RMetS

Royal Meteorological Society

The equivalent-weights particle filter in a high-dimensional system

M. Ades* and P. J. van Leeuwen

Department of Meteorology, University of Reading, UK

*Correspondence to: M. Ades, Department of Meteorology, University of Reading, Earley Gate, PO Box 243, Reading, RG6 6BB, UK.
E-mail: m.ades@pgr.reading.ac.uk

In general, particle filters need large numbers of model runs in order to avoid filter degeneracy in high-dimensional systems. The recently proposed, fully nonlinear equivalent-weights particle filter overcomes this requirement by replacing the standard model transition density with two different proposal transition densities. The first proposal density is used to relax all particles towards the high-probability regions of state space as defined by the observations. The crucial second proposal density is then used to ensure that the majority of particles have equivalent weights at observation time. Here, the performance of the scheme in a high, 65 500 dimensional, simplified ocean model is explored. The success of the equivalent-weights particle filter in matching the true model state is shown using the mean of just 32 particles in twin experiments. It is of particular significance that this remains true even as the number and spatial variability of the observations are changed. The results from rank histograms are less easy to interpret and can be influenced considerably by the parameter values used. This article also explores the sensitivity of the performance of the scheme to the chosen parameter values and the effect of using different model error parameters in the truth compared with the ensemble model runs.

1. Introduction

Numerical models are often used in the geosciences to simulate and predict the evolution of real-life systems. These numerical models need an initial condition from which to start the simulation and generally this cannot be fully determined by observations of the true system. Data assimilation combines the often noisy observations with information from previous numerical predictions to try to estimate the starting state. Since the initial condition is not fully determined, ideally an understanding of the full uncertainty is required in order to assess the full uncertainty in the evolution of the geophysical system. This uncertainty in the initial condition is represented by a probability density function (pdf), called the posterior pdf.

Historically, numerical models have been solved at the large synoptic scale, where relatively simple linear balances tend to dominate the system. This leads to posterior pdfs that are close to unimodal or where the majority of the posterior probability mass is concentrated around a mode of the pdf. Hence, data assimilation methods have been developed that focus on trying to find the mode of the posterior pdf, such as 3DVar and 4DVar (Talagrand and Courtier, 1987). Alternatively the Ensemble Kalman Filter (EnKF: Evensen, 1994; Burgers *et al.*, 1998) searches for the mean of the posterior pdf, since the linear assumptions made as part of the Kalman filter would implicitly lead to a mean close to the mode.

The above methods do give some indication of the uncertainty in the posterior pdf. Variational methods find the mode by exploring the gradient of the log of the posterior pdf. One of the matrices that can be determined in this process, the Hessian, measures the curvature in the posterior pdf local to the mode and so gives some indication of uncertainty. The EnKF makes the assumption that the posterior pdf is Gaussian and provides an estimate for the covariance. This, together with the mean, completely quantifies the Gaussian distribution and so, under this assumption, the posterior pdf is known. However, both the methods rely on the posterior having one clear mode. Geophysical numerical models are becoming ever more complex, with the inclusion of many more nonlinear processes. The grid resolution of the models is also increasing, as computer capabilities are developed. The two combined mean that the posterior pdf is increasingly likely to be multimodal. This is problematic for variational methods, since there is no guarantee that a global, rather than a local, mode will be found via the gradient methods used. Since the curvature is local, it will also only give probability information about the mode found, rather than an understanding of the full uncertainty in a multimodal posterior. The Gaussian assumptions of the EnKF mean it is not possible to quantify a multimodal pdf accurately with this method.

Particle filters are data assimilation methods that provide a representation of the full posterior pdf. Given that the posterior is the conditional pdf of the state of the system given the

observations, Bayes' theorem can be used to calculate the posterior through the multiplication of the prior pdf and the likelihood. The prior pdf represents our prior knowledge about the state of the system and in particle filters is represented by an ensemble of model runs or particles. The likelihood is the probability density of the observations given a specific model state and forms a weight associated with each of the model runs. So, for example, the mean of the posterior pdf now becomes a weighted mean of the ensemble of particles.

Particle filters have been shown to perform well in small-dimensional systems in applications unrelated to geoscience (see Doucet *et al.*, 2001 and references therein). The problem with using them in geophysical applications is that, in general, the dimension of such systems is extremely large. In high-dimensional systems, particle filters suffer from filter degeneracy, where the ensemble of model runs representing the prior pdf become highly unlikely to sample the high-probability region of the posterior pdf simultaneously. This results in the majority of particles having relative weights very close to zero, with only one particle retaining any weight from the likelihood and hence significance in the posterior. All statistical information on the distribution of the posterior is therefore lost and there is no longer any advantage in using particle filters compared with other data assimilation schemes. This is known as the 'curse of dimensionality' and is what prevents particle filters being considered as realistic alternatives for data assimilation (Snyder *et al.*, 2008). Recent geophysical particle filter research has therefore focussed on trying to ensure that, whilst the ensemble of model runs still represents our prior knowledge of the system, they are also samples from the high-probability region of the posterior (Doucet *et al.*, 2000; Chorin and Tu, 2009; Bocquet *et al.*, 2010; Chorin *et al.*, 2010; Morzfeld *et al.*, 2012).

The equivalent-weights particle filter (Van Leeuwen, 2010, 2011) uses a proposal density to steer the particles towards the high-probability region of the posterior pdf. However, just ensuring particles are in the high-probability region is not enough. It can be shown that, when the number of independent observations is large, the relative weights of the particles will still vary significantly, resulting in filter degeneracy (Ades and Van Leeuwen, 2013). In the equivalent-weights particle filter, the proposal density is additionally used to ensure that similar relative weights are obtained for the majority of particles. Hence degeneracy is avoided and the possibility of representing a potentially multimodal posterior pdf is realized.

The ability of the equivalent-weights particle filter to capture the behaviour of a true solution has been shown for the low-dimensional Lorenz (1963) model (Van Leeuwen, 2011; Ades and Van Leeuwen, 2013), the 40- and 1000-dimensional Lorenz (1995) models (Van Leeuwen, 2011) and more recently the barotropic vorticity equation solved over a 65 536 dimensional grid (Van Leeuwen and Ades, 2013). However, in Van Leeuwen and Ades (2013) the system was fully observed. In reality, in most geoscience systems, observations are much scarcer than the dimension of the system and tend to be irregularly distributed over the state.

In this article, the performance of the equivalent-weights particle filter for the same barotropic vorticity equation is explored, but now the effect of changing the number and distribution of observations is examined. A comparison of the extra computational time required by the equivalent-weights particle filter is made to that required by the Sequential Importance Resampling (SIR) filter. In addition, the sensitivity of the scheme to the chosen parameter values is considered, as well as the consequences of using different stochastic error in the ensemble model runs compared with the pseudo truth of a twin experiment. The last is done in an attempt to replicate the realistic situation of using a geophysical model designed without a complete understanding of the behaviour and error statistics of the truth.

The article is organized as follows. In section 2 we discuss particle filters in general, followed by an explanation of the equivalent-weights particle filter in section 3. In section 4 the

numerical model simulating the barotropic vorticity equation is outlined and results for varying numbers and distributions of observations are shown. Computational times are given and the effects of varying the parameter values and changing the stochastic error statistics used in the ensemble are also discussed. Conclusions are drawn in section 5.

2. Particle filters

Particle filters are based on a Bayes theorem expansion of the conditional posterior pdf of the state of the system x^n given a vector of observations y^n:

$$p(x^n|y^n) = \frac{p(y^n|x^n)p(x^n)}{p(y^n)}, \tag{1}$$

where n is the time index. The prior probability density, $p(x^n)$, represents the information coming from the model equations and can be approximated by a discrete set of N delta functions centred on individual model states known as the particles:

$$p(x^n) \approx \frac{1}{N} \sum_{i=1}^{N} \delta(x^n - x_i^n). \tag{2}$$

The particle approximation of the prior pdf can be combined with Eq. (1) so that the conditional posterior pdf becomes the weighted sum of delta functions:

$$p(x^n|y^n) \approx \sum_{j=1}^{N} w_i^n \delta(x^n - x_i^n), \tag{3}$$

with each particle having a weight given by

$$w_i^n = \frac{p(y^n|x_i^n)}{\sum_{j=1}^{N} p(y^n|x_j^n)}. \tag{4}$$

Since the weights w_i^n are calculated using the likelihood, $p(y^n|x_i^n)$, the closer a model state lies to the observations, the greater the weight of the particle and the more significance it has in representing the posterior pdf. For a more comprehensive overview of particle filters, see Doucet *et al.* (2001).

Filter degeneracy occurs when the particles are distant from the majority of observations and so only the few particles that happen to be closest to the greatest number of observations have non-zero weight. Hence only these few particles have significance in the posterior pdf and the representation of the full uncertainty associated with the state of the system given the observations is effectively lost. The higher the dimension of the state or greater the number of independent observations, the more chance there is of this occurring (Snyder *et al.*, 2008; Ades and Van Leeuwen, 2013). It is this issue that prevents the use of particle filters in large-dimensional geoscience applications, despite the solutions proposed so far (see Van Leeuwen (2009) for a review for the geosciences).

3. Equivalent-weights particle filter

The equivalent-weights particle filter ensures that the majority of particles are both close to all the observations and also contributing significant information on the posterior pdf, thus avoiding filter degeneracy. It consists of two separate steps, the relaxation proposal density and the equivalent-weights proposal density.

3.1. Relaxation proposal density

In general, in geophysical applications, there are multiple time steps between observations. This means that the model prior pdf is the probability density of the model trajectory over all time

steps, rather than being the pdf of just the model state at time n. The time step from $j-1$ to j is considered in order to demonstrate in simple fashion how this is incorporated in the particle filter.

Exploiting the Markovian property of the model, the prior pdf at time j can be written as

$$p(x^j) = \int p(x^j, x^{j-1}) \, dx^{j-1} \qquad (5)$$

$$= \int p(x^j|x^{j-1})p(x^{j-1}) \, dx^{j-1}.$$

Given the samples from $p(x^{j-1})$, trajectories are determined using the model transition density $p(x^j|x^{j-1})$, which equates to moving each particle forward in time according to

$$x_i^j = f(x_i^{j-1}) + d\beta_i^j, \qquad (6)$$

where $f(x_i^{j-1})$ is the deterministic model equation and $d\beta_i^j \sim N(0, Q)$ represents model error with model error covariance Q (see Appendix A1 for details).

This formulation of the model prior, is beneficial since it brings with it the freedom to both divide and multiply the model transition density by a proposal transition density:

$$p(x^j) = \int \frac{p(x^j|x^{j-1})}{q(x^j|x^{j-1}, y^n)} q(x^j|x^{j-1}, y^n)p(x^{j-1}) \, dx^{j-1}. \qquad (7)$$

In theory, the proposal transition density $q(x^j|x^{j-1}, y^n)$ can be any density, provided the support is at least equal to that of $p(x^j|x^{j-1})$. However, since the aim is to ensure that the majority of particles are close to all observations, it is logical to choose a proposal density that incorporates observational information at some future time n. Given the samples from $p(x^{j-1})$, trajectories can now be calculated using the proposal transition density $q(x^j|x_i^{j-1}, y^n)$ rather than the model transition density $p(x^j|x_i^{j-1})$. In the equivalent-weights particle filter, this results in each particle being moved forward in time using the adapted model equation

$$x_i^j = f(x_i^{j-1}) + B(\tau)(y^n - h(x_i^{j-1})) + \widehat{d\beta}_i^j, \qquad (8)$$

where $h(x_i^{j-1})$ is the observation operator. This works to provide a slight relaxation $B(\tau)(y^n - h(x_i^{j-1}))$ towards the observation at time n. $B(\tau)$ controls the strength of the relaxation and spreads the observation information from observed to unobserved variables. The stochastic error $\widehat{d\beta}_i^j$ is now drawn from $N(0, \widehat{Q})$.

Sampling from the relaxation proposal transition density in Eq. (7) means that an additional factor is present in the representation of the model prior. This is a value that can be calculated (dependent on the sample chosen from the proposal transition density) and forms an additional weight on the particle. The model prior at time n over multiple time steps now consists of the weighted trajectories,

$$p(x^n) = \sum_{i=1}^N \widehat{w}_i^n \delta(x^n - x_i^n), \qquad (9)$$

with

$$\widehat{w}_i^n = \prod_{j=1}^n \frac{p(x_i^j|x_i^{j-1})}{q(x_i^j|x_i^{j-1}, y^n)}. \qquad (10)$$

These weights can vary considerably between particles and can lead to filter degeneracy before the value from the likelihood in the posterior (see Eq. (4)) is even taken into account (see Appendix A1 for further details).

3.2. Equivalent-weights proposal density

To account for the disparity in weights coming from the model prior, an additional proposal density is introduced. For the majority of time steps, the relaxation proposal density as described above is used. However, in the last time step before an observation, a proposal density is used that ensures the majority of particles have equivalent weights in the posterior pdf.

The final proposal density $q(x^n|x_i^{n-1}, y^n)$ can be used to set

$$p(y^n|x_i^n)p(x_i^n|x_i^{n-1})\widehat{w}_i^{n-1} \approx w^{\text{target}}, \qquad (11)$$

where $\widehat{w}_i^{n-1}$ are the weights from the model prior Eq. (10) taken at time $n-1$ and w^{target} is a target weight. The question is what value w^{target} should have. Since the weight of a particle effectively defines the significance of that particle in the posterior, ideally each particle should hold the maximum possible weight. This equates to a different target weight, $w_i^{\text{target}} = w_i^{\text{max}}$, for each particle. However, it can be shown that even without including the value coming from $-\log(\widehat{w}_i^{n-1})$, w_i^{max} will vary sufficiently to cause filter degeneracy when the number of independent observations is large (Ades and Van Leeuwen, 2013). In the equivalent-weights particle filter, the aim instead is to ensure almost equal weights for the majority of particles. Hence a common value of w^{target} is chosen (see Appendix A2), which will be equal to or less than w_i^{max} for a chosen percentage of particles but for which Eq. (11) can be solved for that percentage. The remaining particles, for which Eq. (11) cannot be solved with the chosen value of w^{target}, will be ignored for now but will return via resampling (Kitagawa, 1996).

In the equivalent-weights proposal density, this leads to the chosen percentage of particles being set equal to

$$x_i^n = f(x_i^{n-1}) + \alpha_i K(y^n - Hf(x_i^{n-1})) + Q^{1/2}\xi_i^n, \qquad (12)$$

where H is the linearization of the observation operator $h(x)$, $K = QH^T(HQH^T + R)^{-1}$ and

$$\alpha_i = 1 + \sqrt{1 - b_i/a_i}, \qquad (13)$$

in which $b_i = 0.5d_i^T R^{-1}d_i - \log(w^{\text{target}}) - \log(\widehat{w}_i^{n-1})$, $a_i = 0.5d_i^T R^{-1}HKd_i$ and $d_i = y^n - Hf(x_i^{n-1})$ (see Ades and Van Leeuwen, 2013 for details). Those particles that are unable to achieve the target weight are simply moved according to the model equations

$$x_i^n = f(x_i^{n-1}) + Q^{1/2}\xi_i^n, \qquad (14)$$

since their weights are so small that they will effectively be discarded in the resampling stage. In both cases, $\xi_i^n \sim (1-\epsilon)\widetilde{U}_k(0, \gamma_U) + \epsilon N(0, \gamma_N I)$ comes from a mixture density, where $\epsilon = 0.001/N$ and $\gamma_U = \gamma_N = 10^{-5}$ (see Appendix A2 for details).

