Twin experiments with the equivalent weights particle filter and HadCM3

Article

Accepted Version

Twin experiments with the equivalent weights particle filter and HadCM3

P.A Browne[a*] and P.J. van Leeuwen[a]

Department of Meteorology, University of Reading, UK.

*Correspondence to: Department of Meteorology, University of Reading, UK. E-mail: p.browne@reading.ac.uk

This paper investigates the use of a particle filter for data assimilation with a full scale coupled ocean-atmosphere general circulation model. Synthetic twin experiments are performed to assess the performance of the equivalent weights filter in such a high-dimensional system. Artificial 2-dimensional sea surface temperature fields are used as observational data every day. Results are presented for different values of the free parameters in the method. Measures of the performance of the filter are root mean square errors, trajectories of individual variables in the model and rank histograms. Filter degeneracy is not observed and the performance of the filter is shown to depend on the ability to keep maximum spread in the ensemble.

Key Words: Equivalent weights particle filter, HadCM3, climate model initialisation, nonlinear data assimilation, uncertainty quantification, model evolution error covariance matrix modelling, EMPIRE

Received . . .

1. Introduction

1.1. The forecasting problem

One of the major goals of modelling in the geosciences is to use numerical models to make predictions of the system in question. There are two major components to such a prediction: firstly, ensuring that the model is representative of the dynamics of the system. Secondly, the model must be initialised to capture the current state, from which the prediction can begin.

To initialise a climate model, observations alone are not enough. Observation networks do not currently (and will probably never) have the capability to observe all of the state variables at every single grid point, and so climate model initialisation is a classic under-determined problem. Data assimilation (DA) is the process by which the observations of the real system will update an initial guess of the state in order to initialise the model.

Bayes' Theorem is the mathematical description of the data assimilation problem. If x represents a model state, and y some observations of it, then

$$p(x \mid y) = \frac{p(x)p(y \mid x)}{p(y)}. \tag{1}$$

This states that the probability density function (PDF) of the state given some observations (the posterior PDF) can be written as the product of the PDF of the state itself (the prior) and the PDF of the observations given the state (the likelihood) normalised by the probability of the observations themselves.

Different data assimilation methods have varying approaches to finding a representation of the posterior PDF. For example, a variational DA method will simply try and find the mode of the posterior (see, for example, Cohn (1997)).

1.2. Particle filters and the posterior PDF

A particle filter is a fully nonlinear data assimilation method which seeks to represent the full posterior PDF. For an overview of particle filters, see for example, van Leeuwen (2009). They represent the prior PDF as an ensemble of delta functions, and propagate this using the dynamical model.

Each of these ensemble members, or particles, has an associated weight referring to its contribution in the posterior PDF, i.e.

$$p(x \mid y) \approx \sum_{i=1}^{m} w_i \delta(x - x_i) \tag{2}$$

where x_i and w_i are a pair denoting the ensemble member and its weight for each of the m ensemble members. The weights must satisfy

$$\sum_{i=1}^{m} w_i = 1, \tag{3}$$

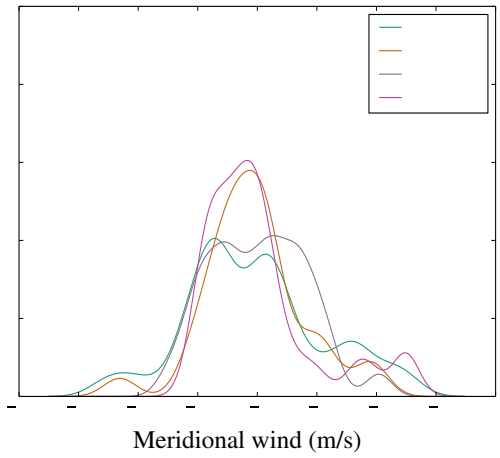
which leads to an issue when $w_j \approx 1$ for some particle $j \in 1, \ldots, m$. This situation is known as filter degeneracy and in such a case, the posterior PDF reduces to effectively being represented by the single ensemble member x_j. Hence the mean of $p(x) \approx x_j$ and all estimated higher order moments of the distribution will be close to 0.

It is well known that for a simple (no proposal density or resampling) particle filter, the number of ensemble members

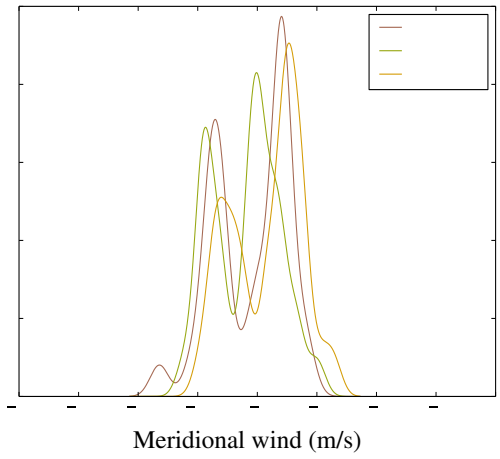
This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/qj.2621

needed to avoid filter degeneracy grows exponentially with the number of independent observations of the system used (Snyder et al. 2008; Bengtsson et al. 2008). This shows that for high dimensional systems, without using a more sophisticated method, these sizes of ensembles are not possible. In this context this drawback is referred to as the *curse of dimensionality*.

Other data assimilation methods make certain assumptions which make the representation of the posterior tractable. Famously, the Kalman filter results from assuming that the posterior PDF is Gaussian, and hence can be represented by its first two moments (Meinhold and Singpurwalla 1983). Since a climate model is strongly nonlinear due to several feedbacks it is unlikely that these Gaussian assumptions are valid in the initialisation of a climate model. For example, Figure 1 shows a reconstruction of the marginal PDF of eastward surface winds in the mid atlantic. In the left panel it can be seen that in the early states of the model evolution the PDF is unimodal and the Gaussian approximation would suffice. However, the panel on the right shows the PDF of the same variable after the model has evolved further in time. These marginal PDFs appear to be bimodal and hence exploring data assimilation methods that do not assume Gaussian PDFs is of interest.



(a) Gaussian-like PDFs in first 20 days of model run



(b) Non-Gaussian-like PDFs in later stages of model run

Figure 1. PDFs of eastward surface wind variable at (1.25N,189.375E)

In this paper we shall consider the equivalent weights particle filter (van Leeuwen 2010; Ades and van Leeuwen 2013, 2015) which makes use of "clever" proposal densities and resampling in order to avoid the curse of dimensionality. The advantages of using such a particle filter scheme include the lack of assumptions about the posterior PDF, its fully nonlinear nature, and the fact that, if it can still avoid filter degeneracy in a system of the size we consider here, it will quantify the uncertainty in the posterior PDF and hence a subsequent forecast.

1.3. Current climate model initialisation strategies

An excellent overview of the initialisation techniques for climate models used for seasonal to decadal prediction was given by Meehl et al. (2014), in relation to those models used as part of CMIP5. Many of the models considered are simply nudged towards some reanalysis product. Others are initialised with a form of 3DVAR in the ocean. In a few cases, the Ensemble Kalman Filter (EnKF) and Ensemble Optimal Interpolation (EnOI) are used. Many of the models are initialised using different methods for the atmosphere and the ocean.

