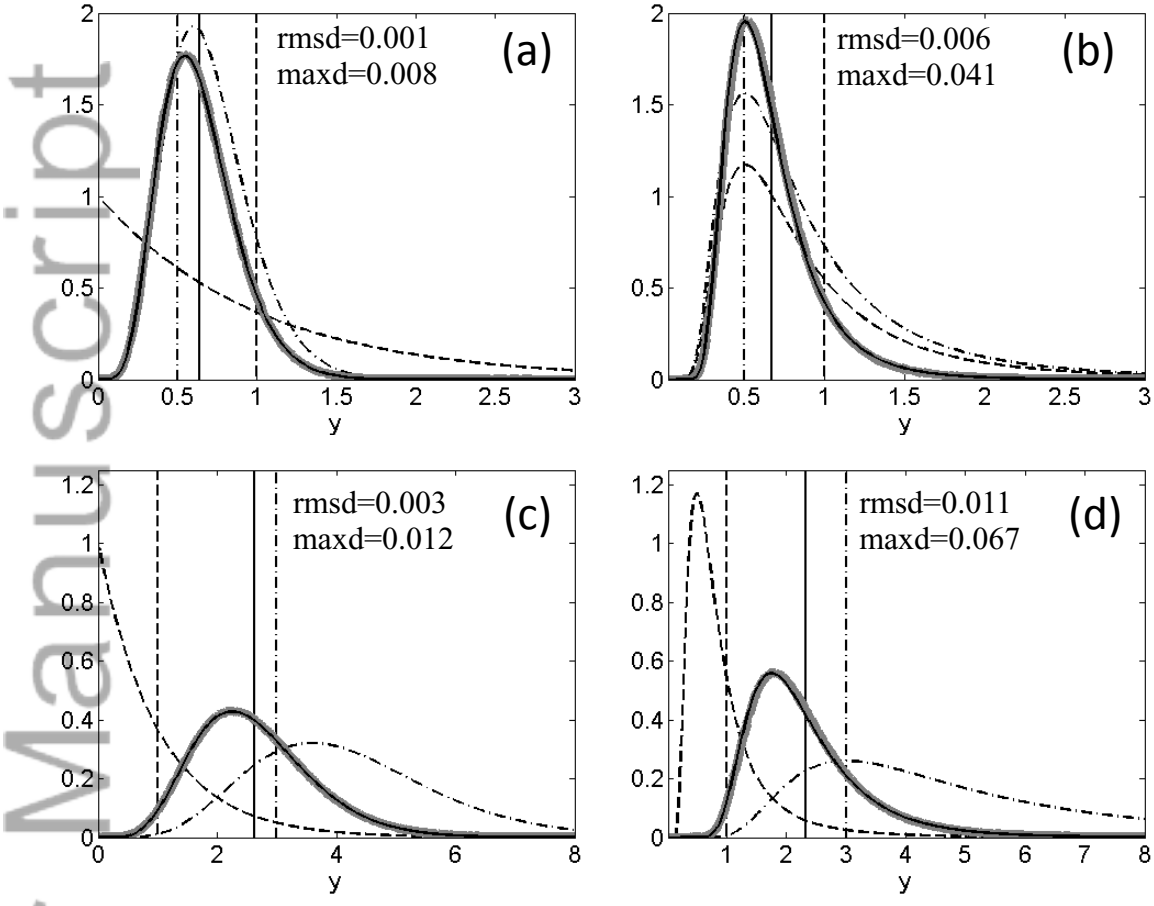

Figure 2: Left two panels pertain to GIG filter, right two panels to IGG filter. In all panels, the mean of the prior is equal to 1 and is indicated by the dashed vertical line while the dashed curve gives the prior pdf. In the top (bottom) two panels, the observed value is equal to 0.5 (3) and is indicated by the dot-dash vertical line while the dot-dash curve gives the likelihood probability density of this observation occurring as a function of the true value of y_j. In all panels, the posterior mean is given by the vertical solid line while the true posterior pdf $\rho\left(y_{ji} \mid y_j^o\right)$ is given by the curved solid line. Note that in all cases the posterior pdf is proportional to the product of the dashed and dot-dashed curves. The corresponding GIG and IGG ensemble based estimates of these pdfs $\rho_{\text{ensemble}}\left(y_{ji} \mid y_j^o\right)$ are given by the thicker grey line. The normalized root mean square

distance (rmsd) and maximum difference (maxd) between the ensemble based and analytical posterior pdfs are written on each panel.