The weight associated with each particle at the final time can now be calculated as

$$w_i^n = \frac{\widehat{w}_i^{n-1}p(x_i^n|x_i^{n-1})p(y^n|x_i^n)}{q(x_i^n|x_i^{n-1}, y^n)}, \qquad (15)$$

where $q(x_i^n|x_i^{n-1}, y^n)$ is evaluated according to Eq. (16):

$$q(x_i^n|x_i^{n-1}, y^n) =$$

$$\frac{1}{|Q^{1/2}|} \begin{cases} \frac{1-\epsilon}{(2\gamma_U)^k} + \frac{\epsilon}{(2\pi)^{k/2}\gamma_N} \exp\left(-\frac{1}{2}\xi_i^{nT}(\gamma_N^2 I)^{-1}\xi_i^n\right), \\ \quad \text{if } \xi_{i,j}^n \in [-\gamma_U, \gamma_U], \qquad \forall j = 1, \ldots, k \\ \frac{\epsilon}{(2\pi)^{k/2}\gamma_N} \exp\left(-\frac{1}{2}\xi_i^{nT}(\gamma_N^2 I)^{-1}\xi_i^n\right), \\ \quad \text{otherwise,} \end{cases}$$

$$(16)$$

and should now be close to equivalent for the percentage of particles chosen. A detailed overview of the steps required to implement the equivalent-weights particle filter is included in Appendix A3.

4. Application to the barotropic vorticity equation

4.1. Model specification

The model used is based on the barotropic vorticity equation, which governs the vorticity in a two-dimensional incompressible homogeneous fluid:

$$\frac{Dq}{Dt} = \frac{\partial q}{\partial t} + \mathbf{u}.\nabla q = d\beta, \tag{17}$$

where $\mathbf{u}$ is the two-dimensional velocity vector (u, v) describing the flow in the $x-y$ plane and q is the vorticity component along the z-axis. We assume that the truth is a random variable and $d\beta$ arises from unknown subgrid sources and sinks of vorticity. In addition, since the fluid is incompressible, a stream function can be defined as

$$u = -\frac{\partial \psi}{\partial y}, \qquad v = \frac{\partial \psi}{\partial x}. \tag{18}$$

Combining these with the relationship between the vorticity and the velocity field,

$$q = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}, \tag{19}$$

leads to the equations

$$\frac{\partial q}{\partial t} - \frac{\partial \psi}{\partial y}\frac{\partial q}{\partial x} + \frac{\partial \psi}{\partial x}\frac{\partial q}{\partial y} = d\beta,$$
$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} = q \tag{20}$$

to be solved at every time step. Vorticity is updated in the first equation using a semi-Lagrangian scheme with time step $\Delta t = 0.04$. The cubic interpolation used as part of this semi-Lagrangian scheme introduces an implicit hyper-diffusion into the system of equations (Durran, 1999). The random component $d\beta$ is then integrated using a simple Euler–Maruyama scheme. The second equation is solved via an inversion of vorticity using FFTs to give the stream function, which in turn can be used to generate the velocity field for the next time step.

The equations are solved over a double periodic domain of 256 by 256 grid points with grid spacing $\Delta x = \Delta y = 1/256$, which gives a state dimension of 65 536. The initial condition is a quasi-random vorticity field, which has spectral coefficients significantly non-zero only in the range of wavenumbers 2–6, with a peak at wavenumber 4. It is then normalized such that the integral of squared vorticity over the domain is equal to unity (Figure 1). This initial condition was chosen, since it results in chaotic flow structures. In the non-stochastic version of the barotropic vorticity equation, energy cascades from small to large scales and the flow becomes more organized as time progresses. The presence of the stochastic term $d\beta$ means that energy continues to be injected at small scales throughout the 1150 time steps of the experiment. Although less chaotic behaviour starts to be seen by the end of the time period, this additional energy means that the flow remains fully turbulent throughout the majority of the experiment.

The stochastic error $d\beta$ is drawn from a multivariate Gaussian with mean zero and covariance matrix $Q = V_\beta \widetilde{Q}$. $\widetilde{Q}$ is a correlation matrix generated using a two-dimensional Second Order Autoregressive (SOAR) spatial correlation based on the distance between grid points with a length-scale of five grid points. At each time step, it is scaled by $V_\beta = 0.025^2 \Delta t$, where $\Delta t = 0.04$. The value of 0.025^2 was chosen as it ensured that the time-averaged l^2 norm of the random error, $\|d\beta\|^2$, was approximately 10% of the vorticity difference caused by the purely deterministic model equations, $\|f(x_i^{n-1})\|^2$. The $\widehat{Q}$ used
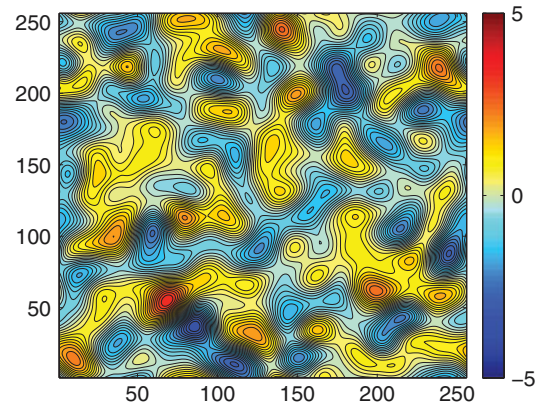

Figure 1. The model state used to initialize the truth run of the barotropic vorticity equation. This figure is available in colour online at wileyonlinelibrary.com/journal/qj

as part of the proposal transition density was chosen to be the same as Q. Using this random error, the decorrelation time-scale of the system was determined to be about 42 time steps. The decorrelation time-scale was calculated by averaging the auto-correlation function,

$$r_i^k = \frac{\sum_{t=1}^{n-k}(x_i^t - \bar{x}_i)(x_i^{t+k} - \bar{x}_i)}{\sum_{t=1}^{n}(x_i^t - \bar{x}_i)^2}, \tag{21}$$

for 100 of the 65 536 variables, where i refers to the individual variable and k is the auto-correlation time difference (Wilks, 1995). The correlation time k at which this averaged correlation function dropped below $1/e$ was taken as the decorrelation time.

In order to ensure a suitably difficult nonlinear data assimilation problem was being solved, observations were made every 50 time steps, which is greater than the calculated decorrelation time. The effect of varying the spatial distribution of the observations forms the main part of this article and will be considered in section 4.2. The observations were obtained from a truth run and independent Gaussian measurement noise with standard deviation 0.05 was added to typical observation vorticity values of $\mathcal{O}(1)$.

In the relaxation proposal transition density, the function $B(\tau)$, which controls the strength and spread of the relaxation term, was chosen as

$$B(\tau) = b\tau Q H^T R^{-1}, \tag{22}$$

with $\tau = (t^j - t^0)/(t^n - t^0)$ being zero at the previous observation time, t^0, and increasing linearly to 1 at the new observation time t^n. The observation-error covariance matrix R is included in $B(\tau)$ both to normalize $y^n - h(x^{j-1})$ by the errors in the observations and to ensure the relaxation term has the correct physical dimension. The b represents a scaling factor, by which the strength of the relaxation towards the observation can be controlled. The effect of changing b is discussed in section 4.4.1, but it is initially set to 0.2.

In the equivalent-weights proposal transition density, 80% of particles were retained at analysis time, although the effect of changing this value is considered in section 4.4.2. The parameters ϵ, γ_U and γ_N were chosen to be: $0.001/N$ (where N is the number of ensemble members), 10^{-5} and 10^{-5} respectively. It was found in the Lorenz (1963) model that increasing these values ultimately leads to filter degeneracy (Ades and Van Leeuwen, 2013). This result is not expected to differ when applying the equivalent-weights particle filter to the barotropic vorticity equation and so these parameter values have been kept throughout. No instances of filter degeneracy were seen at any of the time steps considered.

4.2. Partial observations

The experiment where every variable is observed across the full state every 50 time steps has already been reported on in Van

Leeuwen and Ades (2013). The results showed the effectiveness of the equivalent-weights particle filter in representing the true model state of a twin experiment by the mean of the ensemble. A close-to-uniform rank histogram also indicated that the statistics of the ensemble were appropriate. However, the model error covariance matrix Q not only assumed independent model errors but was also scaled too large. This meant the l^2 norm of the model error was comparable to the l^2 norm of the deterministic model equations. Q has now been scaled appropriately, so that the random error is approximately 10% of the size of the deterministic movement and Q is also now fully correlated. These changes have very little effect on the statistical results found in Van Leeuwen and Ades (2013) and so the fully observed system is not discussed further here.

In order to replicate more realistic observation conditions, two different scenarios have been considered. The first looks at the effect of reducing the number of observed variables uniformly over the grid. The cases where observations are available for every second (1/4 variables observed), fourth (1/16 observed) and eighth (1/64 observed) grid point in both the meridional and zonal directions are shown. The second scenario looks at the effect of first making no observations in the southeast quarter of the domain and then making no observations over the entire east half of the state.

The purpose of using a particle filter as a data assimilation scheme is to enable a representation of a potentially multimodal posterior pdf. For a system as high-dimensional as the barotropic vorticity equation considered here, it is not easy to generate an estimate of the true posterior with which to compare the particle filter representation (Law and Stuart, 2012). Instead, different measures must be used to judge the performance of the equivalent-weights particle filter. One such measure is to compare the mean of the 32 particle ensemble with the true state of the system in a twin experiment (Figure 2). Since the barotropic vorticity equation is nonlinear, a comparison of the mean and truth does not indicate how well the equivalent-weights particle filter represents the posterior. However, if there is no relation between the two then the particles are clearly not sampling from the high-probability region of the posterior and the scheme is not performing as we would wish.

It was found that, as the number of observed variables was decreased, both uniformly and with increasingly larger sections of the state unobserved, the mean of the equivalent-weights particle filter continued to be successful in matching the larger scale features present in the truth. The filament detail was also largely retained when every second and fourth variable was observed (Figure 2(b) and 2(c)) and with every other variable over three quarters of the state (Figure 2(e)), again using the similarity of the mean to the truth. When this is reduced to observed variables every eighth grid point (Figure 2(d)) or every other variable over half the state (Figure 2(f)), the filament behaviour starts to be lost, although the larger structures are still clearly present and the truth is recognisable. The results are shown at one analysis time, but these conclusions hold for all observation times examined.

The success of the equivalent-weights particle filter in representing the truth becomes more apparent if individual particles are considered when only half the state is observed (Figure 3). Although the finer filament structures are still missing in some areas of the full state, in general both the finer detailed and larger scale structures are both present in the individual particles. However, the positioning of these structures differs between particles and compared with the truth. It is this uncertainty in positioning that is the major contributor to the loss of detail in the mean of the ensemble. This demonstrates the weakness of using the similarity to the truth as a 'pseudo' measure with which to judge the ability of the equivalent-weights particle filter to represent the full posterior pdf. The posterior pdf is likely to be broad in areas where there are no observations to constrain it and hence a smooth mean is to be expected.

The uncertainty of the positioning of features can also be seen in the marginal posterior pdfs for individual variables (Figure 4).

The difference in posterior distributions when every other variable is observed over the entire state, compared with only the left half of the state, is apparent. When the entire state is partially observed, the distributions are all unimodal and surround the truth and observations. In contrast, when only half the state is partially observed, the posterior pdfs are much more widely distributed. When the variable is within the observed half of the state (Figure 4(b)), the ensemble correctly finds the observation. This is mainly due to the relaxation term added to the model equations, although the equivalent-weights step also has an effect on the posterior pdf. In the unobserved half the posterior pdfs are much more dispersive (Figure 4(d)) and can show multimodal behaviour (Figure 4(f)). In both cases the absence of filter degeneracy can be verified by the clear spread present in the ensemble.

Another method for judging the spread of the ensemble is to examine the squared difference between the true state and the ensemble mean compared with the ensemble variance. This was done for both the prior and posterior ensemble at time step 600 (Figure 5). In both cases it was found that the variance was higher where there was greater difference between the ensemble mean and the truth. This indicates that the ensemble is behaving correctly by being more dispersive in areas of higher truth minus mean mismatch. In general, the spread around the truth was overestimated at several locations but underestimated elsewhere and the average over the field was close to that of the variance. Since the (truth − mean)^2 is one realization of the statistic represented more generally by the variance, this is acceptable. It is also clear that the prior has a much larger variance than the posterior ensemble, as would be expected; however, similar patterns in variance are observed in both ensembles.

Rank histograms were used to investigate the quality and spread of the ensemble over all observation times (Figure 6). Rank histograms are histograms of the position of the truth in the ranked ensemble (Anderson, 1996; Hamill and Colucci, 1997; Talagrand et al., 1999; Hamill, 2000). A sloped histogram shows that there may be systematic biases, a humped histogram is evidence of too much spread in the ensemble and a U-shaped or concave histogram evidence of too little spread. They can be difficult to interpret, since there can be many reasons why certain patterns in the histograms are seen (Hamill, 2000). However, in the absence of better methods they are used here as an indication of whether the equivalent-weights particle filter is encapsulating the true spread of the posterior pdf. To generate the rank histograms for the barotropic vorticity model, the ensemble values were ranked and compared with the truth every analysis time and every 16th variable in each row and column of the field, assuming they were close to independent.

In all the rank histograms, peaks appear on either side of what remains a relatively flat central histogram. Although these peaks are small when every other variable is observed over the full state, they grow as the number of variables observed is decreased. On first inspection, these results could be taken as an indication of underdispersion. This particular distinctive shape, however, suggests more than just insufficient spread. Instead of the gradual U-shape commonly associated with underdispersive ensembles, here two peaks surround a flat central section. The level central section would suggest a good spread but the peaks imply that there are insufficient outliers in the ensemble, particularly as the number of observations decreases. This effect and the cause is discussed further in section 4.4.2, but the relevant point is that these results suggest that in most cases the high-probability region of the true posterior is being captured by the equivalent-weights particle filter but truncation of the ensemble is preventing it from encapsulating the far-out tails of the pdf.

