# A Particle Filter for Stochastic Advection by Lie Transport (SALT): A case study for the damped and forced incompressible 2D Euler equation[*]

Colin Cotter, Dan Crisan, Darryl D. Holm, Wei Pan[†], and Igor Shevchenko

Department of Mathematics, Imperial College London, London, SW7 2AZ, UK

July 30, 2019

## Abstract

In this work, we apply a particle filter with three additional procedures (model reduction, tempering and jittering) to a damped and forced incompressible 2D Euler dynamics defined on a simply connected bounded domain. We show that using the combined algorithm, we are able to successfully assimilate data from a reference system state (the "truth") modelled by a highly resolved numerical solution of the flow that has roughly $3.1 \times 10^6$ degrees of freedom for 10 eddy turnover times, using modest computational hardware.

The model reduction is performed through the introduction of a stochastic advection by Lie transport (SALT) model as the signal on a coarser resolution. The SALT approach was introduced as a general theory using a geometric mechanics framework from Holm, Proc. Roy. Soc. A (2015). This work follows on the numerical implementation for SALT presented by Cotter et al, SIAM Multiscale Model. Sim. (2019) for the flow in consideration. The model reduction is substantial: The reduced SALT model has $4.9 \times 10^4$ degrees of freedom.

Forecast reliability and estimated asymptotic behaviour of the particle filter are also presented.

# Contents

# 1   Introduction

Data assimilation is the process by which observations (data) are integrated with mathematical models so that inference or prediction of the evolving state of the system can be made. For geoscience applications such as numerical weather prediction, it is an active area of research. There the typical global-scale state space dimension is of order $O(10^9)$, and observation data of dimension $O(10^7)$ are assimilated every 6 – 12 hours. Current established methods used in operation centres include 4DVar, (various extended versions of) ensemble Kalman filter (EnKF) and variational assimilation methods. However, for fully nonlinear systems and complex observation operators these approaches are unsatisfactory. Our work presented in this paper is part of the wider effort to tackle high dimensional nonlinear geoscience problems using particle filters, as can be seen from the survey paper [van Leeuwen et al., 2019] and the references therein.

The idea of modelling uncertainty using stochasticity in geophysical fluid applications is well established, see Buizza et al. [1999], Majda et al. [1999, 2001]. In this paper we work with the stochastic advection by Lie transport (SALT) approach, first formulated in Holm [2015]. It can be thought of as a framework for deriving stochastic partial differential equation (SPDE) models for geophysical fluid dynamics (GFD). The stochasticity is introduced into the advection part of the dynamics via a constrained variational principle called the Hamilton-Pontryagin principle. What results is a stochastic Euler-Poincaré equation, in which the local acceleration part of the transport operator is in the geometric form represented by the Lie derivative of velocity one-form in the direction of a stochastic vector field (in the form of a Stratonovich semimartingale). This approach for adding stochasticity into GFD models is different from the current state-of-the-art in numerical weather prediction (NWP), where stochastic models of uncertainty is introduced into the forcing; for example *stochastic perturbation by physical tendencies* (SPPT) methodology, see Palmer [2018].

By adding stochasticity into the advection operator, one can model uncertain transport behaviour. In particular, the SALT stochastic term can be thought of as a model on the

resolvable scale for the subgrid unresolvable fluid scales for transport. The main advantage of the SALT stochastic term is that it preserves the Kelvin's Circulation Theorem (KCT). However energy is not conserved by SALT because as one can show, an extra term called line stretching results from the application of the Reynolds transport theorem to the time differential of the energy; and the extra term contributes positively to the rate of change of the energy. An alternative, energy conserving stochastic approach called Location Uncertainty (LU) has also been developed Mémin [2014], but LU models do not preserve circulation.

A fundamental ingredient in SPDEs with SALT noise is the spectrum of the *velocity-velocity correlation tensor*, with eigenvectors denoted by $\boldsymbol{\xi_i}$, $i \in \mathbb{N}$, which appears in the Eulerian velocity field

$$\mathsf{d}\boldsymbol{x} = \mathbf{u}(\boldsymbol{x}, t)dt + \sum_i \boldsymbol{\xi}_i(\boldsymbol{x}) \circ dW_t^i. \tag{1}$$

Cotter et al. [2017] showed that taking the diffusive limit of a flow with two timescales leads to a stochastic differential equation in the form of (1), where $\boldsymbol{\xi}_i$ should be rigorously understood as empirical orthogonal functions corresponding to the different modes of the fast flow.

For applications, the vector fields $\boldsymbol{\xi}_i$ need to be supplied *a priori*. A *data driven* calibration methodology for obtaining $\boldsymbol{\xi}_i$ is described in Cotter et al. [2018, 2019], in which the authors numerically investigated two example fluid systems: a damped and forced 2D Euler model with no-penetration boundary condition, and a two-layer 2D quasi-geostrophic model prescribed on a channel. In those works, the SPDE model is interpreted as a parameterisation for the antecedent partial differential equation (PDE) model. Using statistical uncertainty quantification tests, it is shown that by conditioning on a suitable initial prior, an ensemble of SPDE solutions is able to effectively capture the large scale behaviour of the deterministic system on a coarser resolution. It is important to stress the fact that the deterministic system has $O(10^6)$ degrees of freedom, whilst the coarse stochastic system has $O(10^4)$ degrees of freedom. Capturing large scale dynamics on coarse scales enables the reduction of the high resolution PDE system to the coarse scale as an SPDE. This motivates further investigation of the performance of SALT SPDEs using ensemble data assimilation algorithms, where the forecast model is the SPDE prescribed on coarse scales. This is the theme of the present work, where we utilise the calibrated $\boldsymbol{\xi}_i$ described in Cotter et al. [2019] in a data assimilation set-up for the damped and forced 2D Euler dynamics.

For us, sequential data assimilation is mathematically formulated as a nonlinear filtering problem, which can be tackled using a particle filter, see Bain and Crisan [2009], Reich and Cotter [2015]. A particle filter proceeds by alternating between forecast and analysis cycles. In each analysis cycle, observations of the current (and possibly past) state of a system are combined with the results from a prediction model (the forecast) to produce an analysis. The analysis step is typically performed either in the form of a "best estimate" or in terms of approximating conditional distributions. The model is then advanced in time and its result becomes the forecast in the next analysis cycle. However when applied to problems in high dimensions, without additional techniques a basic particle filter algorithm would almost certainly fail. This is due to the fact that in high dimensions the data are too informative.

In this paper, we describe three additional techniques: model reduction, tempering and jittering, which are incorporated into the basic bootstrap particle filter. The combined algorithm is applied to the damped and forced 2D Euler dynamics. These techniques are all necessary for the successful assimilation of data obtained from the true state of the system, which is modelled using a highly resolved numerical solution of $3.1 \times 10^6$ degrees of freedom.

The rest of the paper is structured as follows. In section 2 we describe the damped and forced deterministic system and its SALT version. The deterministic system resolved on a fine resolution spatial grid is viewed as the simulated truth. The reduction to the SALT version is done via the variational approach formulated in Holm [2015]. The numerical calibration of the subgrid parameters $\boldsymbol{\xi}_i$ and the numerical methodologies for solving the two systems are described in Cotter et al. [2019].

In section 3 we formulate sequential data assimilation as a nonlinear filtering problem, in which the SALT equations are used as the signal. We describe in detail each algorithm: bootstrap particle filter, tempering and jittering, which are all required to tackle the high dimensional nonlinear filtering problem.

In section 4 we present and discuss the numerical experiments and results. Two main sets of experiments are considered. In the first set, which we call the *perfect model scenario*, the true underlying state is a realisation of the signal. In the second set, which we call the *imperfect model scenario*, data from the fine resolution true state is assimilated. All experiments were run on a modest workstation which has two Intel Xeon processors totalling 32 logical cores and 64Gb memory. Additionally, an effective method for generating initial ensembles for SALT models is discussed.

Finally, section 5 concludes the present work and discusses the outlook for future research.

The following is a summary of the main numerical experiments contained in this paper:

- Using 100 particles, we ran the particle filter over a period of 10 eddy turnover times (ett, see (46) for definition) separately for observation dimensions $d_y = 81$ and $d_y = 289$, and assimilation intervals $\Delta = 1/25$ ett and $\Delta = 1/5$ ett. Each experiment was repeated 20 times. The mean square error (mse), ensemble spread ($\sigma_{N_p}$), effective sample size (ess, see (31) for definition) and number of tempering steps for the average is shown, in figure 5 for the perfect model scenario and in figure 11 for the imperfect model scenario.

- Forecast reliability rank histograms for the particle filter using 100 particles, observation dimension $d_y = 289$ and assimilation interval $1/25$ ett are shown in figures 6a and 6b for the perfect model scenario, and in figures 12a and 12b for the imperfect model scenario. The reliability rank histograms are also computed for the results from using observation dimension $d_y = 81$; this is shown in figure 7 for perfect model scenario and figure 13 for the imperfect model scenario.

- Asymptotic particle filter behaviour was estimated using an experiment which ran over a period of 50 ett, using 100 particles and assimilation period $\Delta = 1/25$ ett. This amounts to a total of 1250 data assimilation steps. The results are shown in figures 8 – 10 for the perfect model scenario, and figures 14 – 16 for the imperfect model scenario.

# 2  Deterministic and stochastic advection by Lie transport GFD models

In this section, we describe the PDE and the SPDE models with Lie transport type stochastic terms. For the theory on SALT SPDEs see for example Holm [2015], Crisan et al. [2018]. We follow Cotter et al. [2019] (also Cotter et al. [2018]) and use a data-driven approach to numerically model the $\boldsymbol{\xi}_i$'s. Thus information regarding the stochastic dynamics is complete

except for initial and boundary conditions. Viewed as a parameterisation of the subgrid scales, numerically the SPDE shall be prescribed on a coarse resolution grid and the PDE prescribed on a fine resolution grid.

The spread of the SPDE dynamics from using $\boldsymbol{\xi}_i$ parameters calibrated with the data-driven approach described in Cotter et al. [2019] adequately captures the large scale features of the PDE dynamics. Those results indicate the feasibility of the calibrated SPDE as model reduction, thus providing the foundation for the present work where we utilise the SPDE as the *signal* process in a nonlinear filtering formulation. Nonlinear filtering will be the topic of discussion in section 3.

In the following, the domain $D = [0,1]^2$ is assumed for both deterministic and stochastic models.

## 2.1  Deterministic model

We consider the vorticity version of an incompressible Euler flow with forcing and damping. Let $\mathbf{u} : D \times [0,\infty) \to \mathbb{R}^2$, $\mathbf{u}(x,y,t) = (u_1(x,y,t), u_2(x,y,t))$ denote the velocity field. Let $\omega = \hat{z} \cdot \operatorname{curl} \mathbf{u}$ denote the vorticity of $\mathbf{u}$, $\hat{z}$ denotes the $z$-axis. Note that for incompressible flows in two dimensions, $\omega$ is a scalar field. For a scalar field $g : D \to \mathbb{R}$, we write $\nabla^\perp g = (-\partial_y g, \partial_x g) = \hat{z} \times \nabla g$. Let $\psi : D \times [0,\infty) \to \mathbb{R}$ denote the *stream function*. The stream function is related to the fluid velocity and vorticity by $\mathbf{u} = \nabla^\perp \psi$ and $\omega = \Delta \psi$ respectively, where $\Delta = \partial_x^2 + \partial_y^2$ is the Laplacian operator in $\mathbb{R}^2$. The existence of the stream function is guaranteed by the incompressibility assumption.

We now write down the deterministic model,

$$\partial_t \omega + \mathcal{L}_{\mathbf{u}} \omega \ = \ Q - r\omega \tag{2}$$
$$\mathbf{u} \ = \ \nabla^\perp \psi \tag{3}$$
$$\Delta \psi \ = \ \omega. \tag{4}$$

We choose the forcing $Q$ to be given by

$$Q(x,y) = \alpha \sin(\beta \pi x), \qquad (x,y) \in D \tag{5}$$

where $\alpha$, $\beta$ and $r$ are constants having the following roles: $\alpha \geq 0$ controls the strength of the forcing; $\beta$ is an integer interpreted as the number of gyres in the external forcing; and $r > 0$ can be seen as the damping rate. $\mathcal{L}_{\mathbf{u}} \omega$ denotes the Lie derivative of $\omega$ with respect to the vector field $\mathbf{u}$. When applied to scalar fields, $\mathcal{L}_{\mathbf{u}}$ is simply the directional derivative with respect to $\mathbf{u}$, see Chern et al. [1999]

$$\mathcal{L}_{\mathbf{u}} = \mathbf{u} \cdot \nabla.$$

We consider a no-penetration spatial boundary condition

$$\psi|_{\partial \mathcal{D}} = 0 \tag{6}$$

to close the system. This system is a special case of a nonlinear, one-layer quasi-geostrophic (QG) model that is driven by winds above.

## 2.2  Stochastic model

Consider the space $\Omega = C_0([0,\infty), \mathbb{R}^m)$ of continuous function whose value at 0 is zero. It is equipped with the Wiener measure $\mathbb{P}$ and its natural filtration $\{\mathcal{F}_*\}$. Let $\{W_t : t \in [0,\infty)\}$

be the canonical Brownian motion on $\mathbb{R}^m$, that is for $\gamma \in \Omega$, $W_t(\gamma) = \gamma(t)$ is the evaluation map. We write $W_t^i$ to denote the $i$'th component of $W_t$. The SALT version of the Euler fluid equation (2) as derived in Holm [2015]. Cotter et al. [2019] introduced damping and forcing to facilitate statistical equilibrium in the underlying resolved system, leading to the following stochastic partial differential equation (SPDE),

$$\mathsf{d}q + \mathcal{L}_{\mathbf{v}}q \, dt + \sum_{i=1}^{m} \mathcal{L}_{\boldsymbol{\xi}_i}q \circ dW_t^i = (Q - rq) \, dt \tag{7}$$

where the vector fields $\boldsymbol{\xi}_i$ represent scaled eigenvectors of the velocity-velocity correlation tensor $C_{ij} = \boldsymbol{\xi}_i \boldsymbol{\xi}_j^T$.

Equation (7) arises from a time-scale separation assumption for the deterministic Eulerian transport velocity $\mathbf{u}$, leading to the following Stratonovich stochastic differential equation

$$\mathsf{d}\tilde{\boldsymbol{x}}_t(\boldsymbol{x}) = \mathbf{v}(\boldsymbol{x}, t)dt + \sum_{i=1}^{m} \boldsymbol{\xi}_i(\boldsymbol{x}) \circ dW_t^i \tag{8}$$

where $\mathbf{v}$ and $\boldsymbol{\xi}_i$ are divergence free vector fields, from which (7) may be derived. Here one can intuitively think of $\mathbf{v}$ as the "large" scale mean part of $\mathbf{u}$. In this present work since we are interested in the practicality of (7) for data assimilation, we follow Cotter et al. [2019] and make the approximation that the sum in (8) is finite. Hence the stochastic term in (7) also consists of $m$ terms.

Let $\tilde{\psi}$ denote the stream function of $\mathbf{v}$, and let $\zeta_i$ denote the stream function of $\boldsymbol{\xi}_i$, i.e.

$$\boldsymbol{\xi}_i = \nabla^\perp \zeta_i.$$

Note that $\zeta_i$ is constant in time. The $\zeta_i$ can be solved for and stored on the computer after the $\boldsymbol{\xi}_i$ are obtained. For this the boundary condition

$$\zeta_i|_{\partial \mathcal{D}} = 0 \tag{9}$$

is enforced for each $i = 1, \ldots, m$. Then (8) can be expressed in terms of $\tilde{\psi}$ and $\zeta_i$,

$$\mathsf{d}\tilde{\boldsymbol{x}} = \nabla^\perp \left( \tilde{\psi} dt + \sum_{i=1}^{m} \zeta_i \circ dW_t^i \right). \tag{10}$$

Expressing the transport velocity in this form is useful because it allows us to introduce stochastic perturbation (i.e. terms with $\circ \, dW_t^i$) via the stream function when solving the SPDE system numerically, thereby keeping the discretisation of (7) the same as the deterministic equation (2), see Cotter et al. [2019].

The full set of stochastic equations is

$$\mathsf{d}q + \mathcal{L}_{\mathbf{v}}qdt + \sum_{i=1}^{m} \mathcal{L}_{\boldsymbol{\xi}_i}q \circ dW_t^i = (Q - rq)\, dt \tag{11}$$

$$\mathbf{v} = \nabla^\perp \tilde{\psi} \tag{12}$$

$$\Delta \tilde{\psi} = q \tag{13}$$

with boundary condition

$$\tilde{\psi}|_{\partial \mathcal{D}} = 0. \tag{14}$$

The forcing term is the same as the deterministic case.

**Remark 1.** *The Itô form of (7) is obtained from an application of the identity*

$$\int_0^t \mathcal{L}_{\boldsymbol{\xi}_i} q\,(s) \circ dW_s^i = \int_0^t \mathcal{L}_{\boldsymbol{\xi}_i} q\,(s)\, dW_s^i + \frac{1}{2} \left\langle \mathcal{L}_{\boldsymbol{\xi}_i} q, W^i \right\rangle_t \tag{15}$$

*where $\langle .,. \rangle_t$ is the cross-variation bracket and*

$$
\begin{aligned}
\left\langle \mathcal{L}_{\boldsymbol{\xi}_i} q, W^i \right\rangle_t &= \mathcal{L}_{\boldsymbol{\xi}_i} \left\langle q, W^i \right\rangle_t \\
&= \mathcal{L}_{\boldsymbol{\xi}_i} \left\langle \int \{ (Q - rq)dt - \mathcal{L}_{\mathbf{v}} q dt - \sum_{j=1}^{\infty} \mathcal{L}_{\boldsymbol{\xi}_j} q \circ dW_t^j \}, W^i \right\rangle_t \\
&= \mathcal{L}_{\boldsymbol{\xi}_i} \left\langle - \int_0^{\cdot} \mathcal{L}_{\boldsymbol{\xi}_i} q \circ dW_s^i, W^i \right\rangle_t \\
&= \mathcal{L}_{\boldsymbol{\xi}_i} \left( - \int_0^t \mathcal{L}_{\boldsymbol{\xi}_i} q\,(s)\, ds \right) = - \int_0^t \mathcal{L}_{\boldsymbol{\xi}_i}^2 q(s) ds
\end{aligned}
$$

*Hence*

$$\int_0^t \mathcal{L}_{\boldsymbol{\xi}_i} q\,(s) \circ dW_s^i = \int_0^t \mathcal{L}_{\boldsymbol{\xi}_i} q\,(s)\, dW_s^i - \frac{1}{2} \int_0^t \mathcal{L}_{\boldsymbol{\xi}_i}^2 q(s) ds$$

*and (15) is thus*

$$\mathsf{d}q + \mathcal{L}_{\mathbf{v}} q dt + \sum_{i=1}^{m} \mathcal{L}_{\boldsymbol{\xi}_i} q\; dW_t^i = \frac{1}{2} \sum_{i=1}^{m} \mathcal{L}_{\boldsymbol{\xi}_i}^2 q\; dt + (Q - rq)dt \tag{16}$$

*where $\mathcal{L}_{\boldsymbol{\xi}_i}^2 q = \mathcal{L}_{\boldsymbol{\xi}_i} \left( \mathcal{L}_{\boldsymbol{\xi}_i} q \right) = [\boldsymbol{\xi}_i, [\boldsymbol{\xi}_i, q]]$ is the double Lie derivative of $q$ with respect to the divergence free vector field $\boldsymbol{\xi}_i$.*

For the damped and forced stochastic system considered in this section, on the torus $\mathbb{T}^2$ a global wellposedness theorem with the solution space $W^{2,2}(\mathbb{T}^2)$ is proved in Crisan and Lang [2019]. In a forthcoming sequel to this work we also show the the wellposedness of the solution on the bounded domain $D$ with no-penetration boundary conditions. We make the following important assumption.

