# Improving Particle Filter Performance by Smoothing Observations

GREGOR ROBINSON, IAN GROOMS, AND WILLIAM KLEIBER

*Department of Applied Mathematics, University of Colorado Boulder, Boulder, Colorado*

## ABSTRACT

This article shows that increasing the observation variance at small scales can reduce the ensemble size required to avoid collapse in particle filtering of spatially extended dynamics and improve the resulting uncertainty quantification at large scales. Particle filter weights depend on how well ensemble members agree with observations, and collapse occurs when a few ensemble members receive most of the weight. Collapse causes catastrophic variance underestimation. Increasing small-scale variance in the observation error model reduces the incidence of collapse by de-emphasizing small-scale differences between the ensemble members and the observations. Doing so smooths the posterior mean, though it does not smooth the individual ensemble members. Two options for implementing the proposed observation error model are described. Taking a discretized elliptic differential operator as an observation error covariance matrix provides the desired property of a spectrum that grows in the approach to small scales. This choice also introduces structure exploitable by scalable computation techniques, including multigrid solvers and multiresolution approximations to the corresponding integral operator. Alternatively the observations can be smoothed and then assimilated under the assumption of independent errors, which is equivalent to assuming large errors at small scales. The method is demonstrated on a linear stochastic partial differential equation, where it significantly reduces the occurrence of particle filter collapse while maintaining accuracy. It also improves continuous ranked probability scores by as much as 25%, indicating that the weighted ensemble more accurately represents the true distribution. The method is compatible with other techniques for improving the performance of particle filters.

## 1. Introduction

Particle filters are a class of ensemble-based methods for solving sequential Bayesian estimation problems. They are uniquely celebrated because of their provable convergence to the correct posterior distribution in the limit of an infinite number of particles, with minimal constraints on prior and likelihood structure (Crisan and Doucet 2002). Processes that are nonlinear and non-Gaussian can be filtered in this flexible framework, with rigorous assurances of asymptotically correct uncertainty quantification. These advantages stand in contrast to ensemble Kalman filters that lack convergence guarantees for nonlinear or non-Gaussian problems, and to variational methods that provide a point estimate but do not quantify uncertainty in the common case where the Hessian of the objective is unavailable.

The simplest form of a particle filter is descriptively called sequential importance sampling (SIS).

We briefly describe the algorithm here to fix notation and terminology, and recommend Doucet et al. (2001) for a gentler introduction.

SIS begins by approximating the prior probability distribution with density $p(\mathbf{x}_{j-1})$ at discrete time $j-1$ as a weighted ensemble of $N_e$ members $\{[\mathbf{x}_{j-1}^{(i)}, w_{j-1}^{(i)}]\}$, where the weights $w_{j-1}^{(i)}$ are related to the prior probabilities of the corresponding states $\mathbf{x}_{j-1}^{(i)}$. The superscript $(i)$ indexes the collection of particles, and the sum of the weights is one. This kind of approximation, an importance sample, is an ensemble drawn from one distribution that is easy to sample and then reweighted to represent another distribution of interest.

The distribution of interest here is the Bayesian posterior at discrete time $j$, which is proportional to the product of the prior $p(\mathbf{x}_{j-1})$ at time $j-1$, the transition kernel $p(\mathbf{x}_j|\mathbf{x}_{j-1})$, and the likelihood $p(\mathbf{y}_j|\mathbf{x}_j)$. SIS evolves the samples from time $j-1$ to time $j$ according to a proposal kernel that takes the generic form $p[\mathbf{x}_j^{(i)}|\mathbf{x}_{0:j-1}^{(i)}, \mathbf{y}_j]$. The weights are updated to reflect the difference between the proposal kernel and the Bayesian posterior at time $j$:

*Corresponding author*: Ian Grooms, ian.grooms@colorado.edu

$$w_j^i \propto w_{j-1}^i \frac{p\left[\mathbf{y}_j | \mathbf{x}_j^{(i)}\right] p\left[\mathbf{x}_j^{(i)} | \mathbf{x}_{j-1}^{(i)}\right]}{p\left[\mathbf{x}_j^{(i)} | \mathbf{x}_{0:j-1}^{(i)}, \mathbf{y}_j\right]}. \tag{1}$$

The proposal kernel is often set to equal the transition kernel, which simplifies the ratio in (1) so that the weights are proportional to the likelihood: $w_j^i \propto w_{j-1}^i \cdot p = [\mathbf{y}_j | \mathbf{x}_j^{(i)}]$. The proportionality constant is chosen so that the weights sum to one. [Some authors, e.g., van Leeuwen (2010), integrate out dependence on $x_{j-1}$; we instead follow the convention of Doucet et al. (2001).]

Despite its attractive qualities, particle filtering is unpopular in meteorological applications because of an especially vexing curse of dimensionality. The problem is that the importance sampling weights associated with system replicates (particles) have a tendency to develop degeneracy as the system dimension grows. That is to say, a single particle near the observation will have essentially all the sampling weight while the rest of the particles, bearing effectively zero weight, are ignored in the computation of ensemble statistics.

One can quantify the degree of degeneracy with an effective sample size (ESS), which is a heuristic measurement of the importance sample quality defined as

$$\text{ESS}_j = \frac{1}{\sum_{i=1}^{N_e} \left[w_j^{(i)}\right]^2}. \tag{2}$$

The ESS ranges from one if a single weight is nonzero (which is the worst case), to $N_e$ if all weights are equal. If the effective sample size becomes much smaller than the ensemble size, the filter is said to have collapsed. A simple approach to combat collapse is to resample the particles from time to time, eliminating particles with low weight and replicating particles with high weights. There are several common approaches to resampling (e.g., Doucet and Johansen 2011), and by construction of this resampling step, all weights become uniform: $w_j^{(i)} \rightarrow 1/N_e$ [see also the more recent resampling alternatives in Reich (2013) and Acevedo et al. (2017)]. The term "particle filter" commonly implies an SIS filter with a resampling step, also known as sequential importance resampling (SIR).

SIR particle filters are guaranteed to converge to the correct Bayesian posterior in the limit of an infinite number of particles, but the rate of convergence can be prohibitively slow for high-dimensional problems. The number of particles required to avoid collapse is typically exponential in a quantity related to the number of observations, as described by Bengtsson et al. (2008) and Snyder et al. (2008). For example, consider a system with Gaussian prior on $\mathbf{x}_j$ and with likelihood, conditional on $\mathbf{x}_j$:

$$\mathbf{y}_j | \mathbf{x}_j \sim \mathcal{N}(\mathbf{H}\mathbf{x}_j, \mathbf{R}), \tag{3}$$

where $\mathcal{N}(\mu, \mathbf{\Sigma})$ denotes a multivariate normal distribution with mean $\mu$ and covariance $\mathbf{\Sigma}$, $\mathbf{H}$ is a linear observation operator, and $\mathbf{R}$ is the covariance of the additive observation error. For this example Snyder et al. (2008) show that the number of particles $N_e$ required to avoid collapse is on the order of $\exp(\tau^2/2)$, where

$$\tau^2 = \sum_{k=1}^{N_y} \lambda_k^2 \left(\frac{3}{2}\lambda_k^2 + 1\right), \tag{4}$$

in which $N_y$ is the dimension of the observations and $\lambda_k^2$ are eigenvalues of

$$\text{cov}(\mathbf{R}^{-1/2}\mathbf{H}\mathbf{x}_j). \tag{5}$$

Chorin and Morzfeld (2013) also discuss the notion of "effective dimension" and how it relates to particle filter performance. Agapiou et al. (2017) give precise, nonasymptotic results on the relationship between the accuracy of the particle filter, the number of particles, and the effective dimension of the filtering problem in both finite- and infinite dimensional dynamical systems. For simplicity of exposition we rely on the formulas quoted here from Snyder et al. (2008) and Snyder et al. (2015).

A number of methods developed to minimize degeneracy in high-dimensional problems utilize a proposal kernel that is different from the transition prior, using observations to guide proposals. Of all possible proposals that depend only on the previous system state and the present observations, there exists an optimal proposal that minimizes both the variance of the weights and the number of particles required to avoid degeneracy (Doucet et al. 2000; Snyder et al. 2015). It is typically impractical to sample from that optimal proposal. The various methods proposed to minimize weight degeneracy in practice include the implicit particle filter (Chorin and Tu 2009; Chorin et al. 2010; Chorin and Tu 2012; Morzfeld et al. 2012), and the equivalent weights particle filter (van Leeuwen 2010; Ades and van Leeuwen 2013, 2015). Snyder et al. (2015) have shown that improved proposals can reduce the number of particles required to avoid collapse, but the number is still prohibitive for meteorological applications. Another approach to improving the performance of particle filters uses "localization." Localization reduces the effective number of observations (and therefore the required number of particles) by breaking the assimilation into a sequence of smaller subsets. Localization can also improve the performance of particle filters (Penny and Miyoshi 2016; Rebeschini and Van Handel 2015; Poterjoy 2016; Morzfeld et al. 2017), but

breaks convergence guarantees. Other methods improve the filter results by making the observation error model state dependent (Okamoto et al. 2014; Zhu et al. 2016).