It should be stated that the above proposition is a hypothesis, since there could potentially be other reasons why such a shape and relationship is seen in the rank histograms. It should also be noted that, as the number of observations are reduced, the data assimilation problem becomes increasingly difficult. That the equivalent-weights particle filter is able to achieve a flat central
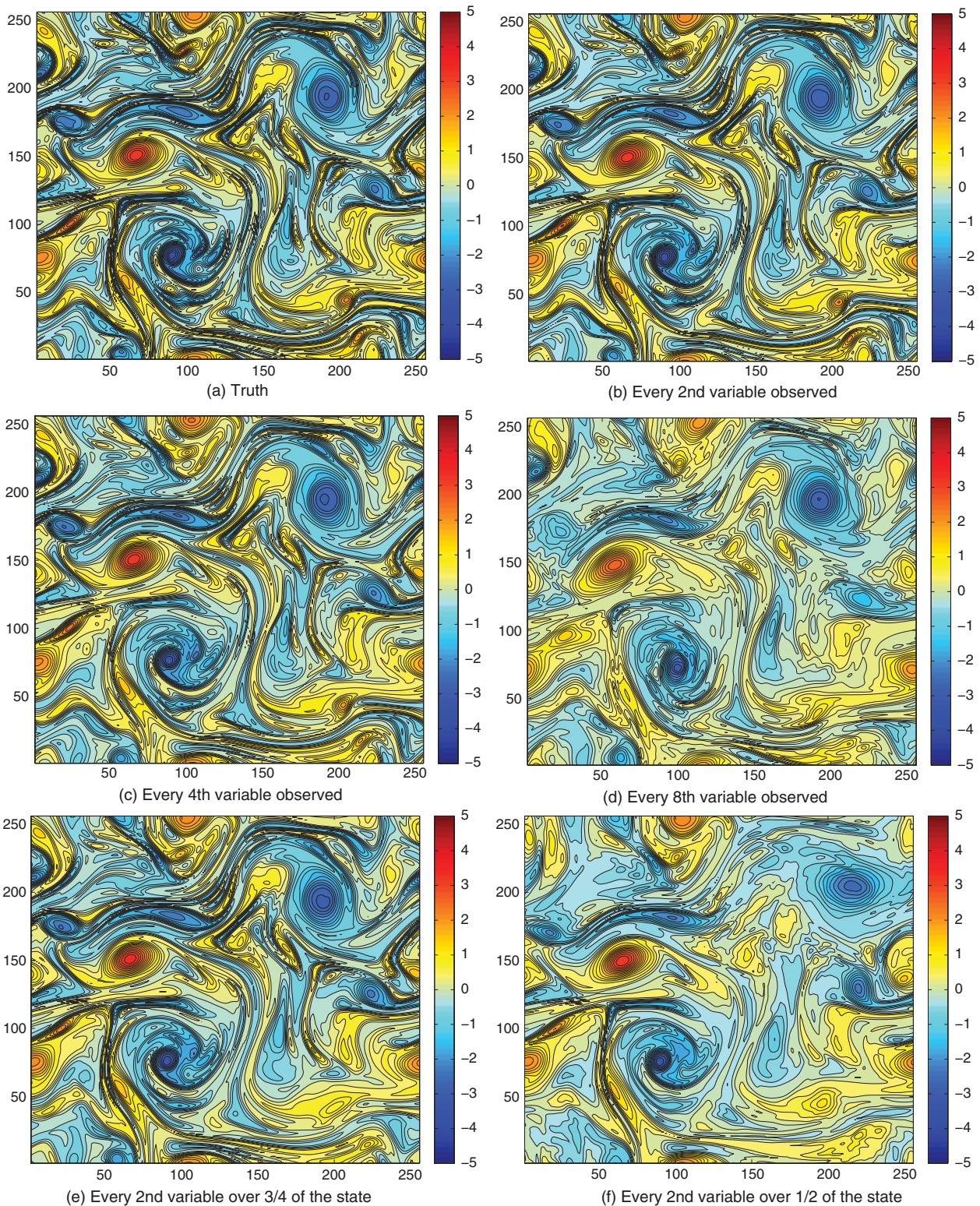
ⓒ 2014 The Authors. *Quarterly Journal of the Royal Meteorological Society*
published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

Q. J. R. Meteorol. Soc. (2014)

Figure 2. (a) The 'pseudo' truth used in the twin experiment compared with the 32-ensemble-member generated mean at time step 600. With observations of every (b) second and (c) fourth variable in each row and column, very little difference is seen between the mean of the ensemble and the truth. When every (d) eighth variable is observed (1024 observed variables of the 65 536), the large-scale structures and some of the filament detail of the truth are still retained. With the bottom right-hand quarter of the state being unobserved (with every other variable observed over the remainder of the state), the details of the filament features are lost in this quarter (see panel (e)). This is more apparent when every other variable is observed only over the left side of the state (see panel (f)). This figure is available in colour online at wileyonlinelibrary.com/journal/qj

section in rank histograms when only every fourth variable is observed (6% of variables) is an important achievement.

Finally, to check that the equivalent-weights proposal density is only affecting the state where there is information from observations, the state of an individual particle can be plotted both before and after the equivalent-weights step (Figure 7). It was found that when the left half of the state was observed, only negligible differences could be seen in the right half of the state when the equivalent-weights step was applied (Figure 7(c)). This

confirms that changes to the state are not being made when there is no information to justify modifications.

4.3. Relative computation times

The particle filter most commonly considered as the basic particle filter is the SIR filter described in section 2. Although filter degeneracy occurs when such a scheme is applied to large-scale systems, such as the barotropic vorticity model discussed here,
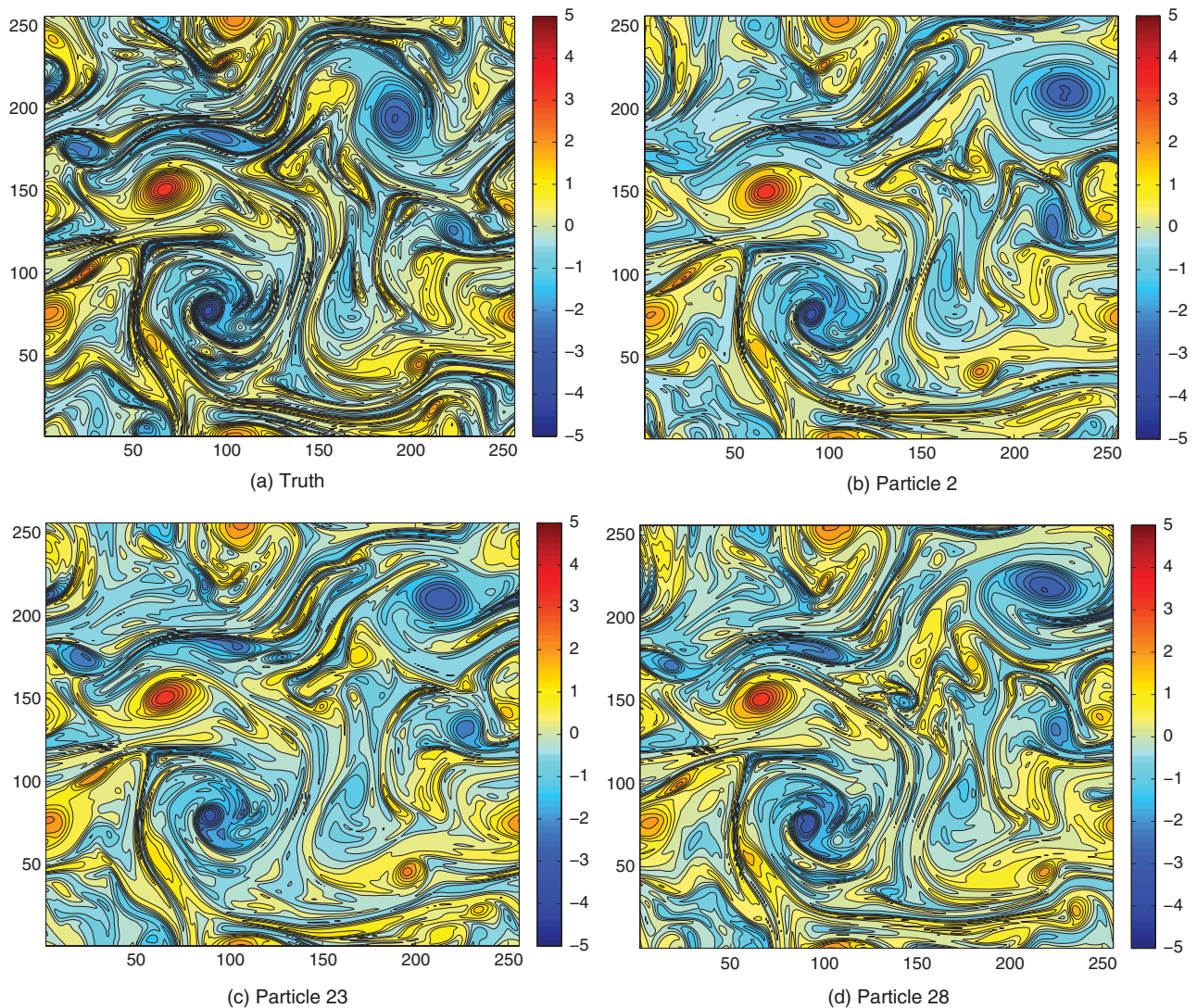
Figure 3. (a) The 'pseudo' truth used in the twin experiment compared with (b) Particle 2, (c) Particle 23 and (d) Particle 28 at time step 600 when every other variable is observed over the left half of the state and no observations are made of the right half. The unobserved right half of the state is comparable to the truth, although the particles are not able to catch the finer filament behaviour and there is some uncertainty over the positioning of the larger scale features (see for example the anti-cyclonic feature in the top right corner). This figure is available in colour online at wileyonlinelibrary.com/journal/qj

it is informative to use it as a standard when assessing the computational costs of a scheme. The equivalent-weights particle filter may theoretically be a viable data assimilation scheme in large-scale systems, but if in reality the computational time is too high then the scheme will never be a possibility for actual geophysical applications. Table 1 shows a comparison of the CPU time for the two separate constituent parts of the SIR filter compared with the equivalent-weights particle filter. The two parts are the integration period between observations, which can be performed in parallel since the individual particles or model runs are independent, and the serial analysis period performed at observation time. Two different observation scenarios have been considered for the equivalent-weights particle filter; a fully observed system (65 536 observed variables) and an 'every other variable observed' scenario (16 384 observed variables). Using the SIR filter, decreasing observed variables will simply decrease the total computational time, as the likelihood needs to be calculated for fewer variables. In contrast, the computational time for the equivalent-weights particle filter will initially increase once the system is not fully observed. This is due to the requirement to invert $(HQH^{\mathrm{T}} + R)$ as part of the equivalent-weights step. Since we use double periodic boundary conditions, a fast Fourier transform (FFT) can be used to do this inversion when the system is fully observed. However, when the system is less than fully observed, an LU decomposition is used, which can result in an increase in the computation time dependent on the number of observed variables.

The additional time seen for the equivalent-weights particle filter in the integration period results from the use of the relaxation proposal density, as opposed to the model transition density used by the SIR filter. In particular, it is due to the need to calculate $p(x_i^j|x_i^{j-1})/q(x_i^j|x_i^{j-1}, y^n)$ at every time step.

In the analysis period, the equivalent-weights particle filter uses the equivalent-weights proposal density, an inherent part of which is calculating $(HQH^{\mathrm{T}} + R)^{-1}(y^n - Hf(x_i^{n-1}))$; see Eq. (A16). This calculation is largely responsible for the increase in CPU time when compared with the SIR filter. However, it should be noted that this is in part due to the current inclusion of this step under the necessarily serial analysis computations, which means the calculation is done for each particle in turn. Since, for this particular calculation, there is no interaction between the particles, it would also be possible to include it under the parallel part of the code and hence reduce the overall CPU time associated with the analysis.

Taking these values over a fully observed model run of 1150 time steps leads to a total wall time of 45 min for the SIR filter and 51 min for the equivalent-weights particle filter. If the number of observed variables is reduced to a quarter of the full state, then the wall time for the equivalent-weights particle filter increases to 61 minutes, assuming that the initial calculations required by the LU decomposition are performed offline. Hence, although the equivalent-weights particle filter does result in additional computation time, it does not represent
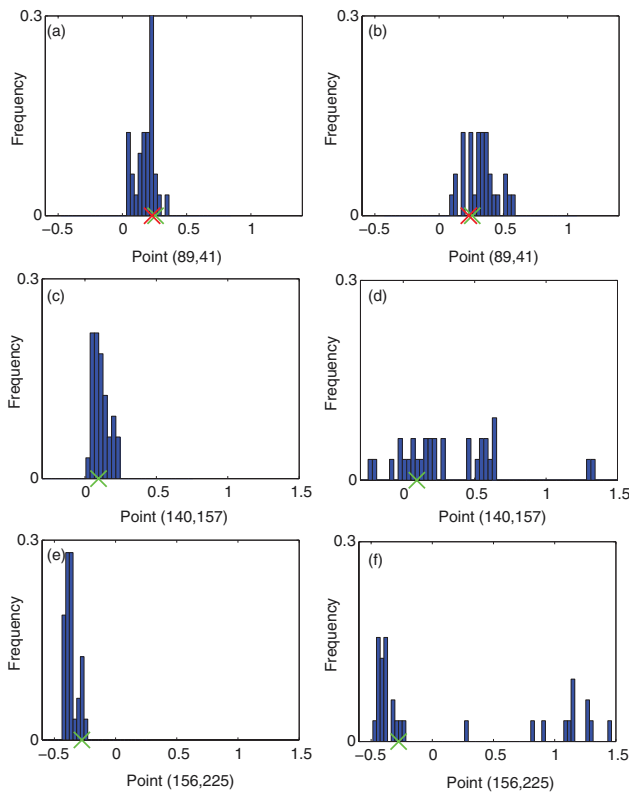
Figure 4. The marginal posterior pdfs of three different variables with observations every other variable over the entire state (a, c, e) compared with the posterior pdfs of the same variable when only half the state is observed (b, d, f). The pseudo true state for each variable is shown by a cross (green in the online article) and the observation by a cross (red in the online article) and the pdf is taken at time step 600 with 32 particles. The top pdf is from a variable (89, 41) that is observed under both scenarios whereas the bottom two points ((140, 157), (156, 225)) are unobserved in both scenarios. More significantly, the unobserved variables are also in the unobserved part of the state in the second scenario. (d) Additional spread and (f) multimodal behaviour caused by not observing whole sections of the state are evident. This figure is available in colour online at wileyonlinelibrary.com/journal/qj

a substantial increase over the SIR filter. Obviously these values will be application-dependent and will depend to a large extent on the structure of the covariance matrix Q. Further comparisons will be made as the equivalent-weights particle filter is explored in different applications in future articles.

4.4. Parameter sensitivities

There are several different parameters that form an integral part of the equivalent-weights particle filter. These are discussed fully in Ades and Van Leeuwen (2013), where specific values for some of the parameters were determined and are detailed at the end of section 4.1. The remaining parameters, b and the percentage of particles retained under the equivalent-weights step, are more dependent on the model the scheme is applied to and can be used to tune the performance of the scheme dependent on the measure chosen. This section looks in more detail at the effects of changing these two parameters in the barotropic vorticity equation.

4.4.1. Relaxation proposal density

The relaxation term has been chosen in this case to be

$$B(\tau)(y^n - Hx_i^{n-1}), \qquad \text{where } B(\tau) = b\tau QH^T R^{-1}, \qquad (23)$$

and the first parameter to be examined is the relaxation factor b, which controls the strength with which each particle is relaxed towards the observations. Hence, changing its value affects the balance between the movement determined by the model equations and the non-physical movement prescribed by the relaxation term.

Table 2 shows the l^2 norm of the movement caused by the model equations, the additive random error and the relaxation term as the strength of the relaxation term grows. The values are taken in the time step immediately prior to an observation, i.e. when the linearly increasing relaxation term is at a maximum, and averaged across all particles and all observation times. It is clear that, regardless of the strength of the relaxation term, the size of the movement associated with the term is an order of magnitude smaller than the movement caused by the addition of the random error. Hence the model equations dominate and the relaxation term can be seen as an adaptation to the random error to ensure we are sampling particles in the locality of the observations. Since the model equation is the controlling term, the majority of the information from the model prior is being retained, despite the relaxation towards the observation.

The size of the movement caused by the relaxation term may be smaller than the model movement, but the effect of increasing b is still clearly evident in the trajectories of the particles. Figure 8 shows the trajectories of all the particles compared with the truth between time steps 550 and 650 for the observed variable at grid point (63 192). With $b = 0.05$, the ensemble is clearly dispersive and follows the truth well. When b is increased to 0.4 and hence a stronger relaxation to the observations is applied, the ensemble has notably less spread and is no longer always able to follow the truth.

The ensemble mean and rank histograms can again be used as a proxy to judge the effect of changing the value of the parameter b on the posterior pdf (Figure 9). Although little difference is evident in the mean of the ensembles as b is increased, the reduction in spread seen in the trajectories is reflected in the rank histograms. When $b = 0.05$, the histogram is slightly humped, suggesting the ensemble is overdispersive. A uniform distribution is seen with $b = 0.2$ and a U-shape distribution when $b = 0.4$. Unlike the peaks present in Figure 6, the rank histogram in Figure 9(c) is the more typical U-shape that indicates an underdispersive ensemble. If the relaxation towards the observation is too strong, then the particles are all overly drawn towards the same point in state space, at a detriment to the variance of the ensemble. These results confirm the behaviour that would be expected by considering Eq. (23) and in addition demonstrate that it occurs even when the size of the relaxation term is smaller than the model error.

The above results refer to the scenario when observations were available for every other variable over the entire state; however, similar conclusions are evident when every other variable is observed only over half the state.