When a variational method is used, only a mode of the PDF is found and so there is very limited information about the uncertainty in the solution. 4DVar applied to a coupled model with components having different timescales has difficulties as it is unclear how long the assimilation window should be. For example, the accuracy of the tangent linear and adjoint models will be much worse for a faster, more nonlinear component when applied over the same period as a slower component in the coupled model. This is the case in a coupled ocean-atmosphere model where the atmosphere evolves substantially quicker than the ocean, and is the subject of much current research (eg (Smith et al. 2015)).

The EnKF methods assume that the posterior can be classified by the mean and covariance only, i.e. the posterior is assumed to be Gaussian. This may not be the case and so it is worth investigating whether this is a valid assumption. Use of the EnKF in a high dimensional system typically requires localisation and inflation of error covariance matrices (see for example (Hamill et al. 2001)), something that is not required in the equivalent weights particle filter which we investigate in this paper. Many researchers are attempting to address these complications with the EnKF, for example Bocquet (2011) has developed a version of the EnKF which does not use inflation.

In this paper we consider the model a fully coupled system. In this sense, we do not treat the ocean and atmosphere separately. Hence we use covariance information from one to influence the other. This is useful in guarding against having unbalanced initial states for the forecast, and hence should reduce the chance of encountering initialisation shock (Chen and Zebiak 1997; Balmaseda et al. 2009).

There has been some work done to apply particle filters to climate models in the field of paleoclimatology. Dubinkina et al. (2011), (Goosse et al. 2012a) and (Goosse et al. 2012b) have applied the Sequential Importance Resampling (SIR) particle filter to LOVECLIM, a more simple climate model than HadCM3 which we shall introduce in Section 2. Using a particle filter on a system the size of LOVECLIM one might expect to encounter filter degeneracy, however as Ades and van Leeuwen (2013) showed, filter degeneracy is a consequence of the number of independent observations, not the size of the state vector. In the experiments with LOVECLIM, the number of independent observations used were small, due to considering spatial averages over very large areas. Apart from this, the weights of the particles had a specified minimal value to avoid filter degeneracy. The negative result of this procedure is that not all information is extracted from the observations. Recently Dubinkina and Goosse (2013) have applied applied a nudging proposal particle filter to

LOVECLIM, i.e. simply not using the weight equalising step of the equivalent weights particle filter (see (8)), with the justification that filter degeneracy does not occur in the low dimensional system. In this paper the size of the observation vector, $27,370$, is much too large to apply the filters above.

1.4. Paper structure

The rest of this paper is organised as follows. In section 2 we discuss the properties of the coupled climate model we use in this study. In section 3 we give a brief description of the particle filter method we consider. In section 4 we describe how we generate two different model evolution error covariance matrices for use in subsequent experiments. section 5 describes the technical details of the setup of the experiments. section 6 gives the results from various assimilation experiments. Finally in section 7 we finish with some conclusions about the efficacy of applying such a particle filter to initialise a coupled climate model.

2. HadCM3

HadCM3 is a coupled ocean-atmosphere general circulation model which has been extensively used in IPCC reports (Gordon et al. 2000; Solomon 2007; Stocker et al. 2013) and does not require flux adjustments (Collins et al. 2001). The atmosphere component has 5 prognostic variables: surface pressure, zonal and meridional velocities, potential temperature and humidity. These are stored on a $3.75^\circ \times 2.5^\circ$ staggered B-grid with 19 vertical levels. The ocean component has 4 distinct prognostic variables: zonal and meridional velocities, temperature and salinity. These are stored on a $1.25^\circ \times 1.25^\circ$ staggered B-grid with 20 vertical levels, horizontally aligned with the atmospheric grid.

The total number of state variables in the model is $2,314,430$. The model's timestepping scheme works on a 24-hour cycle. Firstly it integrates the atmosphere over the 24-hour period with a timestep of 30 minutes. Then the ocean is integrated over the same 24-hour period with a timestep of 60 minutes. The coupling between the ocean and atmosphere occurs once every 24 hours, or equivalently every 72 model timesteps.

We consider the climate model HadCM3 to act as the function f in the equation

$$x^{k+1} = f(x^k) + \beta^k, \tag{4}$$

where x^k represents the model state at time k. β^k is a stochastic term which represents missing or incorrect physics in the model, the distribution of which we will have to specify (see Section 4). We denote the covariance of β^k by Q. If β^k were not present, and the model were purely deterministic, this would restrict the ability to modify the state of the system and hence would limit the data assimilation methods we wish to use. Specifically, β^k is required in order to satisfy the conditions to use the equivalent weights particle filter.

3. The equivalent weights particle filter

In order to describe a method in which the ensemble member in a particle filter can be directed in state space, the concept of a *proposal density* must be introduced. Rewriting Bayes' theorem (1) as

$$p(x \mid y) = \frac{p(x)p(y \mid x)q(x,y)}{p(y)q(x,y)}, \tag{5}$$

where $q(x,y)$ is the proposal density, is possible so long as $q(x,y)$ has support larger than $p(x)$. The key concept is that the proposal density can be a function not just of the model state x, but also (future) observations y. In this sense the ensemble members can be directed towards the future observations by using a proposal density which modifies the increment of the deterministic model to achieve this goal. This is the basis of particle filters such as the Implicit Particle Filter (Chorin and Tu 2009) and the equivalent weights particle filter. When the weights, w_i, in a particle filter are calculated, they can be decomposed into those coming from the prior ($p(x)$), the likeihood ($p(y \mid x)$) and the proposal density $q(x,y)$. The component of the weight associated with the proposal density is weighted by a term involving the distribution of the model error term, β. If $\beta \equiv 0$ then these weights would become infinite, thus there would be no freedom to direct the particles anywhere other than to where the deterministic model suggests.

The equivalent weights particle filter is a fully nonlinear DA method which works in a two-stage process which we describe subsequently. For a comprehensive overview of the equivalent weights particle filter see van Leeuwen (2010) and Ades and van Leeuwen (2013). Suppose that an observation occurs at timestep n. Then for each model timestep k before an observation occurs ($k \in \{0, \ldots, n-2\}$), the model state of each ensemble member, x_i^k, is updated via the equation

$$x_i^{k+1} = f(x_i^k) + A(y^n - H(x_i^k)) + \beta_i^k \tag{6}$$

where β^k is a random vector, y^n is the next observation in time, H is the observation operator that maps the model state onto observation space with associated observation error covariance matrix R and A is a relaxation term. In this work we consider

$$A = pQH^T R^{-1} \tag{7}$$

where the matrices Q and R correspond to the model evolution error covariance and observation error covariance matrices respectively. In this paper we consider p constant, however it is possible to make p a function of the time between observations.

The second stage of the equivalent weights filter involves updating each ensemble member at the observation time n via the formula

$$x_i^n = f(x_i^{n-1}) + \alpha_i QH^T(HQH^T + R)^{-1}(y^n - H(f(x_i^{n-1}))) + \beta_i^n \tag{8}$$

where α_i are scalars computed so as to make the weights of the particles equal. This is done for a given proportion ($0 < \kappa \leq 1$) of the ensemble which can make the desired weight. After this equivalent weights step all ensemble members are resampled from the members with the desired weight using stochastic universal sampling (Baker 1987).

It is important to realise that the covariance of the model state plays no role in the equivalent weights particle filter (apart from the beginning of the experiments, but that is forgotten quickly). Instead, the covariance of the error in the model evolution Q is crucial.