This paper describes a different but compatible approach for improving the dimensional scaling of particle filters by smoothing observations before proceeding as though the observations are uncorrelated; equivalently, we increase the small-scale variance in the error model. The goal of doing so is to achieve more desirable dimensional scaling. Whereas changing the proposal kernel allows particle filtering to sample a given posterior more efficiently, manipulating the observation model changes the posterior itself. This may seem to vitiate convergence guarantees at least as badly as localization does. After all, it is possible that localized particle filters and EnKFs converge to some distribution in the large ensemble limit. However, convergence results are still an open problem for EnKFs and localized particle filters. In any case, the limiting distribution of a localized filter is not the true Bayesian filter, and the nature of the bias in the limiting distribution is unknown. By contrast, we can guarantee convergence to a surrogate distribution with bias that can be described and controlled.

The key insight motivating our approach is evident in (5): increasing the observation error variance for any eigenvector of $\mathbf{R}$ correspondingly decreases the number of particles required. The challenge is to make the problem less expensive to sample with a particle filter, while still accurately incorporating observations on the most physically relevant large scales. This paper describes an analytically transparent and computationally efficient method that reduces the number of particles required to avoid collapse by increasing the observation error variance at small scales.

## 2. Theory

In this section we develop intuition by considering the observation error model in (3) in the special case where $\mathbf{R}$ and $\text{cov}(\mathbf{x}_j)$ are Fourier diagonalizable and $\mathbf{H} = \mathbf{I}$. Writing eigenvalues of $\mathbf{R}$ as $\gamma_k^2$ with $k$ an integer wavenumber from 1 to $N_y$, and the eigenvalues of $\text{cov}(\mathbf{x}_j)$ as $\sigma_k^2$, the matrix in (5) has the following eigenvalues:

$$\lambda_k^2 = \sigma_k^2/\gamma_k^2. \tag{6}$$

The effects of aliasing complicate the Fourier-scale analysis of filtering when observations are not available at every grid point, especially when the observation grid is irregular (Majda and Harlim 2012, chapter 7).

Recall from the introduction that Snyder et al.'s estimate in (4) of the ensemble size required depends on the system covariance, the observing system, and the observation error covariance. Let us ground the theoretical discussion with general comments about the nature of these quantities in operational numerical weather prediction. Typically the model physics are reasonably well known and held fixed, so we take $\text{cov}(\mathbf{x}_j)$ to be given.[1] The observing system, like the dynamical model, is typically given and fixed. The observation error covariance, contrasting both the dynamical model and the observing system, is often a crude heuristic approximation that is easier to modify. Observation error is frequently taken to have no spatial correlation, for example $\mathbf{R} \propto \mathbf{I}$ in the case of distant identical thermometers, in which case $\{\gamma_k\}$ are constant. Otherwise, the observation error may have strong spatial correlations, as may be expected of satellite observations biased by a spatially smooth distribution of unobserved atmospheric particulates, in which case $\gamma_k \to 0$ rapidly for large $k$.

### a. Impact of observation error model on number of particles required

The following hypothetical examples demonstrate how the observation error model can affect the number of particles required for particle filtering. We first use Snyder's asymptotic arguments to estimate the particle filter ensemble size required to reconstruct a Bayesian posterior with a correlated observation error model, whose realizations are continuous with probability one, and contrast this with the ensemble size required under the approximation that observation errors are spatially uncorrelated. Making this approximation decreases the particle filter ensemble size required to reconstruct the Bayesian posterior. This progression is designed to set the stage for our method; we show that using a peculiar choice of $\mathbf{R}$, possessing a growing spectrum, naturally extends the approximation of correlated errors with uncorrelated errors. Our method decreases the number of particles required to approximate the posterior regardless of whether the true errors are correlated or uncorrelated.

---

[1] One can in principle design physical models to make an assimilation problem more tractable to a particle filter, homologous to the approach we describe that alters the observation model. We do not consider that in this article because the theory scantly differs and the praxis is much more problem dependent. The related representation errors, arising from a mismatch between the length scales resolvable by the numerical model and the length scales present in the observations, are difficult to quantify but are presumably spatially correlated.

Fields whose correlations gradually decrease with distance have decaying spectra (i.e., $\gamma_k^2 \to 0$ at small scales). This has a detrimental effect on the effective dimensionality of the problem. Suppose, for example, that observation error variances $\gamma_k^2 = k^{-4}$ and system covariances $\sigma_k^2 = k^{-2}$. Then eigenvalues of (5) are $\lambda_k^2 = k^2$ and

$$\tau^2 \approx \int_1^{N_y} k^2 \left( \frac{3}{2} k^2 + 1 \right) dk \sim \frac{3}{10} N_y^5, \qquad (7)$$

where the sum in (4) has been approximated by an integral. In this example the effective dimensionality of the problem increases extremely rapidly as the number of observations grows. A similar argument can be used to show that if $\sigma_k^2$ decays sufficiently faster than $\gamma_k$ at small scales (large $k$), then the effective dimensionality of the system remains bounded in the continuum limit.

When the spatial correlation of the observation error is unknown, it is not uncommon to use a spatially uncorrelated (i.e., diagonal) observation error model. This approximation is also popular because it is computationally convenient in ensemble Kalman filters, where it enables serial assimilation (Houtekamer and Mitchell 2001; Bishop et al. 2001; Whitaker and Hamill 2002). For observations with correlated errors, such as swaths of remotely sensed data, approximating the errors as spatially uncorrelated changes the posterior relative to a more accurate observation error model with correlations; the approximation seems to work well enough in practice. The spatially uncorrelated approximation, compared to error models with continuous realizations, also makes particle filtering easier. When the error is spatially uncorrelated, $\gamma_k^2$ does not decay to zero at small scales. Repeating the asymptotic argument in the preceding paragraph with constant $\gamma_k^2 = 1$ implies $\lambda_k^2 = k^{-2}$, so

$$\tau^2 \approx \int_1^{N_y} k^{-2} \left( \frac{3}{2} k^{-2} + 1 \right) dk \sim \frac{3}{2} \qquad (8)$$

in the continuum limit. This illustrates that the number of particles required to avoid collapse can be significantly reduced by changing the spatial correlations in the observation error model, and in practice the filter results are still acceptably accurate.

Our proposal is take this approximation a step further: we let observation error covariance grow without bound in the progression to small scales. This model of the observation error, possessing a spectrum bounded away from zero, is called a generalized random field (GRF) and has peculiar properties described in the appendix. Despite those peculiarities of GRFs that complicate

analysis of the continuum limit, the finite-dimensional vector of observational errors can be treated as a multivariate Gaussian random vector.

In sections 2b and 2c, we discuss the impact of this observation error model on the posterior, and various numerical methods for constructing and implementing the associated particle filter. We find the theory to be more intuitive in terms of this covariance framework than working with smoothing operators, but section 2c will make the equivalence precise.

### b. Effect of a generalized random field likelihood on posterior

The performance advantage, described above, does not come for free. Changing the observation error model changes the posterior. To demonstrate how our choice of error model affects the posterior, consider again a fully Gaussian system for which the system covariance $\text{cov}(\mathbf{x}_j)$ has the same eigenvectors as the presumed observation error covariance $\mathbf{R}$, and where the observation operator is the identity. Let $\sigma_k^2$ be eigenvalues of $\text{cov}(\mathbf{x}_j)$ and $\gamma_k^2$ be eigenvalues of $\mathbf{R}$, indexed by $k$ in the diagonalizing basis with index $k$ increasing toward small scales. Let $\hat{\mathbf{x}}_k$ and $\hat{\mathbf{y}}_k$ denote the projection of the prior mean and observations onto the $k$th eigenvector, respectively. Then the posterior mean of $p(\hat{\mathbf{x}}_k | \hat{\mathbf{y}}_k)$ is

$$\hat{\mathbf{x}}_k + \frac{\sigma_k^2}{\sigma_k^2 + \gamma_k^2} (\hat{\mathbf{y}}_k - \hat{\mathbf{x}}_k). \qquad (9)$$

In order for the posterior mean to be accurate at large scales, it will be necessary to design an observation error model with realistic variance at large scales; we return to this point in section 2c. Clearly, if $\gamma_k^2 \to \infty$ at small scales then the posterior mean will equal the prior mean at small scales. If the filter tends to ignore small-scale information, then the small-scale part of the prior mean will eventually tend toward the climatological small-scale mean, which is often zero since climatological means are often large scale. This observation error model can, therefore, be expected to have a smoothing effect on the posterior mean.