4.4.2. Percentage of particles retained

There are two separate effects that occur due to changing the percentage of particles retained under the equivalent-weights step. The first effect relates to the size of the movement required by each particle to ensure that its weight matches the chosen target weight. In the Lorenz (1963) model, an increase in the percentage of particles retained led to greater movement by each particle (Ades and Van Leeuwen, 2013). In turn, this led to an increase in the spread of the ensemble, seen in the rank histograms. Similar results are obtained with the barotropic vorticity equation (Figure 10). As the percentage of particles retained is increased, the ensemble becomes overdispersive (90–100%). Although not shown, the mean of the particles also shows less definition as the percentage increases, as would be expected with a more diverse ensemble.

Table 3 shows the l^2 norm of the average movement caused by the equivalent-weights proposal density as the percentage of particles increases. There is a substantial difference in the size of the movement when only 70% of particles are retained, compared with the much larger movement required with 100% of particles. This table also shows the significant movement caused by establishing equivalent weights for even 70% of particles. As
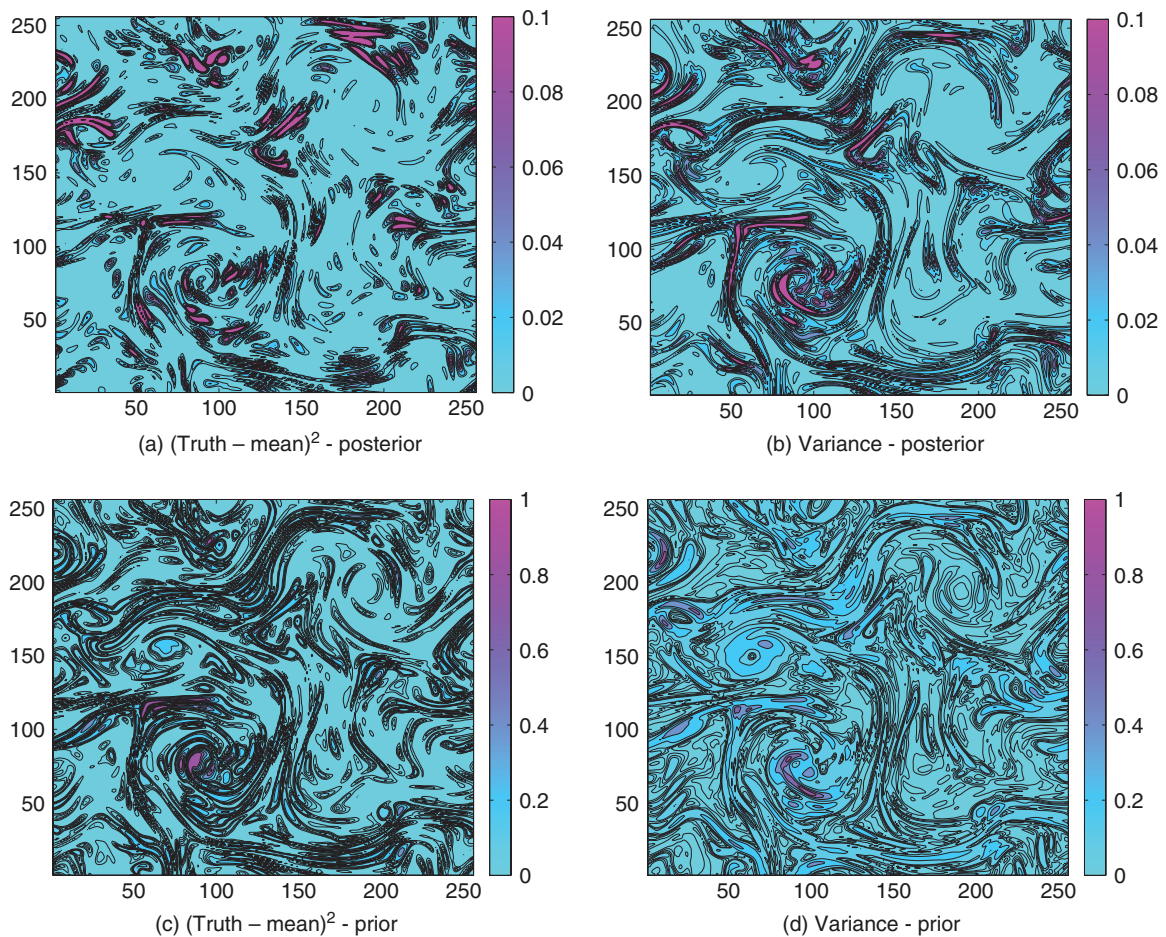
ⓒ 2014 The Authors. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

Q. J. R. Meteorol. Soc. (2014)

Figure 5. (a,c) The truth minus the mean of the ensemble squared compared with (b,d) the variance at time step 600 for both (c,d) the prior and (a,b) the posterior ensemble of 32 particles (the prior was calculated by running the same analysis ensemble forward in time from the previous observation without the relaxation). The patterns in the variance are similar to those observed in the $(\text{truth} - \text{mean})^2$ in both cases, indicating that the ensemble is behaving correctly. The ensemble underestimates the spread at several locations (a maximum variance of 0.466 compared with 1.647 for $(\text{truth} - \text{mean})^2$ for the posterior and 0.994 compared with 1.4088 for the prior), but averaged over the field it is slightly higher (0.018 versus 0.0156 and 0.0996 versus 0.0514). Note that the scale for the posterior is an order of magnitude smaller than the scale for the prior. This figure is available in colour online at wileyonlinelibrary.com/journal/qj

already noted, the movement created by the relaxation term is an order of magnitude smaller than the random error (Table 2). In contrast, the equivalent-weights step creates a change in state at least equal to that caused by the model equations.

The equivalent-weights movement can also be seen if the marginal posterior pdfs are considered (Figure 11). Taking the same variable and looking at the change as the percentage of particles retained is increased shows the change in the variance of the ensemble at this one time step. With 70–80% of particles retained, the ensemble is relatively confident in the position of the observation. Once 90–100% of particles are retained, the spread is noticeably increased and the mode moves away from the observation. Although this is just one variable at one time step, it is indicative of the majority of marginal posterior pdfs observed.

The second effect observed relates to the peaks present on either side of the rank histogram for the majority of experiments shown in this article. For all previous experiments, equivalent weights have been assured for 80% of the ensemble. The remaining 20%, which were unable to achieve the target weight, were discarded and returned as duplicates of the retained 80% of ensemble members in the resampling stage (see section 3.2). Increasing the percentage of particles retained in the equivalent-weights step results in the peaks disappearing and the rank histograms becoming more uniform. This is particularly noticeable if the rank histograms are considered for observations of every other variable only over half the state (Figure 12), but it is also apparent when every other variable over the full state is observed (Figure 10). These results suggest that the distinctive pattern seen in Figure 6 is due to the truncation of the ensemble caused by discarding 20% of the particles. They also demonstrate that it is possible to

achieve an appropriate spread in the ensemble, even when only half of the full state is observed, provided 100% of particles are retained.

The impact of the relaxation factor b and the percentage of particles retained on the spread and mean of the ensemble is apparent from both this section and section 4.4.1. It is also clear that the scheme can be tuned depending on the performance measure of interest, but this may negatively effect a different performance measure. For example, retaining a greater percentage of particles may enable the tails of the distribution to be encapsulated but could lead to a mean less representative of the truth. Since the benefit of the equivalent-weights particle filter is its theoretical ability to represent the true posterior, in this case rank histograms provide the most objective measure for tuning.

Ideally, guidance would be given on appropriate values for both b and the percentage of particles retained, independent of the application. However, since both parameters influence the representation of the posterior, it is difficult to determine the best value for either parameter objectively. For example, increasing the percentage of particles retained in order to capture the tails of the posterior pdf (but as a result increasing the spread of the ensemble) can be countered by decreasing the relaxation factor b. However, since the model equations determine the best knowledge about the system, choices that keep the particles closest to their model trajectories are preferable. Retaining a greater percentage of particles leads to increased movement by individual particles and so a balance is required between the requisite particle movement and the loss of information on the tails of the posterior pdf through resampling. The need to make this choice and the subsequent tuning of the system are both
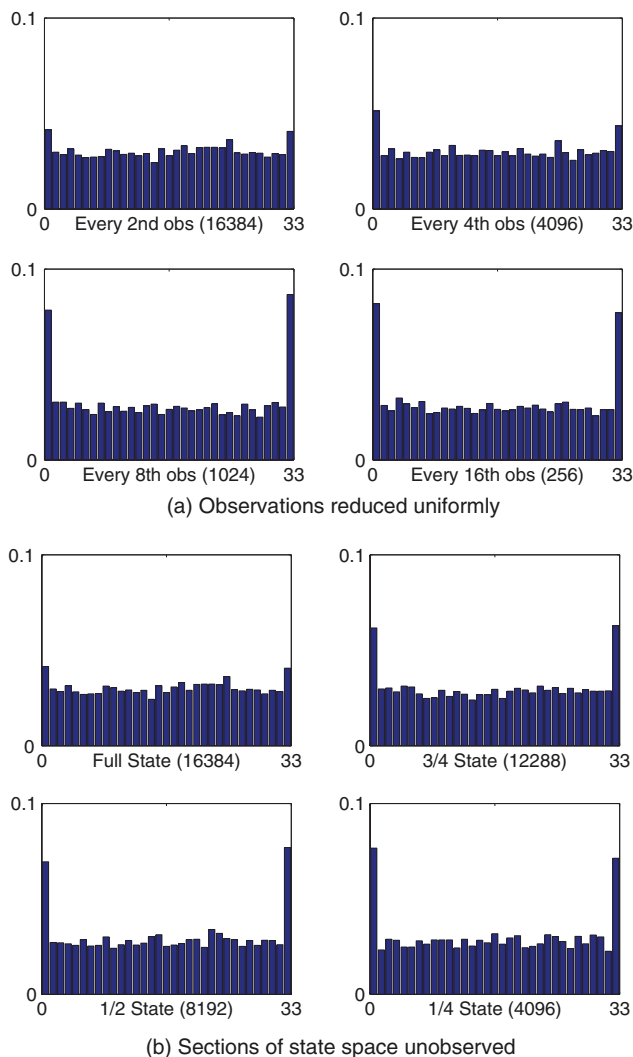
(a) Observations reduced uniformly

(b) Sections of state space unobserved

Figure 6. The rank histograms of where the truth ranks in the ensemble. Independent samples were taken of every 16th variable (beyond the correlation length) at 23 observation times. When (a) the observations are reduced uniformly or (b) only every other variable is observed in certain sections of state space (the observation numbers are shown in the brackets), the two end peaks become larger but the rest remains flat (as discussed in section 4.4.2). This figure is available in colour online at wileyonlinelibrary.com/journal/qj

a strength and a limitation of the equivalent-weights particle filter. For the barotropic vorticity equation, 80% was chosen, but this may change when other models are considered. Fixing the percentage of particles retained at 80% means the factor b becomes a tuning parameter, which can be used to ensure the appropriate spread in the ensemble.

4.4.3. Size of ensemble

As the size of the ensemble is increased, the distribution of the marginal posterior pdf is not significantly impacted (Figure 13). Since all particles relax towards the observation, increasing the number of ensemble members does not lead to the same increase in variance as would be expected when using the SIR filter. This is particularly noticeable when observations are available for every other variable over the entire state. In this case, the marginal posterior pdf when 128 particles are used (Figure 13(c)) is almost indistinguishable from that generated using 512 particles (Figure 13(e)).

4.5. Model error sensitivities

The results discussed in section 4.4 all relate to the sensitivity of the equivalent-weights particle filter to the parameter values chosen as part of the scheme. This section considers the sensitivity
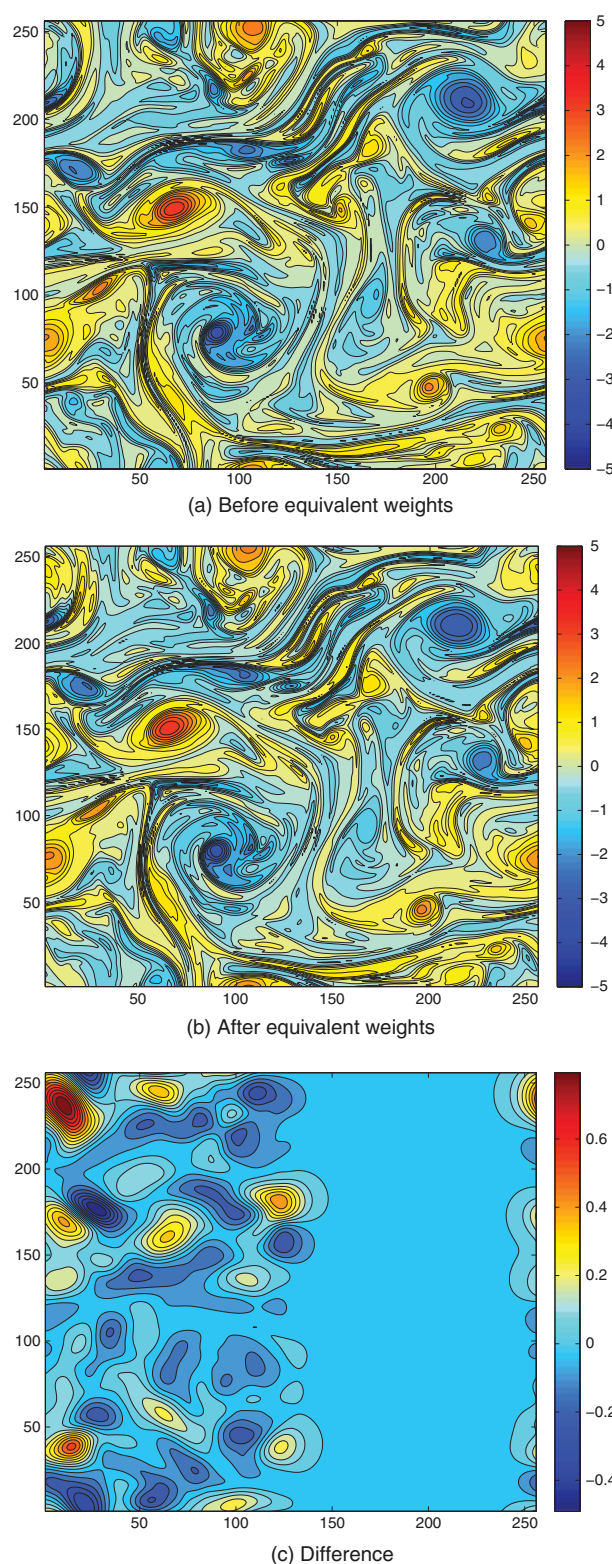

(a) Before equivalent weights

(b) After equivalent weights

(c) Difference

Figure 7. The state of particle 23 (a) before and (b) after the equivalent-weights step with observations of every other variable only on the left side of the state and (c) the difference between the two. As would be expected, there is no difference where there are no observations other than where information is propagating in via the covariance matrix Q from the observed half of the state. The largest differences are seen where features are in the wrong place close to the unobserved part of the state. This figure is available in colour online at wileyonlinelibrary.com/journal/qj

of the filter to differences between the parameters used to generate the 'pseudo' truth and those used to generate the model runs. Since in real life the model is an approximation to the true atmospheric evolution, this allows us an insight into the impact of comparing model runs created with chosen parameters and observations taken from the true atmosphere. In this case, there are no actual parameters in the barotropic vorticity equation that could be altered to give a different model for comparison with

Table 1. A comparison of CPU time for the SIR filter and the equivalent-weights particle filter for the integration period between observations and the analysis step at observation time, as discussed in section 4.3. Both have been averaged over the 23 different instances of a full 1150 time step run. The integration period is over 50 time steps and can be performed in parallel, hence it is independent of the number of particles. In contrast, the analysis step is dependent on the number of particles (32 particles were used to generate these results, with 80% being retained under the equivalent-weights step).