**A 1.** *The stochastic system (11) – (14) is wellposed in some solution space denoted by $\mathbb{S}_{SPDE}$ in the sense that a unique global in time, pathwise distributional solution[1] exists.*

Under assumption 1, it is useful to introduce the following. Let $G : \mathbb{S}_{\mathrm{SPDE}} \times \Omega \to \mathbb{S}_{\mathrm{SPDE}}$ denote the Itô solution map of the stochastic system (11) – (14) so that

$$G(q_0, \gamma)_t = q_t, \quad q_0 \in \mathbb{S}_{\mathrm{SPDE}}. \tag{17}$$

The solution map $G$ is used in the next section where probability measures on the solution space are defined as the push-forward of $\mathbb{P}$ using $G$.

# 3   Nonlinear Filtering

In this section, we formulate data assimilation as a nonlinear filtering problem in which the aim is to utilise observed data to correct the distribution of predictive dynamics. We describe

---

[1]We refer to Crisan and Lang [2019] for the precise definition of these terms.

a particle filter methodology which incorporates three additional techniques that are required to effectively tackle this high dimensional data assimilation problem.

In nonlinear filtering terminology the predictive dynamics is often called the *signal* [2]. The signal in our setting corresponds to the SALT SPDE. Data is obtained via an *observation process* which represents noisy partial measurements of the underlying true system state. The goal is to determine the posterior distribution $\pi_t$ of the signal at time $t$ given the information accumulated from observations. This is known as the *filtering problem*. This is different to *inversion* problems (also called smoothing problems), where one is interested in obtaining the posterior distribution of the system's initial condition, see for example Stuart, A M [2010].

The stochastic filtering framework enables us not just to provide a solution to the data assimilation problem, but also offer a clear language in which to explain the details and the intricacies of the problem. We detail below an elementary introduction to the filtering problem.

Let $\mathbb{S}$ denote a given state space, and let $\mathcal{P}(\mathbb{S})$ denote the set of probability measures on the state space. In what follows the state space will be $\mathbb{S} = \mathbb{R}^{d_x}$, where $d_x$ is the dimension of the space. To avoid technical complications we will assume in the following that time runs discretely $t = 0, 1, \ldots$ . We shall work in a Bayesian setting, in other words we will assume that we know the distribution of the signal $X_t$ for $t = 0, 1, \ldots$, which will be denoted by $p_t$ for $t = 0, 1, \ldots$. We also assume that partial observations, denoted by $Y_t$, of dimension $\mathbb{S}_{\text{obs}} = \mathbb{R}^{d_y}$ with $d_y \leq d_x$ are available to us at times $t = 0, 1, \ldots$ and we wish to approximate the signal $X_t$ given the accumulated observations $Y_1, \ldots, Y_t$. Of course we could aim to approximate $X_t$ using an arbitrary $\mathcal{Y}_t$-adapted estimator $\bar{X}_t$, where $\mathcal{Y}_t$ is the $\sigma$-algebra $\mathcal{Y}_t = \sigma(Y_1, \ldots, Y_t)$. However, the *best* estimator $\hat{X}_t$ is the conditional expectation of $X_t$ given $\mathcal{Y}_t$, $\hat{X}_t = \mathbb{E}[X_t|\mathcal{Y}_t]$. In this context, by the best estimator, we mean the minimiser of the mean square error $MSE(\bar{X}_t) = \mathbb{E}[\|X_t - \bar{X}_t\|^2|\mathcal{Y}_t]$, where $\|\cdot\|$ is the standard Euclidian norm on $\mathbb{R}^{d_x}$. Of course we would not just want to compute/estimate $\hat{X}_t = \mathbb{E}[X_t|\mathcal{Y}_t]$, but also the error that we would make if we approximate $X_t$ with $\hat{X}_t$, i.e., for $t = 0, 1, \ldots$.

$$\mathbb{E}[\|X_t - \hat{X}_t\|^2|\mathcal{Y}_t] = \mathbb{E}[\|X_t\|^2|\mathcal{Y}_t] - \mathbb{E}[\|X_t\||\mathcal{Y}_t]^2.$$

The quantiles of the approximation error will also be of interest. Therefore, in general, the filtering problem consists in determining the condition distribution of the signal given given $\mathcal{Y}_t$ denoted by $\pi_t$. Once $\pi_t$ is determined, then its first moment (the mean vector) will give us $\hat{X}_t$, its covariance matrix can be used to compute the mean square error $MSE(\hat{X}_t)$, etc. So one can adopt one of two different approaches of estimation the signal given partial observations.

- Develop a data assimilation algorithm that results in a a point approximation $\bar{X}_t$ of the signal using the data $Y_1, \ldots, Y_t$. The approximation may or may not be optimal and only, on rare occasions, an estimate of the error $X_t - \bar{X}_t$ will be available.

- Develop a data assimilation algorithm that results in an approximation of $\pi_t$ the conditional distribution of the signal $X_t$ using the data $Y_1, \ldots, Y_t$. This in turn will offer an approximation of the optimal estimator $\hat{X}_t$ as well as the approximation of the error, quantiles, occupation mesures, etc.

Of course, algorithmically we expect the first problem to be a lot easier than the second. The computation, of an estimator $\bar{X}_t$ that is an element of $\mathbb{R}^{d_x}$ would be expected to be a lot

---

[2] Also known as the forecast model in statistics and meteorology literature, [see Reich and Cotter, 2015].

easier that that of a probability measure over $\mathbb{R}^{d_x}$. The first one is a finite dimensional object the latter is an infinite dimensional one. However, in the exceptional case when the signal is a linear time-series and the observation has linear dependence on the signal and they are driven by Gaussian noise the two approaches more or less coincide. The reason is that, in this case $\pi_t$ is Guassian and one can explicitly write the recurrence formula for the pair $(\hat{X}_t, P_t)$, where $P_t$ is the covariance matrix of $\pi_t$. So on one hand one can compute directly the optimal estimator $\hat{X}_t$ and on the other hand the Gaussianity ensure that $\pi_t$ is fully described by $(\hat{X}_t, P_t)$. This is the so-called Kalman-Filter. There are numerous extensions of this method to non-linear filter that attempt a similar methodology for the non-Gaussian conditional distribution. Such approaches are not optimal in the sense that they don't offer a point estimator that is the optimal one and the corresponding "covariance" matrix that is produced is not the covariance matrix of $\pi_t$. The existing literature in this direction is vast, we cite here [Reich and Cotter, 2015, Evensen, 2009, Ljung, 1979].

Particle filters are a class of numerical methods that can be used to implement the second approach. They have been highly successful for problems in which the dimension of the state space $d_x$ has been low to medium. However, in recent works [Kantas et al., 2014, Beskos et al., 2017, 2014] they have been shown to also work in high dimensions $d_x$. In this paper, we tackle a state space with dimension of order $O(10^6)$. For a filtering perspective, we overcome here one other hurdle as we explain below.

Let us denote by $p_t \in \mathcal{P}(\underbrace{\mathbb{R}^{d_x} \times \ldots \times \mathbb{R}^{d_x}}_{(t+1)-\text{times}})$, $t = 0, 1, \ldots$ the (prior) distribution of the signal on the path space $(X_0, X_1, \ldots, X_t)$. The prior distribution of the signal $p_t$ and the observations $Y_s$, $s = 0, 1, \ldots, t$ are the building blocks of $\pi_t$, $t = 0, 1, \ldots$. To be more precise, one can show that there exists a mapping

$$(p, y_0, \ldots y_t) \mapsto \Xi(p, y_0, \ldots y_t) : \mathcal{P}(\underbrace{\mathbb{R}^{d_x} \times \ldots \times \mathbb{R}^{d_x}}_{(t+1)-\text{times}}) \times \underbrace{\mathbb{R}^{d_y} \times \ldots \times \mathbb{R}^{d_y}}_{(t+1)-\text{times}} \mapsto \mathcal{P}(\mathbb{R}^{d_x}), \qquad (18)$$

such that $\pi_t = \Xi(p_t, Y_0, \ldots Y_t)$. Under very general condition on the signal and the observation, this mapping is jointly continuous on the product space $\mathcal{P}(\underbrace{\mathbb{R}^{d_x} \times \ldots \times \mathbb{R}^{d_x}}_{(t+1)-\text{times}}) \times \underbrace{\mathbb{R}^{d_y} \times \ldots \times \mathbb{R}^{d_y}}_{(t+1)-\text{times}}$. This would mean that $\pi_t$ will give a reasonable approximation of the conditional distribution of the signal as long as the distribution $(X_0, X_1, \ldots, X_t)$ does not differ significantly from the one used to construct $\pi_t$. The same will happen when the true law of the observation does not differ significantly from the chosen model. This property of the posterior distribution is crucial, see section 3.1.1 for details.

In the rest of this section we consider only the space-time discretised SPDE signal, of spatial dimension $d_x$. The observation process is given by noisy spatial evaluations of an underlying true system state at discrete time steps. We consider two scenarios for the underlying true system state, henceforth called the *truth*.

In the first scenario, we aim to compute the conditional distribution of the signal given partial observations of a single realised trajectory of the SPDE system. In this case the predictive dynamics and the truth are from the same dynamical system. We call this the *perfect model scenario* (or twin experiment, see Reich and Cotter [2015]).

In the second scenario, we use instead noisy spatial evaluations of a space-time discretised solution corresponding to the PDE system (2) – (6). We call this the *imperfect model scenario*. The truth in this case is computed on a more refined grid than solutions of the SPDE. Nevertheless the solution of the SPDE converges to that of the PDE as the coarser grid converge,

see Cotter et al. [2019]. Similarly the corresponding observations will converge (provided the observation noise does not change). This ensures the successful assimilation of PDE data into the SPDE model, assurance from the uncertainty quantification tests shown in Cotter et al. [2019] is necessary to numerically guarantee that the mis-match between state spaces remains small.

To our knowledge, this is the first application of particle filters to the case where the signal is described by a SALT SPDE system. As we explain below a straight application of the classical bootstrap particle filter algorithm fails. To succeed we implement and incorporate the following procedures.

- Model reduction – approximate a high dimensional system using a low dimensional system via stochastic modelling, the result of which can be further reduced by choosing a projection of the noise process onto a submanifold. This was accomplished in Cotter et al. [2019].

- Tempering – compute a sequence of intermediate measures $\pi_t^k$ parameterised by a finite number of temperatures that control the smoothness of the density of $\pi_t^k$. This procedure eases the problem of highly singular posteriors in high dimensions, which come from the fact that high dimensional observations are too informative.

- Jittering – a Markov chain Monte Carlo (MCMC) based technique for recovering lost population diversity in particle filter algorithms.

These techniques are added to the basic bootstrap particle filter, and are demonstratively necessary, theoretically consistent and rigorously justified. In addition, we shall pay particular attention to the initialisation of the particle filter, though this is discussed in section 4.1.

Before proceeding to the problem formulation, we insert an important remark.

**Remark 2.** *Our spatial discretisations for the PDE and SPDE fields are defined on appropriate finite element spaces, see Cotter et al. [2019] for details of the numerical methods we use for the models under consideration. Under assumption 1, it is important to understand that instead of the finite state space $\mathbb{S} = \mathbb{R}^{d_x}$, the actual problem involves measures defined on infinite dimensional function spaces, in particular it is highly plausible that in theory the state space for the SPDE is Sobolev $W^{k,2}(D)$ for $k \geq 2$. Discussions of these technical complications are not the focus of this work. And since in practise we work with numerical solutions anyhow, we setup our filtering problem in a finite dimensional setting. However, the methods we use are all theoretically consistent in the limit, see Stuart, A M [2010], Dashti and Stuart [2017].*

In light of remark 2, henceforth we drop the word "discretised" when describing the state space, signal and observation processes.

## 3.1 Filtering problem formulation

Consider discrete times $\Lambda = \{t_0, t_1, \ldots, t_n, \ldots\}$. Let $X : \Lambda \times \Omega \to \mathbb{S}$ be a discrete time Markov process called the *signal*. Let $Y : \Lambda \to \mathbb{S}_{\mathrm{obs}}$ be a discrete time process called the *observation* process. We assume $Y(t_0) = 0$ almost surely (a.s.). We consider *Eulerian data assimilation* where the observations correspond to fixed spatial points $Y_t = (Y_t^1, \ldots, Y_t^{d_y}) \in D$, for all $t \in \Lambda$. As already mentioned in the section introduction, we denote the dimensions of $\mathbb{S}$ and $\mathbb{S}_{\mathrm{obs}}$ by $d_x$ and $d_y$ respectively.

We take $X$ and $Y$ to correspond to the *velocity* vector field. Mathematically we could also consider the vorticity field or the stream function, but in real world scenarios those fields may be difficult to observe directly. We denote by $X_{i:j}$ and $Y_{i:j}$ the path of the signal and of the observation process from time $t_i$ to time $t_j$,

$$X_{i:j} = (X_{t_i}, X_{t_{i+1}}, \ldots, X_{t_j}), \quad Y_{i:j} = (Y_{t_i}, Y_{t_{i+1}}, \ldots, Y_{t_j}).$$

Let $x_{i:j}$ and $y_{i:j}$ denote particular trajectories of $X_{i:j}$ and $Y_{i:j}$. For notational convenience, we may write in the subscripts $i$ to mean $t_i$.

It is useful to introduce the following standard notation in the case when $\mu$ is a measure and $f$ is a measurable function, and $K$ is a Markov kernel

$$\mu f \triangleq \int f d\mu, \quad \mu K(A) \triangleq \int K(x, A)\mu(dx), \quad Kf(x) \triangleq \int f(z)K(x, dz).$$

The marginal distribution of the signal changes according to

$$\mathbb{P}(X_t \in A \mid X_{t-1} = x_{t-1}) = \int_A k_t(x_{t-1}, dx_t) \tag{19}$$

for $A \in \mathcal{B}(\mathbb{S})$, and $k_t$ is a probability transition kernel defined by the push-forward of $\mathbb{P}$ using the (discretised) SPDE solution map $G$ from assumption 1.

In standard filtering theory the observation process is defined by

$$Y_t = h(X_t) + \epsilon_t, \quad t \in \Lambda \tag{20}$$

where $h : \mathbb{S} \to \mathbb{S}_{\text{obs}}$ is a Borel-measurable function, and for $t \in \Lambda$, $\epsilon_t : \Omega \to \mathbb{S}_{\text{obs}}$ are mutually independent Gaussian distributed random vectors with mean zero and covariance matrix $\alpha^2$. Thus

$$\mathbb{P}(Y_t \in B \mid X_t = x_t) = \int_B g_t(y_t - h(x_t))dy_t \tag{21}$$

for $B \in \mathcal{B}(\mathbb{S}_{\text{obs}})$ and Gaussian density $g_t$. For convenience, define $g_t^{y_t}(x) \triangleq g_t(y_t - h(x_t))$ which is commonly referred to as the *likelihood* function.

We can now define the filtering problem.

**Problem** (Filtering Problem). *For $t \in \Lambda$, we wish to determine the conditional distribution of the signal given the information accumulated from observations, i.e.*

$$\pi_t \varphi \triangleq \mathbb{E}\left[\varphi(X_t) \mid \mathcal{Y}_t\right], \quad \mathcal{Y}_t = \sigma(Y_{0:t}) \tag{22}$$

*for all bounded measurable functions $\varphi \in B(\mathbb{S})$, with $\pi_0$ being the given initial probability distribution on the state space $(\mathbb{S}, \mathcal{B}(\mathbb{S}))$. In particular when $\varphi = \mathbf{1}_A$ for $A \in \mathcal{B}(\mathbb{S})$ we have $\pi_t \mathbf{1}_A = \pi_t(A) = \mathbb{P}(X_t \in A \mid \mathcal{Y}_t)$.*

In statistics and engineering literature, $\pi_t$ is often called the Bayesian posterior distribution. Note that $\pi_t$ is a random probability measure. For arbitrary $y_{0:t}$, denote

$$\pi_t^{y_{0:t}} \varphi \triangleq \mathbb{E}\left[\varphi(X_t) \mid Y_{0:t} = y_{0:t}\right], \quad \pi_t^{y_{0:t}}(A) = \mathbb{P}(X_t \in A \mid Y_{0:t} = y_{0:t}).$$

We also introduce predicted conditional probability measures $p_t$ and $p_t^{y_{0:t}}$ defined by

$$p_t^{y_{0:t-1}} \varphi \triangleq \mathbb{E}\left[\varphi(X_t) \mid Y_{0:t-1} = y_{0:t-1}\right], \quad p_t^{y_{0:t-1}}(A) = \mathbb{P}(X_t \in A \mid Y_{0:t-1} = y_{0:t-1}).$$

We have $\mathbb{P}$-almost surely the following *Bayes recurrence relation*, see Bain and Crisan [2009]. For $t \in \Lambda$ and $A \in \mathcal{B}(\mathbb{S})$,

$$p_t(A) \triangleq \pi_{t-1} k_t(A) = \int k_t(x_{t-1}, A) \pi_{t-1}(dx_{t-1}) \qquad \text{prediction} \qquad (23)$$

$$\pi_t(A) = C_t^{-1} p_t \, g_t^{Y_t}(A) = C_t^{-1} \int_A g_t^{Y_t}(x_t) p_t(dx_t) \qquad \text{update} \qquad (24)$$

where

$$C_t \triangleq p_t \, g_t^{Y_t} = \int_{\mathbb{S}} g_t^{Y_t}(x_t) p_t(dx_t)$$

is a normalising constant. Due to (24), we may also write $\frac{d\pi_t}{dp_t} \propto g_t^{Y_t}$, thus $\pi_t = p_t \frac{d\pi_t}{dp_t}$.

In the general case for any bounded measurable function $\varphi \in \mathcal{B}(\mathbb{S})$, we have for problem 3.1 the recurrence relation

$$p_t \, \varphi = \pi_{t-1} \, k_t \, \varphi \qquad \text{prediction} \qquad (25)$$

$$\pi_t \, \varphi = C_t^{-1} p_t \, g_t^{Y_t} \, \varphi \qquad \text{update} \qquad (26)$$

Except for a few rare examples of the signal, it is extremely difficult to directly evaluate $\pi_t$ because there are no "simple" expressions. In section 3.2 we shall describe the particle filter methodology that we employ to tackle the filtering problem. Note that in statistics and engineering literature, particle filters are often called sequential monte carlo (SMC) methods.

### 3.1.1 Two observation scenarios

For the numerical experiments, we consider two scenarios for the truth.

1. Perfect model: the observations correspond to a single path-wise solution of the SPDE. In this scenario the filtering problem formulation we have described follows through.

2. Imperfect model: the observations correspond to the solution of the PDE, i.e. (20) is changed to

$$Y_t = h(X_t^\dagger) + \epsilon_t$$

   where $X_t^\dagger$ corresponds to the coarse grained PDE velocity field. Coarse graining is explained in section 4. Here there is mismatch between the truth and the signal. As shown in Cotter et al. [2019], the law of the SPDE discretised on the chosen grid converges to the law of the PDE as the discretisation grid gets refined. Implicity also the law of the sequence of true observations $(Y_0, \dots Y_t)$ is close to the law of the model observations. As stated in (18), $\pi_t$ is a continuous function of the law of the signal and the observations $(Y_0, \dots Y_t)$ so we expect a reasonable approximation of $\pi_t$ even when we don't use the true law of the signal[3] but the model law.[4]

---

[3]The true law of the signal is the push-forward of $\pi_0$, the initial distribution of the signal $X_0$. In the case when $X_0$ is deterministic then the distribution of the signal is a Dirac delta distribution.

[4]For continuous time models, this property is called the robustness of the filter. See Clark and Crisan [2005] for results in this direction.