4. Specification of model evolution error covariance matrices

We must specify a model evolution error covariance matrix Q to use within the equivalent weights particle filter. As in this work we are considering twin experiments, it is safe to assume that there are no biases in the model, and hence we can take the mean of β^k to

be 0. Now we consider the model evolution error to be Gaussian with zero mean and covariance Q so that

$$\beta^k \sim \mathcal{N}(0, Q). \tag{9}$$

In order to estimate Q, we consider a long model run and assume that the variance in the model evolution error will be related to the variance in the model itself. Considering the daily variance in the model over a 5 year control run, we can decompose the covariances in the model as follows.

$$\text{Cov}(x) = \Lambda^{1/2}\Sigma\Lambda^{1/2} \tag{10}$$

where Σ is a correlation matrix and Λ is a diagonal matrix of variances. Due to its large size, we only compute a localised version of $\text{Cov}(x)$, taking covariances of each variable within 2 gridpoints of each other. This has the effect of removing spurious long-range correlations that arise due to under-sampling and makes the calculation computationally feasible. Note that in this coupled system there are no known control variable transforms which allow efficient application of functions of covariance matrices as used in fields such as atmospheric data assimilation.

As we can only compute the highly localised entries in $\text{Cov}(x)$, we seek a method to increase the length-scales that will be present in the model evolution error covariance matrix. Repeatedly applying the dimensionless correlation matrix Σ has this effect.

Hence we model Q as

$$Q \propto \Lambda^{1/2}\Sigma^2\Lambda^{1/2}. \tag{11}$$

The short length-scales present in Q defined in this manner means that the information from observations will be passed over longer length-scales indirectly via the nudging term (7) and the propagation of these terms in the dynamical model (4).

Unfortunately very little previous work has been done in the field to quantify model errors. We have taken a pragmatic approach and ensured the model errors are large enough to influence the evolution of the system, but not too large to destroy the physics represented in the deterministic evolution. Hence the coefficient, or proportionality, is chosen so that the size of the random perturbations for each of the separate variables is, in the L_2-norm, less than 0.1 times that of the deterministic model update. We shall refer to this matrix as Q_1. Note that (11) does directly give the decomposition of Q into standard deviations and a correlation matrix, however such a decomposition is unnecessary for the computations. For efficiency, matrix vector multiplication of a vector by the matrix $Q^{1/2}$ has been implmented using the sparse BLAS library LIBRSB (Martone et al. 2010).

Note that there is very little experience in modelling Q. In this paper we shall test different variants of the model evolution error covariance matrix in order to investigate the sensitivity of the assimilation results to Q. The modelling of Q we have just described assumes that the variance of the errors in the model were related to the variances in the model. However the variances in the model are dominated by the seasonal cycle. It is unrealistic to suggest that the errors in the model also follow directly the seasonal cycle and so we have removed the seasonal cycle from the samples we used to determine it. This was done by subtracting a running mean from each sample, with the mean taken 7 days either side of the sample. This matrix we shall refer to as Q_2.

Figure 2 gives a representation of some of the correlations contained within Σ. For instance we see from Figure 2a and Figure 2b the effect of Ekman transport (see e.g. Price et al. (1987)), noting the change in sign at the equator. Figure 2c and Figure 2d captures the strong relationship between SSTs and near surface humidity, in that increasing the sea surface temperature permits more moisture to be transferred to the atmosphere. As seen in Figure 2d, the areas either side of the equator in the western pacific showing very little correlation may be due to a precipitation feedback mechanism present in HadCM3.

It can be seen from Figure 2 that the correlations contained within Q_1 and Q_2 have broadly the same structure, with those calculated without the seasonal cycle showing stronger correlations. However the variances within these 2 matrices have a somewhat different structure.

Figure 3 shows the unscaled standard deviations in $\Lambda^{\frac{1}{2}}$ that correspond to variability in surface air pressure, surface eastward winds and subsurface sea water temperatures respectively, for both Q_1 and Q_2. In the case of Q_2, Figure 3b showing the standard deviations of the surface pressure variables no longer has the large variance in the southern ocean that is seen with Q_1 in Figure 3a. Figure 3b retains the large variance over the northern Pacific as seen in Figure 3a but the variances over Europe are reduced.

The variances of the eastward winds shown in Figure 3d are not a maximum over the northern Atlantic and northern Pacific as in Figure 3c. Instead the maximum is in the northern Indian ocean and the maritime continent.

Comparing the variances in the ocean temperatures in Figures 3e and 3f we can see that they have broadly the same spatial distribution. However the scaling is very different. In Q_1 which has the seasonal cycle (Figure 3e), the Kuroshio current off the east coast of Japan is the strongest signal and dominates the variance that you can see. In Q_2 where we have removed the seasonal cycle (Figure 3f), the Kuroshio still remains the area of highest variability, but it is markedly less strong in comparison with the flow in the rest of the oceans.

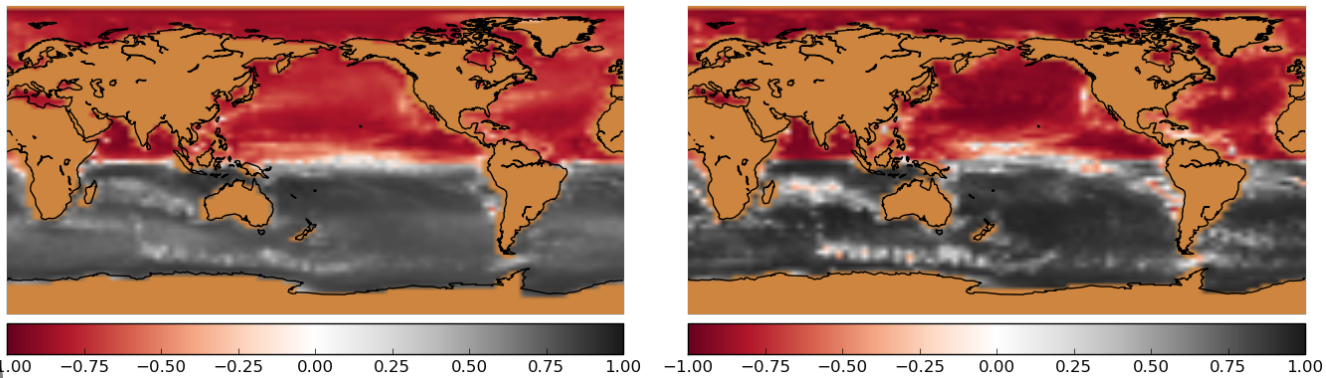
5. Experimental setup

5.1. *EMPIRE coupling*

To allow HadCM3 to be used within the a data assimilation system, the EMPIRE coupling has been used (Browne and Wilson 2015). This involves inserting a small number of MPI commands (Gropp et al. 1996) into the model in order to send the state vector to a different piece of software which acts to implement the equivalent weights particle filter while considering the model as a *"black-box"*. In doing so the climate model remains a separate executable from the data assimilation system.