	SIR filter	Equivalent-weights particle filter	
Observed variables	65 536	65 536	16 384
Integration period (s)	105.7	113.2	113.5
Analysis period (s)	2.4	12.9	31.4

Table 2. The l^2 norm of the model equation, random error and relaxation term movement in the time step preceding an observation (when the relaxation is strongest), averaged across all observation times (23), to explore the effect of changing the parameter b (section 4.4.1). The numbers were generated under the observations of every other variable over the entire state scenario.

b	$E_{i,n}[f(x_i^{n-1})]$	$E_{i,n}[\mathrm{d}\beta_i^{n-1}]$	$E_{i,n}[B(\tau)(y^n - Hx_i^{n-1})]$
0.05	11.41	1.28	0.04
0.2	11.41	1.28	0.09
0.4	11.20	1.28	0.17

the truth. Hence, instead, the effect of changing the spread and size of the random error is examined.

4.5.1. Length-scale in Q

The first parameter to be considered is the length-scale used in the covariance matrix Q defined in section 4.1. If the length-scale of the truth is kept at five grid points and that of the model increased to nine grid points, then the ensemble becomes slightly overdispersive (Figure 14). The mean state at time step 600 fails to capture some of the finer detail of the truth, but is still clearly a good representation. This is consistent with previous results, which indicated that a more dispersive ensemble results in uncertainty in the positioning of the features, which in turn leads to a less defined mean.

4.5.2. Model error size

A greater effect is seen when the magnitude of Q is increased in the ensemble compared with the truth. Logically, it would be expected that increasing the size of the random error would increase the dispersiveness of the ensemble. It was found with the barotropic vorticity equation that changing the scaling of Q from 0.025^2 in the truth run to 0.1^2 in the ensemble actually led to a significant decrease in the spread of the ensemble (Figure 15). However, the mean retained a good approximation to the truth. Similar, although less pronounced, results were seen as the relaxation strength b was increased (Figure 9), indicating that the unexpected results may be due to a change in the relaxation term. Since Q is present in $B(\tau)$ (see Eq. (23)), a change in Q should affect the size of the movement generated by the relaxation term. For an increase from 0.025^2 to 0.1^2, the l^2 norm of the movement caused by the relaxation term would be expected to be of the order of 16 times larger. Similarly, the change to the l^2 norm of the model error, which involves $Q^{1/2}$ rather than Q, would be four times the true size.

Table 4 shows the l^2 norms of the constituent parts of the equivalent-weights particle filter. As expected, the increase in the norm of the model error term is four times larger. However, the increase in the average norm of the relaxation term is much greater than the estimated 16 times. This additional increase must come
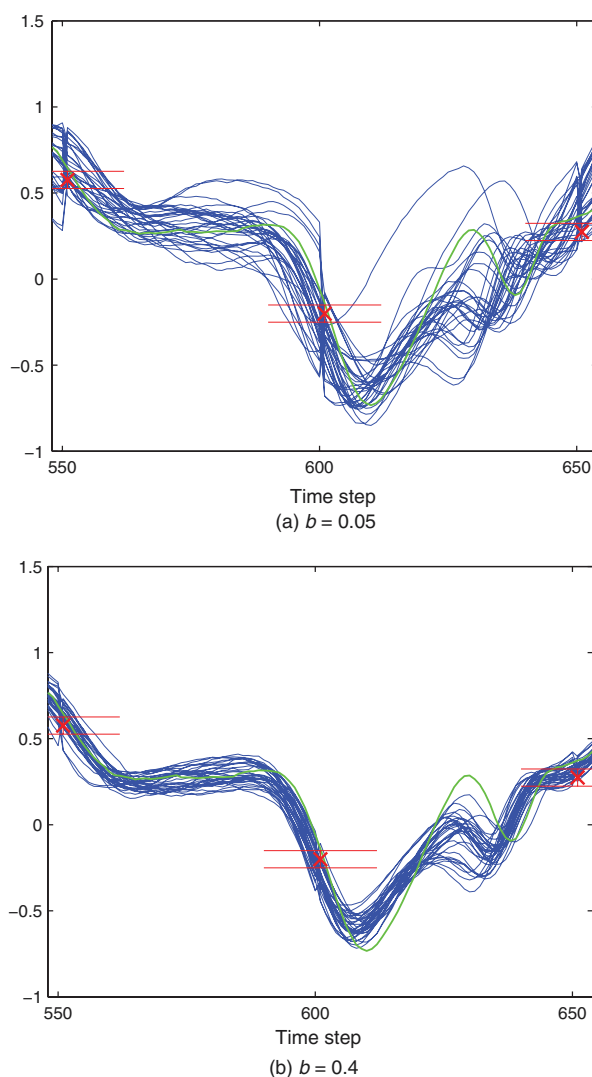

Figure 8. The trajectories of the particles (blue in the online article) for (a) a weaker relaxation with $b = 0.05$ and (b) a stronger relaxation with $b = 0.4$, compared to the truth (green in the online article) and observations with error bars (red in the online article) between time steps 550 and 600 for the observed variable at grid point (65 193). The spread of the ensemble is much less with a stronger relaxation term ($b = 0.4$). This figure is available in colour online at wileyonlinelibrary.com/journal/qj

from $(y^n - Hx_i^{n-1})$, as all the other factors remain fixed. Since y^n is also fixed, the increase must be due to additional movement away from y^n caused by a greater random error. The effect of the change to $(y^n - Hx_i^{n-1})$ is comparable to the effect of changing the factor b, since both increase the relaxation towards the observations. It is more pronounced in Figure 15 compared with Figure 9, since the relaxation term is now approximately 50% of the size of the random error. With $b = 0.4$, the l^2 norm of the relaxation term is still only 13% of the l^2 norm of the model error (Table 2).

The same conclusion can be drawn as was made for increasing the strength of the nudging term through b. Despite expectations to the contrary, inflating the model error can lead to all particles being drawn more strongly towards the same point in state space. This leads to a good representation of the mean but an underdispersive ensemble. It may be possible to achieve an appropriate spread in the ensemble again through tuning of the parameter b, but this would depend on each individual scenario and so has not been considered further here.

4.6. Discussion

As the number of observations is reduced, both uniformly across the state and with whole sections of the state unobserved, the mean of the equivalent-weights particle filter ensemble successfully
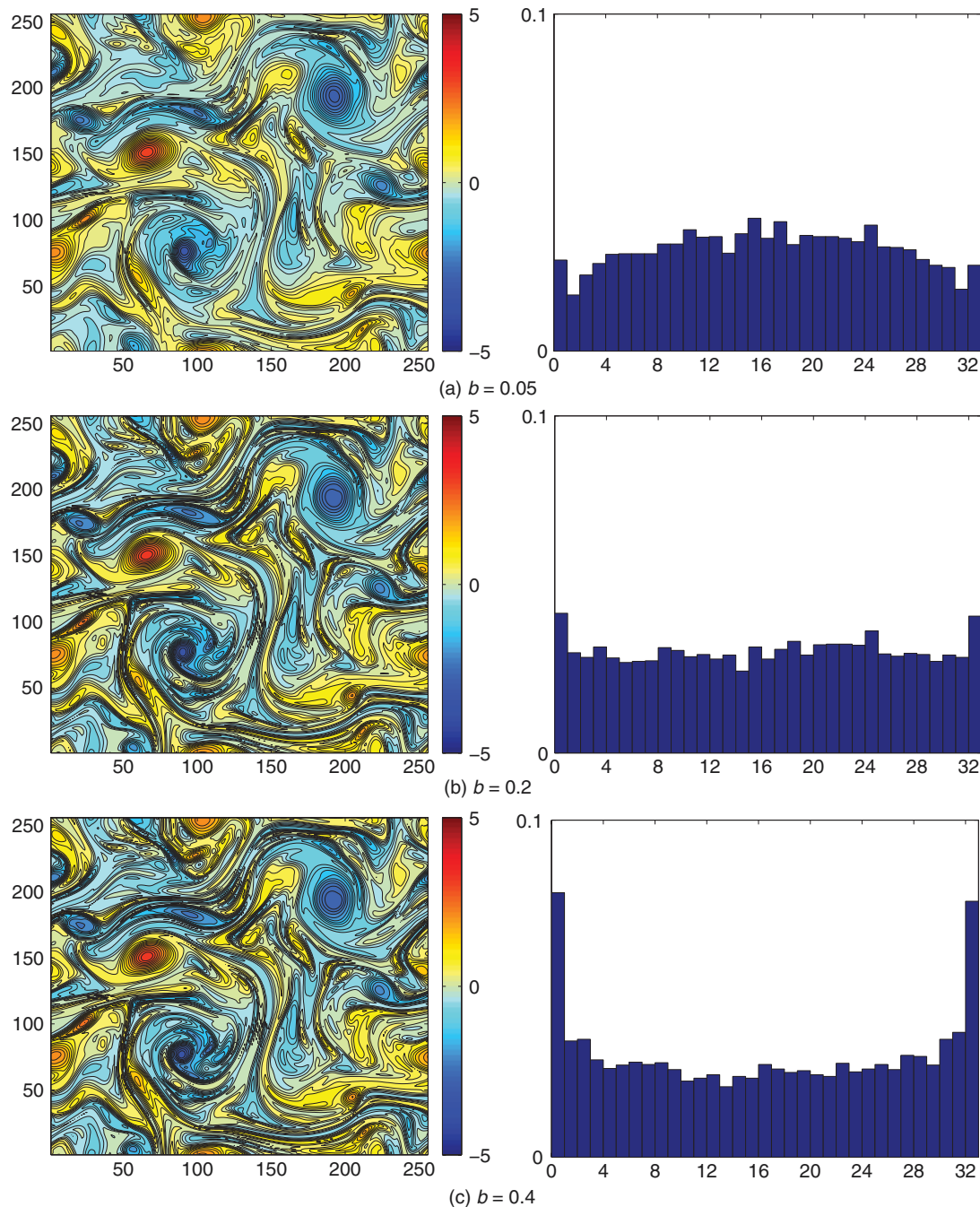
Figure 9. The mean of the 32 particle ensemble (left) at time step 600 and the rank histogram over all 1150 time steps (right) for (a) b = 0.05, (b) b = 0.2 and (c) b = 0.4 in $B(\tau)$. Observations were of every other variable over the entire state. Increasing b has little effect on the mean of the ensemble, but does affect the dispersion of the ensemble, seen in the change from slightly humped to U-shaped histogram. This figure is available in colour online at wileyonlinelibrary.com/journal/qj

captures the truth in a twin experiment. This implies that the ensemble is sampling from the required high-probability region of the posterior pdf. However, closer examination of the rank histograms shows that the observation can lie distinct from the ensemble (Figure 6). The conclusions drawn in this article suggest that this is due to the equivalent-weights particle filter truncating the ensemble under the equivalent-weights step, leading to insufficient knowledge about the far-out tails of the posterior pdf. However, it should be noted that this effect is small when every other variable is observed over half the state and disappears entirely as the percentage of particles retained is increased under all observation scenarios.

In the small-dimensional Lorenz (1963) model, the ability of the equivalent-weights particle filter to represent the posterior was most sensitive to changes in the number of particles retained under equivalent weights (Ades and Van Leeuwen, 2013). In this high-dimensional barotropic vorticity equation, the scheme was much less sensitive to this parameter. There was still a noticeable

increase in the average movement when a larger percentage of particles was retained (Table 3). However, looking at the performance of the ensemble over multiple time steps through the rank histograms (Figure 10) and at a single time step with the marginal posterior pdfs (Figure 11) would seem to indicate less impact when compared with the results from the Lorenz (1963) model. Clearly this is a subjective assessment, since the differing numbers of particles and large difference in the dimensions of the models make a direct comparison impractical.

The parameter b in the relaxation proposal density had as significant an effect as the percentage of particles retained in the barotropic vorticity equation. Although the size of the movement due to the relaxation term was still significantly less than that from the model error, regardless of the strength of relaxation chosen (Table 2), the impact seen in the rank histograms was comparable to the percentage of particles retained. If the inadvertent increase in the relaxation term due to the change in model error is considered (section 4.5.2), then it could be concluded that the
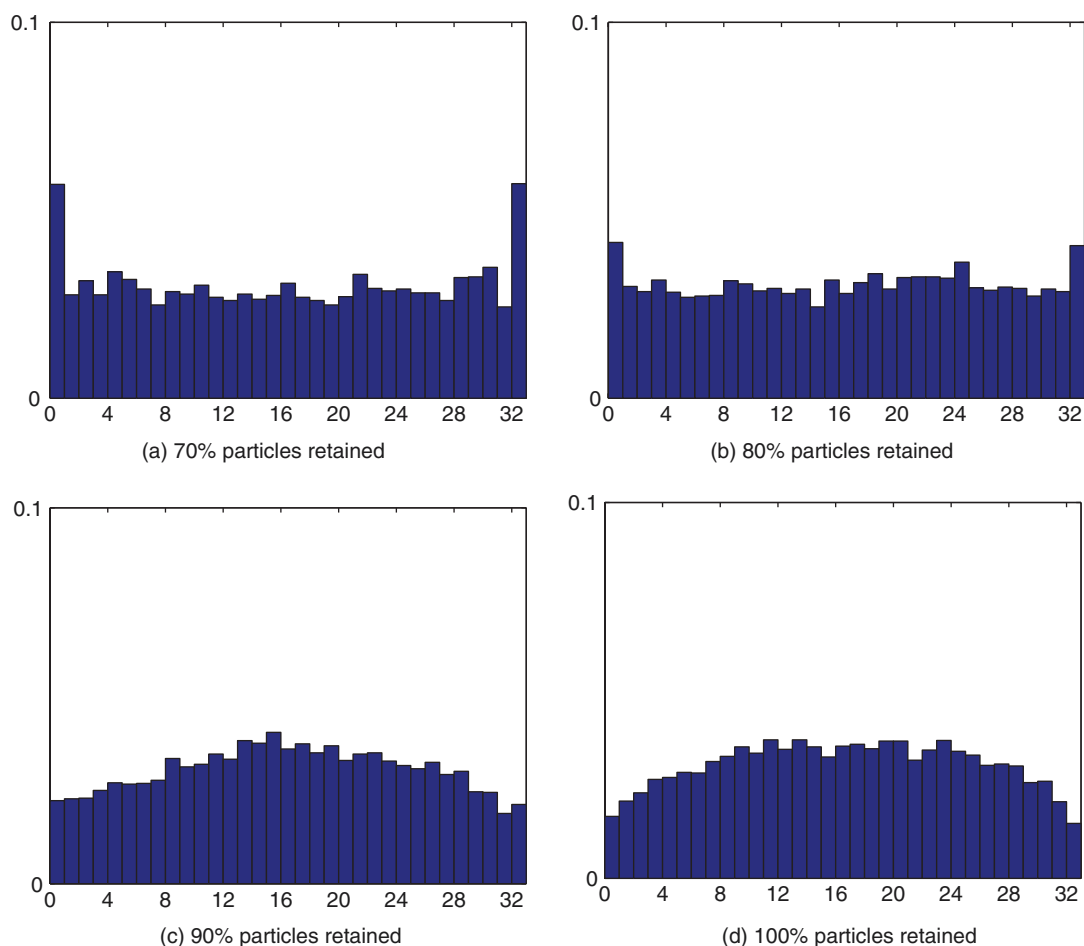
ⓒ 2014 The Authors. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

Q. J. R. Meteorol. Soc. (2014)

Figure 10. The rank histograms of the truth compared with the posterior ensemble over all 1150 time steps as the percentage of particles retained is increased from (a) 70% through (b) 80% and (c) 90% to (d) 100%. Observations were of every other variable over the entire state. The additional movement required to ensure a greater percentage of particles are retained under equivalent weights results in the over dispersion seen in panels (c) and (d). This figure is available in colour online at wileyonlinelibrary.com/journal/qj

Table 3. The l^2 norm of the model equation, random error and relaxation term movement in the time step preceding an observation compared with the movement generated by the equivalent-weights proposal density. The norm is averaged across all observation times (23) and all particles (32) to look at the effect of changing the percentage of particles (section 4.4.2). The numbers were generated under the observations every other variable over the entire state scenario.