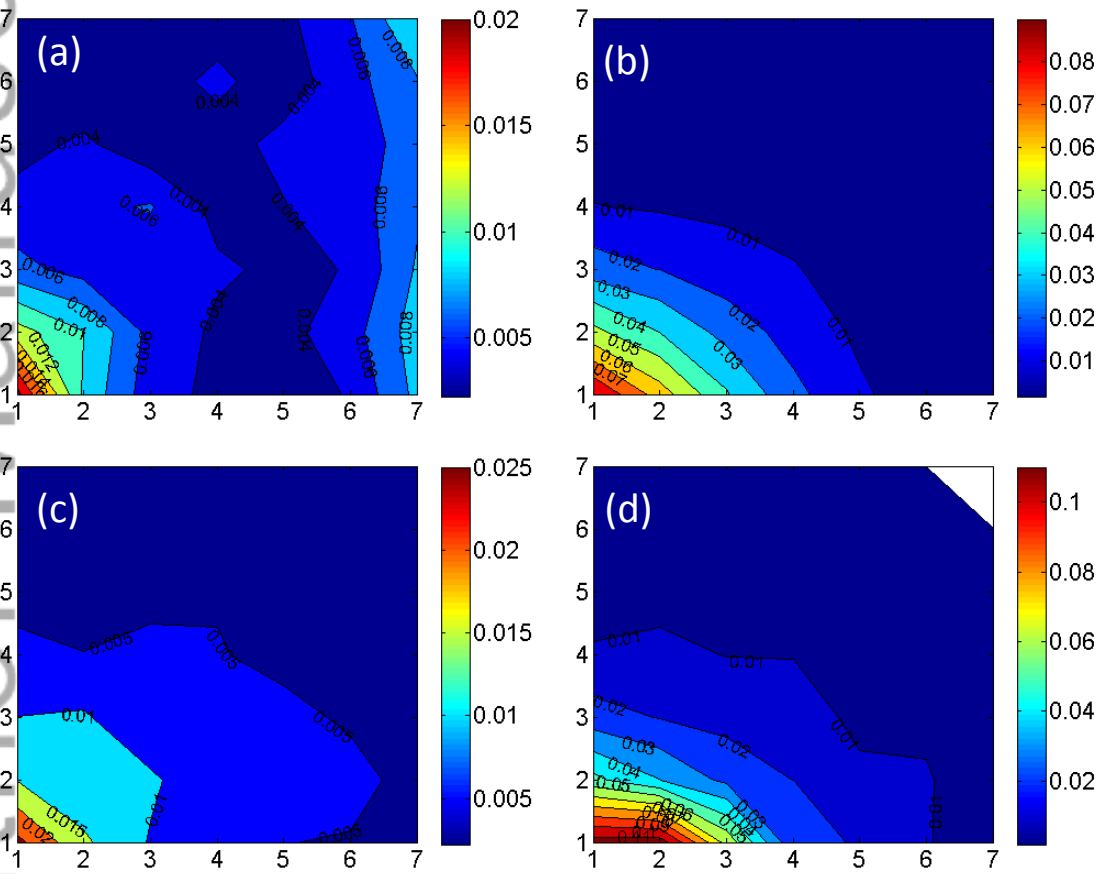

Figure 3: Normalized maximal distance between $\rho\left(y_{ji} \mid y_j^o\right)$ and $\rho_{\text{ensemble}}\left(y_{ji} \mid y_j^o\right)$ as a function of the type 1 relative variances P_j^r and R_j^r. Left two panels pertain to GIG filter,

right two panels to IGG filter. Top two panels pertain to an observed value of 0.5, bottom two panels to an observed value of 3. The prior mean was equal to 1 in all cases. In all panels, the numbers on the abscissa and ordinate axes give the exponents *m* and *n* that define the observation and forecast relative error variance, respectively. Equation (32) {33} defines this relationship for the GIG {IGG} filter.