This is the price to be paid for reducing the effective dimensionality of the system, but the price is not too high. Small scales are inherently less predictable than large scales, so loss of small-scale observational information may not significantly damage the accuracy of forecasts. Practical implementations will need to balance between ignoring enough observational information to avoid particle collapse and keeping enough to avoid filter divergence (i.e., the filter wandering away from the true state of the system).

In the same example as above, the eigenvalues of the posterior covariance are

$$\xi_k^2 = \frac{\sigma_k^2 \gamma_k^2}{\sigma_k^2 + \gamma_k^2}.$$

As noted above, in order for the posterior variance to be accurate at large scales, it will be necessary to design an observation error model with realistic variance at large scales. At small scales we argue that $\xi_k^2$ is small (using the notation $\ll 1$) regardless of the behavior of $\gamma_k^2$. This is because the state $\mathbf{x}$ is associated with a viscous fluid model whose solutions should be continuous. A GRF error model with $1 \ll \gamma_k^2$ will lead to a posterior variance close to the prior variance at small scales: $\xi_k^2 \approx \sigma_k^2 \ll 1$. A more realistic error model with $\gamma_k^2 \ll 1$ will lead to a much smaller posterior variance, but in either case $\xi_k^2 \ll 1$ This argument suggests that the GRF approach should not have a detrimental effect on the posterior variance when applied to atmospheric or oceanic dynamics, provided that the observation error variance at large scales is realistic.

### c. Constructing GRF covariances

In the context of an SIR particle filter using the standard proposal with a nonlinear observation error model of the following form:

$$\mathbf{y}_j = \mathbf{H}(\mathbf{x}_j) + \boldsymbol{\eta}_j,$$

where $\boldsymbol{\eta}_j \sim \mathcal{N}(0, \mathbf{R})$ is the observation error, the incremental weights are computed using

$$w_j^{(i)} \propto w_{j-1}^{(i)} \exp\left\{-\frac{1}{2}\left[\mathbf{y}_j - \mathbf{H}(\mathbf{x}_j^{(i)})\right]^{\mathrm{T}} \mathbf{R}^{-1}\left[\mathbf{y}_j - \mathbf{H}(\mathbf{x}_j^{(i)})\right]\right\}.$$

The goal of this section is to describe two methods for defining an observation error covariance $\mathbf{R}$ that have the increasing variance prescribed above, and that allow for rapid computation of the weights. First, we will suppose that the true observation error variance is known, and we will scale it out so that we are dealing only with the error correlation matrix. If $\mathbf{R}_0$ is a diagonal matrix with elements that are the observational error variances, then we will let

$$\mathbf{R} = \mathbf{R}_0^{1/2} \mathbf{C} \mathbf{R}_0^{1/2},$$

and we will model the matrix $\mathbf{C}$.

There is a well-known connection between stationary Gaussian random fields and elliptic stochastic partial differential equations (Rue and Held 2005; Lindgren et al. 2011) that allows fast approximation of likelihoods. Specifically, the inverse of the covariance matrix of a discretized random field can in some cases be identified with the discretization of a self-adjoint elliptic partial differential equation (PDE). The connection extends in a natural way to generalized Gaussian random fields, with the caveat that the covariance matrix rather than its inverse is identified with the discretization of an elliptic PDE. For example, the matrix $\mathbf{C}$ can be constructed as a discretization of the operator

$$(1 - \ell^2 \Delta)^\kappa, \tag{10}$$

in which $\Delta$ is the Laplacian operator, $\ell > 0$ is a tuning parameter with dimensions of length, and $\kappa > 0$ controls the rate of growth of eigenvalues. Both the continuous differential operator and its discretization have positive spectra with eigenvalues growing in wavenumber. The parameter $\ell > 0$ controls the range of scales with eigenvalues close to 1. For length scales longer than $\ell$ the eigenvalues are close to 1 and the observation error model is similar to the commonly used diagonal, uncorrelated observation error model. The large-scale observation error is correct, meaning that the posterior will also be correct at large scales. For length scales smaller than $\ell$ the observation error variance grows at a rate determined by $\kappa$, rapidly rolling off the influence of small scales.

Taking the matrix $\mathbf{C}$ to be a discretization of an elliptic PDE permits efficient application of the inverse, as required in computing the weights, by means of sparse solvers. It is also possible to construct $\mathbf{C}^{-1}$ directly as the discretization of the integral operator that corresponds to the inverse of this PDE, also enabling fast algorithms that have no limitation to regular observation grids. These kinds of methods will be explored more fully elsewhere.

An alternative to the PDE-based approach for modeling $\mathbf{C}$ is to simply smooth the observations. Let the smoothing operator be a matrix $\mathbf{S}$, and the smoothed observations be denoted $\mathbf{y}_s$. Then the observation model,

$$\mathbf{y}_s = \mathbf{S}\mathbf{R}_0^{-1/2}\mathbf{y}_j = \mathbf{S}\mathbf{R}_0^{-1/2}\mathbf{H}(\mathbf{x}_j) + \boldsymbol{\eta}_s,$$

where the smoothed observation errors are assumed to have independent, unit-variance errors, implies incremental importance weights of the following form:

$$w_j^{(i)} \propto w_{j-1}^{(i)} \times \exp\left\{-\frac{1}{2}\left[\mathbf{y}_j - \mathbf{H}(\mathbf{x}_j^{(i)})\right]^{\mathrm{T}} \mathbf{R}_0^{-1/2} \mathbf{S}^{\mathrm{T}} \mathbf{S} \mathbf{R}_0^{-1/2}\right.$$
$$\left. \times \left[\mathbf{y}_j - \mathbf{H}(\mathbf{x}_j^{(i)})\right]\right\}.$$

If a smoothing operator **S** is available, our proposed method is therefore equivalent to setting $\mathbf{C}^{-1} = \mathbf{S}^T\mathbf{S}$. As long as the smoothing operator leaves large scales nearly unchanged while attenuating small scales, the impact on the effective sample size and on the posterior will be as described in the foregoing subsections. If it is possible to construct **S** to *project* onto a large-scale subspace, it would be equivalent to setting certain eigenvalues of the observation error covariance to infinity.

## 3. Experimental configuration

To illustrate the effects of a GRF likelihood in a simple example, we apply an SIR particle filter to a one-dimensional linear stochastic partial differential equation:

$$\frac{du}{dt} = \left(-b - c\frac{d}{dx} + v\frac{d^2}{dx^2}\right)u + F_t, \qquad (11)$$

where $b$, $c$, $v \in \mathbb{R}^+$ are constant scalars and $F$ is a time-dependent stochastic forcing that is white in time and correlated in space with a form described below. The domain is periodic, with length $2\pi$. Such models have been used to test filtering algorithms by Majda and Harlim (2012). In Fourier space, this model can be represented as the Itô equation:

$$d\hat{u} = -(b + ikc + vk^2)\hat{u}\,dt + \zeta\,dW, \qquad (12)$$

where $\hat{u}$ is the Fourier coefficient at wavenumber $k$, $\zeta$ is the noise amplitude, and $dW$ is a standard circularly symmetric complex white noise. The coefficients are $b = 1$, $c = 2\pi$, and $v = 1/9$. To mimic turbulence in many physical models, we choose a stochastic forcing $F_t$ that decays linearly for large wavenumbers. Specifically, let

$$\zeta^2 = 1/(1 + |k|), \qquad (13)$$

such that the variance of the noise is one-half of its maximum at wavenumber 1. This configuration in (11)–(13) is chosen to possess a fairly limited range of active wavenumbers so that the particle filtering problem is tractable.

The model admits an analytical solution to which we can compare experimental results. Since the dynamic is linear and Fourier coefficients are independent, it follows that each Fourier mode evolves as an Ornstein–Uhlenbeck process independent of all other modes. This means we can efficiently propagate the system by sampling directly from the Gaussian distribution available in closed form for each Fourier coefficient (Øksendal 2003):

$$\hat{u}_{t+\Delta t} = \hat{u}_t e^{-\theta_k \Delta t} + \zeta\sqrt{\frac{1 - e^{-2\theta_{r,k}\Delta t}}{2\theta_{r,k}}}\chi_t, \qquad (14)$$

where $\theta_k = d + ikc + vk^2$, $\theta_{r,k}$ is the real part of $\theta_k$, and $\chi_t$ is a standard circularly symmetric complex normal random variable. The initial condition for the experiment is drawn from the stationary distribution, obtained as the limit $\Delta t \to \infty$ in (14), which for each wavenumber is a circularly symmetric complex normal random number of mean zero and standard deviation $1/[2(1 + |k|)\theta_{r,k}]^{1/2}$.