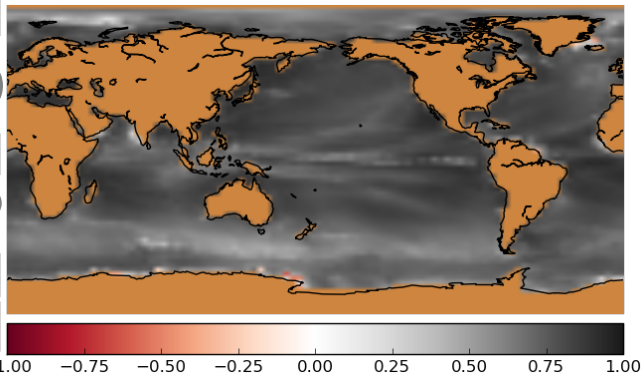
5.2. *Twin experiment truth run*

In order to perform an identical twin experiment, we must first run the model in order to obtain a truth which we wish to recover using the equivalent weights particle filter. For statistical consistency the model is evolved stochastically according to (4). That is, at each step (k) of the truth run, stochastic noise (β^k) is added to the deterministic model where $\beta^k \sim \mathcal{N}(0, Q)$. Synthetic observations from this truth run will be taken as defined in subsection 5.3 and used within the data assimilation method to try to constrain the system.

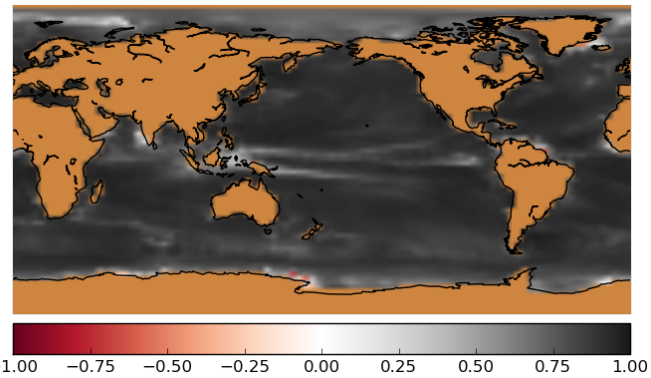


(a) Correlations between atmospheric U and ocean V for $Q = Q_1$

(b) Correlations between atmospheric U and ocean V for $Q = Q_2$



(c) Correlations between atmospheric humidity and ocean θ for $Q = Q_1$



(d) Correlations between atmospheric humidity and ocean θ for $Q = Q_2$

Figure 2. Representative point correlations between a variable in the atmosphere at the lowest level and the variable in the top layer of the ocean directly below it. Dark red regions correspond to strong negative correlations, white regions no correlation and dark grey corresponds to regions of strong positive correlations.

When assimilating real data, the model used to evolve the state would then be *imperfect*. That is, the approximation made in defining the model error evolution covariance Q would likely be incorrect, and so the model and reality would evolve in different ways. For using the equivalent weights particle filter with real data, the approximation of Q becomes crucial, in a similar way to approximating the background error covariance matrix B and the observation error covariance matrix R required for many classical data assimilation techniques.

5.3. Observing network

We observe only the sea surface temperature (SST) in the coupled system. The SSTs are stored as the top level ocean temperature variable. As we are considering only twin experiments we are free to choose full coverage of SST observations, and so the observation operator will be simply the linear operator that restricts the whole state space to the upper level ocean temperature variables.

We assume that the observation errors are Gaussian, so that for a given model state x, the corresponding observation y will be of the form

$$y = Hx + \eta. \tag{12}$$

Thus $\eta \sim \mathcal{N}(0, R)$ is a realisation of observation noise. Note here that, written in this form, the observation operator is assumed unbiased and linear. For simplicity this work will consider only uncorrelated observation errors, i.e. R is assumed diagonal. This is unlikely to be true in practice, and future works will be devoted to investigating the effect of assimilating correlated observations using this model together with the equivalent weights particle filter.

Moreover, we assume that the variance in the observations are spatially uniform so that

$$R = \sigma^2 I \tag{13}$$

with σ taken to be $\sqrt{0.3}$K. This gives a total of $27,370$ independent observations every day.

5.4. Creating an initial ensemble

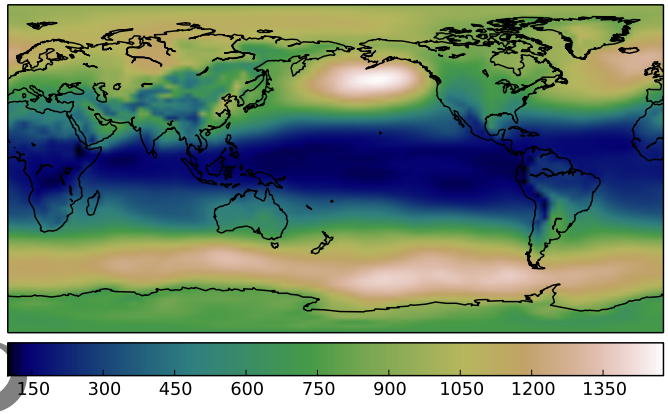
The generation of the initial ensemble is *ad hoc*. However, as the particle filter is a sequential method, an initial ensemble has to be generated only once, after which the ensemble will just evolve over time, updated whenever observations are available, year after year. Because of the random forcing the memory of the system is limited and the exact details of the initial ensemble become irrelevant.

A typical method of creating an initial ensemble for a twin experiment would be to have the ensemble perturbed around the truth. Assuming that the initial PDF is Gaussian with covariance B then given some truth state x_t, each of the m ensemble members x_i would be created as follows.

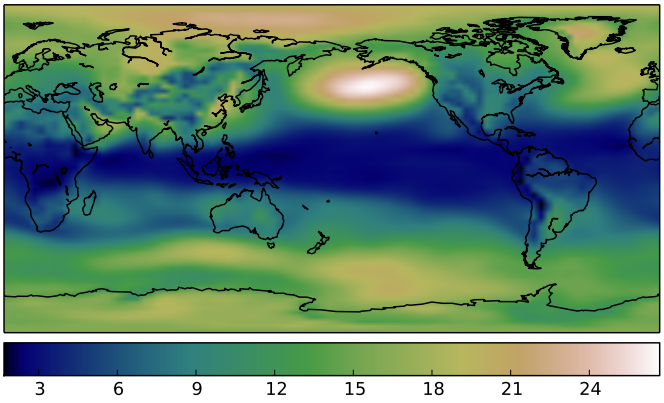
$$x_i = x_t + B^{1/2}\xi_i \tag{14}$$

for all ensemble members $\forall i \in 1, \ldots, m$, where ξ_i are realisations of an uncorrelated random variable. A more statistically consistent manner of generating the initial ensemble would be to ensure that the initial distribution is not centred on the truth, but rather the truth follows the same stochastic distribution as each ensemble member. This requires some reference state x_r then