## 3.2 Particle filter

Particle filter methods are among the most successful and versatile methods for numerically tackling the filtering problem. A basic algorithm implements the Bayes recurrence relation by approximating the measure valued processes $\pi_t$ and $p_t$ by $N$-particle empirical distributions. The position of each particle is updated using the signal's transition kernel. At the same time, individual weights are kept up-to-date in accordance with the updated particle positions. It is in the weights updating step that we take into account the information provided by the observations: particles are reweighted using the likelihood function. A new set of particle positions can be sampled based on the updated weights and the procedure iterates.

Due to the high dimensional nature of the systems in consideration, additional techniques are necessary in order to make the basic algorithm work effectively. We provide a concise presentation of the algorithms employed, and note that these methods are all mathematically rigorous. For more thorough discussions we refer the reader to Bain and Crisan [2009], Reich and Cotter [2015], Dashti and Stuart [2017], Kantas et al. [2014], Beskos et al. [2014].

### 3.2.1 Bootstrap particle filter

The basic algorithm, called the *bootstrap particle filter* or the *sampling importance resampling* (SIR) algorithm, proceeds in accordance with the Bayes recurrence relation (23) – (24) by repeating *prediction* and *update* steps. To define the method, we write an $N$-particle empirical approximation of $\pi_t$. Thus at each $t = t_i \in \Lambda$, we have

$$\pi_t \approx \pi_t^N \triangleq \frac{1}{\sum_{m=1}^N w_t^{(m)}} \sum_{n=1}^N w_t^{(n)} \delta(x_t^{(n)}) = \sum_{n=1}^N \bar{w}_t^{(n)} \delta(x_t^{(n)}) \tag{27}$$

where $\delta$ denotes Dirac measure. The discrete measure $\pi_t^N$ is completely determined by particle positions $x_t^{(n)} \in \mathbb{S}$ and weights $w_t^{(n)} \in \mathbb{R}$, $n = 1, \ldots, N$. We define the update rule

$$\{x_t^{(n)}, w_t^{(n)}\}_{n=1}^N \to \{x_{t'}^{(n)}, w_t^{(n)}\}_{n=1}^N, \quad t' = t_{i+1}, \, t' \in \Lambda$$

for advancing $\pi_t^N$ to $p_{t'}^N$ to be given by the numerical implementation of the SPDE solution map G, see (17),

$$x_{t'}^{(n)} = G(x_t^{(n)}, \omega^{(n)})_{t'}, \quad \omega^{(n)} \in \Omega. \tag{28}$$

Note that each particle position $x_t^{(n)}$ is updated independently.

For the weights, suppose the particles $x_t^{(n)}$, $n = 1, \ldots, N$ are independent samples from $\pi_t$ then we have equal weighting for each particle

$$\pi_t^N = \frac{1}{N} \sum_{n=1}^N \delta(x_t^{(n)}).$$

This does not change in the prediction step, thus

$$p_{t'}^N = \frac{1}{N} \sum_{n=1}^N \delta(G(x_t^{(n)}, \omega^{(n)})_{t'}) = \frac{1}{N} \sum_{n=1}^N \delta(x_{t'}^{(n)}). \tag{29}$$

To go from $p_{t'}^N$ to $\pi_{t'}^N$, the weights $w_{t'}^{(n)}$ need to be updated to take into account the observation data $y_{t'}$ at time $t'$. This is done using the likelihood function (21),

$$\bar{w}_{t'}^{(n)} \propto g_{t'}^{y_{t'}}(x_{t'}^{(n)}), \quad \sum_n \bar{w}_{t'}^{(n)} = 1. \tag{30}$$

Using (27) but with the collection of updated particle positions and normalised weights $\{x_{t'}^{(n)}, \bar{w}_{t'}^{(n)}\}_{n=1}^{N}$ we obtain $\pi_{t'}^{N}$.

In the above we assumed to have started with independent samples from $\pi_t$ before proceeding with prediction and update. Thus after we obtain $\pi_t^N$ we have to generate independent (approximate) samples from $\pi_{t'}$ in order to iterate the above prediction and update steps for future times. This is done via *selection* and *mutation* steps. Otherwise the non-uniform weights are carried into future iterations until resampling is required.

**Selection** In high dimensions, $\pi_{t'}^{N}$ can easily become *singular* due to the observations being too informative. This means after the update step, most of the normalised weights are very small. Thus with a finite support, $\pi_{t'}^{N}$ does not have enough particle positions in around the concentration of the true distribution $\pi_{t'}$. Therefore it is desirable to add a *resampling* step so that particles with low weights are discarded, and replaced with (possibly multiple copies of) higher weighted particles. This selection is done probabilistically; for example, one could draw uniform random numbers in the unit interval and select particles based on the size of $\bar{w}_{t'}^{(n)}$, see Bain and Crisan [2009], Reich and Cotter [2015].

**Mutation** Since the resampling step can introduce duplicate particle positions into the ensemble, without reintroducing the lost diversity, repeated iterations of resampling will eventually lead to a degenerate distribution (i.e. measures whose support are singletons). To tackle this issue we apply *jittering* after every resampling step. Jittering is based on Markov Chain Monte Carlo (MCMC) whose invariant measure is the target $\pi_{t'}^{N}$. The jittering step shifts duplicate particle positions whilst preserving the target distribution. We discuss this in section 3.3.

After resampling is applied, we obtain a new ensemble $\hat{x}_{t'}^{(n)}$, $n = 1, \ldots, N$ with equal weights $1/N$, i.e.

$$\pi_{t'}^{N} = \frac{1}{N} \sum_{n=1}^{N} \delta(\hat{x}_{t'}^{(n)}).$$

When we do not resample, then the particles in the ensemble keep the weights given by $\bar{w}_{t'}^{(n)}$, and use (27) for $\pi_{t'}^{N}$.

The resampling step should be done only when necessary to reduce computational cost, because the jittering step requires evaluating the solution map $G$. Therefore we employ a test statistic to quantify the non-uniformity in the weights and only resample when the non-uniformity becomes unacceptable. For this we use the *effective sample size* (ess) statistic. It is defined by the inverse $l^2$-norm of the normalised weights $\bar{\boldsymbol{w}} = (\bar{w}^{(1)}, \ldots, \bar{w}^{(N)})$,

$$\text{ess}(\bar{\boldsymbol{w}}) \triangleq \|\bar{\boldsymbol{w}}\|_{l^2}^{-2} = \frac{1}{\sum_n (\bar{w}^{(n)})^2}. \tag{31}$$

The ess statistic measures the variance of the weights. If the particles have near uniform weights then the ess value is close to $N$. On the other hand if only a few particles have large weights then the ess value is close to 1. In practice we resample whenever (31) falls below a given threshold

$$\text{ess} < N_{\text{thresh}}.$$

Algorithm 1 summarises the bootstrap particle filter. The algorithm starts with an empirical approximation of the initial prior $\pi_{t_0}$ and steps forward in time, assimilates observation data in repeating cycles of prediction-update steps. The ess statistic is employed. When resampling is required, selection-mutation steps are applied.

**Algorithm 1** Bootstrap particle filter

Draw independent samples $x_0^{(n)} \sim \pi_0$, $i = 1, \ldots, N$ and set weights $\bar{w}_0^{(n)} = 1/N$

1: **for** $j = 1, 2, \ldots$ **do**
2:    Compute $x_j^{(n)} = G(x_{j-1}^{(n)}, \omega^{(n)})_j$, $n = 1, \ldots, N$ with $t_j, t_{j-1} \in \Lambda$.
3:    Obtain observation data $y_j$ and compute weights $\bar{w}_j^{(n)} \propto \bar{w}_{j-1}^{(n)} g_j^{y_j}(x_j^{(n)})$, $n = 1, \ldots, N$ using (30).
4:    **if** ess $< N_{\text{threshold}}$ **then**
5:       Sample $\hat{x}_j^{(n)}$, $n = 1, \ldots, N$ according to the weights $\bar{w}_j^{(n)}$.
6:       Set the weights to be $\bar{w}_j^{(n)} = 1/N$.
7:       Apply jittering if there are duplicates in $\{\hat{x}_j^{(n)}\}_{n=1}^N$ to obtain the jittered set $\tilde{x}_j^{(n)}$, $n = 1, \ldots, N$.
8:       Set $x_j^{(n)} = \hat{x}_j^{(n)} = \tilde{x}_j^{(n)}$, $n = 1, \ldots, N$.
9:    **end if**
10:   Move on to the next For loop cycle.
11: **end for**

## 3.3 MCMC and jittering

In this section we describe an effective Metropolis-Hastings MCMC based method called *jittering* with the proposal step chosen specifically for our signal. Jittering reintroduces lost diversity due to resampling by replacing an ensemble of samples that contain duplicates $x_t^{(n)} \sim \pi_t$, $n = 1, \ldots, N$ with a new ensemble $\hat{x}_t^{(n)}$, $n = 1, \ldots, N$ without duplicates, such that the distribution $\pi_t$ is preserved.

MCMC is a general iterative method for constructing ergodic time-homogeneous Markov chains $u(m)$, $m \geq 0$ with transition kernel $K(u, \cdot)$, that are invariant with respect to some target distribution $\pi$, i.e.

$$\pi K(\cdot) = \int K(u, \cdot) \pi(du) = \pi(\cdot).$$

By the Birkhoff's ergodic theorem, we have the following identity

$$\int f(u) \pi(du) = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^n f(u_k) \qquad \text{a.s.}$$

for any integrable and measurable function $f$. Practically, this means starting from an initial $u(0)$, each $u(m)$ with $m \in \mathbb{N}$ can be treated as samples from the target distribution $\pi$.

A generic Metropolis-Hastings MCMC algorithm is described in algorithm 2. A Markov transition kernel $K$ defined on the state space is used to generate proposals. Together with the right conditions on the acceptance probability function $a$ to guarantee *detailed-balance*, the algorithm produces a Markov chain with kernel that is reversible with respect to the target measure $\mu$, see Dashti and Stuart [2017]. In the Gaussian case, a classic and widely used choice for $K$ and $a$ is

$$K(u(m), ) = \rho u(m) + \sqrt{1 - \rho^2} \zeta, \quad \zeta \sim \mathcal{N}(0, \mathcal{C})$$
$$a(u, v) = 1 \wedge \frac{\exp(-\Phi(v))}{\exp(-\Phi(u))} \tag{32}$$

for any appropriate covariance operator $\mathcal{C}$ and log likelihood function $\Phi$, see Kantas et al. [2014]. The parameter $\rho$ controls the local exploration size of the Markov chain. In practise for high

dimensional problems $\rho$ needs to be very close to 1 in order to achieve a reasonable average acceptance probability (an acceptance probability of 0.234 is reasoned to be optimal in Neal et al. [2006]). For bad choices of $\rho$ the MCMC chain may mix very slowly and would require a burn-in step size that makes the whole algorithm computationally unattractive.

---

**Algorithm 2** Generic Metropolis-Hastings MCMC, see Dashti and Stuart [2017]

---

Let $\mu$ be a given measure on the state space. Let $u(0) \sim \mu$. Generate a $\mu$-invariant Markov chain $u(m)$, $m > 0$ as follows

  1: Propose

$$\tilde{u} \sim K(u(m), du) \tag{33}$$

  2: Accept $u(m+1) = \tilde{u}$ with probability

$$a(u(m), \tilde{u}), \tag{34}$$

    otherwise $u(m+1) = u(m)$.

  3: $m \to m+1$ and repeat.

---

With (32), algorithm 2 is known as the Preconditioned Crank Nicolson (pCN) and is well-posed in the mesh refinement limit, see Dashti and Stuart [2017], Kantas et al. [2014]. Thus when applied to discretised problems the algorithm is robust under mesh-refinement. It is commonly applied in Bayesian inverse problems where the posterior is absolutely continuous with respect to a Gaussian prior on Banach spaces. It is important to note that here the design of the algorithm is important because in high dimensions measures tend to be mutually singular, but for Metropolis-Hastings algorithms the acceptance probability is defined as the Radon-Nikodym derivative given by the stationary Markov chain transitions.

Our choice (55) for the prior is not Gaussian. The distribution (19) is also not Gaussian for any $t \in (0, T)$. The distribution of the SPDE solution is investigated numerically in Cotter et al. [2019], in which it is noted that non-Gaussian scaling is interpreted as intermittency in turbulence theory. Because of this, pCN is not feasible here. Therefore it is important to choose $K$ and $a$ such that the following properties hold.

  (i) Robustness under mesh refinement. Although we are considering finite dimensional state spaces, in the limit the state spaces under assumption 1 are infinite dimensional function spaces.

  (ii) The chain should mix and stablise sufficiently quickly so that the number of burn-in steps required is reasonable.

Then with the appropriately chosen $K$ and $a$, we apply algorithm 2 as a jittering step to shift apart duplicate particles introduced into the ensemble by the resampling step.

Given these considerations, we use directly the SPDE solution map $G$ (17) to define the transition kernel $K$. Let the target distribution be the posterior distribution $\pi_{t_k}$, $t_k \in \Lambda$. With a slight abuse of the notation introduced in (17), we write

$$G(u, W_{k-1:k})_{t_k}$$

to mean the solution of the SPDE at time $t_k$ along a realised Brownian trajectory over the time interval $[t_{k-1}, t_k]$ starting from position $u \in \mathbb{S}$. When we consider $u \in \mathbb{S}$ and interval $\Delta_k := t_k - t_{k-1}$ as fixed, then we view $G_{u, \Delta_k}(W) := G(u, W_{k-1:k})$ as a function on $\Omega$.

Let $\pi_{t_k}^N$ be the empirical approximation of $\pi_{t_k}$ with $N$ particles $u_k^{(n)}$, $n = 1, \ldots, N$. We consider each particle $u_k^{(n)}$ a child of some parent $u_{k-1}^{(n)}$ at time $t_{k-1} \in \Lambda$ for a realised Brownian trajectory $W$ over the interval $[t_{k-1}, t_k]$, i.e.

$$u_k^{(n)} = G_{u_{k-1}^{(n)}, \Delta_k}(W)_k.$$

To jitter $u_k^{(n)}$, set $W^{(0)} = W$ and $u(0) = u_k^{(n)}$ (see algorithm 2). At the $m$-th MCMC iteration, $m \geq 1$, propose

$$\tilde{u} = G_{u_{k-1}^{(n)}, \Delta_k}(\rho W^{(m-1)} + \sqrt{1 - \rho^2} Z^{(m-1)})_k \tag{35}$$

where $Z^{(m-1)}$ is a Brownian trajectory over $[t_{k-1}, t_k]$ generated independently from $W^{(m-1)}$.

We use the canonical Metropolis-Hastings accept-reject probability function

$$a(u(m-1), \tilde{u}) = 1 \wedge \frac{g^y(\tilde{u})}{g^y(u(m-1))} \tag{36}$$

where $g^y$ is the likelihood function, see (24). The proposal (35) is accepted with probability (36) independently of $(\tilde{u}, u(m-1))$. In this case set

$$u(m) = \tilde{u} \quad \text{and} \quad W^{(m)} = \rho W^{(m-1)} + \sqrt{(1 - \rho^2)} Z^{(m-1)}.$$

Otherwise the proposal is rejected, in which case set

$$u(m) = u(m-1) \quad \text{and} \quad W^{(m)} = W^{(m-1)}$$

and go to the next iteration in algorithm 2.

Algorithm 3 summarises our MCMC procedure. The algorithm includes tempering scaling $\phi_k$ of the accept-reject function (36). *Tempering* is explained in the next subsection. Practically, to save computation, we may apply jittering to just the duplicated particles after resampling, and run each jittering procedure for a fixed number of steps.

**Proposition 1.** *With the proposal* (35) *and accept-reject function* (36), *the Markov chain generated by algorithm 2 is reversible with respect to $\pi_t^{y_{0:t}}$.*

*Proof.* The generic Metropolis-Hastings algorithm 2 defines the following Markov transition kernel

$$Q(u, dv) = K(u, dv) a(u, v) + \delta_u(dv) \left( \int (1 - a(u, w)) K(u, dw) \right). \tag{37}$$

Since $\pi_t = p_t \frac{d\pi_t}{dp_t}$, if $K$ is such that it satisfies the detailed balance condition with respect to $p_t$

$$p_t(du) K(u, dv) = p_t(dv) K(v, du), \tag{38}$$

then using the accept-reject function (36)

$$a(u, v) = 1 \wedge \frac{g_t^{y_t}(v)}{g_t^{y_t}(u)}$$

we have $Q(u, \cdot)$ is a Markov kernel that is $\pi_t$-invariant, see Dashti and Stuart [2017].

Let $\gamma$ denote the Wiener measure. Note that for a Brownian path $W \sim \gamma$ it is standard to show that the noise proposal in (35)

$$W' := \rho W + \sqrt{1 - \rho^2} Z \sim \gamma$$

for $Z \sim \gamma$ independent of $W$. Thus due to the prediction formula (23) and Markov transition (19) for the signal, conditioned on the value $u_{t-1} \in \mathbb{S}$, we have for $u = G_{u_{t-1}, \Delta}(W)_t \sim \pi_{t-1} k_t = p_t$, a sample $v$ obtained using the proposal (35) is thus

$$v = G_{u_{t-1}, \Delta}(W')_t \sim \pi_{t-1} k_t = p_t, \quad \text{with } u_{t-1} \sim \pi_{t-1}$$

i.e. conditioned on $u_{t-1}$, we have

$$p_t(du)K(u, dv) = p_t(du)p_t(dv)$$

which is symmetric in the pair $(u, v)$ giving us detailed balance (38).

$\square$

---

**Algorithm 3** MCMC jittering for the 2D damped and forced SALT Euler dynamics

Let $t_i \in \Lambda$, $\Delta_i = t_i - t_{i-1}$. Given the ensemble of equal weighted particle positions $\left\{ x_{t_i}^{(n),k} \right\}_{n=1,\dots,N}$ corresponding to the $k$'th tempering step with temperature $\phi_k$, and proposal step size $\rho \in [0,1]$, repeat the following steps.

1: **for** $n = 1, \dots, N$ **do**
2:     Let particle $x_{t_i}^{(n),k}$ be such that $x_{t_i}^{(n),k} = G_{x_{i-1}^{(n)}, \Delta_i}(W)$, for an initial condition $x_{t_{i-1}}^{(n)} \in \mathbb{S}$ and a realised Brownian path $W(t_{i-1} : t_i)$ over the time interval $[t_{i-1}, t_i]$.
3:     Set $u^0 = x_{t_i}^{(n),k}$ and $W^0(t_{i-1} : t_i) = W(t_{i-1} : t_i)$.
4:     **for** $m \geq 1$ **do**
5:         Propose $v \sim K(u(m-1), \cdot)$ given by

$$v = G_{x_{i-1}^{(n)}, \Delta_i}(W')$$

        where

$$W' = \rho W^{m-1}(t_{i-1} : t_i) + \sqrt{1 - \rho^2} Z(t_{i-1} : t_i)$$

        for $Z$ a Brownian path independent of $W^{m-1}$.
6:         Accept $v$ with probability

$$a(u(m-1), v) = 1 \wedge \left( \frac{g_{t_i}^{y_{t_i}}(v)}{g_{t_i}^{y_{t_i}}(u(m-1))} \right)^{\phi_k} \tag{39}$$

        where $g_{t_i}^{y_{t_i}}$ is the likelihood function and $y_{t_i}$ is the observation at time $t_i$, in which case set $u^m = v$ and $W^m = W'$. Otherwise set $u^m = u^{m-1}$ and $W^m = W^{m-1}$.
7:     **end for**
8: **end for**

---

## 3.4 Tempering

Empirical approximations of $\pi_t$ defined on high dimensional space can very quickly become degenerate, which is indicated by low effective sample size (ess) statistic. In order to facilitate smoother transitions between posteriors, so that ensemble diversity is improved, we employ the *tempering* technique, see Neal [2001], Kantas et al. [2014], Beskos et al. [2017, 2014]. Use of other techniques such as *nudging* and *space-time particle filter* (see Beskos et al. [2017]) will be explored in future research work.