A particular solution, hereafter called the "true system state" solution is computed at 2048 equally spaced points in the $2\pi$-periodic spatial domain, and at 101 equally spaced points in the time interval [0, 4] (the initial condition being at $t = 0$). From this solution, synthetic observations are generated at every 32nd spatial location (except as otherwise noted) by adding samples from a stationary zero-mean multivariate normal distribution with variance 0.36 and correlations of the form $\exp(-|\delta/0.06|)$, where $\delta$ is the distance between observations. There are thus $64 \times 100$ total observations (there are no observations of the initial condition).

The standard deviation of the observational error is 0.6, while the pointwise climatological standard deviation of the system is about 0.8. This is a very high observational noise level; we set the observational noise this high because the theoretical estimates of the required ensemble size are extremely large for smaller observational noise. Observational noise levels in meteorological applications are not usually this high relative to the climatological variability of the system. Despite this high level of noise, the observing system is dense enough in space and time that the filter is able to recover an accurate estimate of the system.

The GRF observation error covariance, used only for assimilation, is constructed as the periodic tridiagonal matrix formed by the second-order-centered finite-difference approximation to the operator $0.36(1 - \ell^2\partial_x^2)$. The diagonal elements (the observation error variance) are all $0.36[1 + 2(\ell/\delta)^2]$, where $\delta$ is the distance between observations; the elements corresponding to nearest-neighbor covariances are all $0.36[1 - (\ell/\delta)^2]$. When $\ell = 0$, the observation error covariance is diagonal. The local observation error variances increase when $\ell$ increases, and the nearest-neighbor covariances decrease and can even become negative. The eigenvectors of this matrix are discrete Fourier modes. When $\ell$ increases, the variance increases for all Fourier modes except the constant mode, which remains at this baseline variance of 0.36. Experiments are run with 101 values of $\ell^2$ equally spaced in the interval [0, 1]. The GRF

observation error covariance is not used to generate the synthetic observations.

Assimilation experiments are run with an SIR particle filter to test how the GRF observation error model impacts its performance. An ensemble size of $N_e = 400$ is used, except as noted otherwise. The SIR particle filter is configured to resample using the standard multinomial resampling algorithm in Doucet et al. (2001). The ESS is tracked before resampling. Resampling reduces the information content of the ensemble by eliminating some particles and replicating others; to avoid unnecessary loss of information, resampling is only performed whenever the ESS falls below $N_e/2$.

Two quantities are used to evaluate the effect of the GRF error model on the particle filter's performance. The first is the root-mean-square error (RMSE) between the particle filter's posterior mean and the true system state, where the mean is taken over the spatial domain. The second is the continuous ranked probability score (CRPS; Hersbach 2000; Gneiting and Raftery 2007). This measures the accuracy of the posterior distribution associated with the particle filter's weighted ensemble. The score is nonnegative; a score of zero is perfect, and smaller scores are better. It is more common to compare the RMSE to the ensemble spread, a function of the ensemble covariance trace (Fortin et al. 2014), but the CRPS is a more precise way to describe the quality of a probabilistic estimate. The CRPS is computed at every point of the spatial and temporal grid of $2048 \times 100$ points. We compute the CRPS for a range of different $N_y \in (16, 32, 64, 128)$ in order to probe the effects of changing the number of observations. All assimilation runs with the same $N_y$ use the same observations.

We will gauge particle filter performance with the GRF likelihood by comparing it to the reference case of a particle filter computed using a spatially uncorrelated likelihood. In some cases we will also want to compare the particle filter estimate to the true Bayesian posterior. Though one of the main reasons for using a particle filter is that it works in nonlinear, non-Gaussian problems, a benefit of experimenting with a linear Gaussian problem is that the exact solution to the optimal filtering problem can be computed for this comparison using the Kalman filter. In particular, the Kalman filter provides the exact posterior covariance $\mathbf{P}_k$:

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}^{\mathrm{T}}(\mathbf{R} + \mathbf{H}\mathbf{P}_{k|k-1}\mathbf{H}^{\mathrm{T}})^{-1}$$
$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k\mathbf{H})\mathbf{P}_{k|k-1},$$

which allows us to estimate the number of particles required to avoid filter degeneracy a priori (without

running the particle filter) using (4) and (5). The prior covariance at time $k$ is denoted $\mathbf{P}_{k|k-1}$ in the above formulas.

## 4. Results

We compute $\tau^2$ from the Kalman filter results at $t = 4$, the end of the assimilation window. This gives an approximation to the steady-state filtering problem because the posterior covariance converges exponentially to a limiting covariance (Chui and Chen 2009). This process is repeated for each of 11 $\ell^2$ values linearly distributed between 0 and 1 and the results are plotted in the first panel of Fig. 1. Note that the $\ell^2 = 0$ case is a spatially uncorrelated observation error model. We observe a dramatic reduction in the theoretical number of particles required to avoid filter collapse. The theory of Bengtsson et al. (2008) and Snyder et al. (2008) predicts that the spatially uncorrelated noise model requires on the order of $10^{26}$ particles to avoid collapse in this simple one-dimensional PDE with 2048 Fourier modes. As $\ell^2$ increases from 0 to 1, the number of required particles drops rapidly to about 8000. In fact, as shown below, the SIR particle filter performs well with $\ell^2 = 1$ for an ensemble size of 400.

Reducing $\tau^2$ by increasing $\ell^2$ is a result of increasing the observation variance, and the chosen form of the surrogate observation error model is designed to increase the variance primarily for small scales while leaving large scales intact. The impact on the posterior is visualized in the second panel of Fig. 1. This panel shows the time-averaged RMSE of the particle filter mean of the first 50 Fourier modes, normalized by the climatological standard deviation of each Fourier coefficient, for $\ell^2 \in (0, 0.04, 0.4)$. Here we observe that increasing $\ell^2$ primarily increases the posterior variance at small scales, as designed.

The distribution of ESS throughout the 100 assimilation cycles is plotted in Fig. 2 for various values of $\ell^2$. The boxplots are constructed from the time series of ESS over all 100 assimilation cycles. In this proxy for the quality of uncertainty quantification achieved by the particle filter, we observe approximately a tenfold increase in median ESS with $\ell^2 = 0.3$ and a thirtyfold increase in median ESS with $\ell^2 = 1$ compared to $\ell^2 = 0$. The ESS averages only 10%–20% of $N_e$ when $\ell^2 = 1$, with occasional collapses. This is not inconsistent with the theory, which requires $N_e$ of about 8000 to avoid collapse, yet still shows the significant improvements from using a GRF likelihood with relatively small ensembles. The results below suggest that the particle filter can give an accurate probabilistic estimate of the system
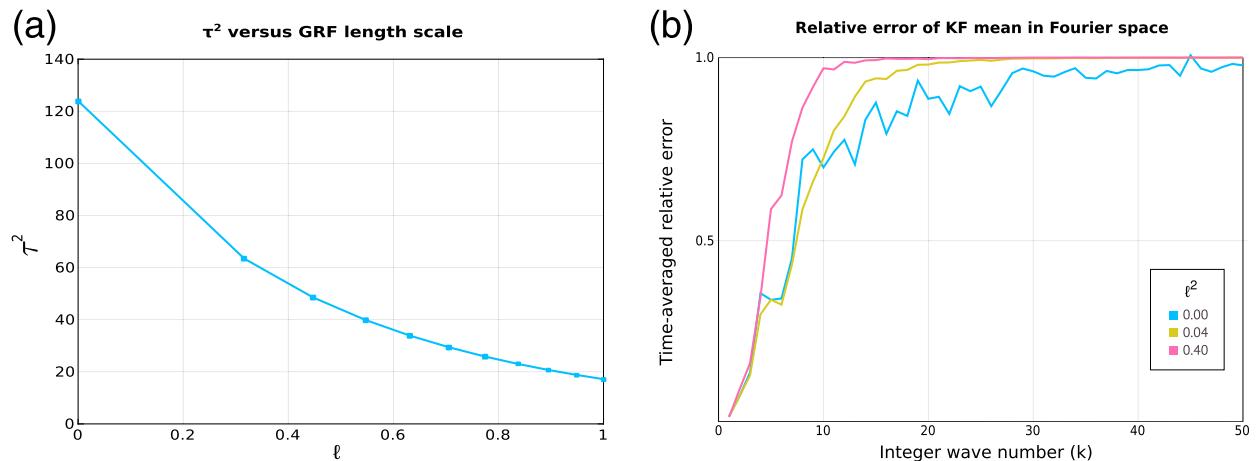
FIG. 1. (a) The $\tau^2$ in (4) is shown for different values of GRF length scale $\ell$. Because the number of particles required to avoid degeneracy increases exponentially in $\tau^2/2$, the observed decrease in $\tau^2$ as we roll off scales greater than $\ell$ indicates a reduced computational burden in using particle filtering for uncertainty quantification. Similarly, the decrease suggests that for fixed computation cost one may be able to mitigate the variance underestimation that tends to plague particle filters in high dimensions. Although the ordinate in this figure is $\ell$ to make direct contact with the length scale, all other figures are given in terms of $\ell^2$ to relate more directly to the spectrum of the GRF likelihood. (b) The RMSE in the Kalman filter's posterior mean, in Fourier space, normalized by the climatological standard deviation of each Fourier coefficient for different values of $\ell^2$: 0.00 (blue), 0.04 (yellow), and 0.40 (red). Here we see how the error in the posterior mean, considered as a function of wavenumber, approaches the climatological standard deviation more rapidly when $\ell^2$ is larger. It is exactly this posterior variance increase at small scales that underpins our approach: a posterior with larger total variance is easier for a particle filter to sample, while keeping the posterior accurate at large scales is key in making a forecast.

state even when the ESS is a small percentage of the ensemble size.