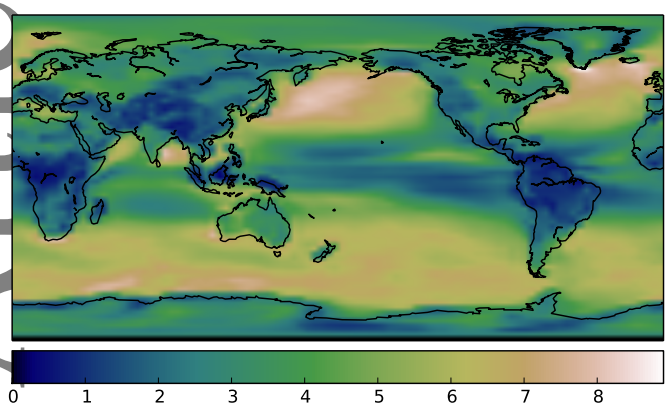
$$x_t = x_r + B^{1/2}\xi_t \tag{15}$$



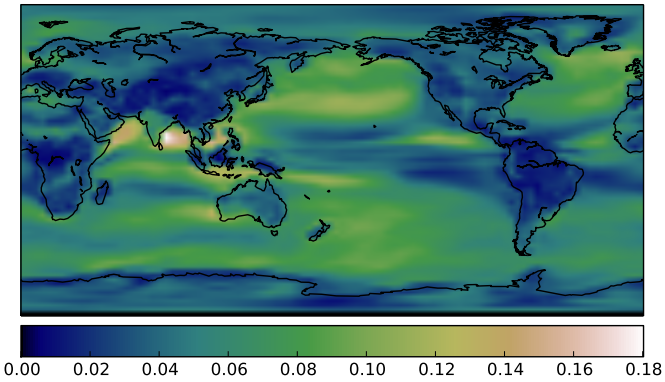
(a) Unscaled standard deviations in $\Lambda^{\frac{1}{2}}$ corresponding to variability in atmospheric pressure variables for $Q = Q_1$



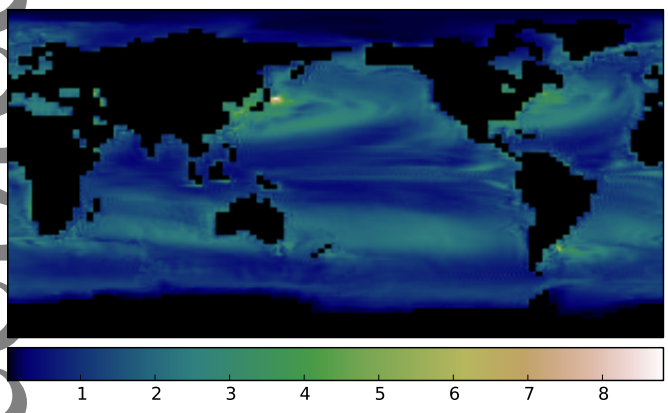
(b) Unscaled standard deviations in $\Lambda^{\frac{1}{2}}$ corresponding to variability in atmospheric pressure variables for $Q = Q_2$



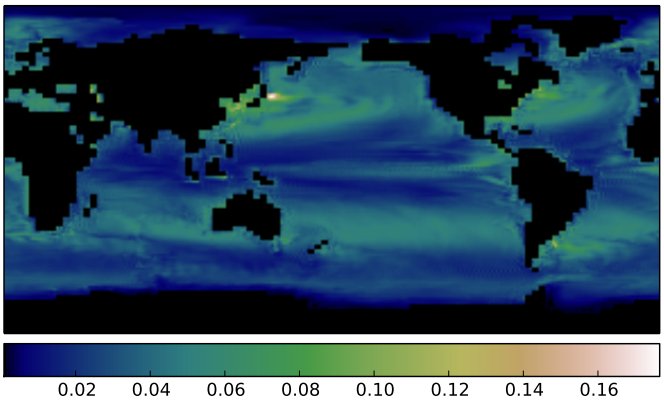
(c) Unscaled standard deviations in $\Lambda^{\frac{1}{2}}$ corresponding to variability in eastward surface winds for $Q = Q_1$



(d) Unscaled standard deviations in $\Lambda^{\frac{1}{2}}$ corresponding to variability in eastward surface winds for $Q = Q_2$



(e) Unscaled standard deviations in $\Lambda^{\frac{1}{2}}$ corresponding to variability in sea water temperature at the 4^{th} depth level for $Q = Q_1$



(f) Unscaled standard deviations in $\Lambda^{\frac{1}{2}}$ corresponding to variability in sea water temperature at the 4^{th} depth level for $Q = Q_2$

Figure 3. Standard deviations in $\Lambda^{\frac{1}{2}}$. In the left column are variables from $Q = Q_1$ showing the seasonal variability and in the right column variables from $Q = Q_2$ with seasonal variability removed.

and

$$x_i = x_r + B^{1/2}\xi_i \tag{16}$$

where $\xi_t, \xi_i \sim \mathcal{N}(0, I)$ for all $i \in 1, \ldots, m$. In this way the truth would be statistically indistinguishable from any of the ensemble members as $x_t, x_i \sim \mathcal{N}(x_r, B)$

With HadCM3 we do not have B, instead we experimented with the approximation that $B = \beta^2 Q$, for some scalar β. However this technique has been unable to produce a realistic amount of spread in the initial ensemble. If the coefficient β is large enough to create sufficient spread, the perturbation to the reference state $\beta Q^{1/2}\xi_i$ destroys the delicate balances present in the model (eg hydrostatic and geostrophic balance) causing the model to crash.

The strategy which we adopt in this paper is to take each initial ensemble member to be an instantaneous state of a long control run. In order for each ensemble member to be plausibly representative of the truth, they are all sampled at the same time of year. The advantage of this procedure is to eliminate the seasonal cycle in our initial conditions. Initialising the ensemble in such a way gives each ensemble member a dynamically consistent (balanced) state, and the desired spread is achieved by ensuring the samples are taken enough years apart within the control run. Note that the initial state for the truth is also taken as one of the instantaneous states from the control run.

5.5. *Further technical aspects*

The linear solve required in the equivalent weights step,

$$\text{find } x \text{ such that } \quad (HQH^T + R)x = b, \tag{17}$$

is performed with HSL_MA87 (Hogg et al. 2009), a direct linear solver optimized for shared memory systems.

6. Results

In this section we compare the results of assimilation runs with different parameters chosen in the equivalent weights particle filter, namely the strength of the nudging p and the proportion of particles, κ, kept by the equivalent weights step, for the more physically realistic Q matrix Q_2. Corresponding results for Q_1, i.e. the model error covariance matrix dominated by the seasonal cycle. can be found in the supplementary material online.

We have run the model for 6 months, beginning on 1 Dec 1859, assimilating artificial SST data at the end of each day. Hence there were 180 assimilation cycles. We have arbitrarily chosen an ensemble of 32 members as this was as many as could be practically afforded for the experiments. Ideally, to reduce statistical noise in the results, we would run all the experiments multiple times with different random forcing. However due to the computational complexity of these runs we have been limited to a single instance. For example, running 32 ensemble members required the use of 1152 processors for approximately 3 hours on ARCHER, the UK national supercomputer. For full details on the timings of the run see Browne and Wilson (2015).

6.1. *Root Mean Squared Errors*

Throughout this section we shall use root mean squared error (RMSE) to mean the square root of the spatial mean of the square of the error between the ensemble mean and the truth.

Figure 4 shows root mean square errors for 3 different variables with different parameters used in the EWPF. In green is shown the RMSE for a stochastically forced ensemble, i.e. simply forecasting from the initial conditions. Figure 4a shows the root mean squared errors for the observed variables. We can see that sufficient nudging brings the ensemble closer to the truth than the stochastic ensemble. The stronger the nudging the more reduced the RMSE is, until the nudging term becomes too large and the model breaks. In this case we were unable to increase the size of the nudging beyond what is shown here without the model crashing. This shows the sensitive nature of the model to perturbations.

The long term performance of the assimilation is improved when $\kappa = 1$ when compared to $\kappa < 1$. In Figure 4b and 4c only the assimilation runs with $\kappa = 1$ give RMSEs comparable with the stochastic ensemble.

