# Particle Flow for a Quasi-Geostrophic Model

20

This chapter discusses the application of a particle-flow filter to a two-layer quasi-geostrophic model. The reasons for including this example are twofold. First, it shows that it is possible to apply fully nonlinear data assimilation to high-dimensional systems, even without localization. Second, we introduce how to set up such an experiment in more detail and discuss the choices one must make.

20.1 Introduction

We start by demonstrating how the evolution of the particles in pseudo-time progresses in a highly idealized model to understand the basic ideas behind particle-flow methods better. We look at one specific gridpoint in a 1000-dimensional Lorenz 1996 model for an observation that is the square of the state variable at that gridpoint, so $d = x_{\text{true}}^2 + \epsilon$. The value of $d = 7.3$ with observation-error standard deviation equal to 0.2. The prior is a wide Gaussian with mean 0.5 and standard deviation 1, represented by 100 particles as depicted by the lower red dots in Fig. 20.1. The blue lines denote the movement of the particles in the one-dimensional state space of this gridpoint. The vertical axis is pseudo time, scaled between 0 and 1. We made many iterations with small steps to accurately illustrate the movement in this part of the state space.

Figure 20.1 shows that the particles flow towards the posterior pdf, centered on the possible gridpoint values corresponding to $x^2 = d = 7.3$, hence $x = \pm\sqrt{7.3} = \pm 2.7$, with a standard deviation of order 0.1. The pseudo-time trajectories seem to cross each other, e.g., in the lower right corner of the plot. An actual crossing of trajectories would lead to the failure of the method, indicating the use of too large pseudo time steps. However, this crossing is not actual because we only plot

Fig. 20.1 The plot shows the evolution of 100 particles (colored lines) in pseudo time (horizontal axis) at one gridpoint in a 1000-dimensional Lorenz 1996 data-assimilation experiment. Note the motion from the relatively narrow prior distribution to the two modes
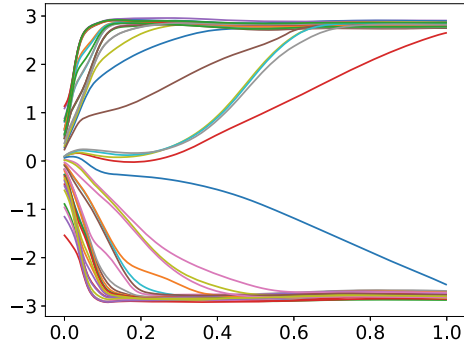
the flow of the 1000-dimensional particles in a one-dimensional projection. In the 1000-dimensional full space, the particles do not cross.

The particle-flow filter demonstrates behavior that is impossible to obtain with an ensemble Kalman filter, not even an iterative ensemble Kalman filter that uses the ensemble gradient for the adjoint of the observation operator. The reason is that the gradients will have different signs for the two posterior modes, while the ensemble provides only one average gradient. Only iterative methods that use either the adjoint of H or different ensemble gradients for different state-variable values can accurately find the two modes. Variational methods will converge to one of the modes, dependent on the first guess. Finally, a standard particle filter will not move the particles, only their weights. The relatively narrow prior in Fig. 20.1 does not cover the two posterior modes, and no particles will end up in these modes. Resampling would produce two artificial modes at the extremes of the prior. Only a particle-flow filter can produce these modes in its standard configuration.

20.2   Application to the QG Model

The quasi geostrophic (QG) model solves the following equations for a 2-layer system

$$\frac{\partial p_1}{\partial t} + J(\psi_1, p_1) = A \Delta q_1, \tag{20.1}$$

$$\frac{\partial p_2}{\partial t} + J(\psi_2, p_2) = A \Delta q_2, \tag{20.2}$$

where the potential vorticity p_i in each layer is the sum of the relative vorticity, the planetary vorticity, and a stretching term,

$$p_1 = \nabla^2 \psi_1 + f - F_1(\psi_1 - \psi_2), \tag{20.3}$$

$$p_2 = \nabla^2 \psi_2 + f + F_2(\psi_1 - \psi_2). \tag{20.4}$$

Here ψ_1 and ψ_2 are the stream functions in the two model layers, and **A** is the horizontal diffusion or mixing coefficient. The Jacobian $J(\psi, p) = \frac{\partial \psi}{\partial x} \frac{\partial p}{\partial y} - \frac{\partial \psi}{\partial y} \frac{\partial p}{\partial x}$

denotes the advection of potential vorticity. The Coriolis parameter is $f = f_0 + \beta y$ in which y is the meridional coordinate (the so-called β-plane approximation), and the F_i are constants related to the densities and height of the two layers.