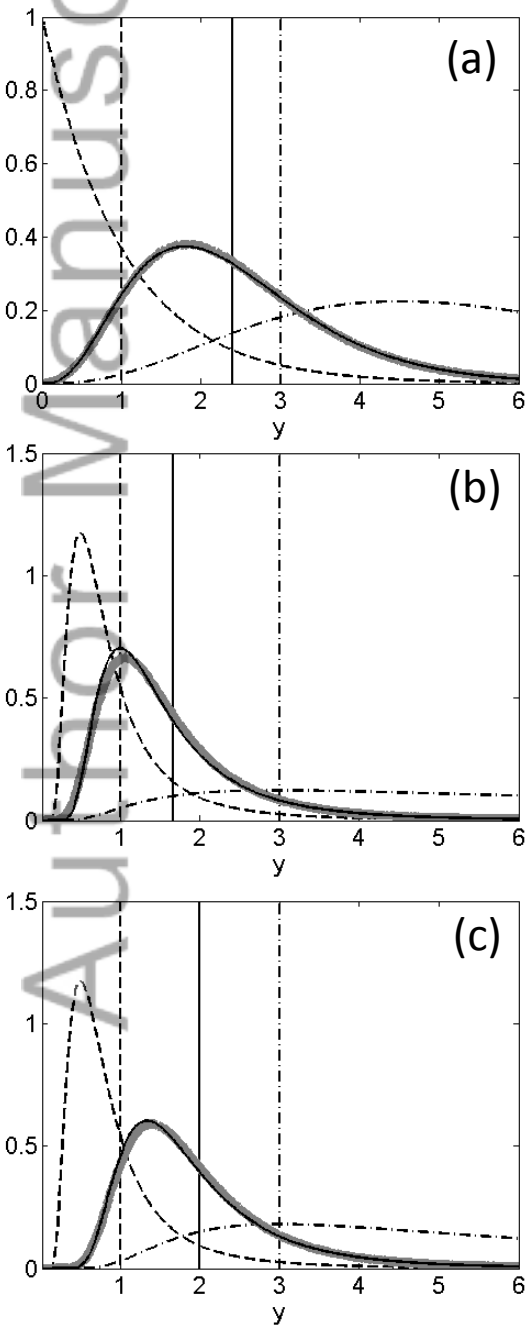

Figure 4: As in Figure 2, the difference between the curve given by the thin black curve $\rho\left(y_{ji} \mid y_j^o\right)$ and the thick grey line $\rho_{ensemble}\left(y_{ji} \mid y_j^o\right)$ indicates the error in the new ensemble filters presented in this paper. In (a) and (b) are shown the worst cases for the GIG and IGG filters, respectively. These worst cases correspond to highly uninformative prior and likelihood distributions for which the type 1 prior and observation relative error variance is given by unity. In (c) is shown, the second worst case for the IGG filter in which the type 1 relative observation error variance is equal to 0.5.