Figure 4b shows the root mean squared errors of the specific humidity at the surface. Note that in this variable the assimilated ensemble behaves in a similar manner to the stochastic ensemble. This is indicative of the fact that the atmosphere is poorly constrained by the sea surface temperatures. Subsequent results will provide more evidence that the atmosphere is not well constrained and is behaving like the stochastic ensemble.

Figure 4c shows the root mean squared errors of the zonal sea water flow at the surface. In this case we can see that only assimilation runs with strong nudging and $\kappa = 1$ follow the stochastic ensemble. In the other case, when $\kappa < 1$ the ensembles lose their spread as the resampling step replaces some particles with copies of other, better performing particles. This has been observed for many other cases of $\kappa < 1$ which we do not include in Figure 13b. With the stochastic error term that we have introduced to HadCM3, this does not provide a sufficient spread in the remaining ensemble to encompass the truth. In this way, the RMSE increases markedly as the truth then lies outside of the ensemble. This is evidenced further by the rank histograms shown later in Section 6.3.

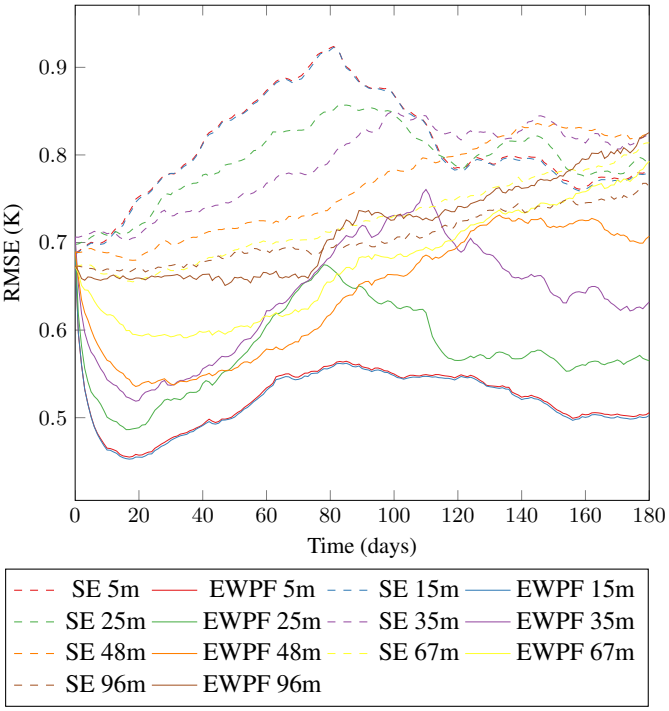


Figure 5. RMSE of sea water temperature at various model levels. Note that at 96m, the RMSE of the assimilation becomes larger than that of the stochastic ensemble (SE).

To show the effect of using the EWPF for assimilating SSTs on the other variables in the coupled ocean-atmosphere system, Figures 5 to 9 are plots of RMSEs for different levels in the model across different physical variables. For all these plots we are displaying the results of the experiment with $p = 3.5d7$ and $\kappa = 1$.

Figure 5 shows the seawater temperatures for levels of increasing depth. It can be see that the SSTs are significantly constraining these variables for at least the top 50m of the ocean, at which point the influence of the observations becomes much less significant. At the 96m level, the seawater temperatures appear to behave in a similar manner to the stochastic ensemble.

Figure 6 shows the seawater meridional flow for levels of increasing depth. The effect of the assimilation on the near surface fields is clearly stronger than those deeper in the ocean, with visible reductions in RMSE seen to a depth of 35m. Below this depth the RMSEs of the EWPF and the stochastic ensemble are qualitatively similar.

Figure 7 shows the seawater zonal flow for levels of increasing depth. If we compare this to Figure 6 we can see that the zonal flow is much less constrained by the SSTS. The reasons for this are currently not well understood and require further investigation into the physical reasons for this. The RMSE plots shown in

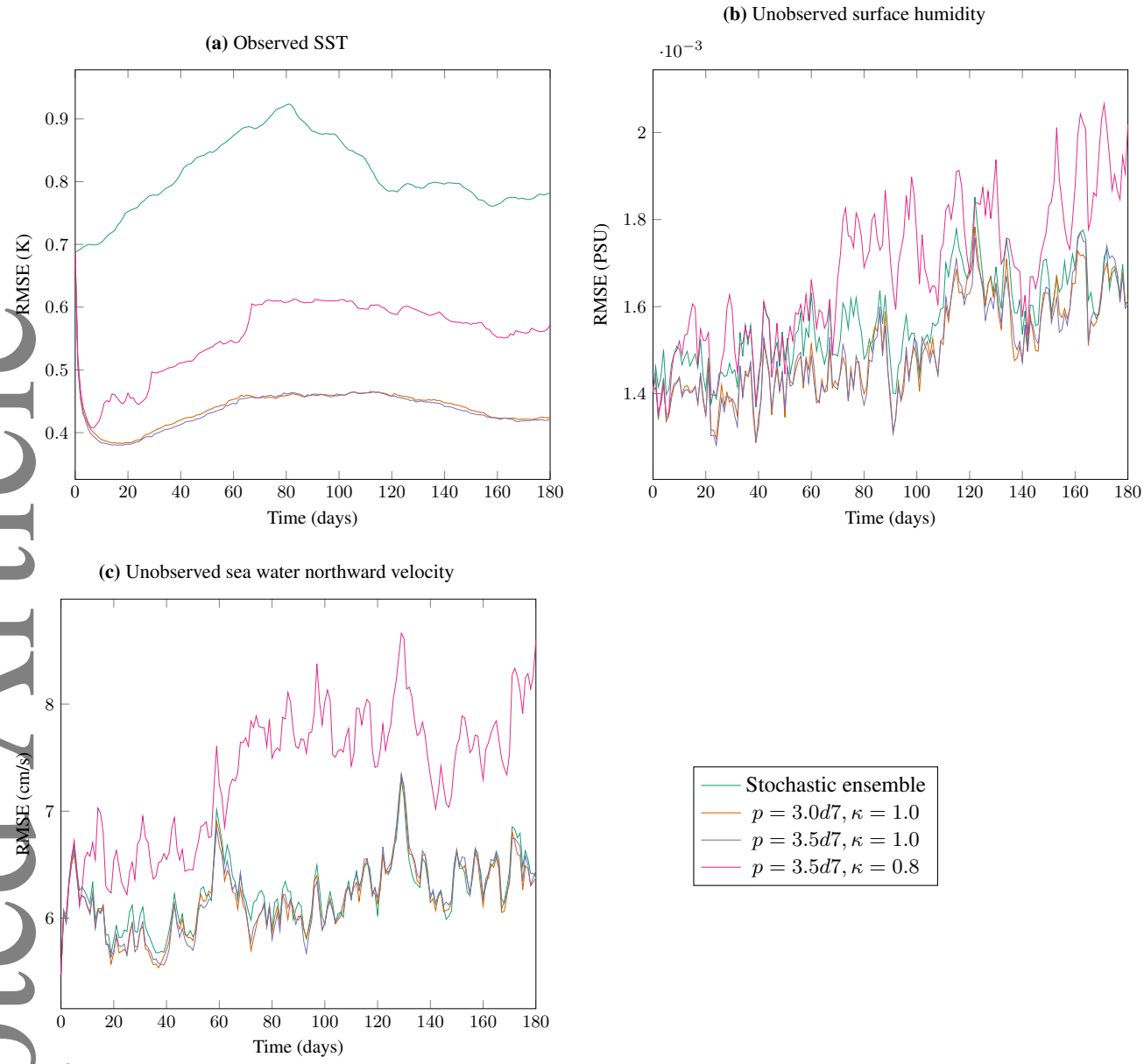


Figure 4. Root mean squared errors through time for different types of variable in the coupled climate model for different parameters, with the model evolution error covariance matrix Q_2 – the model evolution error covariance matrix *not* containing the seasonal cycle.