We employ tempering when the ess value for an posterior ensemble, falls below an apriori threshold $N_{\text{threshold}}$. The idea of tempering is to artificially scale the log likelihoods by a number $\phi \in (0,1]$ called the *temperature*, which in effect increases the variance of the distribution so that the apriori ess threshold is attained. Once this done resampling can be applied (with MCMC if required) which leads to a more diverse ensemble. Of course particles in this new ensemble are samples of the altered distribution which is not what we desire, therefore the procedure is repeated by finding the next temperature value in the range $(\phi, 1]$. This is repeated until the temperature scaling is 1 so that the original distribution is recovered.

More precisely, let

$$0 = \phi_0 < \phi_1 < \cdots < \phi_R = 1 \tag{40}$$

be a sequence of *temperatures*. Let

$$\pi_{t,r}(A) \triangleq C_{t,r}^{-1} p_t (g_t^{Y_t})^{\phi_r}(A) \tag{41}$$

be called the tempered posterior at the $r$-th tempering step or simply the $r$-th tempered posterior, where $C_{t,r} = p_t(g_t^{Y_t})^{\phi_r}$ (compare with the recurrence formula (24)). Note that $\pi_{t,R} = \pi_t$ and $\pi_{t,0} = p_t$. Thus with

$$\frac{d\pi_{t,r}}{d\pi_{t,r-1}} \propto (g_t^{Y_t})^{\phi_r - \phi_{r-1}}$$

we have

$$\pi_t = p_t \frac{d\pi_{t,1}}{d\pi_{t,0}} \cdots \frac{d\pi_{t,R}}{d\pi_{t,R-1}}$$

which suggests the iterative procedure,

$$\pi_{t,r-1} \longmapsto \pi_{t,r} \propto \pi_{t,r-1}(g_t^{Y_t})^{\phi_r - \phi_{r-1}}, \quad r = 1, \ldots, R. \tag{42}$$

Empirically this means, for each $r = 1, \ldots, R$, assume we have equal weighted particle positions $\{x_t^{(n)}\}_{n=1,\ldots,N}$ that give us the empirical $(r-1)$-th tempered posterior

$$\pi_{t,r}^N = \frac{1}{N} \sum_{n=1}^N \delta(x_t^{(n)}),$$

we compute unnormalised tempered weights

$$w_t^{(n),r}(\phi_r) \triangleq (g_t^{y_t}(x_t^{(n)}))^{\phi_r - \phi_{r-1}}, \quad n = 1, \ldots, N \tag{43}$$

to obtain the empirical $r$-th tempered posterior

$$\pi_{t,r}^N = \sum_{n=1}^N \bar{w}_t^{(n),r}(\phi_r)\delta(x_t^{(n)}).$$

Then we resample according to $\pi_{t,r}^N$ and apply the MCMC jittering algorithm 3 (see remark 3) to separate apart any duplicated particles before going to the $r+1$'th iteration step.

The sequence of temperatures $\phi_r$ is chosen so that at each tempering iteration $r$, the empirical tempered distribution $\pi_{t,r}^N$ attains the apriori ESS threshold $N_{\text{threshold}}$, i.e.

$$\text{ess}(\bar{\boldsymbol{w}}^r(\phi_r)) \geq N_{\text{threshold}} \tag{44}$$

where $\bar{\boldsymbol{w}}^r(\phi_r) = (\bar{w}^{(1),r}, \ldots, \bar{w}^{(N),r})(\phi_r)$ are the normalised weights corresponding to (43). This way the choice for the temperatures can be made on-the-fly by using search algorithms such as *binary search* at each tempering iteration.

**Remark 3.** *Proposition 1 shows the MCMC jittering algorithm preserves the target distribution $\pi_t$ with the accept-reject function (36). The same argument shows the algorithm preserves the tempered posteriors as long as the accept-reject function is chosen to be (39). The Markov transition kernel $K$ satisfies the detailed balance condition with respect to $p_t$ independent of tempering.*

Using tempering to smooth out the transition between consecutive filtering measures (i.e. from $\pi_{t_k}$ to $\pi_{t_{k+1}}$) ensures that the importance weights in (41) exhibit low variance, so that no small group of particles are favoured much more than the rest when resampling, see Kantas et al. [2014], thus leading to a more diverse population.

In algorithm 4 we summarise the complete procedure for one filtering step, i.e. from $\pi_{t_{i-1}}^N$ to $\pi_{t_i}^N$, incorporating adaptive tempering and MCMC jittering for SALT into the bootstrap particle filter.

---

**Algorithm 4** One step particle filter for SALT with adaptive tempering and MCMC jittering

Consider the $i$'th filtering step corresponding to $t_i \in \Lambda$. Given the ensemble of equal weighted particle positions $\{x_{i-1}^{(n)}\}_{n=1,\ldots,N}$ that defines the empirical posterior $\pi_{i-1}^N$, we wish to assimilate observation data $y_i$ at time $t_i$ to obtain a new equally weighted ensemble $\{x_i^{(n)}\}_{n=1,\ldots,N}$ that defines $\pi_i^N$. Define

$$\text{ess}(\phi., \{x_i^{(n)}\}_{n=1,\ldots,N}) \triangleq |\bar{w}_i^{(n),\cdot}(\phi.)|_{l^2}^{-1}$$

for $\phi \in (0,1]$, and $\bar{w}_i^{(n)}$ are the normalised values of the unnormalised tempered weights (43).

1: Compute $x_i^{(n)} = G(x_{i-1}^{(n)}, W)_{i-1}$, $n = 1, \ldots, N$.
2: Set $\phi_0 = 1$, $r = 1$, $x_i^{(n),0} = x_i^{(n)}$ each $n$.
3: **while** $\text{ESS}(\phi_{r-1}, \{x_i^{(n),r-1}\}) < N_{\text{threshold}}$ **do**
4:     Find (using e.g. binary search) the largest $\phi_r \in (\phi_{r-1}, 1)$ such that $\text{ESS}(\phi_r, \{x_i^{(n),r-1}\}) \geq N_{\text{threshold}}$.
5:     Resample according to $\bar{w}^{(n),r}(\phi_r)$ to obtain a new set $\{x_i^{(n),r}\}$.
6:     Apply jittering algorithm 3 to any duplicated particles.
7: **end while**
8: Set $R = r$. Do steps 5. and 6. with $\phi_R = 1$ and set $x_i^{(n)} = x_i^{(n),r}$, $n = 1, \ldots, N$ to obtain $\pi_{t_i}^N$.

---

# 4   Numerical setup and experiment results

The setup for the numerical experiments follow on the $\boldsymbol{\xi}$ calibration and uncertainty quantification work presented in Cotter et al. [2019]. Thus the parameter choices for the models are as follows: forcing strength $\alpha = 0.1$, number of gyres $\beta = 8$ and damping rate $r = 0.01$.

The PDE (2) and SPDE (7) are prescribed on mesh of size $512 \times 512$ cells and $64 \times 64$ cells respectively for the spatial domain. We use a Galerkin finite element discretisation for the spatial variable and a third order stability preserving Runge-Kutta for the time stepping, see Cotter et al. [2019] for details. This means spatially each mesh cell contains six grid points. Thus the PDE and SPDE velocity fields are of 3145728 and 49152 degrees of freedom respectively. Henceforth we shall refer to the PDE spatial dimension as *fine resolution* and the SPDE spatial dimension as *coarse resolution*[5].

---

[5]However since we are using an explicit in time method for solving the SPDEs, the coarse time step may

The time step for the fine resolution is chosen in accordance with the CFL condition and in this case is $\Delta_{\mathrm{f}} = 0.0025$. The CFL time step for the coarse resolution is $\Delta_{\mathrm{c}} = 0.02$.

The reference fine resolution PDE trajectory was spun-up from the initial configuration

$$
\begin{aligned}
\omega_{\mathrm{spin}} = {} & \sin(8\pi x)\sin(8\pi y) + 0.4\cos(6\pi x)\cos(6\pi y) \\
& + 0.3\cos(10\pi x)\cos(4\pi y) + 0.02\sin(2\pi y) + 0.02\sin(2\pi x)
\end{aligned}
\tag{45}
$$

until some energy equilibrium state, see Cotter et al. [2019]. We call the equilibrium state's corresponding time point the initial time $t_0$.

We use *eddy turnover time* (abbrev. ett) as the time dimension for the PDE system. It describes the time scale of flow features correponding to a given length scale, and is defined by

$$
\tau_l \triangleq \frac{l}{|\bar{\mathbf{u}}|}
\tag{46}
$$

where $|\bar{\mathbf{u}}|$ is the magnitude of the stablised mean velocity[6], and $l \in [0, L]$ a length scale. Here $L = 1$ corresponds to the axis length of the domain $D$. For our experiments, we choose $l = \frac{1}{2}$. It is estimated that 1 ett roughly equals to 2.5 numerical time units, or 1000 (fine resolution) CFL numerical time steps. Since the SPDE is thought of as a stochastic parameterisation for the PDE, we shall use the same eddy turnover time dimension for the SPDE. Thus 1 ett is 125 coarse resolution CFL numerical time steps.

For the SPDE model, we use the calibrated EOFs $\boldsymbol{\xi}_i, i = 1, \ldots, N_\xi$, from Cotter et al. [2019] with $N_\xi$ corresponding to 50% of the total spectrum. This choice is informed by uncertainty quantification tests and amounts to $N_\xi = 51$ when the SPDE is prescribed on a mesh of size $64 \times 64$ cells.

We consider two scenarios for the filtering problem, a perfect model scenario and an imperfect model scenario, see section 3.1.1. In the imperfect model scenario, since the SPDE solution is meant to capture the large-scale features of the deterministic fine resolution dynamics that are resolvable at the coarse resolution, we should obtain observations from the coarse grained PDE solution. For coarse graining, we use the inverse Helmholtz operator

$$
H \triangleq \left( \mathrm{Id} - \frac{1}{k^2}\Delta \right)^{-1}
\tag{47}
$$

and apply $H$ to the PDE stream function (4) to average out its small scale features. The boundary condition we impose on the coarse grained stream function is the same Dirichlet condition as for (4). The value $k$ in the definition of $H$ corresponds to the coarse resolution, in this case $k = 64$. To obtain the coarse grained PDE velocity field, apply the linear operator $\nabla^\perp$ to the coarse grained stream function. The coarse grained PDE velocity field is then used to generate the observation data in the imperfect model filtering scenario. It is important to note that this coarse graining procedure is only applied when we obtain observation data, the underlying fine resolution dynamics is unchanged.

In both scenarios the observations are defined as noisy point measurements of the truth's velocity field. The observation locations (thought of as "weather stations") are given by a uniform regular grid of dimension $d_y = 81$; see section 3.1 for the problem's mathematical formulation. We also investigate the impact of increasing the number of weather stations to $d_y = 289$. For this paper we only consider fixed uniform geometry for the weather stations.

---

need to be smaller to accommodate the fact that Brownian increments are unbounded.

[6]Our PDE system is spun-up from (45) to an energy stable state, thus $|\bar{\mathbf{u}}|$ is constant in time.

Further the 81 weather stations are a subset of the 289 weather stations. Figure 1a visually illustrates a snapshot of the coarse grained numerical PDE solution velocity vector field overlaid with the positions of the 81 weather stations.

**Remark 4.** *The dimension of the observation space compared to the dimension of the underlying truth is minute. Using* 289 *weather stations amounts to* 1.18% *of the overall degrees of freedom in the perfect model scenario, and* 0.01837% *of the overall degrees of freedom in the imperfect model scenario. These parameter choices are the best we can do given our computational hardware, so that we can obtain numerical results in a reasonable amount of time. For reference, all numerical experiments for this paper were run on a workstation equipped with two Intel Xeon CPUs totalling* 32 *logical processors, and* 64GB *of memory.*

The observation error covariance $\alpha$ (see (20) for definition) is calibrated by computing the standard deviation of the fine resolution PDE velocity field within coarse cells, and then averaged along the time axis. More precisely, let $\mathbb{S}_{\text{PDE}}$ denote the discretised PDE state space. Let $X_t \in \mathbb{S}_{\text{PDE}}$ denote a snapshot in time of the PDE velocity field. Let superscript indices denote vector component. Define $\bar{X}_t \in \mathbb{S}_{\text{PDE}}$ by

$$\bar{X}_t^{i_k} = \frac{1}{\text{card}(j : X_t^j \in \text{coarse cell } k)} \sum_{j: X_t^j \in \text{coarse cell } k} X_t^j, \qquad i_k \in \{i : X_t^i \in \text{coarse cell } k\}$$

for coarse cells corresponding to the coarse resolution mesh. Thus $\bar{X}_t^i$ are the local coarse cell averages of $X_t$. Then we define $\alpha$ by

$$\alpha = \lambda \frac{1}{M} \sum_{i=1}^{M} \left( \left( X_{t_i} - \bar{X}_{t_i} \right) \otimes \left( X_{t_i} - \bar{X}_{t_i} \right) \right)^{1/2} \tag{48}$$

where $\lambda$ is a weight we can use to add control over $\alpha$. Throughout the rest of the paper, we set $\lambda = 10$. The idea is $\alpha$ at the observation locations represent the local variability of the truth at the observation locations. The $\alpha$ computed this way is a vector field defined on the fine resolution grid. It is evaluated at the observation locations when used as part of (20).

Figure 1b visually illustrates the magnitude of $\alpha$ overlaid with the observation locations. We use the same calibrated $\alpha$ in both problem scenarios.

In the perfect model scenario, the truth from which we obtain the observations is a single simulated realisation of the SPDE. The initial condition for the SPDE truth is a particular sample from $\pi_0$, see figure 2 for a visualisation of the SPDE truth without observation noise at the initial time $t_0$. We discuss what $\pi_0$ is and how we sample from it in section 4.1. In the imperfect model scenario the truth is the coarse grained PDE velocity field. Figure 3 shows a visualisation of the PDE truth without observation noise at the initial time $t_0$.

We use an ensemble size of $N_{\text{p}} = 100$ particles. Each particle's initial condition is a sample from the initial distribution $\pi_0$. Also we do not assimilate at $t_0$ since the initial distribution is assumed given, see section 3.1. The error metrics we shall look at for the numerical experiments are

- the *mean square error* (mse) between the ensemble mean $\hat{X}_t = \frac{1}{N_p} \sum_{i=1}^{N_p} X_t^i$, which approximates the posterior mean $\mathbb{E}(X_t | \mathcal{Y}_t)$, and the truth $X_t^\dagger$,

$$\text{mse}(t) = \|\hat{X}_t - X_t^\dagger\|_{L^2}^2, \tag{49}$$
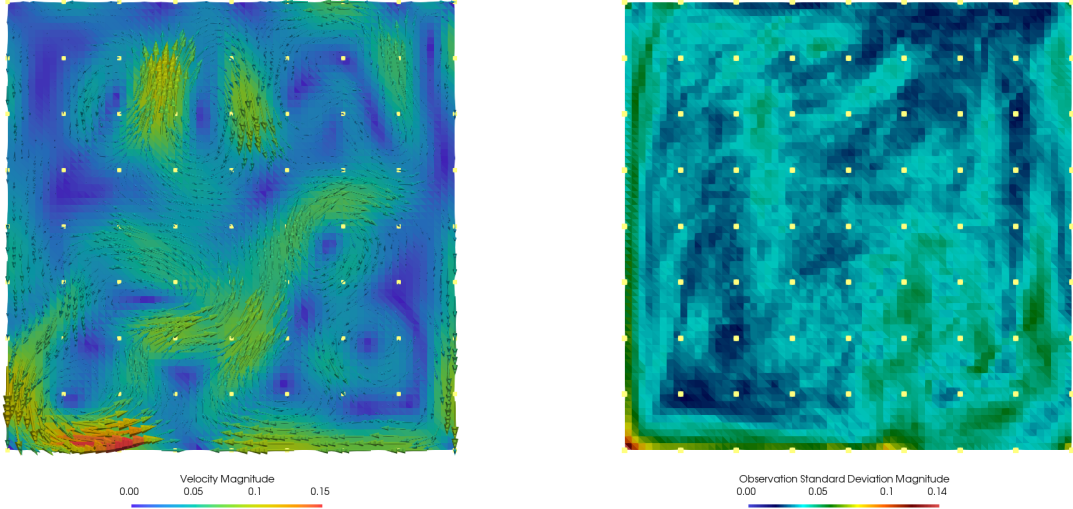
(a) Snapshot of the coarse grained PDE velocity field at a $t \in \Lambda$ overlaid with observation locations (yellow dots) which are defined by a grid of $(8 \times 8)$ cells.

(b) Magnitude of the calibrated observation error $\alpha$ (see (48)) overlaid with observation locations (yellow dots) which are defined by a grid of $(8 \times 8)$ cells.

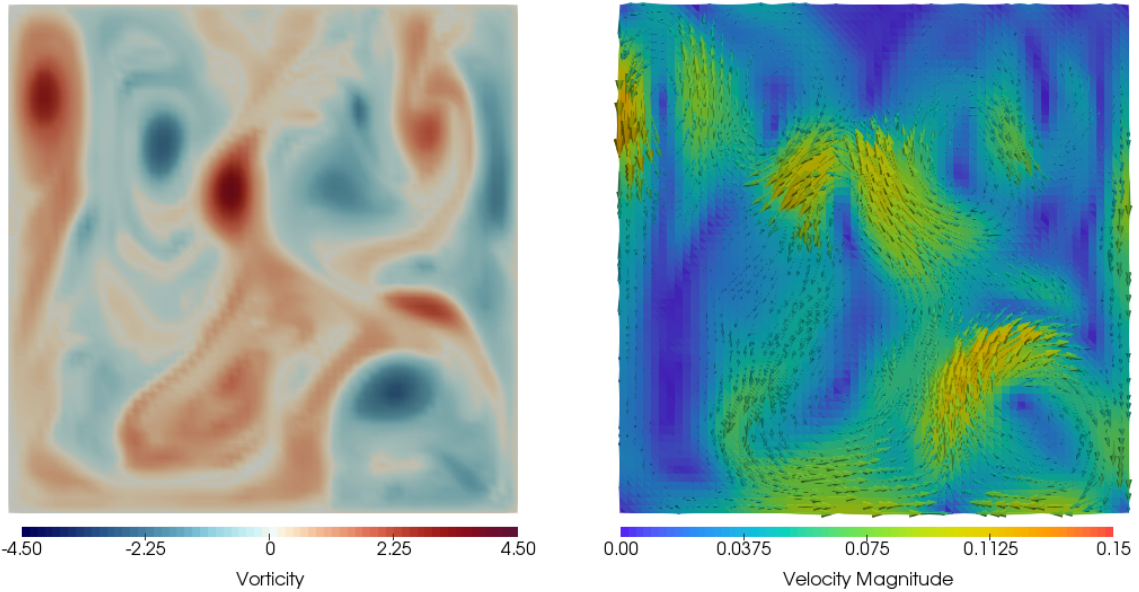Figure 1: Observation locations and observation error magnitude.