Next we consider how the RMSE of the particle filter posterior mean from the true system state depends on $\ell$. Figure 3 shows boxplots of the RMSE as a function of $\ell^2$. The boxplots are constructed from the RMSE time series for the final 90 assimilation time steps in each

experiment. The RMSE appears fairly insensitive to $\ell^2$. The median RMSE for all cases remains below the observation error standard deviation of 0.6. These results demonstrate that the particle filter remains a fairly accurate point estimator—both when the filter is collapsed while $\ell$ is small, and when the posterior is over-dispersed as a result of large $\ell$. The Kalman filter using
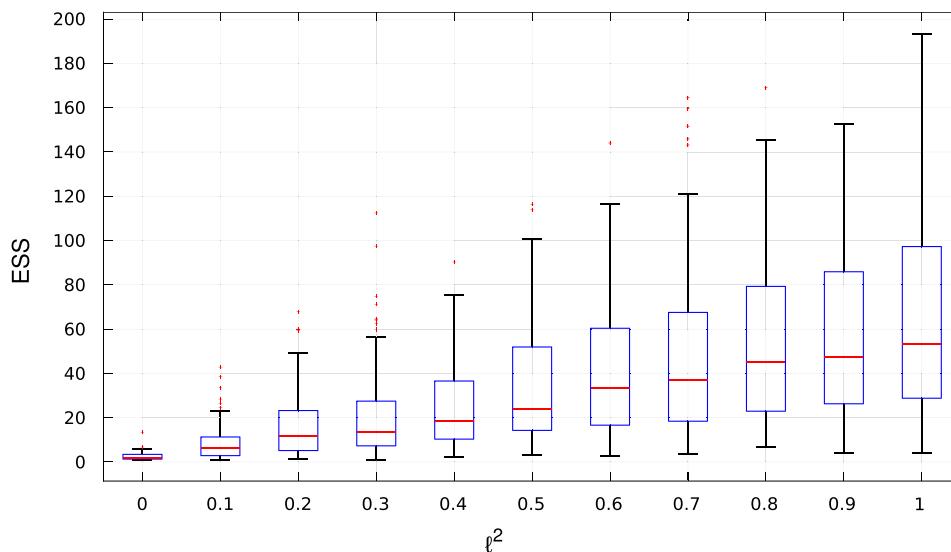


FIG. 2. Effective sample size distributions for different values of $\ell^2$ from 0 to 1. Each box represents the middle 50% quantile, a central line represents the median, and the whiskers span the data not considered outliers by the $1.5 \times$ IQR rule.
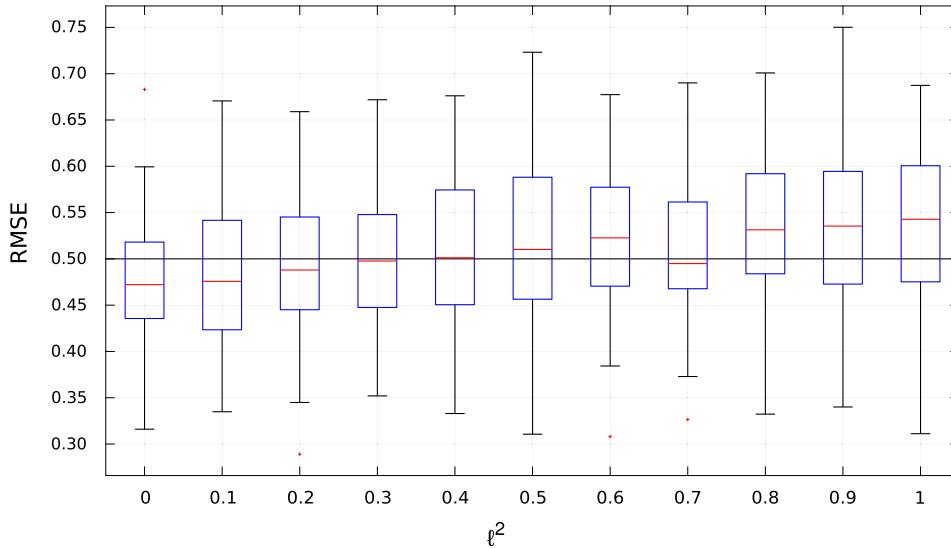
FIG. 3. RMSE between the truth and the posterior mean, using 11 different values of $\ell^2$ from 0 to 1. The first category, with $\ell^2 = 0$, corresponds to the uncorrelated observation error model. The RMSE using GRF likelihoods (i.e., $\ell^2 > 0$) does not dramatically suffer in comparison to that of the white likelihood that is more common in operational practice. In exchange for this small cost in RMSE, using the GRF likelihood comes with notable gain in the accuracy of uncertainty quantification. Each box represents the middle 50% quantile, a central line representing the median, and the whiskers span the data not considered outliers by the $1.5 \times$ IQR rule. The horizontal line at 0.5 serves only to guide the eye.

the true observation model, which is the optimal filter in the best-case scenario for this problem, achieves a median RMSE of 0.32.

The use of a GRF likelihood clearly reduces the incidence of collapse in the particle filter, with mild detriment to the RMSE. The RMSE measures a spatially integrated squared error, which can mask errors at small scales. The arguments of section 2b suggest that the GRF posterior mean will be inaccurate primarily at small scales. [We visualize the severity of this effect in Fig. 4, which compares the true state (red) to the posterior mean (blue) and to ensemble members (gray) for four different values of $\ell^2$: 0 (diagonal error model), 0.2, 0.4, and 0.6.] The ensemble members are shaded according to their weight: weights near 1 yield black lines while weights near 0 yield faint gray lines. At $\ell^2 = 0$ there are few ensemble members visible, reflecting the fact that the ESS is small. Nevertheless, the posterior mean is reasonably close to the true state. As $\ell^2$ increases, the number of visible ensemble members increases (reflecting increasing ESS), and the posterior mean becomes smoother. Although the posterior mean at $\ell^2 = 0.6$ is smoother than the true system state, the individual ensemble members are not overly smooth; they are instantiations of the dynamical model and are, as such, qualitatively similar to the true state.

The foregoing results have shown that the GRF observation error model improves the ESS without substantially damaging the RMSE, and that the posterior

mean is smoother than the true state, but also that the individual ensemble members (particles) are not too smooth. We finally test whether the uncertainty quantification afforded by the particle filter is improved by using a GRF observation error model. To this end we compute the CRPS at each point of the spatiotemporal grid of $2048 \times 100$ points. The median CRPS is computed using all 204 800 spatiotemporal grid points for 101 values of $\ell^2$ equally spaced between 0 and 1. The result is shown in Fig. 5. Median CRPS with $N_y = 64$ improves from about 0.27 at $\ell^2 = 0$ to 0.22 at $\ell^2 = 0.3$, and then remains steady or slightly increases at larger $\ell^2$.[2] Some sampling variability is still evident in the median CRPS, with occasional values as low as 0.21.

Varying the number of observations, also shown in Fig. 5, displays additional interesting behavior about the distributional estimate the particle filter provides. In each $N_y$ case we explored, there is a choice of $\ell^2$ that improves the particle filter CRPS. The differences in optimal $\ell^2$ emphasizes that the optimal parameter depends not only on the active scales in the underlying physics, but also on the resolution of the data.

---

[2] For comparison, the ensemble spread simultaneously improves by a factor of about 2, going from a time-averaged 36% of RMSE when $\ell^2 = 0$ to 71% RMSE when $\ell^2 = 1$.