A practical scheme to solve the QG model equations is the following. First, calculate the potential vorticity from the stream-function fields. Next, propagate the potential-vorticity fields over one time step. Then solve the Helmholtz equations for the new stream-function fields, which the advection terms use to propagate the potential vorticity over the next time step.

The model setup uses two layers of 257 by 129 gridpoints with a grid spacing of 100 km. The dimension of the state vector is 66306. The time step is 30 min, and $F_1 = F_2 = 2.8 \times 10^{-12}$ m^{-2}. The Coriolis parameter in the middle of the domain is $f_0 = 7.28 \times 10^{-5}$ s^{-1} and $\beta = 2.0 \times 10^{-11}$ m^{-1}s^{-1}.

20.3 Data-Assimilation Experiment

We initialized the model with a meandering jet of wavenumber four in the upper layer with maximum stream-function value 5×10^7 m^{2}s^{-1}, and the stream function in the lower layer was taken as a factor 0.03 times that of the upper layer. This model state was spun up for 250 time steps, approximately five days.

Figure 20.2 gives examples of the true model stream function in the upper layer at different time steps during the data-assimilation experiment. The plots show different stages of the evolution of the flow field, with the Jet Stream flowing from East to West at the boundary between reddish and greenish colors. We observe several eddies (low- and high-pressure cells) north and south of the meandering Jet Stream. The three plots show the shedding of high-pressure cells for a little more than two days.

An initial ensemble of 100 members was created by adding Gaussian random noise with a decorrelation length scale of 20 gridpoints to a similarly perturbed true spun-up state. The standard deviation of the perturbations was 100 m^{2}s^{-1}. At every time step, we added model errors drawn from a Gaussian with zero mean and the same decorrelation length scale and a standard deviation of 0.005 times the one in the true initial fields.

We assimilated observations every 10 time steps, corresponding to 5 h. We observed the stream function at 600 equally-distributed gridpoints in each layer. This number corresponds to a fraction of 0.036 of the total number of gridpoints. Observation errors were uncorrelated, with standard deviation 5×10^5 m^{2}s^{-1}.

To understand what else is needed, we show the evolution equation for the particles in pseudo time s,

$$\frac{d\mathbf{x}_j}{ds} = \mathbf{D} \frac{1}{N} \sum_{l=1}^{N} \Big(\mathbf{K}(\mathbf{x}_j, \mathbf{x}_l) \nabla_\mathbf{x} \log f(\mathbf{x}_l | \mathbf{d}) + \nabla_{\mathbf{x}_l} \mathbf{K}(\mathbf{x}_j, \mathbf{x}_l) \Big). \tag{20.5}$$