Figure 2: Perfect model scenario. Visualisation of the *vorticity* scalar field (left) and the *velocity* vector field (right) of the SPDE truth at the initial time $t_0$. The different colours of the vorticity field reflect the clockwise and anti-clockwise directions of the velocity vectors. The velocity vectors are visualised using arrows, the size of which reflect their magnitude. This initial condition is obtained by applying the *deformation* procedure described in section 4.1 to the PDE truth shown in figure 3, for 104 fine resolution numerical time steps.
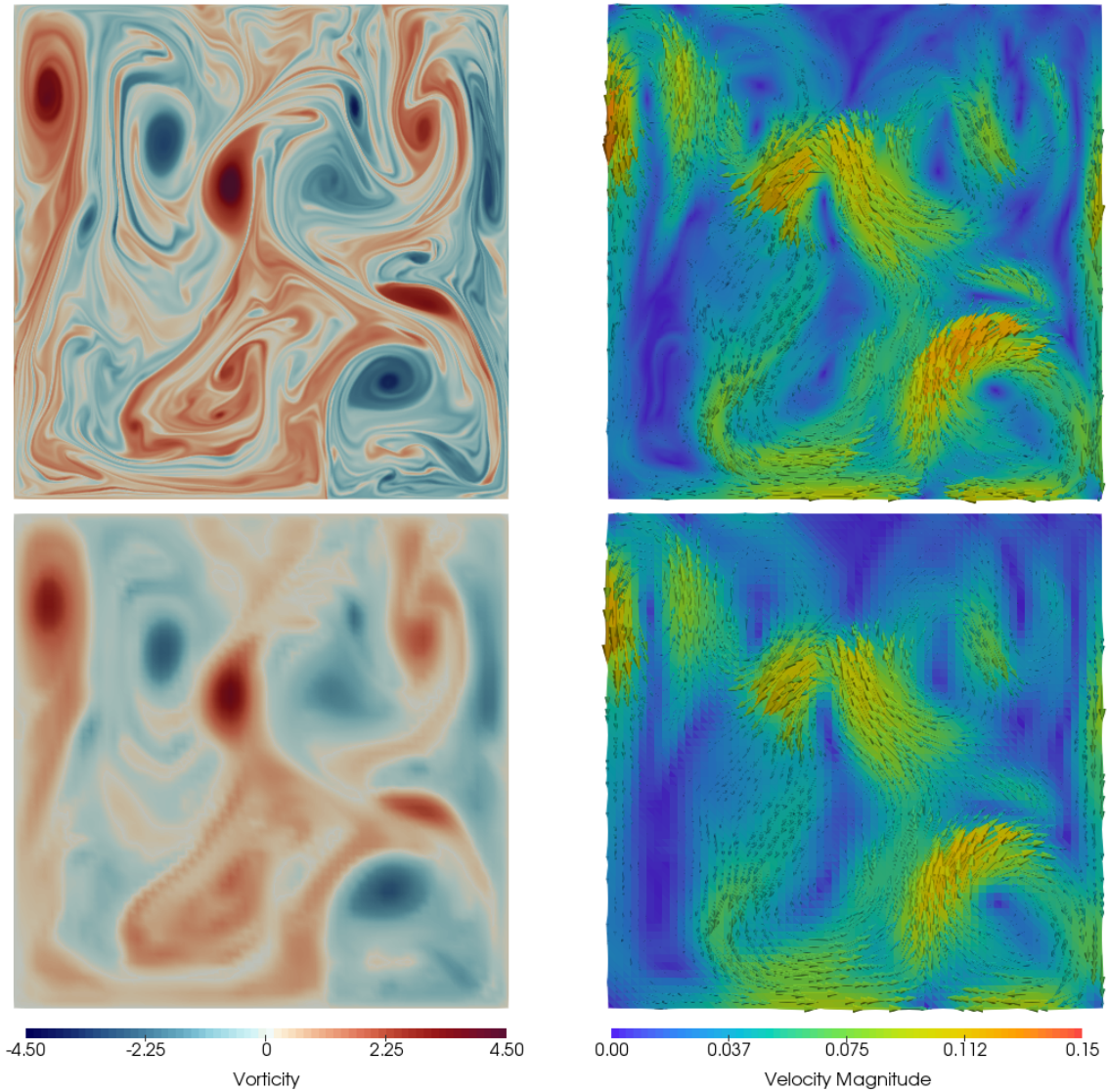
Figure 3: Imperfect model scenario. Visualisation of the PDE truth at the initial time $t_0$. Here the top row correspond to the fine resolution visualisation of the *vorticity* scalar field (left) and *velocity* vector field (right). The bottom row correspond to the *coarse grained* version of the top row using the coarse graining operator (47). The different colours of the vorticity fields reflect the clockwise and anti-clockwise directions of the velocity vectors. The velocity vectors are visualised using arrows, the size of which reflect their magnitude.

- the *ensemble standard error* $\sigma_{N_p}$

$$\sigma_{N_p}^2(t) = \frac{1}{N_{\mathrm{p}} - 1} \sum_{i=1}^{N_{\mathrm{p}}} \|X_t^i - \hat{X}_t\|_{L^2}^2, \tag{50}$$

- the effective sample size (ess) statistic (31) for measuring the variance of the ensemble weights,

- and *rank histograms* for assessing the reliability of the particle filter, see Broecker [2018], Reich and Cotter [2015]. This is a standard measure of ensemble reliability. At any reference grid location, given the ensemble values $\{x_t^i\}_{i=1,\dots,N_p}$ that corresponds to the forecast distribution $p_t^{N_p}$ (23), and an observation value $y_t$, define the *rank* function

$$R(y_t, \{x_t^i\}_{i=1,\dots,N_p}) = k \quad \text{if } x_t^j \le y_t \text{ for } j < k, \text{ and } x_t^j > y_t \text{ for } j \ge k. \tag{51}$$

The rank function $R$ takes values in $\{0, 1, \dots, N_p\}$. If the ensemble forecast is reliable then $R$ is a uniform random variable, meaning the verification and the ensemble members are indistinguishable. Thus collecting the rank values over time $t \in \Lambda$, we should obtain a "flat" histogram plot if the particle filter gives reliable results. Further it is shown in Broecker [2018] that the rank statistic $R$ is of $\chi^2$ distribution with $N_p$ degrees of freedom.

Table 1 shows the mean square error, ensemble standard error and ess values for the 100 initial ensemble particles generated using the deformation procedure in section 4.1. We choose the ess threshold $N_{\mathrm{threshold}}$ to be 80% of the ensemble size $N_p$. For the initial ensemble, the ess values are computed for reference only. We do not weight the initial ensemble using observations.

| | ess | mse | $\sigma_{N_p}$ | $\|X^\dagger\|_{L^2}^2$ |
|---|---|---|---|---|
| perfect model scenario | 84.04% | $2.3585 \times 10^{-8}$ | $1.2269 \times 10^{-3}$ | $2.2043 \times 10^{-3}$ |
| imperfect model scenario | 83.83% | $1.2385 \times 10^{-8}$ | $1.2269 \times 10^{-3}$ | $2.2033 \times 10^{-3}$ |

Table 1: An initial ensemble is generated using the deformation procedure described in section 4.1 and contains $N_p = 100$ particles. In this table we report the effective sample size (according to 81 observation locations), mean square error and ensemble standard error statistics, averaged over 50 independent repeats, for both the perfect model scenario and the imperfect model scenario. The truth $X^\dagger$ in the perfect model scenario at time $t_0$ is also generated using the deformation procedure. We use this initial ensemble for both scenario experiments. Percentage values are shown in the ess column. We have chosen the ess threshold to be 80% of the ensemble size. Thus this initial ensemble attain the chosen threshold. For reference, the squared $L^2$ norm of the truths are shown in the last column of the table. The ensemble standard errors (50) are the same in both scenarios since $\sigma_{N_p}$ does not depend on $X^\dagger$.

## 4.1 Initial distribution

The initial distribution $\pi_0$ comes from the following construction which we call *deformation*, see Cotter et al. [2019]. Let $\omega_{\mathrm{truth}}$ be a fine resolution PDE vorticity field. Using the coarse

graining operator $H$ (defined in (47)), define operator $K : \mathbb{S}_{\text{PDE}} \times \mathbb{R} \to \mathbb{S}_{\text{SPDE}}$ by

$$K(\mathbf{u}, \beta) = \nabla^\perp H(\Delta^{-1}\omega^{\mathbf{u},\beta}) \tag{52}$$

where $\omega^{\mathbf{u},\beta}$ is the (vorticity) solution of the linear PDE

$$\partial_t \omega + \beta \, \mathbf{u} \cdot \nabla \omega = 0 \tag{53}$$
$$\omega_0 = \omega_{\text{truth}} \tag{54}$$

$\beta \sim \mathcal{N}(0, \epsilon)$, is a centered Gaussian weight with an apriori variance parameter $\epsilon$, and $\mathbf{u} \sim \mathcal{U}(\mathbb{S}_{\text{PDE}})$ is random draw from a uniform distribution on $\mathbb{S}_{\text{PDE}}$. $\beta$ and $\mathbf{u}$ are independent. Then

$$\pi_0(A) \triangleq \mathbb{P}\left(K(\mathbf{u}, \beta) \in A\right) \qquad A \in \mathcal{B}(\mathbb{S}_{\text{SPDE}}) \tag{55}$$

**Remark 5.** *Practically, we randomly draw a vorticity field from the energy stable period prior to the initial data assimilation time point $t_0$. The drawn vorticity state is then used to compute its corresponding stream function by inverting the Laplacian and using the same Dirichlet condition as for (4). The velocity field $\mathbf{u}$ in (53) is then obtained from the stream function. Thus for the linear system (53) the boundary condition is supplied via the sampled $\mathbf{u}$.*

For the initial ensemble (see table 1 for diagnostics), the deformation equation (53) parameters are set as follows. $\omega_{\text{truth}}$ is set to be the imperfect model scenario initial condition (see figure 3 for visualisations). For the weighting parameter $\beta$, we set $\epsilon = 0.25$. Equation (53) is solved independently 100 times to obtain $N_p = 100$ ensemble members. Each time, equation (53) is solved over a time period that attains the chosen ess threshold. For the results shown in table 1, this amounts to a time period of 0.26 or equivalently 104 fine resolution CFL time steps.

In Hamiltonian mechanics, the conservation laws associated with relabelling symmetries are called Casimirs. In lemma 2 we show our choice for the prior distribution is physical in the sense that any sample generated by the procedure $K(\mathbf{u}, \beta)$ preserves the Casimirs of the truth $\omega_{\text{truth}}$.

**Definition 1** (Casimir, see Gay-Balmaz and Holm [2013]). *For 2D incompressible ideal fluid motion, the Casimirs are*

$$C_\Phi = \int_{\mathcal{D}} \Phi(\omega) dx$$

*for any $\Phi \in C^\infty(\mathbb{R}, \mathbb{R})$.*

**Lemma 2** (Preservation of Casimirs). *Let the domain $\mathcal{D}$ be bounded with piecewise smooth boundary. Assume the sampled vector field $\mathbf{u} \in \mathbb{S}_{PDE}$ is divergence free and $\mathbf{u} \cdot \hat{n} = 0$ with $\hat{n}$ being the normal to the boundary $\partial\mathcal{D}$, then $\omega^{\mathbf{u},\beta}$ preserves the Casimir values of $\omega_{truth}$.*

*Proof.* We have

$$\frac{d}{dt}C_\Phi = \int_{\mathcal{D}} \frac{d}{dt}\Phi(\omega) dx$$
$$= \int_{\mathcal{D}} \Phi'(\omega)\partial_t\omega \, dx = -\int_{\mathcal{D}} \Phi'(\omega)\beta \, \mathbf{u} \cdot \nabla\omega \, dx$$
$$= -\int_{\mathcal{D}} \beta \, \mathbf{u} \cdot \nabla\Phi(\omega) \, dx = 0$$

where the last equality follows from integration by parts and the conditions assumed on $\mathbf{u}$. $\quad\square$
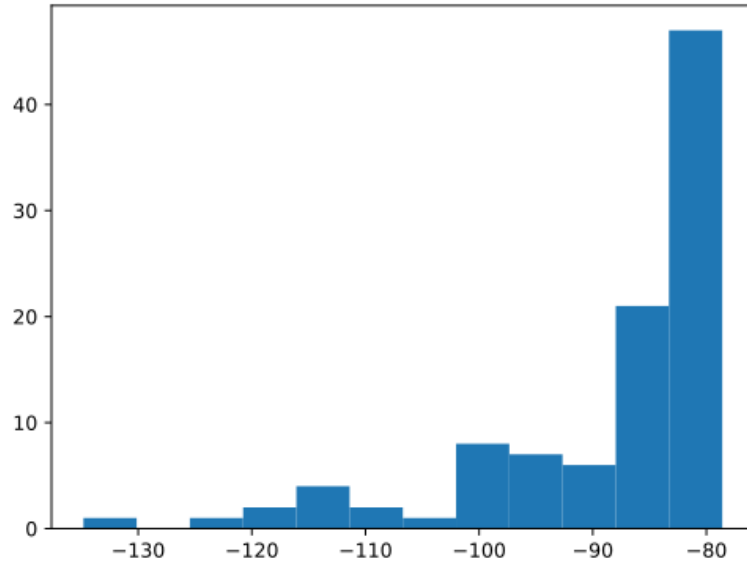
Figure 4: Histogram showing log-likelihood $\ln g^{Y_t}(\cdot)$ weights for an ensemble of 100 particles that define the forecast distribution at time $t_1$ with $t_1 - t_0 = 1$ eddy turnover time. The highly right skewedness of the bins and the numerical range of log-likelihood values demonstrate the singular nature of $\pi_{t_1}^{N_p}$ and the requirement of tempering and MCMC jittering techniques to make the basic bootstrap particle filter work. Otherwise resampling would lead to degenerate posteriors $\pi_t^{N_p}$.

Before discussing experiment results, figure 4 shows a histogram of the log-likelihood $\ln g^{Y_t}(\cdot)$ values for an ensemble of 100 particles that defines the forecast distribution $p_{t_1}^{N_p}$, with $t_1 - t_0 = 1$ ett. It shows straightforwardly the singular nature of $\pi_{t_1}^{N_p}$ and that without tempering and MCMC jittering, a plain bootstrap particle filter algorithm would fail in the sense that particle diversity would be lost very quickly, leading to degenerate posteriors $\pi_t^{N_p}$.

## 4.2   Perfect model scenario

A single realisation of the SPDE was used as the truth for the experiments in this scenario. The data assimilation experiments are defined by the following parameters: time interval between assimilations $\Delta$ and observation data dimension $d_y$. We ran experiments for the following parameter sets using ensemble size 100 and experiment period 10 ett:

1. $\Delta = 1/25$ ett, $d_y = 81$;

2. $\Delta = 1/25$ ett, $d_y = 289$;

3. $\Delta = 1/5$ ett, $d_y = 289$.

Note that 10 eddy turnover times is equivalent to 10000 fine resolution time steps. We compare the results of parameter set 1) with those of 2) to see the impact of increased observation data. We compare the results of parameter set 2) with those of 3) to see the impact of more frequent assimilations.

For each parameter set, the data assimilation experiment was repeated 20 times, but all using the same truth. However the observation errors were independent between repeats. Figure 5 shows the results of mse, $\sigma_{N_p}$, number of tempering steps and ess for the averages.
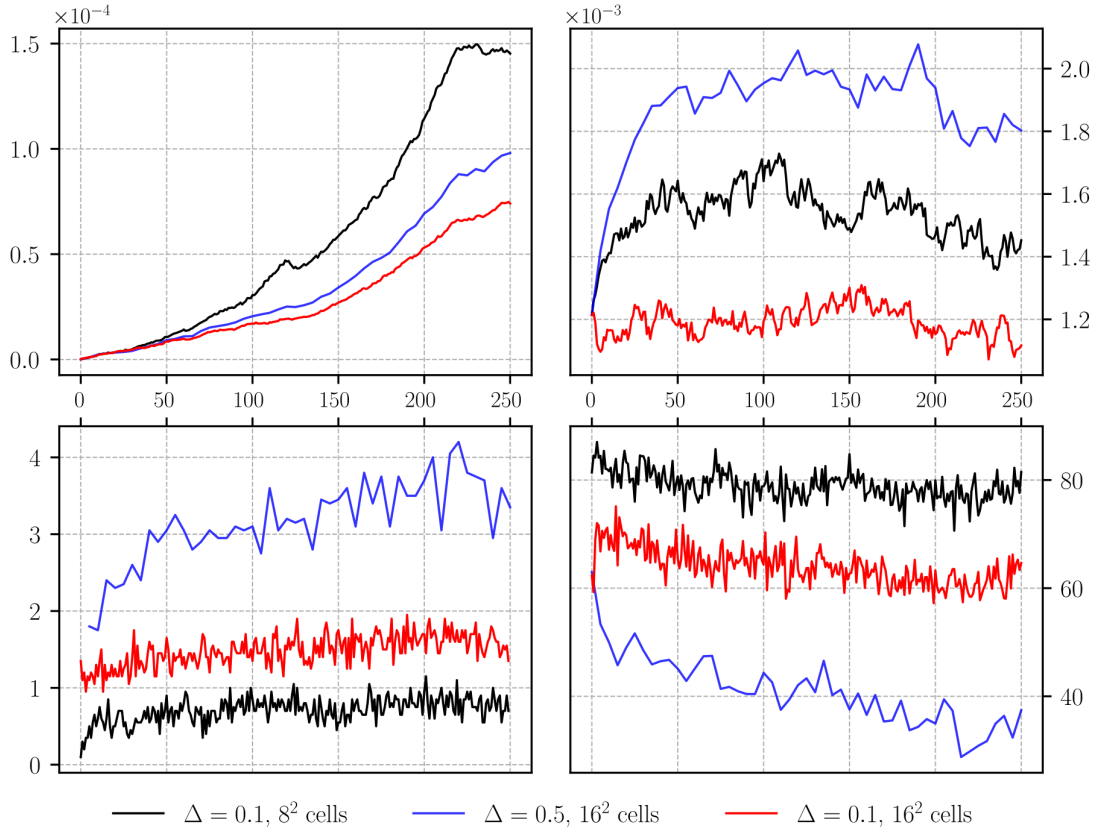
Figure 5: Perfect model scenario. The four figures show the mean square error (*tl*: top-left), ensemble standard error (*tr*: top-right), number of tempering steps (*bl*: bottom-left) and ess (*br*: bottom-right), averaged over 20 independent experiment runs, conditioned on the same truth. However we did not condition on the same observation noise. The x-axis shows the number of data assimilation steps corresponding to the assimilation interval $\Delta = 1/25$ ett (0.1 time units). In all experiments the ensemble size is 100. Three plots are shown in each figure. The black lines correspond to particle filter parameter choices of $\Delta = 1/25$ ett and 81 weather stations. The red lines correspond to $\Delta = 1/25$ ett and 289 weather stations. The blue lines correspond to $\Delta = 1/5$ ett and 289 weather stations. For the black and red lines, the parameter choices mean a total of 250 data assimilation steps. For the blue lines, it means a total of 50 data assimilation steps. See section 4.2 for discussion.

The experiments were run for a period of 10 ett. For parameter sets 1 and 2 this means a total of 250 data assimilation steps. For parameter set 3 this means a total of 50 data assimilation steps.

The same initial ensemble was used in all experiments. Thus the ess and $\sigma_N$ plots all start from the same values (c.f. table 1). The ess statistic depends on the observation dimension $d_y$. The initial ensemble was sampled to have an ess value above 80% for $d_y = 81$. For $d_y = 289$ the ess values are expectly worse and start between 60% and 65%.

The mse results show that by increasing the number of fixed weather stations to 289 from 81, the resulting mse is reduced by 50%. But we get less accurate results if we assimilate less frequently (parameter set 3). However parameter set 3 still gives 1/3 error reduction over parameter set 1. Note that 81 weather stations amount to 0.3296% of the total degrees of freedom, compared to 1.176% from using 289 weather stations, see remark 4.

Taking into account the $\sigma_{N_p}$ results, we see although parameter set 3 (observe more data, but less frequently) gives better mse results than parameter set 1, its ensemble spread is larger than parameter set 1's ensemble. This suggests that, whilst over a longer assimilation interval the posterior ensemble spreads out more, using more observation data gave more accurate results. Parameter set 2 gives the best results out of the three sets.