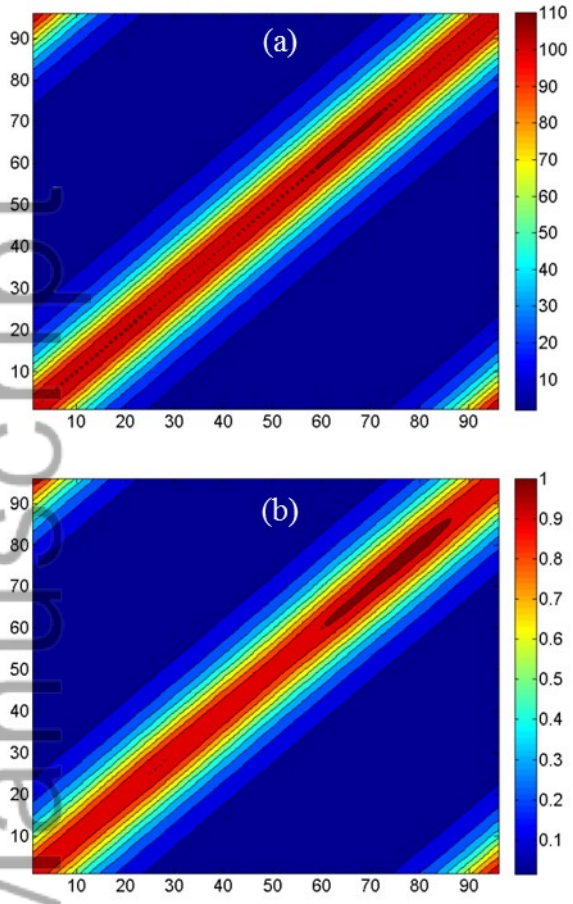

Figure 5: Covariance matrices of random vectors used to create prior distributions of the zonal wind. The prior mean is a random vector with mean zero and covariance matrix given by 5a. For the zonal wind *u* mean depicted by the thick green line in figure 6, the individual prior ensemble members of *u* depicted as colored lines in figure 6 were obtained by adding perturbations with mean zero and covariance matrix given by 5b to the zonal mean.

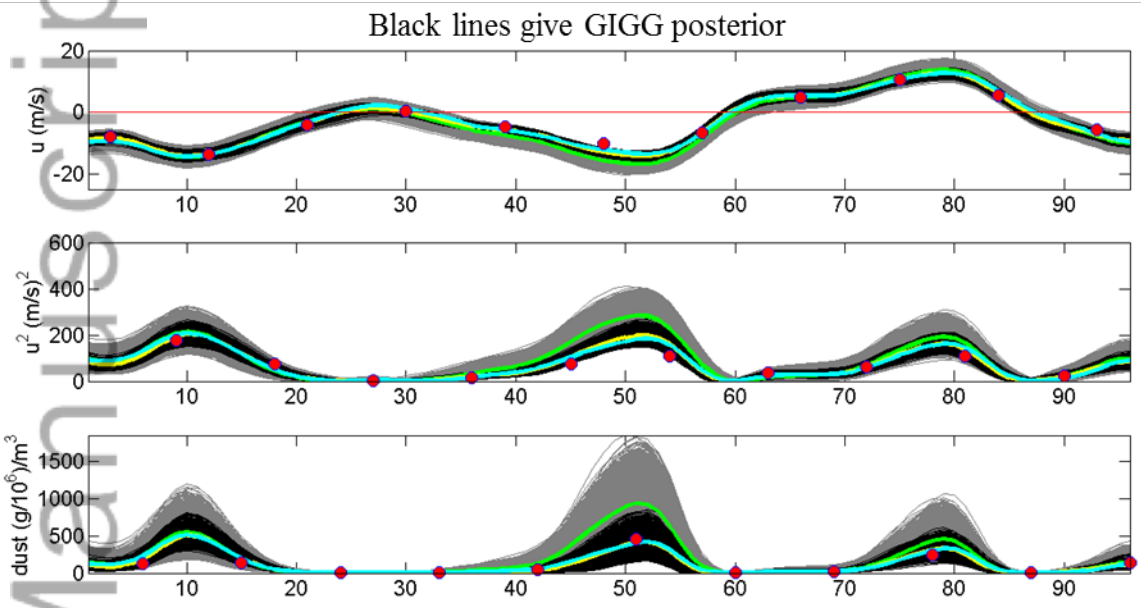

Figure 6: Example of system used to examine filter performance. Abscissa gives the index of the grid point on the periodic model domain. Ordinate gives the value of the model variable. Grey lines depict the 10000 members of the prior ensemble forecast. Thick green line gives the prior forecast ensemble mean. Cyan line depicts the truth. Red dots indicate the error prone observations of the truth. Black lines depict the posterior analysis ensemble obtained from the GIGG filter. Yellow line depicts the posterior analysis ensemble mean. (The yellow line is largely obscured by the truth (cyan line)). Top, middle and bottom panels refer to zonal wind u, u^2 and dust, respectively.