This equation shows we need to provide three ingredients: the likelihood $f(\mathbf{d}|\mathbf{x})$, a continuous version of the prior $f(\mathbf{x})$, and the matrix-valued kernel $\mathbf{K}$. We assume the likelihood is known, and we do have a representation of the prior by a set of particles.
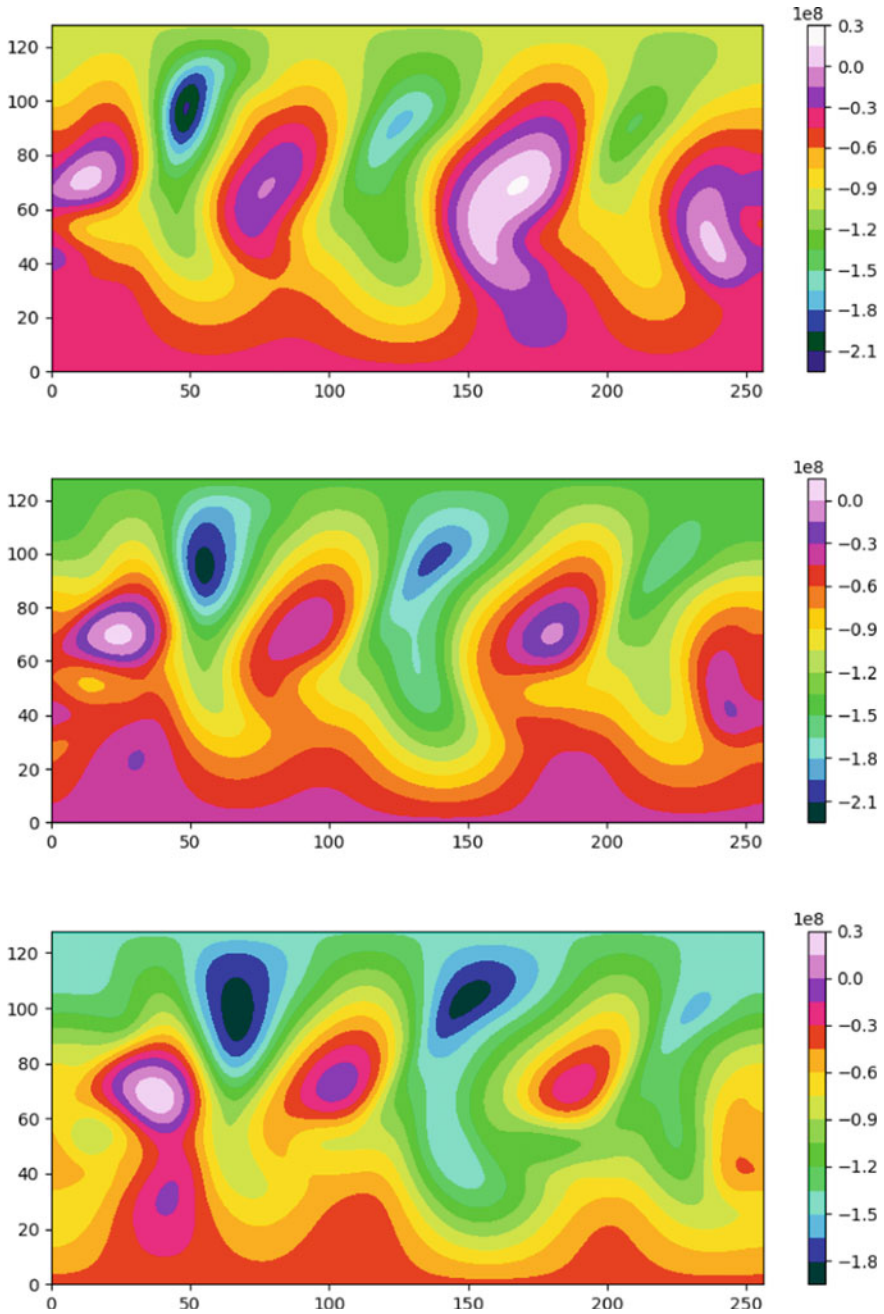
Fig. 20.2 The plots show the upper-layer stream-function fields from the true run 25 h apart from top to bottom

In the evolution equation for the particles, we need to take the gradient of the prior pdf to the state $\mathbf{x}$, so a representation in terms of delta functions is not sufficient. Several possibilities for approximations are possible. One is to assume the prior is a Gaussian mixture model centered on the particle positions. Another is to use a single Gaussian as the prior pdf. Note that this approximation is only needed to find an approximate gradient of the prior. The prior particles can still represent a non-Gaussian pdf. This situation is similar to EnKF, which updates each ensemble member separately. In the EnKF, the posterior pdf can retain non-Gaussian structures present in the prior ensemble even though it uses a Gaussian approximation to define the update. In this application, we assumed that the prior particles represent a Gaussian pdf, defined by the ensemble mean and ensemble covariance.

We use a matrix-valued kernel with off-diagonal entries equal to zero, and on the diagonal a scalar Gaussian kernel,

$$k_{ii}\left(\mathbf{x}_j, \mathbf{x}_l\right) = \exp\left(-\frac{1}{2}\frac{\left(x_j^i - x_l^i\right)^2}{\sigma_i^2}\right), \qquad (20.6)$$

where σ_i^2 is the prior variance in state variable i. In the limit of an infinite number of ensemble members, theory tells us that any smooth, symmetric kernel will result in the prior particles converging to the posterior pdf. In practice, with a small number of particles, care has to be taken to ensure fast convergence.