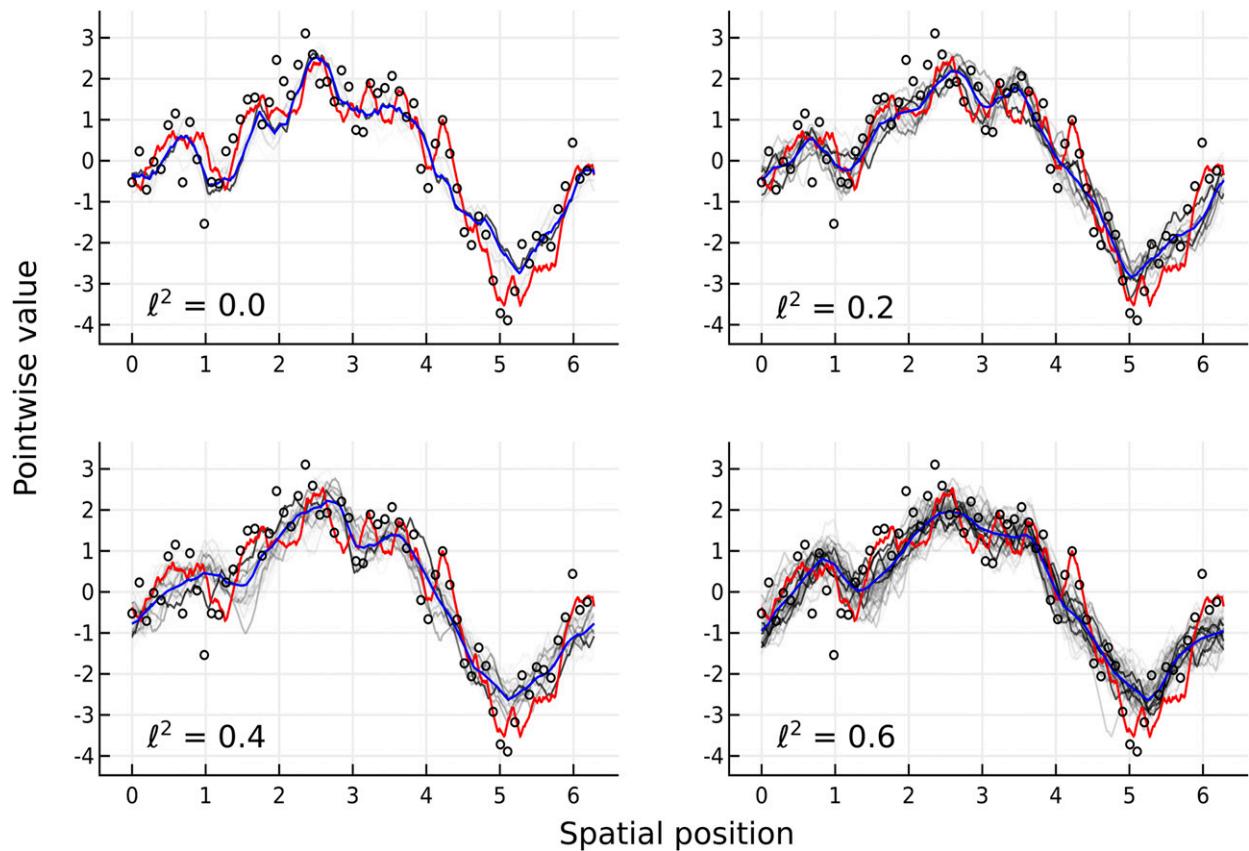## Weighted samples for different $\ell^2$



FIG. 4. (left to right and top to bottom) The true state (red trace), PF mean (blue trace), observations (black circles), and samples from the posterior visually weighted with darkness proportional to sample weight (gray traces) for different values of $\ell^2 \in (0.0, 0.2\ 0.4, 0.6)$. This figure demonstrates again how a small change to the likelihood can substantially improve the problem of underestimating variance, and that this effect comes with diminishing marginal returns as the surrogate model yields progressively smoother estimates of the posterior mean. Observe also that the samples are all realistic instantiations of the physical process, rather than overly smooth estimates. The assimilation time shown here was chosen to exhibit monotonic improvement in $\ell^2$, which is the time-averaged behavior; due to the probabilistic nature of particle filtering, there is an abundance of times when there is not such monotonic improvement.

There is less information to spare about physically important scales when observations are sparse (cf. $N_y = 16$), in which case there is only a narrow window of suitable choices for $\ell^2 \approx 0.12$ before the smoothing effect deteriorates the predictive quality of the particle filter by oversuppressing active scales in the observations.

On the other hand, dense observations provide more abundant small-scale information that makes the particle filtration more susceptible to collapse. This necessitates a larger choice of $\ell^2$ to achieve optimal particle filter performance. In this case, the more abundant information in denser observations can compensate for the injury we do to the surrogate posterior by more aggressively smoothing away small scales. Indeed the most dramatic improvement in the particle filter's uncertainty

quantification occurs for $N_y = 12$. Here the particle filter greatly struggles for small $\ell^2$, where we observe a CRPS over 0.29; however, when $\ell^2 \approx 0.7$ the CRPS dips under 0.22, competitive with that of all other observation models considered here. This suggests that smoothing is particularly helpful in improving the particle filter's overall probabilistic estimate when observations are dense.

The CRPS results show that the particle filter's uncertainty quantification is improved by the GRF likelihood: a 25% decrease (improvement) in CRPS is comparable to the improvement achieved by various statistical postprocessing techniques for ensemble forecasts (Kleiber et al. 2011a,b; Scheuerer and Büermann 2014; Feldmann et al. 2015). Somewhat surprisingly, the CRPS significantly improves moving from $\ell^2 = 0$ to
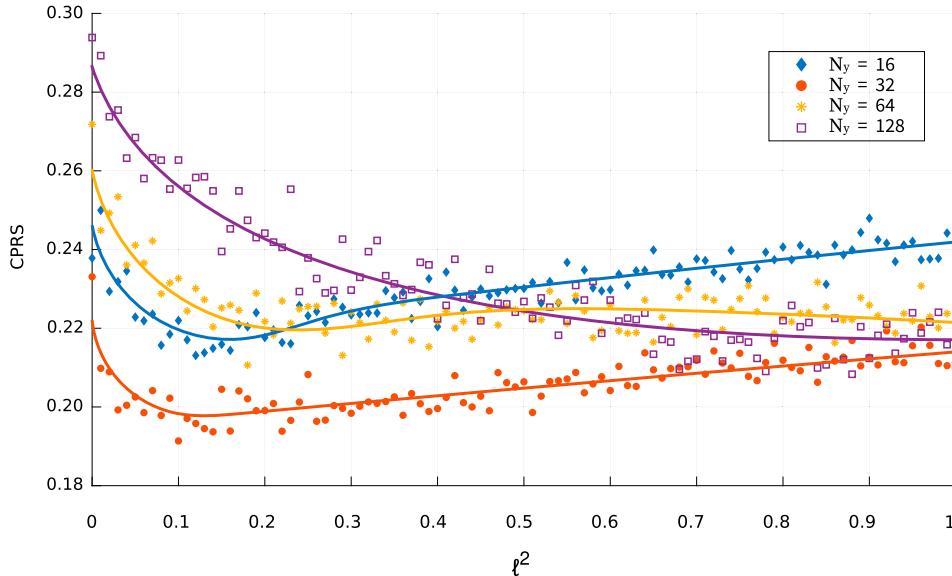
FIG. 5. CRPS median over all time steps and grid locations, shown as a function of $\ell^2$. Each point plotted represents a particle filter assimilation run, with the same true and observed data, for different values of squared GRF length scale $\ell^2$. Each marker style represents different numbers of observations, demonstrating how the particle filter is sensitive to the number of observations: 16 (blue diamonds), 32 (red circles), 64 (yellow asterisks), and 128 (clear squares). The traces are spline approximations of the data that serve to guide the eye. In each $N_y$ case we explored, there is a choice of $\ell^2$ that improves the particle filter CRPS. This plot emphasizes that the optimal choice of $\ell^2$ depends not only on the active scales in the underlying physics, but also on the resolution of the data. There is less information to spare about physically important scales when observations are sparse (cf. $N_y = 16$), in which case there is only a narrow window of suitable choices for $\ell^2 \approx 0.12$ before the smoothing effect deteriorates the predictive quality of the particle filter. On the other hand, dense observations provide more abundant small-scale information that necessitates a larger choice of $\ell^2$ to achieve optimal particle filter performance. Fortunately, the more abundant information in denser observations can compensate for the injury we do to the surrogate posterior by more aggressively smoothing away small scales.

$\ell^2 = 0.1$ despite the fact that the ESS remains quite small. Overall, these CRPS results suggest that even small improvements in ESS can substantially improve the quality of the probabilistic state estimate. They also confirm that improving the ESS as a result of increasing $\ell^2$ must be considered in balance against the consequent departure from the true posterior; the CRPS does not improve at large $\ell^2$, even though the ESS improves, because the surrogate posterior becomes less realistic.