%	$E_{i,n}[f(x_i^{n-1})]$	$E_{i,n}[\mathrm{d}\beta_i^{n-1}]$	$E_{i,n}[B(\tau)(y^n - Hx_i^{n-1})]$	$E_{i,n}[\mathrm{equ.weights}]$
70	11.43	1.28	0.09	11.89
80	11.40	1.28	0.09	13.94
90	11.29	1.28	0.13	25.83
100	11.08	1.28	0.22	53.95

representation of the posterior pdfs is actually more sensitive to this parameter. It may be possible to find an appropriate relationship between the strength of the relaxation term and the model errors, but this would be extremely difficult to quantify and has not been considered here.

Interestingly, similar to the results found with the Lorenz (1963) model, increasing the number of particles has little effect on the overall distribution of the marginal posterior pdfs. This is extremely encouraging, since it would indicate that it is possible to have a representation of the posterior pdf with many fewer particles than the size of the dimension, a result that could not be stated with the smaller dimensional Lorenz (1963).

The response of the scheme to different parameter choices used in the ensemble compared with the truth was most sensitive to the magnitude of model error, although change was also seen when a different length-scale was used. The response was not specific enough to draw any generic conclusions about the behaviour of

the scheme when the true atmosphere is modelled. However, it does show the importance of understanding the true model error statistics. This is an issue for data assimilation in general, rather than a particular issue for the equivalent-weights particle filter.

5. Conclusion and discussion

The aim of developing a scheme such as the equivalent-weights particle filter is to allow a representation of the high-probability region of a potentially multimodal posterior pdf. However, there are two major factors that must be addressed before the equivalent-weights particle filter can be considered as a serious possibility for real-life applications.

The first is the ability of the scheme to perform in high-dimensional settings. The failure of the SIR particle filter even in relatively small-dimensional systems is a known and well-understood problem. Although the 65 500 dimensional barotropic vorticity equation still has a state significantly smaller than some geophysical applications, it is sufficiently high-dimensional to provide a rigorous assessment of the performance of the scheme in a large dimension. Van Leeuwen and Ades (2013) showed the success of the scheme when the system was fully observed. This article looks at the more realistic situation of only a certain proportion of variables being observed. In particular, it looks at the effect of not observing whole segments of the state and the impact this has on the ability of the scheme to represent the posterior pdf.

A comparison of the mean of 32 particles with the truth in a twin experiment shows that the equivalent-weights particle filter continues to be successful at sampling from the high-probability region of the posterior pdf, even when only a small percentage of variables are observed. The rank histograms are
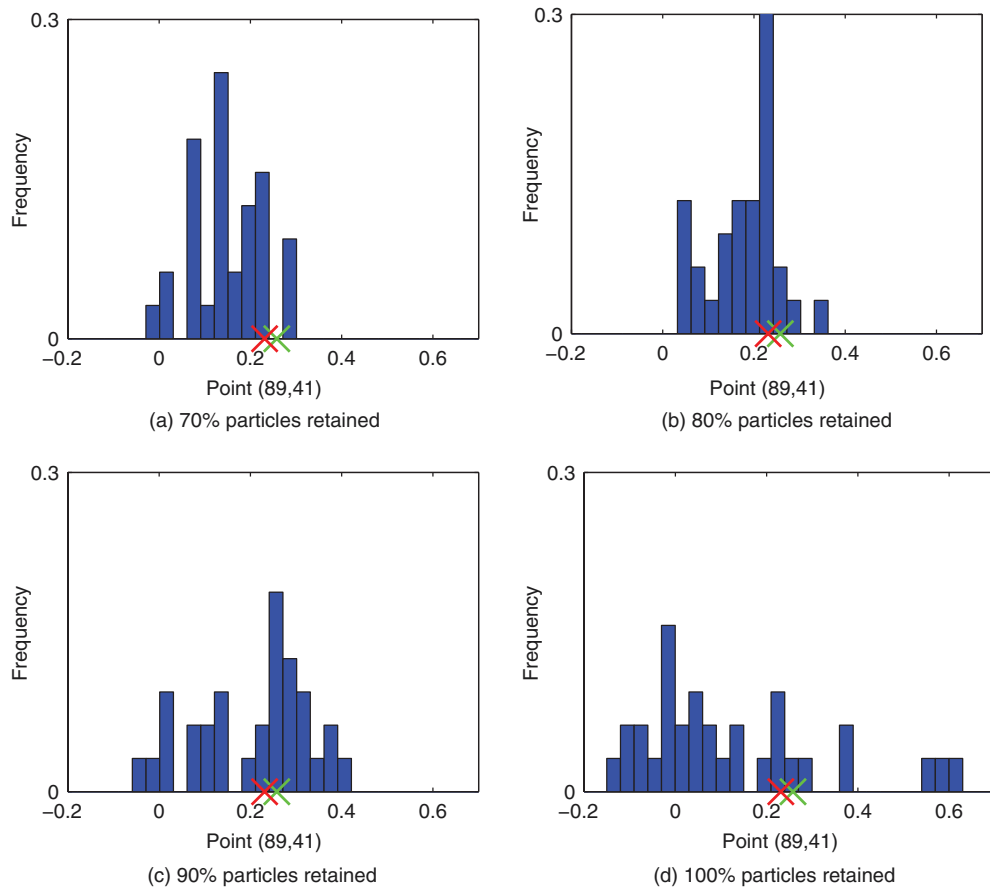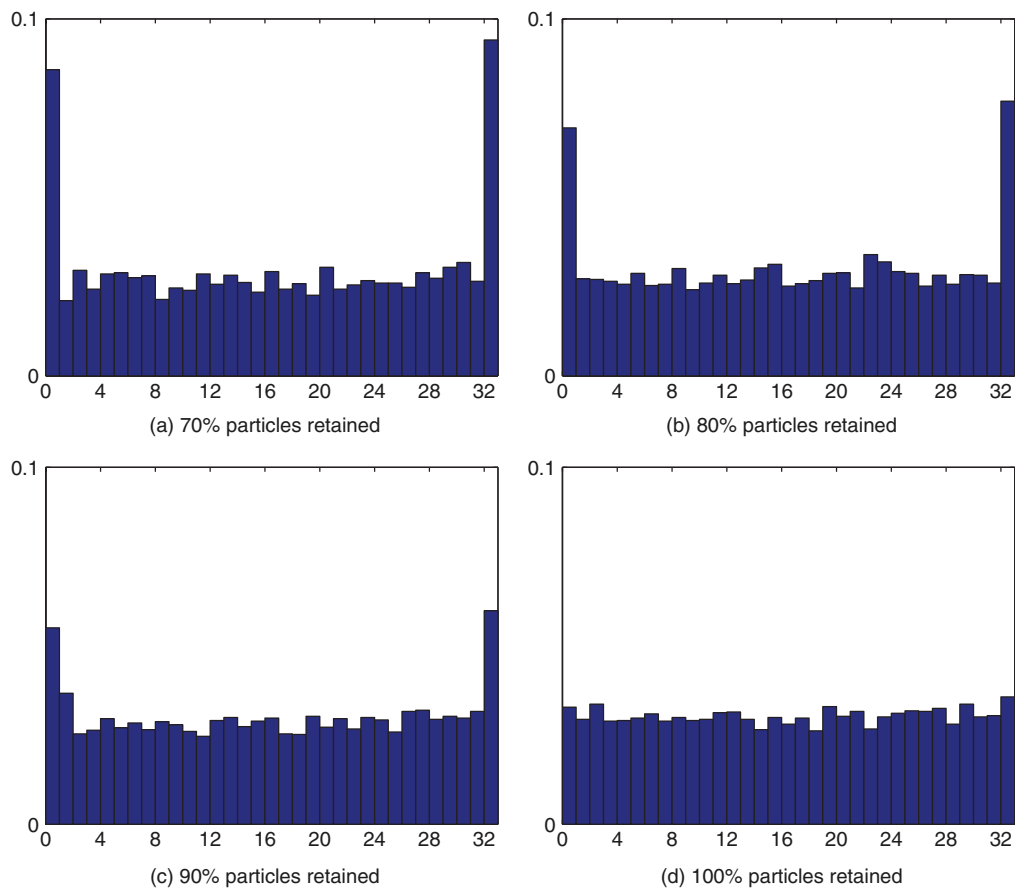
Figure 11. The marginal posterior pdf of the variable at location (89,41) and at time step 600 with 32 particles as the percentage of particles retained is increased from (a) 70% through (b) 80% and (c) 90% to (d) 100%. Observations were of every other variable over the entire state. The additional movement required to ensure a greater percentage of particles are retained under equivalent weights results in additional spread in the posterior pdfs as the percentage increases. This figure is available in colour online at wileyonlinelibrary.com/journal/qj

Figure 12. The rank histograms of the truth compared with the posterior ensemble over all 1150 time steps as the percentage of particles retained is increased from (a) 70% through (b) 80% and (c) 90% to (d) 100%. Observations were of every other variable over the left half of the state. The additional movement required to ensure a greater percentage of particles are retained under equivalent weights results in the reduction of the peaks on either side of the histogram. This figure is available in colour online at wileyonlinelibrary.com/journal/qj
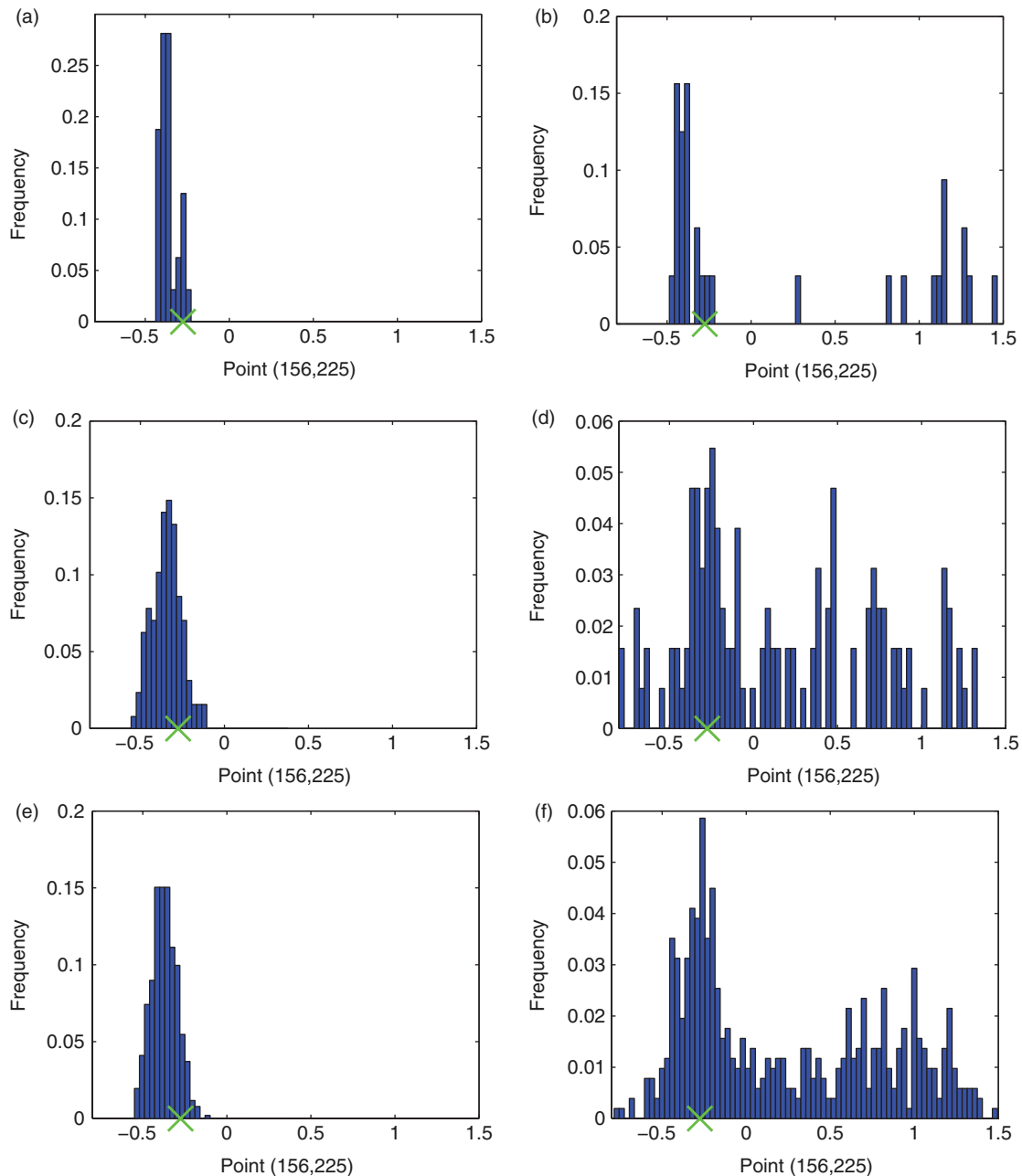
Figure 13. The marginal posterior pdf of the variable at location (156,225) and at time step 600 with (a,b) 32 particles, (c,d) 128 particles and (e,f) 512 particles. Observations are of every other variable over the entire state (panels (a), (c), (e)) and every other variable only over the left half of the state (panels (b), (d), (f)). In both cases, the marginal posterior pdf is for an unobserved variable, which is in the unobserved half of the state when only half the state is observed. A cross (green in the online article) shows the truth and 80% of particles were retained. Increasing the number of particles increases the density of the pdf, but the similarity between the posterior generated using 32 particles and that with 512 particles is clear. This figure is available in colour online at wileyonlinelibrary.com/journal/qj

harder to interpret, partly because they can be influenced by the parameter values chosen. In general, under the different observation scenarios (section 4.2), they have a flat central section with higher bars on either side (Figure 6), rather than the smooth U-shape that would provide clear evidence of underdispersion. The flat central section implies that, in the majority of cases, the ensemble members are samples from the true posterior pdf, although the higher bars indicate a difficulty in encapsulating the far-out tails. A direct comparison with the true marginal posterior pdfs is not possible for such a high-dimensional system, but the ensemble marginal posterior pdfs show the ability of the scheme to represent multimodal distributions. Of particular significance for operational data assimilation is the fact that the shape of the marginal posterior pdf remains consistent even as the number of particles is increased. Also of relevance for operational systems is the fact that this representation is achieved without adding significantly to the computational time of the SIR filter. These results show that the equivalent-weights particle filter ensures filter degeneracy does not occur even in high-dimensional

systems. However, providing conclusive evidence that the scheme also gives a good representation of the true posterior pdf without one for comparison is harder. These results show the potential of the scheme within the limitations of the measures used, but further investigation using the gold standard of Markov-Chain Monte-Carlo (MCMC) (Law and Stuart, 2012) would be beneficial.