20.4   Results

Figure 20.3 shows the prior mean, the truth, and the posterior mean of the lower layer stream-function fields at day 10 of the assimilation experiment as an example of the outputs. The posterior mean is indeed much closer to the truth than the prior mean, as expected. The data assimilation manages to deepen low-pressure areas, make high-pressure regions less deep, and generate a more accurate splitting of the Jet Stream around the gridpoint (200, 70).

Figure 20.4 compares the time evolution of the spatially averaged mean-square errors of the ensemble mean and the ensemble variance. We see the typical decrease of errors at assimilation times and the growth of the actual and predicted errors between assimilation times. The two curves closely follow each other, showing that the ensemble spread is a realistic estimate of the true error (defined as the square of the difference between ensemble mean and the truth run).

The particle-flow filter is an iterative scheme that reduces the KL-divergence at every time step. To illustrate this property, the right plot in Fig. 20.4 shows the mean square error spatially averaged between the ensemble mean and the truth as a function of the iteration number. Each line corresponds to a different observation time. The error converges to a fixed value, mainly determined by the observation error. For practical reasons, we limited the number of iterations to 50, but additional iterations could have reduced the divergence even further.
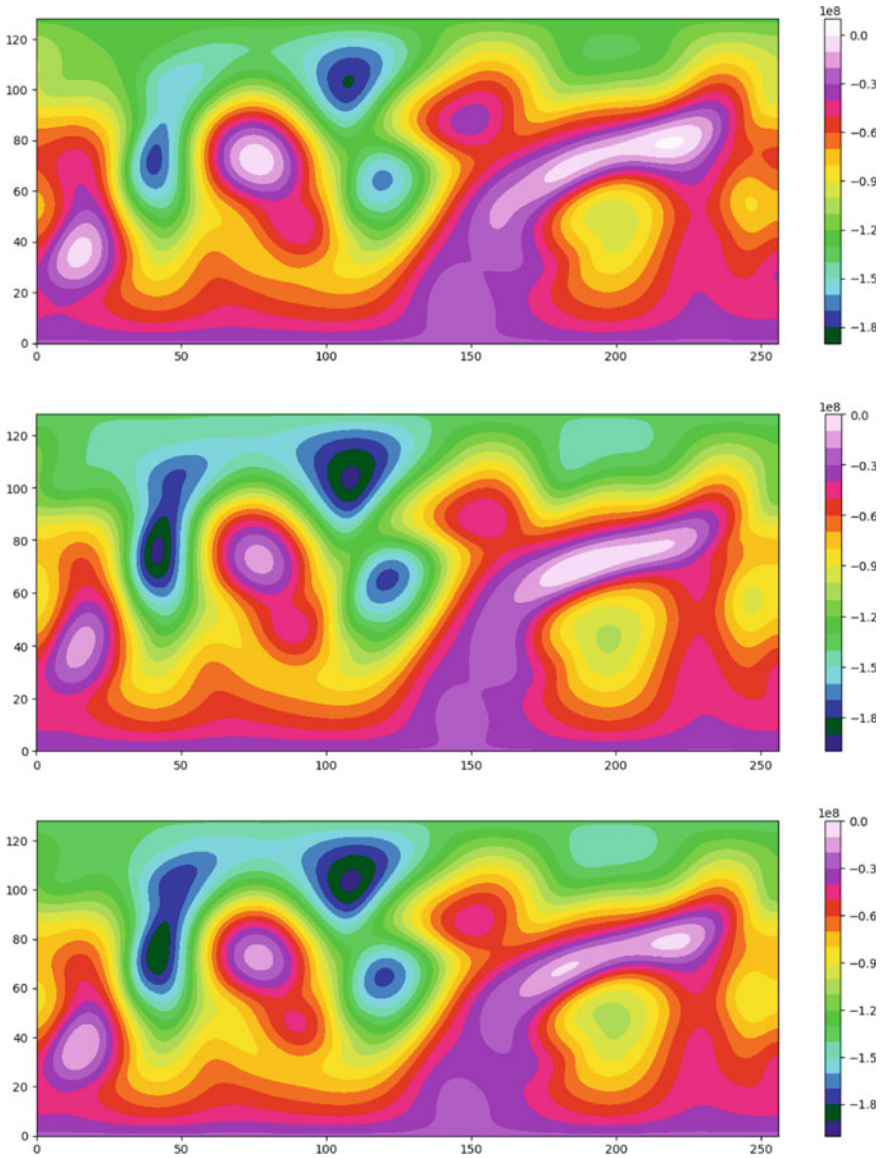
Fig. 20.3 The plots show the lower-layer stream-function fields' prior mean, truth, and posterior mean at day ten from top to bottom