Figure 6 demonstrates how SIR uncertainty quantification depends on ensemble size. The figure shows a kernel density estimate of CRPS over all 2048 grid points and all 100 time steps, for varying number of particles $N_p \in (9100, 200, 400, 800, 1600)$. The CRPS mode remains unchanged, but the mean decreases as the distribution concentrates around the mode primarily at the expense of mass in the tail. The weak dependence of CRPS on ensemble size underscores the appeal of improving uncertainty quantification (UQ) by other means.

## 5. Conclusions

We have demonstrated theoretically [in the framework of Bengtsson et al. (2008) and Snyder et al. (2008)] and in a simple experiment that the number of particles required to avoid collapse in a particle filter can be significantly reduced through a judicious construction of the observation error model. This observation error model has large observation error variance at small scales, which reduces the effective dimensionality and focuses attention on the more dynamically relevant large scales. This observation error model is equivalent to smoothing observations before proceeding as though the observations are uncorrelated. The cost of this approach is that it alters the posterior, leading to a smoother posterior mean. In practice, a balance will need to be found between avoiding collapse and retaining as much observational information as possible.

An observation error model whose variance increases at small scales is associated with a so-called generalized random field (GRF). This connection allows for rapidly
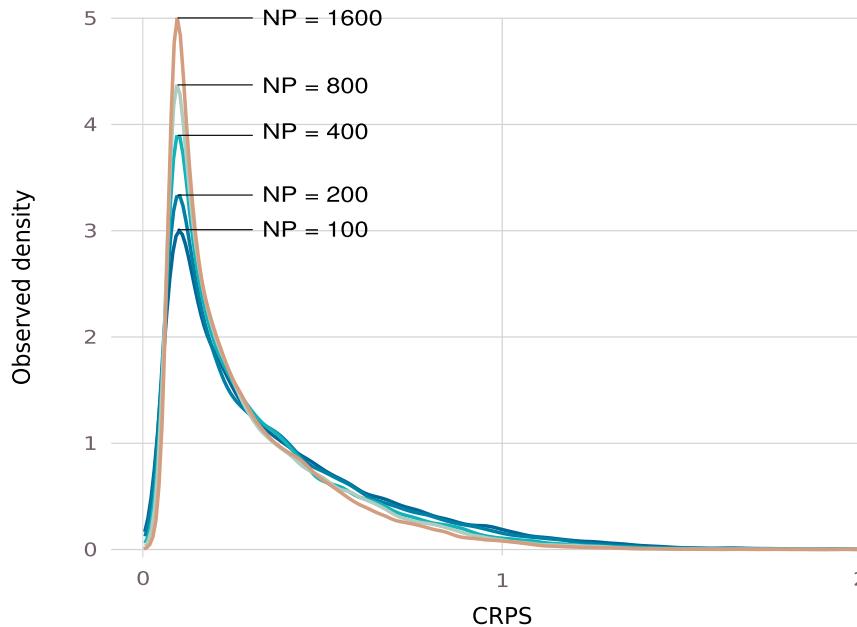
FIG. 6. Kernel density estimates (KDE) of the CRPS observed for different numbers of particles demonstrate the concentration of probability as the number of particles increases while $\ell^2 = 0.30$ and $N_y = 64$ are held fixed, for a fixed simulation and fixed observations thereof. Each KDE is built from the CRPS computed for each of 2048 grid cells and all 100 time steps. The slow convergence in the number of particles is one of the reasons it is attractive to seek other means of making the particle filter more effective in sampling high-dimensional distributions.

applying the covariance matrix's inverse (which is required to compute the particle weights) using fast numerical methods for self-adjoint elliptic partial differential equations. The method can also be implemented by smoothing the observations before assimilating them, and then assimilating the smoothed observations with an assumption of independent errors. Both of these avenues are amenable to serial processing of observations, as required by certain parallel implementations (e.g., Anderson and Collins 2007). All of these approaches are compatible with periodic or aperiodic domains.

The results of the experiment using a one-dimensional stochastic partial differential equation show that this approach improves the effective sample size (ESS), which measures how well the weights are balanced between the particles, by an order of magnitude. The root-mean-square error of the particle filter's posterior mean is not significantly impacted by the approach. One of the main motivations for using particle filters is that they provide meaningful uncertainty estimates even in problems with nonlinear dynamics and observations, and non-Gaussian distributions. Thus, the continuous ranked probability score (CRPS) is used to test the quality of the particle filter's associated probability distribution. The GRF observation error model improves the CRPS by as much as 25%, which is a large improvement, comparable to results obtained by statistical postprocessing of the ensemble (e.g., Kleiber et al. 2011a,b; Scheuerer and Büermann 2014; Feldmann et al. 2015). This improvement in CRPS is obtained even when the ESS is less than 20 out of 400, which shows that good probabilistic state estimation can be achieved even with ESS much less than the ensemble size. The theoretical results suggest that an ensemble size on the order of 8000 is required to avoid collapse in this example problem. Good results are obtained with an ensemble size of 400, even though the ensemble does collapse from time to time.

The theory of Snyder et al. (2008) estimates the ensemble size required to avoid collapse, which is unrealistically large for typical meteorological applications using standard observation error models. Using a GRF observation error model increases the ESS for a fixed ensemble size, making it easier to achieve the goal of avoiding collapse. The approach advocated here may still prove insufficient to enable particle filtering of weather, ocean, and climate problems; the minimum required ensemble size will be reduced, but may still be impractically large. Happily, the method is entirely compatible with approaches based on altered proposals (Chorin and Tu 2009; van Leeuwen 2010; Ades and van Leeuwen 2015) and with localization methods (Penny and Miyoshi 2016; Rebeschini and Van Handel 2015;

Poterjoy 2016). The method is also compatible with ensemble Kalman filters and with variational methods, but it is not clear whether the approach would yield any benefit there.

Indeed, dynamics of extratropical synoptic scales are often assumed to be approximately linear and are easily estimated with an ensemble Kalman filter. But ensemble Kalman filters do not provide robust uncertainty quantification in the face of nonlinear observation operators or nonlinear dynamics (e.g., at synoptic scales in the tropics). In contrast, the method proposed here has the potential to provide robust uncertainty quantification even with nonlinear dynamics and observations. However, it is still unknown in what contexts our peculiar error model damages the posterior more severely than approximating the system as linear and Gaussian for the sake of assimilating data with ensemble Kalman filters. We expect performance comparison to be context dependent, and hope future work will help reveal how to balance advantages and disadvantages that are relevant in practice.

## APPENDIX

### Generalized Random Fields

Generalized random fields (GRFs) are discussed at length in Yaglom (1987), and a few extra details can be found in Gelfand and Vilenkin (1964). A GRF whose Fourier spectrum is not integrable at small scales has infinite variance. The prototypical example is a spatially uncorrelated field, whose spectrum is flat.

A GRF is not defined pointwise. Rather than being defined pointwise, or "indexed by spatial location," it is indexed by rapidly decaying test functions (often taken to be elements of a Schwartz space). This is perhaps best explained by reference to an ordinary random field. If $Z(\mathbf{x})$ is a random field that is defined pointwise and $\phi(\mathbf{x})$

is a test function then we can define a new, "function indexed" random field $Z(\phi)$ using the following expression:

$$Z(\phi) = \int Z(\mathbf{x})\phi(\mathbf{x})\,dx.$$

If the field $Z$ is not defined pointwise, it may still be indexed by test functions.

The concept of a covariance function for an ordinary random field can be generalized to a GRF. The resulting object is a "covariance kernel," which can be a generalized function (i.e., an element of the dual of a Schwartz space). The prototypical covariance kernel is the so-called Dirac delta function which is not, in fact, a function.

The observation error covariance model advocated in this article can be conceptualized in two ways. It can be thought of as an approximation to a GRF where the spectrum has been truncated at the smallest resolvable scale on the grid. Alternatively, one can assume that observations are not taken at infinitesimal points in space, but rather that the observing instrument senses over a small region of space via some test function $\phi$. The value of the GRF for an observation is thus indexed by the allowed test functions $\phi$ rather than the spatial location of the observation.

## REFERENCES

Acevedo, W., J. de Wiljes, and S. Reich, 2017: Second-order accurate ensemble transform particle filters. *SIAM J. Sci. Comput.*, **39**, A1834–A1850, https://doi.org/10.1137/16M1095184.

Ades, M., and P. J. van Leeuwen, 2013: An exploration of the equivalent weights particle filter. *Quart. J. Roy. Meteor. Soc.*, **139**, 820–840, https://doi.org/10.1002/qj.1995.

——, and ——, 2015: The equivalent-weights particle filter in a high-dimensional system. *Quart. J. Roy. Meteor. Soc.*, **141**, 484–503, https://doi.org/10.1002/qj.2370.