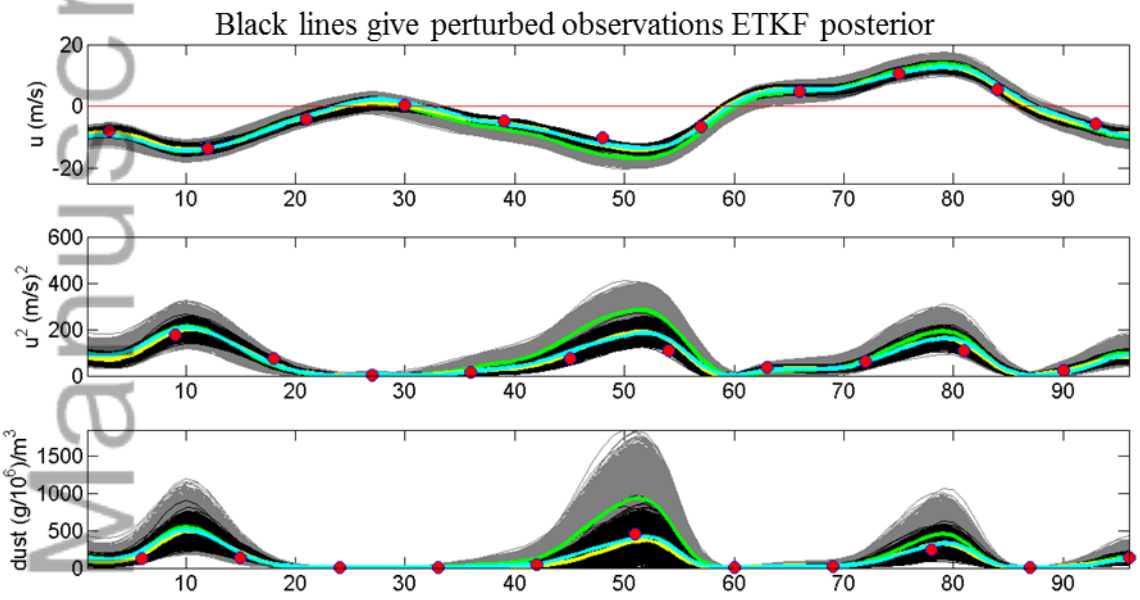

Figure 7: As in Figure 6, but for the perturbed observations ETKF.

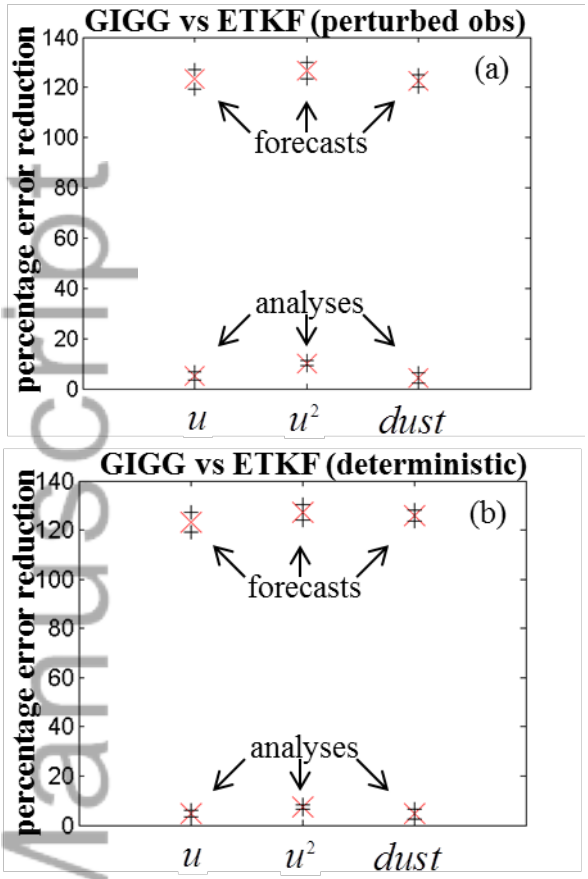

Figure 8: Percentage reduction in average square error of analysis and forecast ensemble mean over the 3584 independent trials. In (a) and (b) is shown the reduction due to using GIGG instead of the perturbed observations ETKF and the deterministic ETKF, respectively. The red crosses give the mean percentage error reduction while the upper and lower "+" signs give the upper and lower 99% confidence intervals for each result.