Figures 6 and 7 are global averages, and to understand this behaviour would mean to look into regional differences, which is beyond the scope of this paper.

Figure 8 shows the atmospheric temperature for levels of increasing height. Note that the variables are stored in pressure levels within HadCM3. These RMSE results indicate that only assimilating SSTS with the EWPF is not constraining the atmospheric temperatures. However for this specific experiment ($\kappa = 1$) there remains sufficient ensemble spread so as not to make the errors worse than the stochastic ensemble.

Figure 9 shows the atmospheric humidity for levels of increasing height. At the surface, there is improvement in the RMSE over the stochastic ensemble. As we might expect, this improvement is reduced as we consider levels higher up in the atmosphere further from the SST observations.

6.2. *Trajectories of individual state variables*

Figure 10 shows trajectories of two different observed variables and one unobserved variable for the same parameter configurations as shown in Figure 4. Figures 10a to 10c show a sea surface temperature variable in the southern hemisphere and Figures 10d to 10f show a sea surface temperature variable in the northern hemisphere. Figures 10g to 10i show a surface wind variable that is not observed. In yellow is the stochastic ensemble, blue the results of the assimilation with the specified parameters, and red the truth run.

Considering Figures 10a to 10c we can see that all the assimilation runs are constraining the ensemble to be a lot closer the the truth than the stochastic ensemble. However, when $\kappa = 0.8$ in Figure 10a we see that the truth is quite often outside of the ensemble and the spread of the ensemble is poor at times, when compared with σ, the observation error standard deviation. In Figure 10c we can see that there is a better spread but the ensemble does not completely track the truth as it is not nudged as strongly as the other two assimilation runs shown. Figure 10b is performing

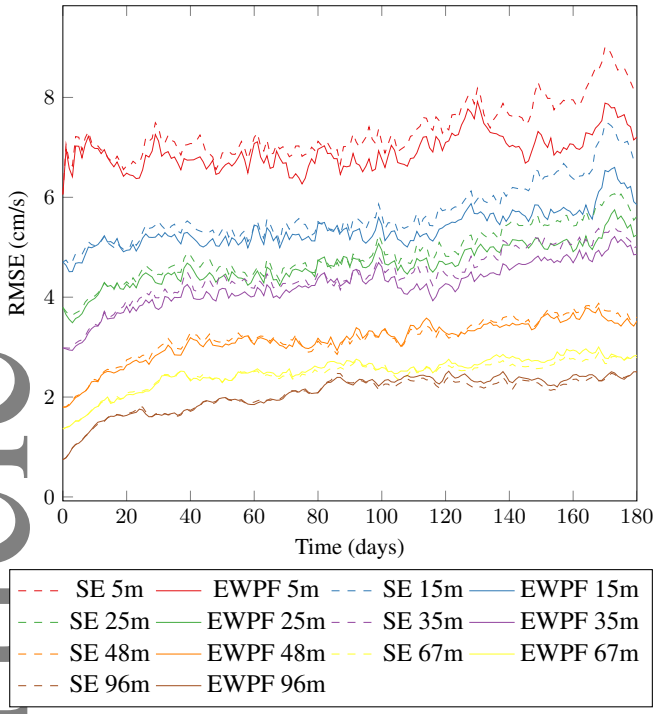


Figure 6. RMSE of sea water meridional flow at various depths. Note that the absolute errors reduce at lower depths as there is less variability in the deeper ocean. Note also that at levels deeper than 67m, the EWPF is performing similarly to the stochastic ensemble.

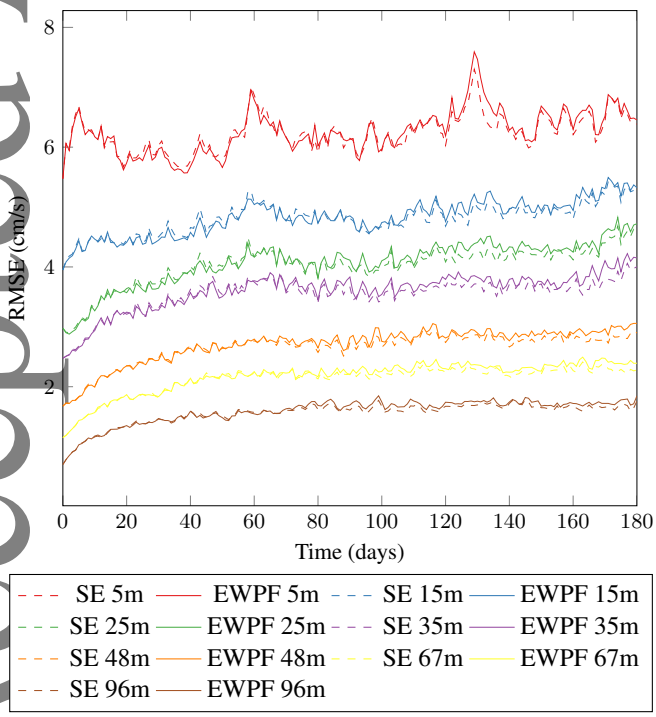


Figure 7. RMSE of sea water zonal flow at various depths.

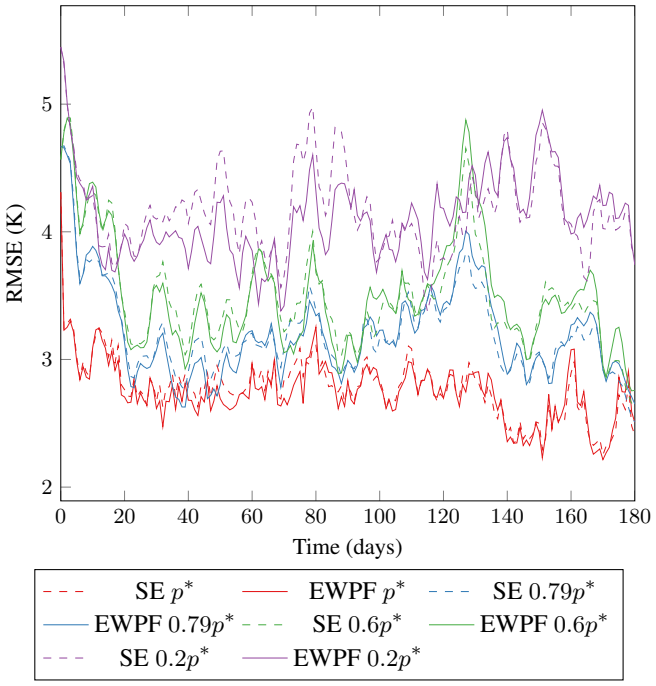


Figure 8. RMSE of atmosphere temperature at various heights.