The second important issue relates to model balances. Model balances are known relationships between variables, which, if they are not upheld, can lead to unbalanced states, e.g. spurious gravity waves in the model states. In the equivalent-weights particle filter, both the relaxation and the equivalent-weights proposal densities have the potential to affect these balances. The known relationships between variables are implicitly contained in the model equations, leading to balanced states at each time step. The relaxation term has no such constraints and so has the potential to induce artificial gravity waves by its addition to the model equations. However, the need to include model error in the geophysical equations in data assimilation schemes is
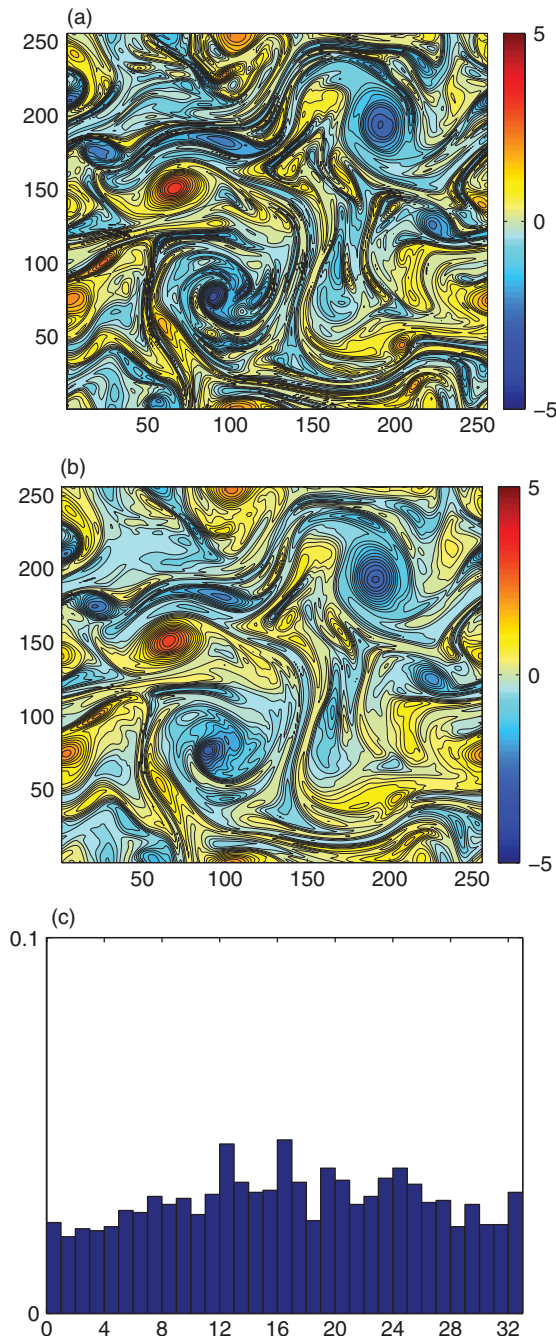
Figure 14. (a) The true state compared with (b) the mean of the 32 particle ensemble at time step 600, with (c) the rank histogram over all 1150 time steps when every other variables is observed over the entire state. Using a longer length-scale for the ensemble compared with the truth results in a more dispersive ensemble (section 4.5.1). This figure is available in colour online at wileyonlinelibrary.com/journal/qj

Figure 15. (a) The true state compared with (b) the mean of the 32 particle ensemble at time step 600, with (c) the rank histogram over all 1150 time steps when every other variables is observed over the entire state. Using an increased model error for the ensemble compared with the truth results in an underdispersive ensemble (section 4.5.2). This figure is available in colour online at wileyonlinelibrary.com/journal/qj

increasingly being considered important. This forms an inherent part of particle filters through the model transition density and will have an effect on the balances maintained through the model equations. Provided the relaxation term is of a similar magnitude to the model error, or smaller, it should not affect balances additionally beyond what is already considered acceptable in other data assimilation schemes.

The equivalent-weights proposal density also has the potential to impact balances through the artificial movement required to ensure equivalent weights for the majority of particles. Again, the size of the movement could be restricted but this has the potential to be at the expense of ensuring almost equal weights. The significance of the majority of particles in estimating the posterior is a key factor in the equivalent-weights particle filter and so compromising this would be undesirable. An alternative would be to apply initialization, as is currently done in the EnKF
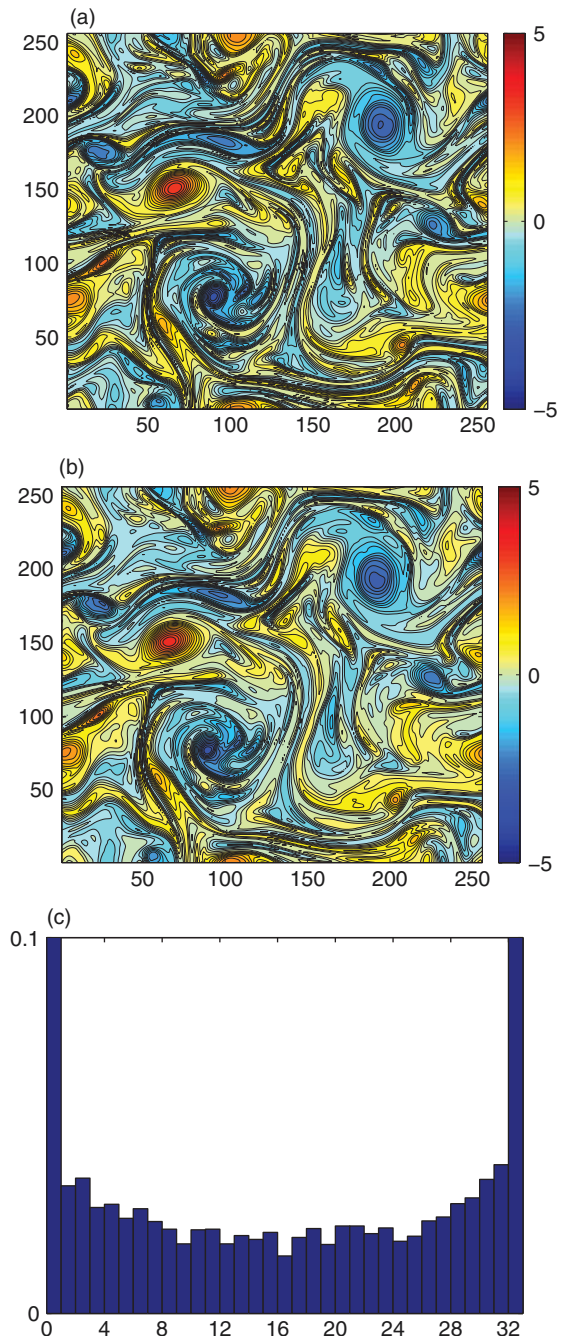
Table 4. The l^2 norm of the model equation, random error and relaxation term movement in the time step preceding an observation compared with the movement generated by the equivalent-weights proposal density. The norm is averaged across all time steps and all observation times as the percentage of particles retained is changed. The numbers were generated under the observations every other variable over the entire state scenario.

σ_m	$E_{i,n}[f(x_i^{n-1})]$	$E_{i,n}[d\beta_i^{n-1}]$	$E_{i,n}[B(\tau)(y^n - Hx_i^{n-1})]$
0.025	11.40	1.28	0.09
0.1	10.23	5.12	2.43

(Houtekamer and Mitchell, 2005; Buehner *et al.*, 2010). The effect of the equivalent-weights particle filter on model balances is a critical issue that needs to be addressed before the scheme

can be considered for real-life geophysical applications. Work is currently being undertaken to answer this question in a one-layer primitive equation model and will be the focus of future articles.

Appendix A: Equivalent-weights particle filter

This appendix [j1]outlines the link between the theory of the equivalent-weights particle filter and the practical application of the scheme in more detail. Although the majority of the information is also contained in Ades and Van Leeuwen (2013), some of the specific details have subsequently been developed. This is particularly the case in relation to the equivalent-weights proposal density and its application in higher dimensional systems. The appendix also includes a summary of the full scheme for clarity.

A1. Relaxation proposal density

In the SIR filter, a particle is propagated forward in time according to the model transition density $p(x^j|x_i^{j-1})$. The model transition density specifies how likely a state x^j is given the state x_i^{j-1}. Since the prognostic model equations $f(x_i^{j-1})$ are deterministic, this relates to the random error $d\beta_i^j$ added to the deterministic movement to represent unknown processes:

$$x_i^j = f(x_i^{j-1}) + d\beta_i^j. \tag{A1}$$

If the model error is assumed to be multivariate Gaussian, with known covariance matrix Q, then the model transition density is given by

$$p(x^j|x_i^{j-1}) \sim N(f(x_i^{j-1}), Q). \tag{A2}$$

The probability of samples from this density can be directly calculated according to

$$p(x_i^j|x_i^{j-1})$$
$$\propto \exp\left[-\frac{1}{2}\left(x_i^j - f(x_i^{j-1})\right)^T Q^{-1} \left(x_i^j - f(x_i^{j-1})\right)\right]. \tag{A3}$$

It should be reiterated that this does not mean any linear or Gaussian assumptions are being made about the model equations: the $f(x_i^j)$ may be fully nonlinear. It is the additive model error that is taken to be Gaussian and leads to the model transition density given in Eq. (A2).

In a similar manner, the relaxation proposal density of the equivalent-weights particle filter is chosen as

$$q(x^j|x_i^{j-1}, y^n) = N(f(x_i^{j-1}) + B(\tau)[y^n - h(x_i^{j-1})], \widehat{Q}). \tag{A4}$$

It works to provide a slight relaxation $B(\tau)(y^n - h(x_i^{j-1}))$ towards the observation at time n, in addition to the model equations.

$B(\tau)$ controls the strength of the relaxation and spreads the observation information from observed to unobserved variables. $\widehat{Q}$ is now the error covariance associated with both the model equations and this additional deterministic movement. This leads to each particle being moved forward in time according to

$$x_i^j = f(x_i^{j-1}) + B(\tau)(y^n - h(x_i^{j-1})) + \widehat{d\beta}_i^j. \tag{A5}$$

Again, no assumptions are being made about the linearity or otherwise of the model equations or the relaxation term. The distribution of the relaxation proposal density is related to the additive model error. A sample from this relaxation proposal density is evaluated as

$$q(x_i^j|x_i^{j-1}, y^n)$$
$$\propto \exp\left[-\frac{1}{2}\left(x_i^j - [f(x_i^{j-1}) + B(\tau)(y^n - h(x_i^{j-1}))]\right)^T\right.$$
$$\left. \times \widehat{Q}^{-1}\left(x_i^j - [f(x_i^{j-1}) + B(\tau)(y^n - h(x_i^{j-1}))]\right)\right]$$
$$= \exp\left[-\frac{1}{2}(\widehat{d\beta}_i^j)^T\widehat{Q}^{-1}\widehat{d\beta}_i^j\right]^{-1}. \tag{A6}$$

Sampling from the relaxation proposal density instead of the model transition density leads to the additional factor

$$\frac{p(x_i^j|x_i^{j-1})}{q(x_i^j|x_i^{j-1}, y^n)} \tag{A7}$$

to be calculated at every time step. Using Eqs (A3) and (A6), this is given by

$$\frac{p(x_i^j|x_i^{j-1})}{q(x_i^j|x_i^{j-1}, y^n)}$$
$$\propto \exp\left[-\frac{1}{2}\left(B(\tau)(y^n - h(x_i^{j-1})) + \widehat{d\beta}_i^j\right)^T\right.$$
$$\times Q^{-1}\left(B(\tau)(y^n - h(x_i^{j-1})) + \widehat{d\beta}_i^j\right)$$
$$\left. + \frac{1}{2}(\widehat{d\beta}_i^j)^T Q^{-1}\widehat{d\beta}_i^j\right], \tag{A8}$$

where it has been assumed that $\widehat{Q} = Q$.

In a high-dimensional system, the matrix Q^{-1} has the potential to cause computational cost issues in the practical application of the scheme. It is possible, however, to simplify this calculation and remove the inverse of Q. First, it must be recalled that the standard way to produce a sample of random error, $\widehat{d\beta}_i^j \sim N(0, Q)$, with the correct covariance structure is via $\widehat{d\beta}_i^j = Q^{1/2}\widehat{\xi}_i^j$, where $\widehat{\xi}_i^j \sim N(0, I)$. In Eq. (A6), this leads to

$$(\widehat{d\beta}_i^j)^T Q^{-1}\widehat{d\beta}_i^j = (Q^{1/2}\widehat{\xi}_i^j)^T Q^{-1}(Q^{1/2}\widehat{\xi}_i^j)$$
$$= (\widehat{\xi}_i^j)^T Q^{1/2T} Q^{-1} Q^{1/2}\widehat{\xi}_i^j$$
$$= (\widehat{\xi}_i^j)^T\widehat{\xi}_i^j, \tag{A9}$$

removing the need for the inverse in $q(x_i^j|x_i^{j-1}, y^n)$. Similarly, the factor $B(\tau)$ in the relaxation term is chosen as

$$B(\tau) = b\tau QH^T R^{-1} \tag{A10}$$

(see section 4 for details) and so includes the matrix Q. This means that

$$B(\tau)(y^n - h(x_i^{j-1})) + \widehat{d\beta}_i^j$$
$$= Q^{1/2}\left[b\tau Q^{1/2}H^T R^{-1}(y^n - h(x_i^{j-1})) + \widehat{\xi}_i^j\right] \tag{A11}$$

and hence Q^{-1} can also be removed when calculating $p(x_i^j|x_i^{j-1})$, at the cost of additionally generating $b\tau Q^{1/2}H^TR^{-1}(y^n - h(x_i^{j-1}))$.

The product of Eq. (A8), or alternatively the sum of the exponents, must be stored for all but the last time step to give

$$\widehat{w}_i^{n-1} = \prod_{j=1}^{n-1} \frac{p(x_i^j|x_i^{j-1})}{q(x_i^j|x_i^{j-1}, y^n)}. \qquad (A12)$$

A2. Equivalent-weights proposal density

Including the weights $\widehat{w}_i^{n-1}$ from the model prior Eq. (A12) in the full weight of a particle in the posterior gives

$$w_i^n \propto p(y^n|x_i^n)\frac{p(x_i^n|x_i^{n-1})}{q(x_i^n|x_i^{n-1}, y^n)}\widehat{w}_i^{n-1}. \qquad (A13)$$

The final proposal density $q(x^n|x_i^{n-1}, y^n)$ can be used to set

$$p(y^n|x_i^*)p(x_i^*|x_i^{n-1})\widehat{w}_i^{n-1} = w^{\text{target}}, \qquad (A14)$$

where w^{target} is a target weight to be determined. For Gaussian likelihood and model error, this is equivalent to solving the quadratic expression for x_i^*:

$$\frac{1}{2}(y^n - Hx_i^*)^TR^{-1}(y^n - Hx_i^*)$$
$$+ \frac{1}{2}(x_i^* - f(x_i^{n-1}))^TQ^{-1}(x_i^* - f(x_i^{n-1}))$$
$$- \log(\widehat{w}_i^{n-1}) = -\log(w^{\text{target}}).$$

If the model error or likelihood are not Gaussian, then this equation will change. Provided that it is still possible to evaluate a sample from the chosen density, however, then this equation can still be solved, although iterative methods may be required. The value of w^{target} can be found by first establishing the maximum weight it is possible for each particle to achieve through

$$w_i^{\max} = \widehat{w}_i^{n-1} \exp\left[-\frac{1}{2}\left(y^n - Hf(x_i^{n-1})\right)^T\right.$$
$$\left. \times \left(HQH^T + R\right)^{-1}\left(y^n - Hf(x_i^{n-1})\right)\right]. \qquad (A15)$$

These maximum weights can then be ranked from highest to lowest. The target weight is chosen equal to the percentage chosen lowest maximum weight. For example, if 80% of particles are required then the weight closest to 80% down the list of ranked weights should be chosen as the target weight. For clarity, it should be noted here that a linear H or a linearization of a nonlinear h is being used. If h is nonlinear then iterative methods will be required for the solution of these equations.