Agapiou, S., O. Papaspiliopoulos, D. Sanz-Alonso, and A. Stuart, 2017: Importance sampling: Intrinsic dimension and computational cost. *Stat. Sci.*, **32**, 405–431, https://doi.org/10.1214/17-STS611.

Anderson, J. L., and N. Collins, 2007: Scalable implementations of ensemble filter algorithms for data assimilation. *J. Atmos. Oceanic Technol.*, **24**, 1452–1463, https://doi.org/10.1175/JTECH2049.1.

Bengtsson, T., P. Bickel, and B. Li, 2008: Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems. *Probability and Statistics: Essays in Honor of David A. Freedman*, D. Nolan and T. Speed, Eds., Institute of Mathematical Statistics, 316–334, https://doi.org/10.1214/193940307000000518.

Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129**, 420–436, https://doi.org/10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2.

Chorin, A. J., and X. Tu, 2009: Implicit sampling for particle filters. *Proc. Natl. Acad. Sci. USA*, **106**, 17 249–17 254, https://doi.org/10.1073/pnas.0909196106.

——, and ——, 2012: An iterative implementation of the implicit nonlinear filter. *ESAIM:M2AN*, **46**, 535–543, https://doi.org/10.1051/m2an/2011055.

——, and M. Morzfeld, 2013: Conditions for successful data assimilation. *J. Geophys. Res.*, **118**, 11 522–11 533, https://doi.org/10.1002/2013JD019838.

——, ——, and X. Tu, 2010: Implicit particle filters for data assimilation. *Commun. Appl. Math. Comput. Sci.*, **5**, 221–240, https://doi.org/10.2140/camcos.2010.5.221.

Chui, C., and G. Chen, 2009: *Kalman Filtering: With Real-Time Applications.* 4th ed. Springer, 230 pp., https://doi.org/10.1007/978-3-540-87849-0.

Crisan, D., and A. Doucet, 2002: A survey of convergence results on particle filtering methods for practitioners. *IEEE Trans. Signal Process.*, **50**, 736–746, https://doi.org/10.1109/78.984773.

Doucet, A., and A. M. Johansen, 2011: A tutorial on particle filtering and smoothing: Fifteen years later. *Oxford Handbook of Nonlinear Filtering*, D. Crisan and B. Rozovskii, Eds., Oxford University Press, 656–704.

——, S. Godsill, and C. Andrieu, 2000: On sequential Monte Carlo sampling methods for Bayesian filtering. *Stat. Comput.*, **10**, 197–208, https://doi.org/10.1023/A:1008935410038.

——, N. de Freitas, and N. Gordon, Eds., 2001: An introduction to sequential Monte Carlo methods. *Sequential Monte Carlo Methods in Practice*, Springer, 3–14, https://doi.org/10.1007/978-1-4757-3437-9_1.

Feldmann, K., M. Scheuerer, and T. L. Thorarinsdottir, 2015: Spatial postprocessing of ensemble forecasts for temperature using nonhomogeneous Gaussian regression. *Mon. Wea. Rev.*, **143**, 955–971, https://doi.org/10.1175/MWR-D-14-00210.1.

Fortin, V., M. Abaza, F. Anctil, and R. Turcotte, 2014: Why should ensemble spread match the rmse of the ensemble mean? *J. Hydrometeor.*, **15**, 1708–1713, https://doi.org/10.1175/JHM-D-14-0008.1.

Gelfand, I., and N. Vilenkin, 1964: *Generalized Functions.* Vol. 4, *Applications of Harmonic Analysis*, AMS Chelsea Publishing, 384 pp.

Gneiting, T., and A. E. Raftery, 2007: Strictly proper scoring rules, prediction, and estimation. *J. Amer. Stat. Assoc.*, **102**, 359–378, https://doi.org/10.1198/016214506000001437.

Hersbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Wea. Forecasting*, **15**, 559–570, https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2.

Houtekamer, P. L., and H. L. Mitchell, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123–137, https://doi.org/10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2.

Kleiber, W., A. E. Raftery, J. Baars, T. Gneiting, C. F. Mass, and E. Grimit, 2011a: Locally calibrated probabilistic temperature forecasting using geostatistical model averaging and local Bayesian model averaging. *Mon. Wea. Rev.*, **139**, 2630–2649, https://doi.org/10.1175/2010MWR3511.1.

——, ——, and T. Gneiting, 2011b: Geostatistical model averaging for locally calibrated probabilistic quantitative precipitation forecasting. *J. Amer. Stat. Assoc.*, **106**, 1291–1303, https://doi.org/10.1198/jasa.2011.ap10433.

Lindgren, F., H. Rue, and J. Lindström, 2011: An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *J. Roy. Stat. Soc.*, **73B**, 423–498, https://doi.org/10.1111/j.1467-9868.2011.00777.x.

Majda, A. J., and J. Harlim, 2012: *Filtering Complex Turbulent Systems.* Cambridge University Press, 368 pp.

Morzfeld, M., X. Tu, E. Atkins, and A. J. Chorin, 2012: A random map implementation of implicit filters. *J. Comput. Phys.*, **231**, 2049–2066, https://doi.org/10.1016/j.jcp.2011.11.022.

——, D. Hodyss, and C. Snyder, 2017: What the collapse of the ensemble Kalman filter tells us about particle filters. *Tellus*, **69A**, 1283809, https://doi.org/10.1080/16000870.2017.1283809.

Nystrom, N. A., M. J. Levine, R. Z. Roskies, and J. R. Scott, 2015: Bridges: A uniquely flexible HPC resource for new communities and data analytics. *Proc. 2015 XSEDE Conf.: Scientific Advancements Enabled by Enhanced Cyberinfrastructure*, St. Louis, MO, ACM, https://doi.org/10.1145/2792745.2792775.

Okamoto, K., A. McNally, and W. Bell, 2014: Progress towards the assimilation of all-sky infrared radiances: An evaluation of cloud effects. *Quart. J. Roy. Meteor. Soc.*, **140**, 1603–1614, https://doi.org/10.1002/qj.2242.

Øksendal, B., 2003: *Stochastic Differential Equations: An Introduction with Applications.* 6th ed. Springer, 379 pp., https://doi.org/10.1007/978-3-642-14394-6.

Penny, S. G., and T. Miyoshi, 2016: A local particle filter for high-dimensional geophysical systems. *Nonlinear Processes Geophys.*, **23**, 391–405, https://doi.org/10.5194/npg-23-391-2016.

Poterjoy, J., 2016: A localized particle filter for high-dimensional nonlinear systems. *Mon. Wea. Rev.*, **144**, 59–76, https://doi.org/10.1175/MWR-D-15-0163.1.

Rebeschini, P., and R. Van Handel, 2015: Can local particle filters beat the curse of dimensionality? *Ann. Appl. Probab.*, **25**, 2809–2866, https://doi.org/10.1214/14-AAP1061.

Reich, S., 2013: A nonparametric ensemble transform method for Bayesian inference. *SIAM J. Sci. Comput.*, **35**, A2013–A2024, https://doi.org/10.1137/130907367.

Rue, H., and L. Held, 2005: *Gaussian Markov Random Fields: Theory and Applications.* Chapman & Hall Press, 280 pp.

Scheuerer, M., and L. Büermann, 2014: Spatially adaptive postprocessing of ensemble forecasts for temperature. *J. Roy. Stat. Soc.*, **63C**, 405–422, https://doi.org/10.1111/rssc.12040.

Snyder, C., T. Bengtsson, P. Bickel, and J. Anderson, 2008: Obstacles to high-dimensional particle filtering. *Mon. Wea. Rev.*, **136**, 4629–4640, https://doi.org/10.1175/2008MWR2529.1.

——, ——, and M. Morzfeld, 2015: Performance bounds for particle filters using the optimal proposal. *Mon. Wea. Rev.*, **143**, 4750–4761, https://doi.org/10.1175/MWR-D-15-0144.1.

Towns, J., and Coauthors, 2014: XSEDE: Accelerating scientific discovery. *Comput. Sci. Eng.*, **16**, 62–74, https://doi.org/10.1109/MCSE.2014.80.

van Leeuwen, P. J., 2010: Nonlinear data assimilation in geosciences: An extremely efficient particle filter. *Quart. J. Roy. Meteor. Soc.*, **136**, 1991–1999, https://doi.org/10.1002/qj.699.

Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924, https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2.

Yaglom, A. M., 1987: *Correlation Theory of Stationary and Related Random Functions.* Springer, 526 pp.

Zhu, Y., and Coauthors, 2016: All-sky microwave radiance assimilation in NCEP's GSI analysis system. *Mon. Wea. Rev.*, **144**, 4709–4735, https://doi.org/10.1175/MWR-D-15-0445.1.