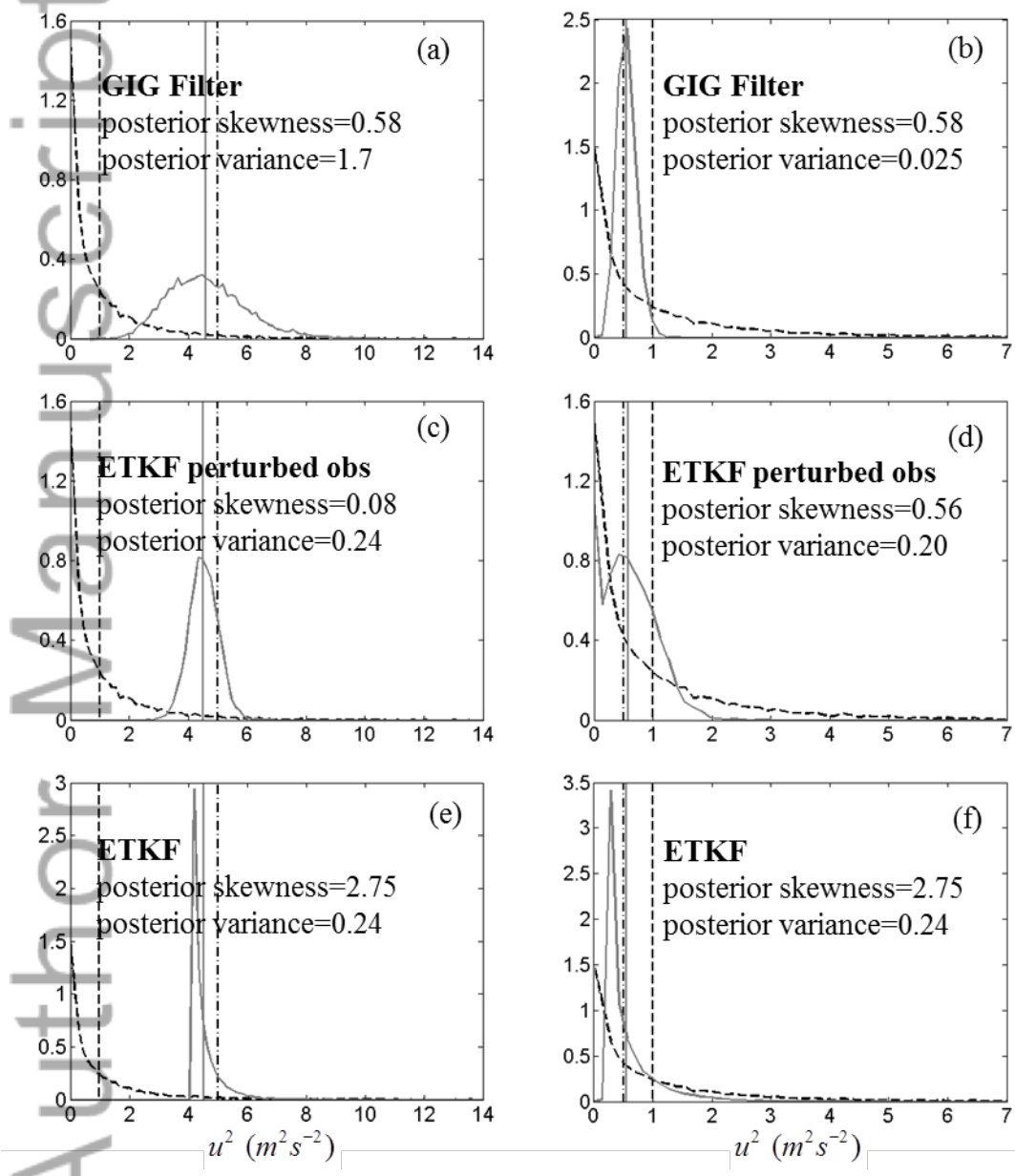

Figure 9: Illustration of the stark differences between the higher order moments of the posterior ensemble distributions produced by the GIGG (panels a and b), perturbed observations ETKF (panels c and d) and deterministic ETKF (panels e and f). In all panels, the black dashed line gives the prior ensemble density, the vertical black dashed line gives the value of the prior ensemble mean (equal to 1 in all cases), the vertical dot-dashed line gives the value of the observation and the solid grey line is the pdf of the posterior ensemble, while the vertical solid grey line gives the posterior mean. The three left (right) panels pertain to an observation of 5 (0.5).