Once w^{target} is chosen, then one possible solution to Eq. (A15) for each particle is given by

$$x_i^* = f(x_i^{n-1}) + \alpha_i K(y^n - Hf(x_i^{n-1})), \qquad (A16)$$

where $K = QH^T(HQH^T + R)^{-1}$ and

$$\alpha_i = 1 + \sqrt{1 - b_i/a_i}, \qquad (A17)$$

in which $b_i = 0.5d_i^TR^{-1}d_i - \log(w^{\text{target}}) - \log(\widehat{w}_i^{n-1})$, $a_i = 0.5d_i^TR^{-1}HKd_i$ and $d_i = y^n - Hf(x_i^{n-1})$ (see Ades and Van Leeuwen (2013) for details). This choice for x_i^* can again be used to remove Q^{-1} in Eq. (A15) through the Q included in K.

This is a purely deterministic solution of Eq. (A15), which leads to a proposal transition density $q(x_i^n|x_i^{n-1}, y^n)$ given by a delta function. A proposal transition density given by a delta function

can be zero where the model transition density is non-zero, leading to division by zero as can be seen in Eq. (A13). Hence the final step in the equivalent-weights proposal density is to add some random error to the deterministic state:

$$x_i^n = f(x_i^{n-1}) + \alpha_i K(y^n - Hf(x_i^{n-1})) + \widehat{d\beta}_i^n$$
$$= x_i^* + Q^{1/2}\xi_i^n. \qquad (A18)$$

If $\xi_i^n \sim N(0, I)$ then the random error is Gaussian. Unfortunately, it can be shown that adding Gaussian random error in Eq. (A18) can lead to filter degeneracy (Ades and Van Leeuwen, 2013). Ideally the random error would not move the deterministic state far from x_i^* and would avoid filter degeneracy by ensuring the weights of the particles remain close to equivalent. One distribution that would satisfy these criteria is the uniform distribution over a small interval. A sample from a k-dimensional multivariate uniform distribution can be defined as

$$\xi_i^n \sim \widetilde{U}_k(0, \gamma_U) \implies \xi_{i,j}^n \sim U[-\gamma_U, \gamma_U], \quad \forall j = 1, \dots, k, \qquad (A19)$$

where $\xi_{i,j}^n$ represents the jth element of the vector ξ_i^n. So each element of ξ_i^n is picked from the standard uniform distribution

$$U[-\gamma_U, \gamma_U] = \begin{cases} \frac{1}{2\gamma_U} & \text{if } \xi_{i,j}^n \in [-\gamma_U, \gamma_U], \\ 0, & \text{otherwise,} \end{cases}$$

with a width determined by γ_U. Keeping the parameter γ_U small ensures that each uniform distribution has a small non-zero interval. Since each element of ξ_i^n is independent, the probability of a sample $p(\xi_i^n)$ coming from $\widetilde{U}_k(0, \gamma_U)$ is given by $p(\xi_i^n) = p(\xi_{i,1}^n)p(\xi_{i,2}^n) \dots p(\xi_{i,k}^n)$ and hence

$$\widetilde{U}_k(0, \gamma_U) = \begin{cases} \frac{1}{(2\gamma_U)^k} & \text{if } \xi_{i,j}^n \in [-\gamma_U, \gamma_U], \forall j = 1, \dots, k, \\ 0, & \text{otherwise.} \end{cases}$$

The vector ξ_i^n is still multiplied by $Q^{1/2}$ to introduce covariances between these now uniformly distributed elements.

In order to calculate the final weights, we are interested in the proposal density $q(x^n|x_i^{n-1}, y^n)$. The transformation of probability densities,

$$p_z(z) = p_x(x(z))\left|\frac{\mathrm{d}x}{\mathrm{d}z}\right|, \qquad (A20)$$

can be used to establish $q(x^n|x_i^{n-1}, y^n)$, where $p_z(z)$ represents the probability density function of variable z over the space spanned by z and similarly for $p_x(x(z))$. Hence

$$p(\xi^n) = q(x^n|x_i^{n-1}, y^n)\left|\frac{\partial(x_i^* + Q^{1/2}\xi^n)}{\partial \xi^n}\right|$$
$$= q(x^n|x_i^{n-1}, y^n)\left|Q^{1/2}\right|$$

$$\implies q(x^n|x_i^{n-1}, y^n) = \frac{1}{|Q^{1/2}|}p(\xi^n). \qquad (A21)$$

Choosing $q(x^n|x_i^{n-1}, y^n)$ as a uniform density, however, will again lead to division by zero, given that the model transition density $p(x^n|x_i^n)$ is Gaussian. The solution is to generate a sample ξ_i^n from a mixture density

$$\xi_i^n \sim (1 - \epsilon)\widetilde{U}_k(0, \gamma_U) + \epsilon N(0, \gamma_N I). \qquad (A22)$$

Using a mixture density means the error is drawn from a combination of a uniform distribution and a Gaussian

distribution, with the proportion controlled by the value of ϵ. Choosing ϵ to be small ensures that in general ξ_i^n are picked from the part of the proposal where the uniform distribution is non-zero. In particular, relating ϵ to the size of the ensemble, for example $\epsilon = 0.001/N$, means that drawing ξ_i^n from the Gaussian tails is very unlikely even as the ensemble size increases and hence equivalent weights can be assured for all particles. However, unlike the use of a purely uniform distribution, the possibility of picking from the Gaussian tails ensures continuous support across the entire space of x_i^n.

A3. Overview of the equivalent-weights particle filter

To conclude this appendix, the full equivalent-weights particle filter scheme is summarised here.

(1) The prior pdf $p(x^0)$ is represented by N individual model states or particles.
(2) Each particle is run forward to time $n-1$ (the time step immediately preceding the next available observation vector) using the relaxation transition density. If it is assumed that the proposed relaxation model error is additive and Gaussian, at each time step this relates to the following steps.
(a) Establishing the new model states using

$$x_i^j = f(x_i^{j-1}) + B(\tau)(y^n - h(x_i^{j-1})) + \widehat{d\beta}_i^j, \qquad (A23)$$

where $\widehat{d\beta}_i^j = \widehat{Q}^{1/2}\widehat{\xi}_i^j \sim N(0, \widehat{Q})$.
(b) Calculating the corresponding weights associated with time step j according to

$$
\begin{aligned}
w_i^j \propto \exp\Big[& -\frac{1}{2}\Big(b\tau Q^{1/2}H^T R^{-1}(y^n - h(x_i^{j-1})) + \widehat{\xi}_i^j \Big)^T \\
& \times \Big(b\tau Q^{1/2}H^T R^{-1}(y^n - h(x_i^{j-1})) + \widehat{\xi}_i^j \Big) \\
& + \frac{1}{2}(\widehat{\xi}_i^j)^T (\widehat{\xi}_i^j) \Big].
\end{aligned}
\qquad (A24)
$$

The product of these weights up to time step $n-1$ is stored for use as $\widehat{w}_i^{n-1}$ at the new observation time:

$$\widehat{w}_i^{n-1} = \prod_{j=n-r+1}^{n-1} w_i^j. \qquad (A25)$$

(3) The particles are moved forward to the new observation time n in the last time step according to the equivalent-weights proposal density. The actual steps to calculate the model state for each particle at time n are as follows.
(a) Calculate the maximum weight it is possible for each particle to achieve via

$$
\begin{aligned}
w_i^{\max} = \widehat{w}_i^{n-1} \exp\Big[& -\frac{1}{2}\left(y^n - Hf(x_i^{n-1})\right) \\
& \times (HQH^T + R)^{-1}\left(y^n - Hf(x_i^{n-1})\right)\Big].
\end{aligned}
\qquad (A26)
$$

(b) Order the maximum weights for all the particles and then choose w^{target} such that a certain percentage of particles can achieve it.
(c) For the chosen percentage of particles where $w_i^{\max} \geq w^{target}$, select the deterministic model state according to

$$x_i^* = f(x_i^{n-1}) + \alpha_i K(y^n - Hf(x_i^{n-1})), \qquad (A27)$$

where α_i is defined according to Eq. (A17) and $K = QH^T(HQH^T + R)^{-1}$. Otherwise the model state is given by

$$x_i^* = f(x_i^{n-1}). \qquad (A28)$$

(d) Add stochastic error $Q^{1/2}\xi_i^n$ to the deterministic model state x_i^* to find the final particle model state x_i^n at time n. This error comes from the mixture density given by

$$\widehat{q}(\xi^n|x_i^*) = (1-\epsilon)\widetilde{U}_k(0, \gamma_U) + \epsilon N(0, \gamma_N^2 I). \qquad (A29)$$

To sample from the mixture density, a value u is sampled from $u \sim U[0,1]$. If $u < \epsilon$ then $\xi_i^n \sim N(0, \gamma_N^2 I)$, otherwise $\xi_i^n \sim \widetilde{U}_k(0, \gamma_U)$.
(e) The final weight of each particle is then calculated according to

$$w_i^n = \frac{\widehat{w}_i^{n-1} p(x_i^n|x_i^{n-1}) p(y^n|x_i^{n-1})}{q(x_i^n|x_i^{n-1}, y^n)}, \qquad (A30)$$

where

$$
\begin{aligned}
& \widehat{w}_i^{n-1} p(x_i^n|x_i^{n-1}) p(y^n|x_i^{n-1}) \\
& = \widehat{w}_i^{n-1} \exp\Big[-\frac{1}{2} \\
& \times \Big(\alpha_i Q^{1/2}H^T(HQH^T + R)^{-1}(y^n - Hf(x_i^{n-1})) + \xi_\psi \Big)^T \\
& \times \Big(\alpha_i Q^{1/2}H^T(HQH^T + R)^{-1}(y^n - Hf(x_i^{n-1})) + \xi_{\hat\psi} \Big) \\
& -\frac{1}{2}\left(y^n - Hx_i^n\right)^T R^{-1}\left(y^n - Hx_i^n\right)\Big]
\end{aligned}
\qquad (A31)
$$

and

$$
\begin{aligned}
\frac{1}{q(x_i^n|x_i^{n-1}, y^n)} = \Big[& 1 + \frac{\epsilon}{(1-\epsilon)}\left(\frac{2}{\pi}\right)^{k/2}\frac{\gamma_U^k}{\gamma_N} \\
& \times \exp\left(-\frac{1}{2}\xi_i^{nT}(\gamma_N^2 I)^{-1}\xi_i^n \right) \Big]^{-1}
\end{aligned}
\qquad (A32)
$$

if $-\gamma_U \leq \xi_{i,j} \leq \gamma_U, \forall j = 1, \ldots, k$ and

$$
\begin{aligned}
\frac{1}{q(x_i^n|x_i^{n-1}, y^n)} = \Big[& \frac{\epsilon}{(1-\epsilon)}\left(\frac{2}{\pi}\right)^{k/2}\frac{\gamma_U^k}{\gamma_N} \\
& \times \exp\left(-\frac{1}{2}\xi_i^{nT}(\gamma_N^2 I)^{-1}\xi_i^n \right) \Big]^{-1}
\end{aligned}
\qquad (A33)
$$

otherwise. The factor of $|Q^{1/2}|$ necessary in the conversion from $\widehat{q}(\xi_i^n|x_i^*)$ to $q(x_i^n|x_i^{n-1}, y^n)$, as well as the common normalization factor

$$\frac{(1-\epsilon)}{(2\gamma_U)^k}, \qquad (A34)$$

are not required in the calculations, since they are present in the weight of every particle and hence would cancel in the following step.
(4) The weights of the particles are normalized by the sum of all particle weights. The ensemble of particles, together with their weights, now represents the posterior pdf $p(x^n|y^n)$.
(5) Finally, the particles are resampled so that they once again all have weight equal to $1/N$. This step is required, since only the percentage of particles that are able to achieve the target weight will have almost equal weights under the equivalent-weights step. The particles evolved according to Eq. (A28) will have smaller weights and so will be resampled as duplicate copies of the equivalently weighted particles. The resampled ensemble of particles can now be run forward to the next observation vector in time.

References

Ades M, Van Leeuwen PJ. 2013. An exploration of the equivalent weights particle filter. *Q. J. R. Meteorol. Soc.* **139**: 820–840, doi: 10.1002/qj.1995.

Anderson J. 1996. A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J. Clim.* **9**: 1518–1530.

Bocquet M, Pires CA, Wu L. 2010. Beyond Gaussian statistical modeling in geophysical data assimilation. *Mon. Weather Rev.* **138**: 2997–3023.

Buehner M, Houtekamer P, Charette C, Mitchell HL, He B. 2010. Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I: Description and single-observation experiments. *Mon. Weather Rev.* **138**: 1550–1566.

Burgers G, van Leeuwen PJ, Evensen G. 1998. Analysis scheme in the ensemble Kalman filter. *Mon. Weather Rev.* **126**: 1719–1724.

Chorin AJ, Tu X. 2009. Implicit sampling for particle filters. *Proc. Natl. Acad. Sci.* **106**: 17 249–17 254.

Chorin AJ, Morzfeld M, Tu X. 2010. Interpolation and iteration for nonlinear filters. *Commun. Appl. Math. Comput. Sci.* **5**: 221–240.

Doucet A, Godsill S, Andrieu C. 2000. On sequential Monte Carlo sampling methods for Bayesian filtering. *Stat. Comput.* **10**: 197–208.

Doucet A, de Freitas N, Gordon N. 2001. *Sequential Monte-Carlo Methods in Practice*. Springer-Verlag: Berlin.

Durran DR. 1999. *Numerical Methods for Wave Equations in Geophysical Fluid Dynamics*. Springer: Berlin.

Evensen G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**: 10 143–10 162, doi: 10.1029/94JC00572.

Hamill TM. 2000. Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Weather Rev.* **129**: 550–560.

Hamill TM, Colucci S. 1997. Verification of Eta-RSM short-range ensemble forecasts. *Mon. Weather Rev.* **125**: 1312–1327.

Houtekamer PL, Mitchell HL. 2005. Ensemble Kalman filtering. *Q. J. R. Meteorol. Soc.* **131**: 3269–3289.

Kitagawa G. 1996. Monte-Carlo filter and smoother for non-Gaussian nonlinear state-space models. *J. Comput. Graph. Stat.* **10**: 253–259.

Law KJH, Stuart AM. 2012. Evaluating data assimilation algorithms. *Mon. Weather Rev.* **140**: 3757–3782.

Lorenz EN. 1963. Deterministic nonperiodic flow. *J. Atmos. Sci.* **20**: 130–141, doi:10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.

Lorenz EN. 1995. Predictability: A problem partly solved, Proc. Sem. Predictability, Vol. 1, ECMWF, Reading, UK; 1–18.

Morzfeld M, Tu X, Atkins E, Chorin AJ. 2012. A random map implementation of implicit filters. *J. Comput. Phys.* **231**: 2049–2066.

Snyder C, Bengtsson T, Bickel P, Anderson J. 2008. Obstacles to high-dimensional particle filtering. *Mon. Weather Rev.* **136**: 4629–4640.

Talagrand O, Courtier P. 1987. Variational assimilation of meteorological observations with the adjoint voriticity equation. I. Theory. *Q. J. R. Meteorol. Soc.* **113**: 1311–1328.

Talagrand O, Vautard R, Strauss B. 1999. Evaluation of probabilistic prediction systems. In *Proceeding of Workshop on Predictability*, 1–25. European Centre for Medium-Range Weather Forecasts: Reading, UK.

Van Leeuwen PJ. 2009. Particle filtering in geophysical systems. *Mon. Weather Rev.* **137**: 4089–4114.

Van Leeuwen PJ. 2010. Nonlinear data assimilation in geosciences: An extremely efficient particle filter. *Q. J. R. Meteorol. Soc.* **136**: 1991–1999.

Van Leeuwen PJ. 2011. Efficient nonlinear data-assimilation in geophysical fluid dynamics. *Comput. Fluids* **46**: 52–58.

Van Leeuwen PJ, Ades M. 2013. Efficient fully nonlinear data assimilation for geophysical fluid dynamics. *Comput. Geosci.* **55**: 16–27, doi: 10.1016/j.cageo.2012.04.015.

Wilks DS. 1995. *Statistical Methods in the Atmospheric Sciences*. Academic Press: London.