Both the ess and the number of tempering steps results reflect the features seen in the mse results. As the error increased, a drop in ess value is observed. This led to the tempering procedure increasing the number of intermediate tempering steps. Further, the initial ess value can be preserved, only if one assimilates adequately often. Otherwise the spread gets "too" large, which makes it difficult for tempering and jittering to maintain a diverse ensemble in and around the concentration of the posterior measure. In such cases, one may need to introduce additional techniques such as nudging to obtain better ensembles. We look to investigate such techniques in future work.

For one particular experiment run using 100 particles and parameter set 2, figures 6a and 6b show the rank histogram plots at nine reference grid locations for the velocity x-component and y-component respectively. The plots do not show features of strong bias, under-dispersion or over-dispersion indicating that the filter is producing reliable results over the experiment period of 10 ett.

However, when using a fewer number of weather stations (parameter set 1), in practise we observed it is more likely that the particle filter will give less reliable results. Figure 7 shows the result from an experiment using 100 particles and parameter set 1. We see evidence of skew at some of the plotted grid locations.

## Estimated asymptotic results and discussion

To estimated the asymptotic behaviour of the particle filter, we ran an experiment using 100 particles and parameter set 2 for 50 eddy turnover times. This period length is equivalent to 50000 fine resolution time steps, and for the filter it amounts to 1250 data assimilation steps. The results are shown in figure 8. There, we have compared the results produced by the filter with that of the $N$-particle approximation of the prior distribution, i.e. $\mathbb{P}(X_t)$ without any data assimilation, c.f. section 3.1. The results show the asymptotic behaviour of the normalised mse stays bounded. However over this extremely long period of 50 ett, based on only the normalised mse results, the prior ensemble over the long run gives just as good results as the particle filter posterior ensemble. This is probably because this amount of data only makes a small change
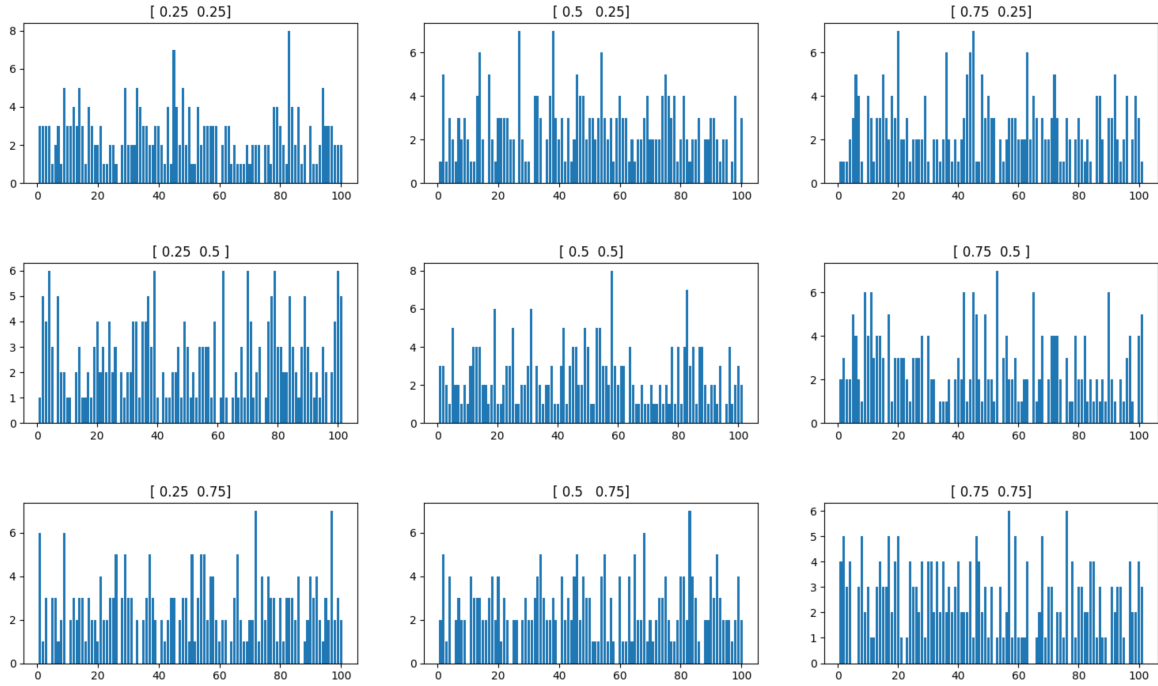
(a) Velocity x-component
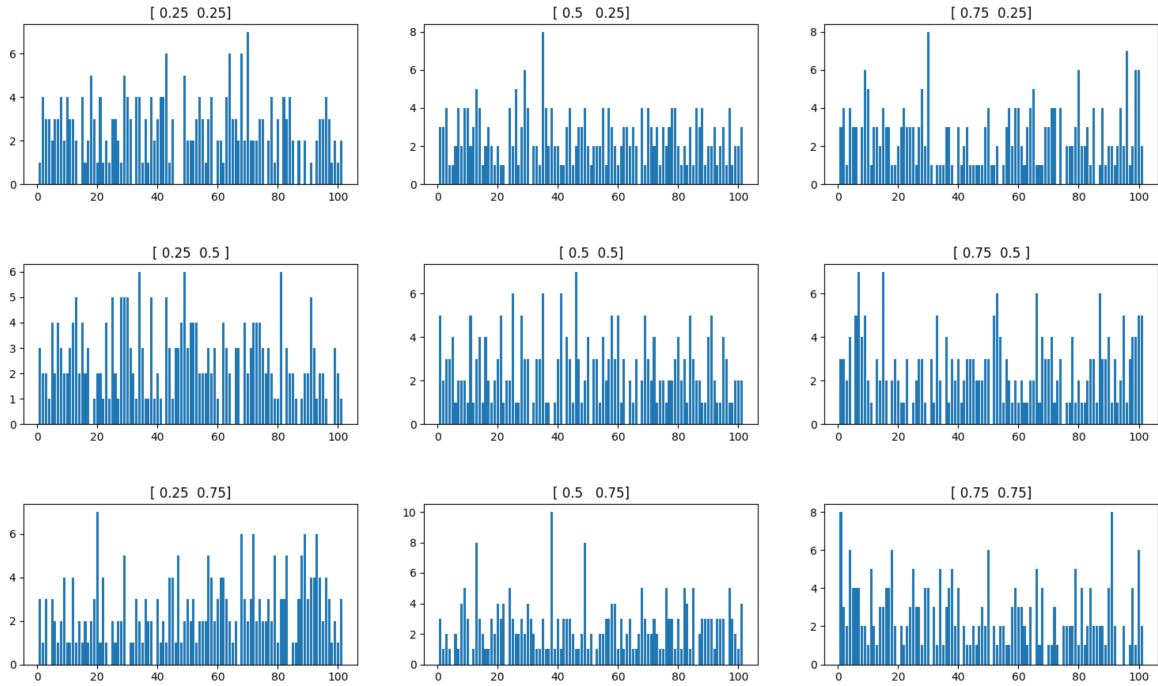


(b) Velocity y-component

Figure 6: Perfect model scenario. Forecast reliability rank histogram plots at nine grid locations, for a single run using the parameters: 100 particles, assimilation period $\Delta = 1/25$ ett, and 289 weather stations. Experiment period: 10 ett. Grid locations are shown above the plots. See section 4.2 for discussion.
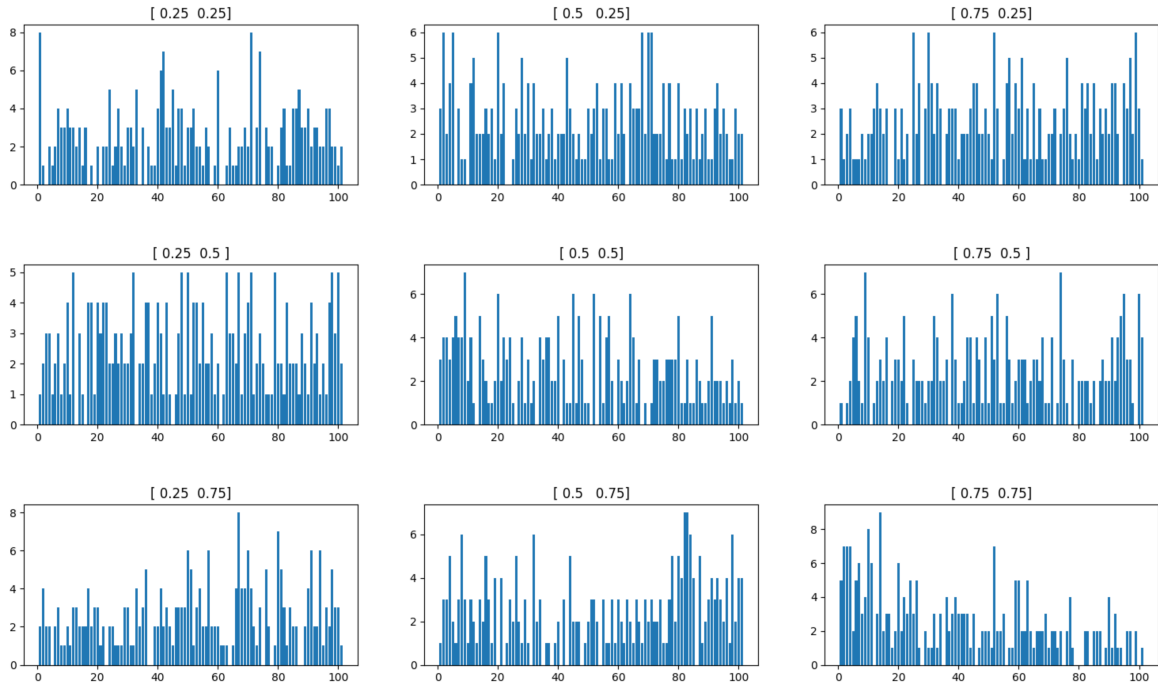
Figure 7: Perfect model scenario. Velocity x-component rank histogram plots at nine grid locations, for a single run using the parameters: 100 particles, assimilation period $\Delta = 1/25$ ett, and 81 weather stations. Experiment period: 10 ett. Grid locations are indicated above the plots. See section 4.2 for discussion.

to the equilibrium distribution of the posterior distribution versus the prior. However, over the intermediate long term (pre step 500, $\sim 20$ ett) though, the results show the particle filter working well, bearing in mind that we are controlling the system using information from only 1.18% of the total degrees of freedom, see remark 4. We see that assimilating data slows down the convergence to equilibrium.

**Remark 6.** *For the approximate asymptotic results shown in figure 8 we compute the prior ensemble using only the SPDE solver, starting from the same initial ensemble used for the particle filter. The ess values for the prior ensemble is computed using the Gaussian likelihood function (21) at each assimilation time. This is done to gauge the "quality" of the prior ensemble members against observation data at each assimilation step.*

From the results we see that the filter's ensemble has much higher ess values than the prior's ensemble. However, the ess value for the filter's ensemble is not maintained after 400 steps. This could be because of a number of issues. Not enough observation data is being assimilated, i.e. $d_y$ is too small. We have seen from the results shown in figure 5 how a small increase from 81 weather stations to 289 is able to provide a big improvement in terms of error reduction. Additional techniques can be introduced to the particle filter, such as nudging, space-time particle filter and localisation. One of the aims of this paper is to provide a working benchmark algorithm for high dimensional data assimilation using SALT parameterisation. With additional techniques, we aim to improve upon the results we have shown. Further, impact of numerical errors have so far been ignored and omitted. When looking at the asymptotic behaviour, we cannot ignore the impact of numerical errors. The SALT formulation is special in that the stochastic term preserves the dynamics' underlying geometric structure. In the non ideal setup (i.e. with forcing and damping) considered in this paper, the important Casimir values

which define the coadjoint orbits[7] of the system are no longer preserved. However, it is not difficult to show that the Casimirs in our setup satisfies a noise independent equation. Since the methodology used in the generation of the initial ensemble is Casimir preserving, this means that all the particles in theory should have the same Casimir values throughout.

As supplement to the long term results shown in figure 8, we show in figures 9 and 10 the Eulerian trajectories of the velocity x-component at four randomly sampled spatial locations. The results together show the empirical mean and the ensemble spread of the prior stablise beyond 500 data assimilation steps, which suggest the time for the SPDE to reach its invariant measure. Figure 9 shows that the filter lost track of the truth at certain Eulerian locations. Comparing figure 9 with figure 8, the size of the realised observation errors further suggests we do not observe enough data, i.e. the observed data are not informative enough asymptotically. However increasing the number of weather stations will require much more computational effort.

## 4.3   Imperfect model scenario

We repeated the numerical experiments in section 4.2 for the fine resolution PDE truth. As discussed in the beginning of section 4, in this scenario the observations come from the coarse grained PDE solutions.

The same particle filter parameter sets from section 4.2 were used. In this scenario observation data from 289 weather stations correspond to only 0.018% of the truth state space. To have the same observation to state space dimension ratio of 1.18% as in the perfect model scenario, the observation state space dimension in this scenario would need to be $d_y = 36992$. To have comprehensive numerical results using this many weather stations is not feasible given our computational resources.

Figure 11 shows the results of mse, $\sigma_{N_p}$, number of tempering steps and ess for each parameter set, averaged over 20 independent repeats. The same PDE truth was used in all repeats, but for independent observation noises. As in figure 5 for the perfect model scenario, for parameter sets 1 and 2 there are 250 data assimilation steps, for parameter set 3 there are only 50 data assimilation steps stretched over the same experiment period of 10 ett.

All experiments started with the same initial ensemble generated using the deformation procedure described in section 4.1, where $\omega_{\text{truth}}$ in (53) is the imperfect model scenario PDE truth vorticity field at the initial time $t_0$. The uncertainty quantification results from Cotter et al. [2019] indicate that the initial ensemble covers the coarse grained truth. See table 1 for a summary of the statistics for the initial ensemble. Note that the ess value of 83.83% is computed using 81 weather stations. For $d_y = 289$ the initial ess drops to around 60% see figure 11. Please note that the initial ess value is only used as a reference, we do not assimilate at $t_0$. All ensemble members thus have equal weights at $t_0$.

It is to be expected that the results would not be comparable to those from the previous section. In addition to the sources of errors discussed in section 4.2, in this scenario the truth is from a different dynamical system (PDE) to the signal process (SPDE), see the discussion around (18). Additional degrees of freedom are thus introduced into the error of the particle filter algorithm, see Clark and Crisan [2005]. Further, using 81 weather stations (parameter set 1) amounts to observing 0.00515% of the truth state space in this scenario. This percentage is improved to 0.01837% when using 289 weather stations in parameter sets 2 and 3. In either case,

---

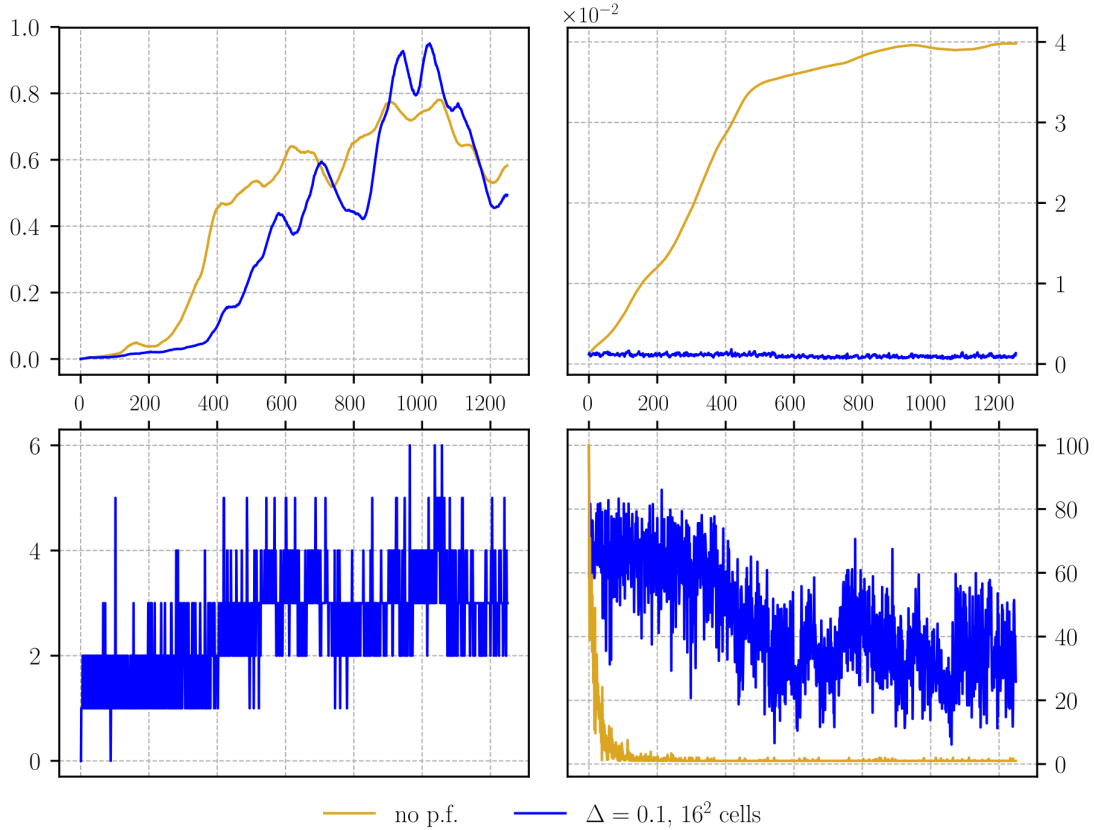[7]See Marsden and Ratiu [2013] for an exposition of geometric mechanics.