We obtained these experimental results by observing the stream-function value directly at 600 points in each layer at each observation time. It is interesting to see what happens when using a nonlinear observation operator. We also performed an experiment where we observed the square of the stream function at each observation point. This situation typically leads to a skewed posterior pdf when all prior particles
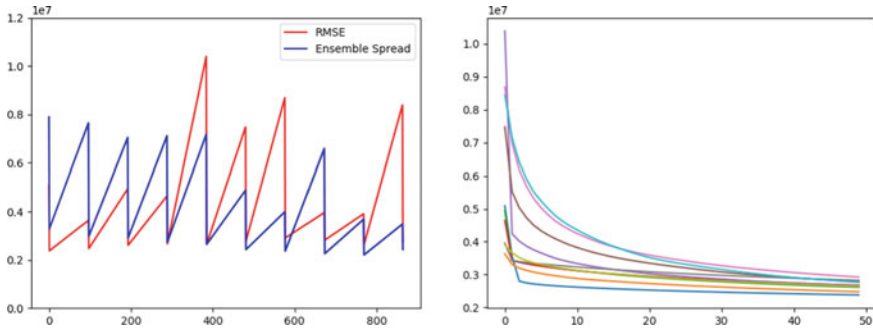
Fig. 20.4 The left plot shows the time evolution of the true error (red line) and the ensemble variance (blue) over 19 d. The right plot shows the convergence of the ensemble mean to the truth as a function of the pseudo time step (or iteration). The different curves correspond to the observation times in the left plot

are positive. The likelihood is bimodal, but the prior only sees one of the modes. The skewness arises because of the nonlinear transformation between state and observation space. The more exciting situation appears in observed gridpoints where prior particles have different signs for the stream-function value. In that case, both the positive and the negative root of the observation are covered by the prior. Thus, the likelihood will be bimodal in the domain where the prior is non-zero. Figure 20.5 depicts what can happen in such a case. Since the observation is the square of the stream-function value, it points to two possible solutions, one positive and one negative. The blue histogram represents the prior pdf. The red bars indicate the possible values of the observation at this gridpoint, and the orange histogram represents the
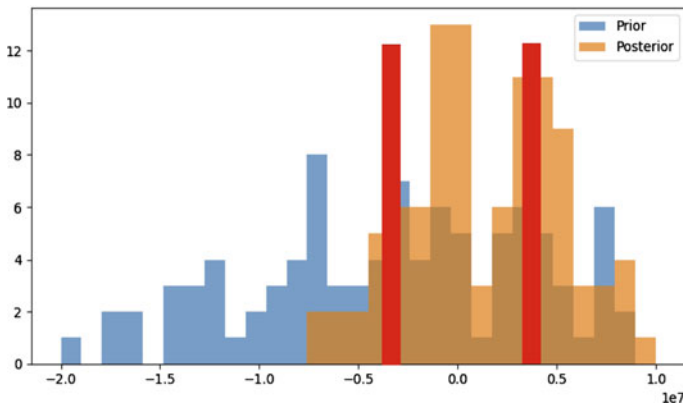
Fig. 20.5 The figure shows the QG model's particle-flow filter results in a selected gridpoint using a quadratic observation operator. The two red bars denote the two possible positions of the observation in state space. The prior (blue) and posterior (orange) histograms represent the distribution of the 100 prior and posterior particles

posterior pdf. The prior pdf is a wide pdf with no particular structure. The likelihood in state space is bimodal, and the posterior pdf is indeed bimodal as expected.

This example demonstrates how to set up a particle-flow filter in a large-dimensional system. Localization is not needed explicitly. Research on these methods is still in its infancy, but fully nonlinear data assimilation seems to have come within reach.