Figure 8: Perfect model scenario, approximate asymptotic diagnostics. The four figures correspond to the mean square error (tl:top-left), ensemble standard error (tr: top-right), number of tempering steps (bl:bottom-left) and ess (br: bottom-right). The x-axis shows the number of data assimilation steps corresponding to the assimilation interval $\Delta = 1/25$ ett (0.1 time units). The blue plots correspond to the mean of $\pi_t^N$, with ensemble size 100, $\Delta = 1/25$ ett and 289 weather stations. Experiment period is 50 ett. The gold plots correspond to the mean of the empirical prior distribution. Both sets of ensemble start from the same initial ensemble. The prior distribution does not assimilate data. The bl subfigure thus contains only the blue plot. However, ess values are computed for the prior distribution ensemble for reference. No observation data was used in the ess calculations at the initial time $t_0$ which is why the initial ensemble ess is showing 100%. See section 4.2 for discussion of these results.
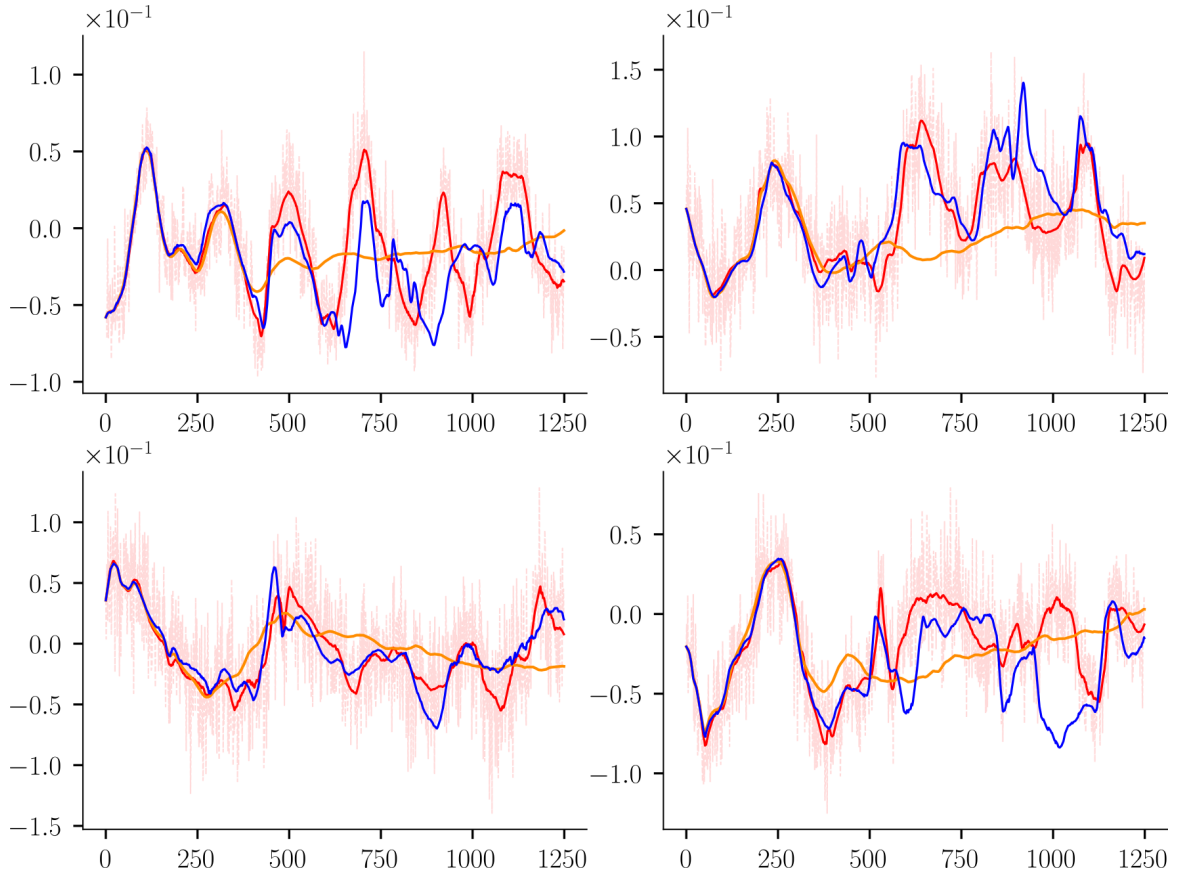
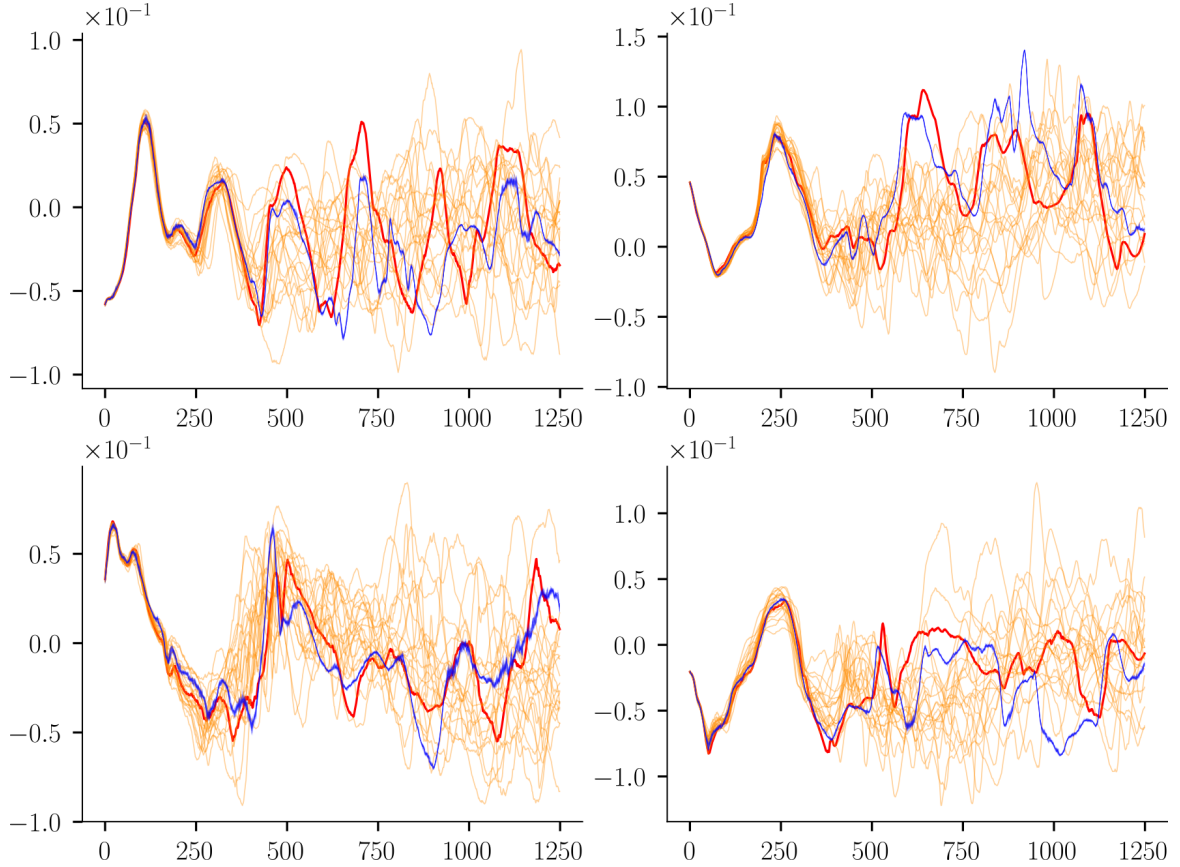Figure 9: Perfect model scenario, approximate asymptotic trajectory at four randomly sampled Eulerian grid locations from a single experiment run. The solid red plots correspond to the truth. The pink regions correspond to the realised observation values, i.e. truth plus realised observation noise. The blue plots are of the mean corresponding to $\pi_t^N$, with 100 ensemble members, $\Delta = 1/25$ ett and 289 weather stations. Experiment period is 50 ett, totalling 1250 data assimilation steps. The gold plots are of the mean corresponding to the empirical prior distribution. Both sets of ensemble start from the same initial ensemble. See section 4.2 for discussion.

Figure 10: Perfect model scenario, approximate asymptotic trajectory at four randomly sampled Eulerian grid locations from a single experiment run. The solid red plots correspond to the truth. The pink regions correspond to the realised observation values, i.e. truth plus realised observation noise. The blue plots are of the mean corresponding to $\pi_t^N$, using 100 ensemble members, $\Delta = 1/25$ ett and 289 weather stations. Experiment period is 50 ett, totalling 1250 data assimilation steps. The gold plots are the trajectories of 25 empirical prior distribution ensemble members, out of an ensemble of size 100. Both sets of ensemble start from the same initial ensemble. See section 4.2 for discussion.
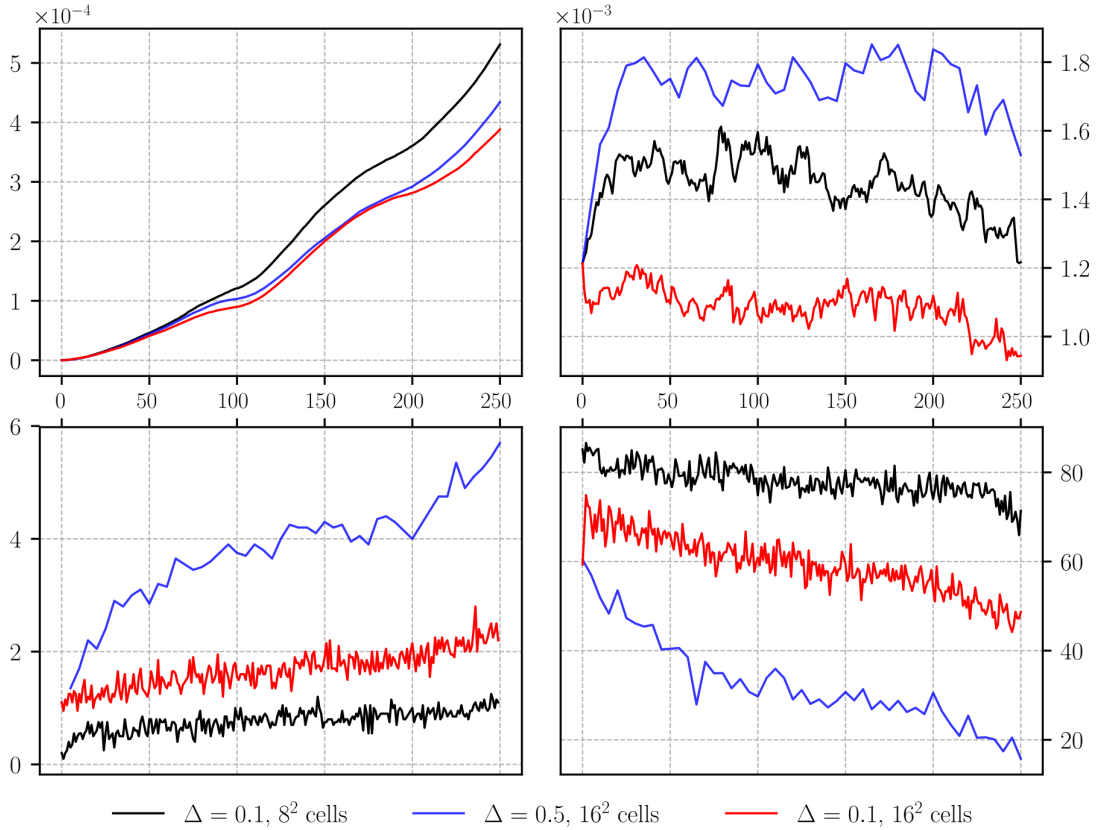
Figure 11: Imperfect model scenario. The four figures show the mean square error (*tl*: top-left), ensemble standard error (*tr*: top-right), number of tempering steps (*bl*: bottom-left) and ess (*br*: bottom-right), averaged over 20 independent experiment runs, conditioned on the same truth. However we did not condition on the same observation noise. The x-axis shows the number of data assimilation steps corresponding to the assimilation interval $\Delta = 1/25$ ett (0.1 time units). In all the experiments the ensemble size is 100. Three plots are shown in each figure. The black lines correspond to particle filter parameter choices of $\Delta = 1/25$ *ett* and 81 weather stations. The red lines correspond to $\Delta = 1/25$ *ett* and 289 weather stations. The blue lines correspond to $\Delta = 1/5$ *ett* and 289 weather stations. For the black and red lines, the parameter choices mean a total of 250 data assimilation steps. For the blue lines, it means a total of 50 data assimilation steps. See section 4.3 for discussion.

we have two orders of magnitude less information about the truth state than the experiments in section 4.2.

The features in the results from the perfect model scenario are observed here too. In figure 11, the results show increasing the observation state space dimension led to improvement in mse. A reduction of about 25% is observed when comparing parameter set 1 mse to parameter set 2. Parameter set 3 results show that if we assimilate more information but less frequently, the improvement in mse is less substantial, and the ensemble spread is larger. As the performance of the filter deteriorates over time, the tempering procedure introduced more intermediate tempering steps. However the algorithm was not able to maintain the initial ess levels. The ess values were better maintained for parameter sets 1 and 2. For parameter set 3 the ess started at about 60% and ended at below 20%, suggesting the filter lost track of the truth. Without using additional techniques such as nudging, the results highlight the need to assimilate enough information about the truth as often as possible in order to have reliable results from the filter. Using the results here as a basis, we will investigate the capability of additional techniques in future work.

Figures 12a and 12b show the rank histogram plots at nine reference grid locations of the velocity x-component and y-component respectively for one particular experiment run using 100 particles and parameter set 2. There are evidence suggesting bias, and under-dispersion at some of the plotted grid locations. The problem becomes more prominent if we use even fewer observations (parameter set 1). Figure 13 shows the velocity x-component rank histogram for an experiment using parameter set 1. Strong evidence for bias is observed at grid location [0.75, 0.75].

## Estimated asymptotic results and discussion

Since we assimilated so little information about the truth state at each step in this scenario, the asymptotic behaviour of the particle filter is also estimated in order to understand the stability of the mse. We ran an experiment using 100 particles and parameter set 2 for 50 eddy turnover times. This length of period is equivalent to 50000 fine resolution time steps, and for the filter it amounts to a total of 1250 data assimilation steps. The results are shown in figure 14. As in section 4.2, we have compared the results produced by the filter with that of the $N$-particle approximation of the prior distribution, i.e. $\mathbb{P}(X_t)$ without any data assimilation.

The results in figure 14 show that the estimated asymptotic mse is bounded. However over this extremely long period of 50 ett, the long term prior ensemble mean gives better mse results than the particle filter posterior ensemble mean. Over the intermediate long term (pre step 500, $\sim 20ett$) the filters results show good promise. Bearing in mind that we are controlling the system using information from only 0.018% of the total degrees of freedom. The ess results suggest that the posterior ensemble lost track of the truth beyond step 500. But the quality of the posterior ensemble, measured by the ess statistic, is still better than the prior ensemble.

The discussion about the asymptotic results in section 4.2 applies here too. This scenario is further complicated by the presence of model error. One could also aim to base the choice for the $\boldsymbol{\xi}$ parameters on the data assimilation results, but this is beyond the scope of this paper. In the bottom left subfigure in figure 15 we see an example of an Eulerian location at which the truth moves outside the support of the empirical prior. Thus this imperfect model test case really is pushing the algorithm beyond its theoretical limit.

As supplement to the estimated asymptotic results shown in figure 14, we show in figures 15
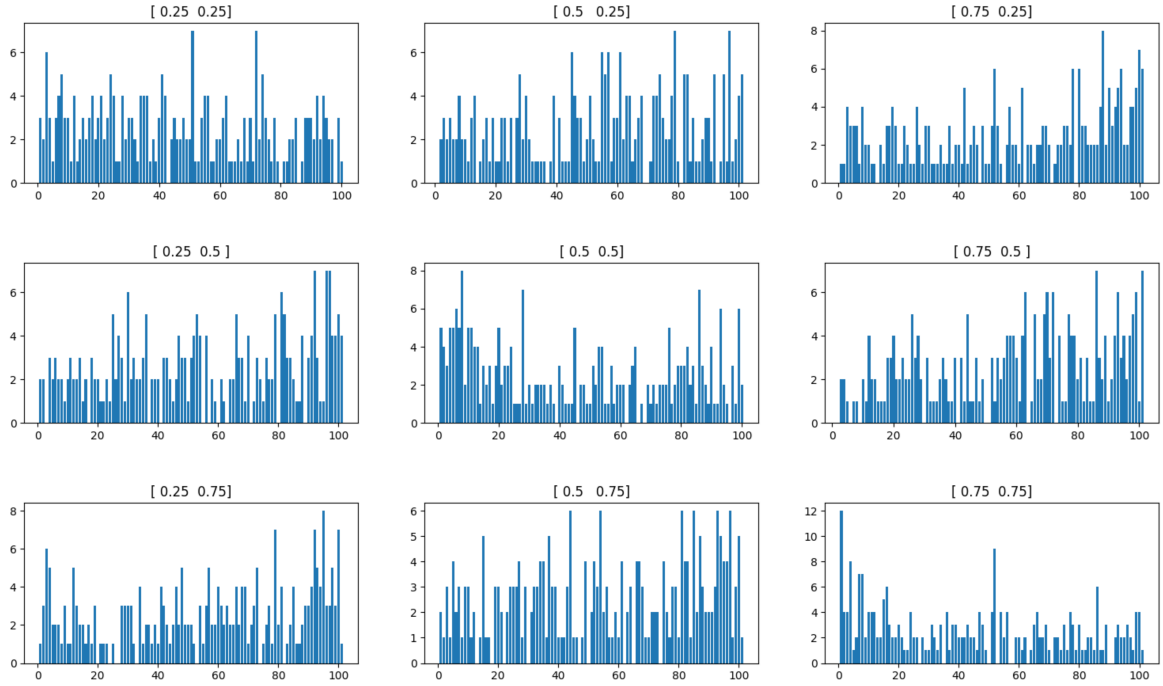
(a) velocity x-component
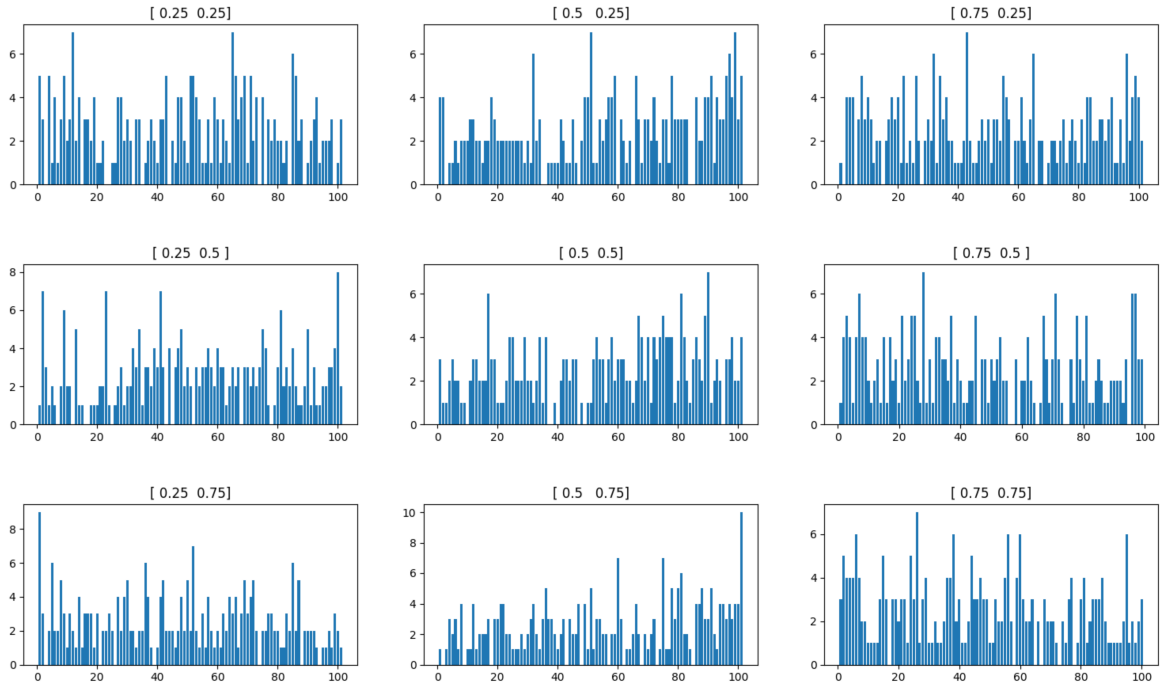


(b) velocity y-component

Figure 12: Imperfect model scenario. Forecast reliability rank histogram plots at nine grid locations, for a single run using the parameters: 100 particles, assimilation period $\Delta = 1/25$ ett, and 289 weather stations. Experiment period: 10 ett. Grid locations are shown above the plots. See discussion in section 4.3.
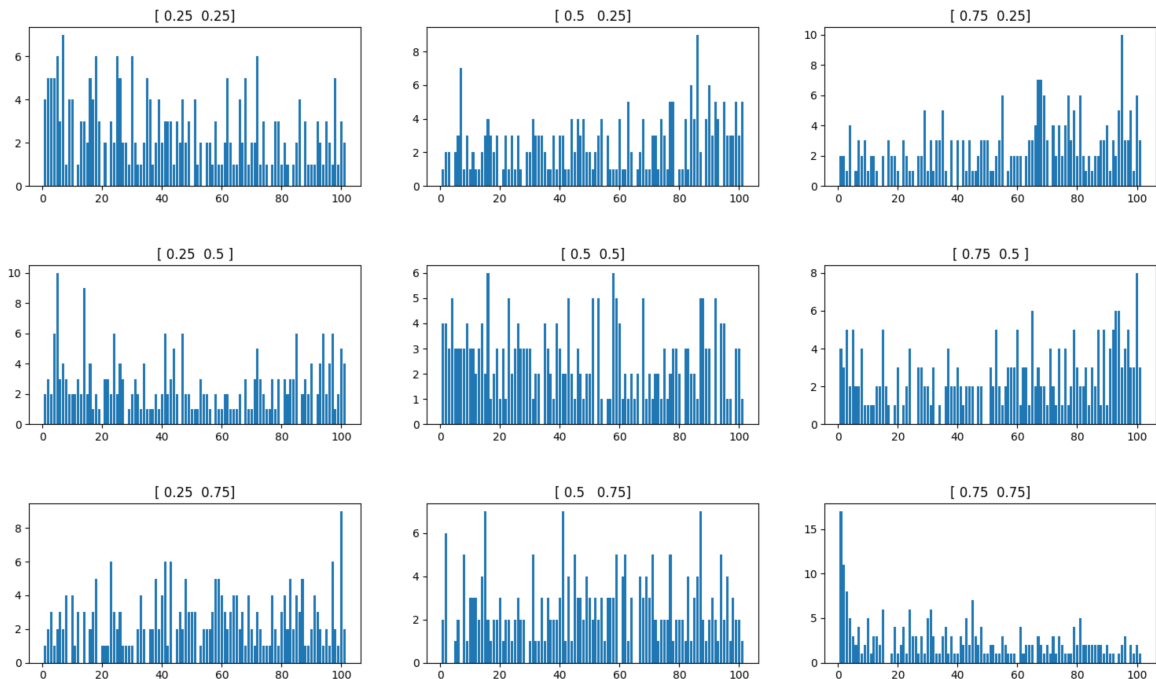
Figure 13: Imperfect model scenario. Velocity x-component rank histogram plots at nine grid locations, for a single run using the parameters: 100 particles, assimilation period $\Delta = 1/25$ ett and 81 weather stations. Experiment period 10 ett. Grid locations are indicated above the plots. See discussion in section 4.3.

and 16 the Eulerian trajectories of the velocity x-component at four randomly chosen spatial locations. In three of the subfigures we see that the ensemble mean of the filter lost track of the truth at around step 750, but were able to recover afterwards. This time point coincides with the peak mse error shown in figure 14. The time point also coincides with the time point in figure 16 at which the truth moves outside the uncertainty region defined by the prior ensemble.

# 5 Conclusion

In this work we used a particle filter which included three additional procedures (model reduction, tempering and jittering) in a high dimensional data assimilation (DA) case study. We interpreted the task as solving a filtering problem with a continuous time signal via discrete observations. The "truth" was modelled by a highly resolved numerical solution of a damped and forced incompressible 2D Euler equation which had roughly $3.1 \times 10^6$ degrees of freedom. The data consisted of a time series of 250 discrete observations of the fluid velocity measured on a sparse spatial grid which varied in size from 81 to 289. The model reduction involved the addition of a stochastic parametrisation of the above equation solved on a coarser grid of about $4.9 \times 10^4$ degrees of freedom. We showed that the data assimilation can be successfully carried out using modest computational hardware by employing all of the above additional procedures to a system of 100 particles. The resulting approximations compared favourably with the true trajectory. We also tested the reliability of the resulting forecast with no bias, under dispersion or over dispersion observed.

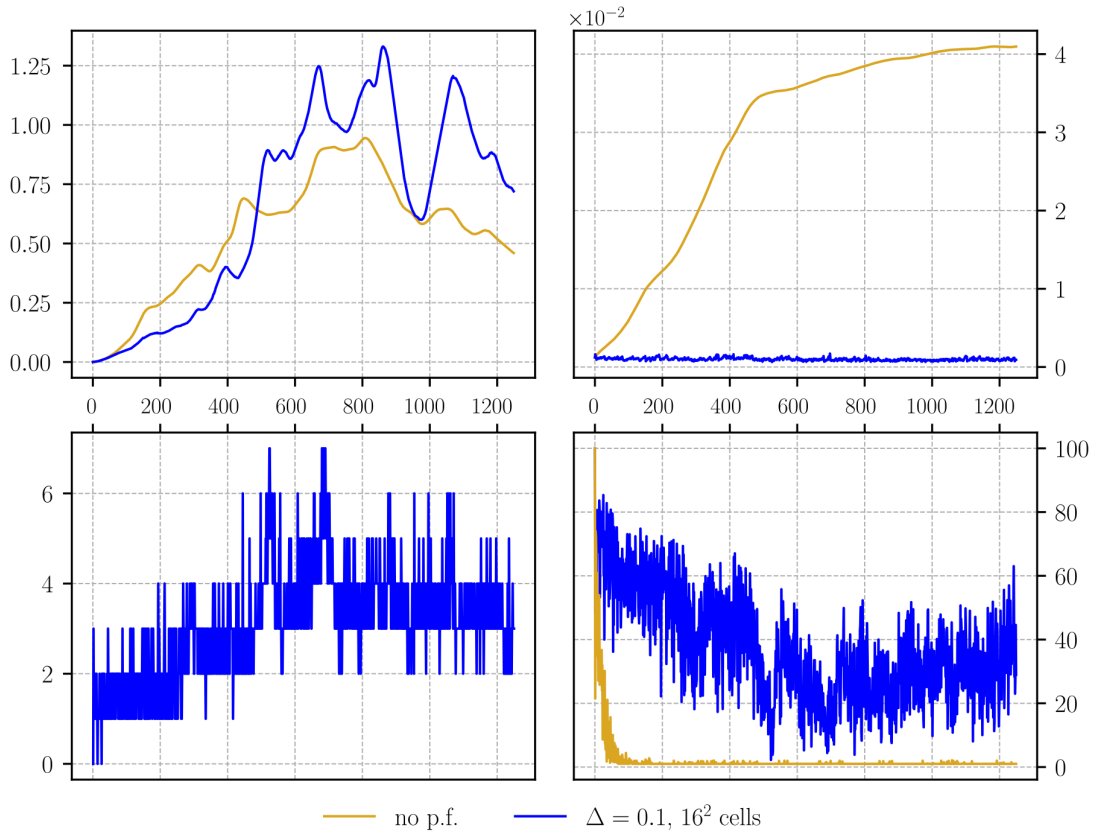In a sequel to this work we aim to incorporate additional procedures (nudging, space-time

Figure 14: Imperfect model scenario, approximate asymptotic diagnostics. The four figures correspond to the mean square error (tl:top-left), ensemble standard error (tr: top-right), number of tempering steps (bl:bottom-left) and ess (br: bottom-right). The x-axis shows the number of data assimilation steps corresponding to the assimilation interval $\Delta = 1/25$ ett (0.1 time units). The blue plots correspond to the mean of $\pi_t^N$, with an ensemble of size 100, $\Delta = 1/25$ ett and 289 weather stations. Experiment period is 50 ett. The gold plots correspond to the mean of the empirical prior distribution. Both sets of ensemble start from the same initial ensemble. The prior distribution does not assimilate data. The bl subfigure thus contains only the blue plot. However, ess values are computed for the prior distribution ensemble for reference. See section 4.3 for discussion.
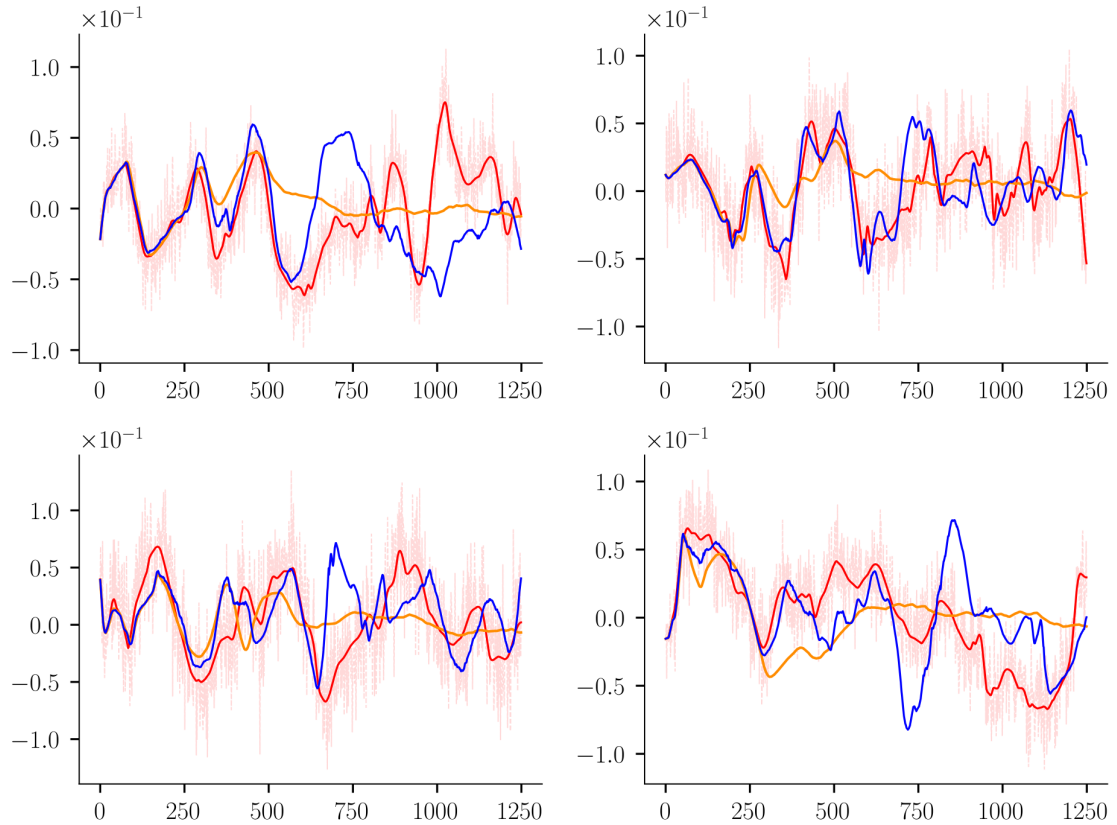
Figure 15: Imperfect model scenario, approximate asymptotic trajectory at four randomly sampled Eulerian grid locations from a single experiment run. The solid red plots correspond to the truth. The pink regions correspond to the realised observation values, i.e. truth plus realised observation noise. The blue plots are of the mean corresponding to $\pi_t^N$, using 100 ensemble members, $\Delta = 1/25$ ett and 289 weather stations. Experiment period is 50 ett, totalling 1250 data assimilation steps. The gold plots are of the mean corresponding to the empirical prior distribution. Both sets of ensemble start from the same initial ensemble. See section 4.3 for discussion.
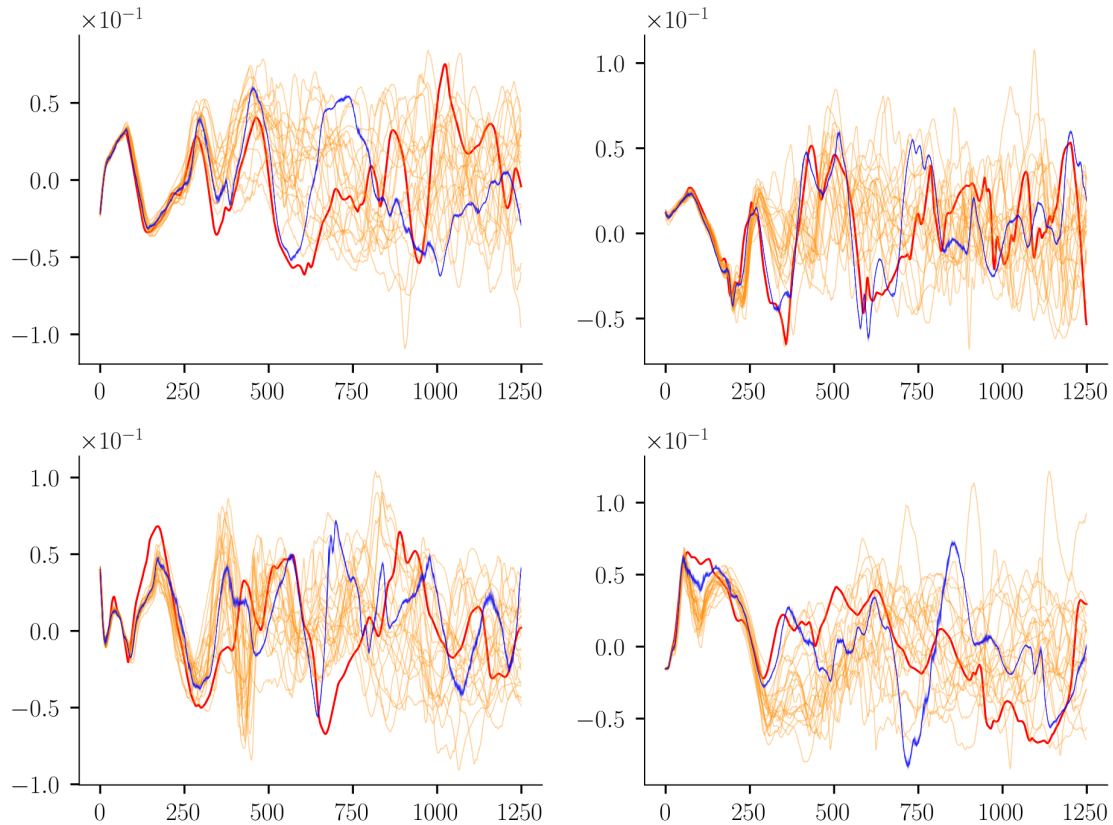
Figure 16: imperfect model scenario, approximate asymptotic trajectory at four randomly sampled Eulerian grid locations from a single experiment run. The solid red plots correspond to the truth. The pink regions correspond to the realised observation values, i.e. truth plus realised observation noise. The blue plots are of the mean corresponding to $\pi_t^N$, using 100 ensemble members, $\Delta = 1/25$ ett and 289 weather stations. Experiment period is 50 ett, totalling 1250 data assimilation steps. The gold plots are the trajectories of 25 prior distribution ensemble members, out of an ensemble of size 100. Both sets of ensemble start from the same initial ensemble. See section 4.3 for discussion.

data assimilation) to the ones discussed here to further refine the performance of the particle filter discussed here.

**Code and data availability**

All Python implementation code and experiment data files are available from the corresponding author upon request.

# Acknowledgements

# References

Alan Bain and Dan Crisan. *Fundamentals of stochastic filtering*, volume 3. Springer, 2009.

Alexandros Beskos, Dan O Crisan, Ajay Jasra, and Nick Whiteley. Error bounds and normalising constants for sequential Monte Carlo samplers in high dimensions. *Advances in Applied Probability*, 46(1):279–306, 2014.

Alexandros Beskos, Dan Crisan, Ajay Jasra, Kengo Kamatani, and Yan Zhou. A stable particle filter for a class of high-dimensional state-space models. *Advances in Applied Probability*, 49 (1):24–48, 2017.

Jochen Broecker. Assessing the reliability of ensemble forecasting systems under serial dependence. *Quarterly Journal of the Royal Meteorological Society*, 144(717):2666–2675, October 2018. doi: 10.1002/qj.3379. URL `http://centaur.reading.ac.uk/78201/`.

Roberto Buizza, M Milleer, and Tim N Palmer. Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, 125(560):2887–2908, 1999.

S S Chern, W H Chen, and K S Lam. *Lectures on Differential Geometry*. WORLD SCIENTIFIC, 1999. doi: 10.1142/3812. URL `https://www.worldscientific.com/doi/abs/10.1142/3812`.

J.M.C. Clark and D. Crisan. On a robust version of the integral representation formula of nonlinear filtering. *Probability Theory and Related Fields*, 133(1):43–56, Sep 2005. ISSN 1432-2064. doi: 10.1007/s00440-004-0412-5. URL `https://doi.org/10.1007/s00440-004-0412-5`.

C. Cotter, D. Crisan, D. Holm, W. Pan, and I. Shevchenko. Numerically modeling stochastic Lie transport in fluid dynamics. *Multiscale Modeling & Simulation*, 17(1):192–232, 2019. doi: 10.1137/18M1167929. URL `https://doi.org/10.1137/18M1167929`.

C. J. Cotter, G. A. Gottwald, and D. D. Holm. Stochastic partial differential fluid equations as a diffusive limit of deterministic Lagrangian multi-time dynamics. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 473(2205):20170388, 2017. doi: 10.1098/rspa.2017.0388. URL `https://royalsocietypublishing.org/doi/abs/10.1098/rspa.2017.0388`.

Colin Cotter, Dan Crisan, Darryl D Holm, Wei Pan, and Igor Shevchenko. Modelling uncertainty using circulation-preserving stochastic transport noise in a 2-layer quasi-geostrophic model. *arXiv preprint arXiv:1802.05711*, 2018.

Dan Crisan and Oana Lang. Well-posedness for a stochastic 2D Euler equation with transport noise. *arXiv preprint arXiv:1907.00451*, 2019.

Dan Crisan, Franco Flandoli, and Darryl D. Holm. Solution properties of a 3D stochastic Euler fluid equation. *Journal of Nonlinear Science*, Oct 2018. ISSN 1432-1467. doi: 10.1007/s00332-018-9506-6. URL `https://doi.org/10.1007/s00332-018-9506-6`.

Masoumeh Dashti and Andrew M Stuart. The Bayesian approach to inverse problems. *Handbook of Uncertainty Quantification*, pages 311–428, 2017.

Geir Evensen. *Data assimilation: the ensemble Kalman filter.* Springer Science & Business Media, 2009.

François Gay-Balmaz and Darryl D Holm. Selective decay by Casimir dissipation in inviscid fluids. *Nonlinearity*, 26(2):495, 2013.

Darryl D. Holm. Variational principles for stochastic fluid dynamics. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 471 (2176), 2015. ISSN 1364-5021. doi: 10.1098/rspa.2014.0963. URL http://rspa.royalsocietypublishing.org/content/471/2176/20140963.

Nikolas Kantas, Alexandros Beskos, and Ajay Jasra. Sequential Monte Carlo methods for high-dimensional inverse problems: A case study for the Navier–Stokes equations. *SIAM/ASA Journal on Uncertainty Quantification*, 2(1):464–489, 2014.

Lennart Ljung. Asymptotic behavior of the extended Kalman filter as a parameter estimator for linear systems. *IEEE Transactions on Automatic Control*, 24(1):36–50, 1979.

Andrew J Majda, Ilya Timofeyev, and Eric Vanden Eijnden. Models for stochastic climate prediction. *Proceedings of the National Academy of Sciences*, 96(26):14687–14691, 1999.

Andrew J Majda, Ilya Timofeyev, and Eric Vanden Eijnden. A mathematical framework for stochastic climate models. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 54(8):891–974, 2001.

Jerrold E Marsden and Tudor S Ratiu. *Introduction to mechanics and symmetry: a basic exposition of classical mechanical systems*, volume 17. Springer Science & Business Media, 2013.

E. Mémin. Fluid flow dynamics under location uncertainty. *Geophys. Astro. Fluid*, 108:119–146, 2014.

Peter Neal, Gareth Roberts, et al. Optimal scaling for partially updating MCMC algorithms. *The Annals of Applied Probability*, 16(2):475–515, 2006.

Radford M Neal. Annealed importance sampling. *Statistics and computing*, 11(2):125–139, 2001.

Tim Palmer. The ECMWF ensemble prediction system: Looking back (more than) 25 years and projecting forward 25 years. *Quarterly Journal of the Royal Meteorological Society*, 2018.

Sebastian Reich and Colin Cotter. *Probabilistic forecasting and Bayesian data assimilation.* Cambridge University Press, 2015.

Stuart, A M. Inverse problems: A Bayesian perspective. *Acta numerica.*, 19:451–559, 2010. ISSN 0962-4929.

Peter Jan van Leeuwen, Hans R. Künsch, Lars Nerger, Roland Potthast, and Sebastian Reich. Particle filters for high-dimensional geoscience applications: A review. *Quarterly Journal of the Royal Meteorological Society*, 2019. ISSN 1477870X. doi: 10.1002/qj.3551.