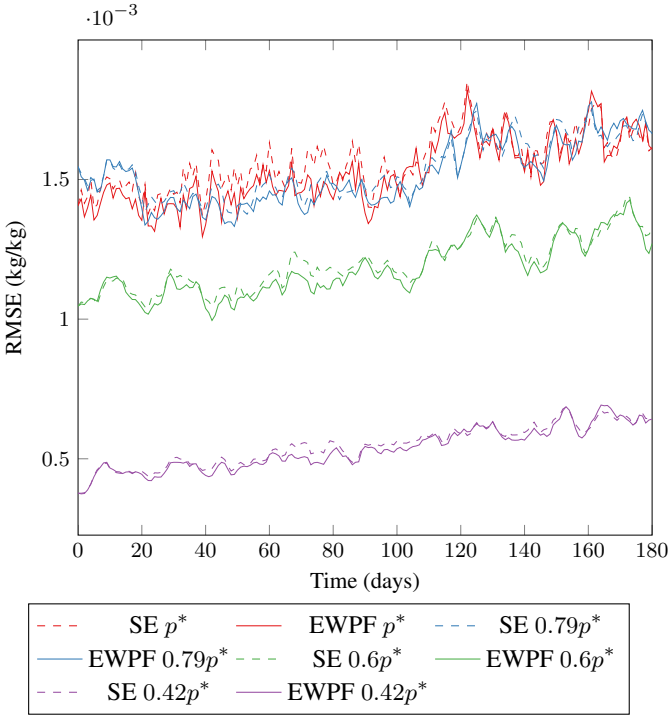


Figure 9. RMSE of atmosphere humidity at various heights.

best in this case as the ensemble is following the truth while retaining a good degree of spread.

For the other observed sea surface temperature variable shown in Figures 10d to 10f we can see that the final two ensembles with $\kappa = 1$ are performing in a broadly similar manner. Figure 10d however, with $\kappa = 0.8$ again shows too small a spread in the ensemble and for a disproportionately large number of timesteps the truth falls outside of the ensemble.

Considering now the trajectories of ensembles with $\kappa = 1$, we can see that the observed variables follow the truth well. However, after day 150, when the truth diverges from the stochastic ensemble in the second observed SST trajectory (Figures 10e and 10f) it can be seen that the spread in the ensemble increases. This shows an increase in the uncertainty of the assimilation which we may not see with other assimilation methods that do not use an ensemble.

In the final row, Figures 10g to 10i show trajectories of an unobserved zonal wind variable at the surface. Consistent with the previous trajectories, Figure 10g shows an ensemble with reduced spread compared to the others. However, none of these assimilated ensembles appear to accurately track the truth, further showing that the atmospheric component of the model is not well constrained by the sea surface temperature observations.

6.3. Rank histograms

In Figure 11, the columns correspond to the same columns as in Figure 10, i.e. the different parameter values in the equivalent weights particle filter. Figures 11a to 11c are rank histograms of the observed variables, Figures 11d to 11f rank histograms of eastward sea water flow in the deep ocean and Figures 11g to 11i rank histograms of northward wind in the high atmosphere. The histograms are formed by summing the rank histograms at a number of well separated grid points through all the timesteps of the experiment. In the ocean these are 10 latitude and 19 longitude grid points apart, getting increasingly sparse towards the poles. In the atmosphere these are 4 latitude and 16 longitude grid points apart, again getting increasingly sparse towards the poles.

Figures 11a, 11d and 11g when $\kappa = 0.8$ show clearly that the ensemble is under-dispersive in all the variables. This is vastly improved as shown in the other plots with $\kappa = 1$. Figures 11h and 11i appear to be flat. This shows again that, in the atmosphere, the truth is indistinguishable from the ensemble members. This is consistent with the atmosphere being unconstrained by the observations.

In the experiments with $\kappa = 1$, the observed SST variables (Figures 11b and 11c) the rank histograms appear slightly underdispersive. The remaining plots (Figures 11e and 11f) show that the ensembles are underdispersive in other ocean variables. This is no doubt compounded by the small amount of stochastic noise we were able to introduce to the system.

6.4. Delta function representation of marginal PDFs

In this section we show the delta function representation of the (marginal) posterior distribution as defined in (2). The height of each particle in Figure 12 correspond to the weights of the particles. These marginal pdfs are shown after resampling, so by definition these particles have weight $\frac{1}{m}$. Had filter degeneracy occured, no spread would be seen in the ensemble. This is not the case as, by construction, the equivalent weights step (8) avoids this and hence keeps the ensemble well spread.

Figure 12 shows the delta function representation of marginal PDFs arising from the assimilation run with $p = 3.5d7$ and $\kappa = 1$ in the experiment with $Q = Q_2$. Figures 12a, 12b and 12c show the representations of specific humidity at the surface, sea water temperature at 47m depth and 3347m depth respectively. Figures 12d to 12f show the same ocean eastward velocity variable at 3 different points in time.

These marginal PDFs are the main motivation for using the nonlinear particle filter to initialise the climate model. From these PDFs we can get an estimate of the uncertainty in the posterior PDF at and given point, not just a *maximum a posteriori* estimate that we would get from using a variational method. With only 32 particles it is unclear if any of these PDFs are truly multimodal, however Figures 12d to 12f suggest they may be bimodal.

7. Conclusions

In this paper we have applied a fully nonlinear data assimilation method to a coupled climate model. The equivalent weights particle filter was applied to a system with 2,314,430 variables and 27,370 independent observations per timestep. The particle filter did not degenerate at any stage thanks to the equivalent weights step. This model is over 35 times larger than any other model that the method has been applied to before (Ades and van Leeuwen 2015).

We have been able to constrain the model well in the observed variables. This constraining of model variables extended into the ocean temperatures a number of model levels below the observed SST fields, but had little effect in the atmosphere and the non-temperature prognostic variables in the ocean. This suggests that, to further constrain the other variables, we require different observations to be used. For instance, observations from ARGO profiles to have observations in the deep ocean may be of great benefit for stronger uncertainty quantification of a subsequent prediction.

The best results we have obtained have come in the cases in which we applied the strongest nudging. However, the amount of nudging we could use was restricted by the numerical sensitivity of the model. Too much nudging would lead to catastrophic failure of the model and hence the assimilation process could not continue. Applying the EWPF to a model with less numerical sensitivity may allow for stronger nudging to be employed, although this should still be smaller than the increment from the dynamical model.

We found it hard to attain an appropriate spread in the ensemble. Indeed, simply to achieve a reasonable spread in a stochastically forced ensemble required a non-standard method of initialising it. Because of this lack of spread available from the stochastic forcing, we achieved the best results when we kept all of the ensemble members in the equivalent weights step. This is different behaviour than has been observed in the equivalent weights filter when compared to its application in previous models (Ades and van Leeuwen 2013, 2015). We conjecture that this is due to an increase in degrees of freedom in this larger model, allowing equivalent weights to be achieved with a relatively smaller change to the model state, however this requires much more thorough investigation.

Using the equivalent weights particle filter, being a sequential data assimilation scheme, avoided any potential difficulties associated with the different timescales in the various components of the coupled climate system. For instance, in a 4DVar data assimilation method, the time period over which observations of the ocean and atmosphere are assimilated would typically be different. Naively one would think that the window of observations in the relatively fast moving atmosphere should be shorter than that of the relatively slow moving ocean. This can bring challenges to solving the fully coupled variational data assimilation problem, leading to "weakly-coupled" variational solution methods (Smith et al. 2015). Other sequential methods such as the EnKF would similarly avoid the issues of multiple timescales that variational methods face.

The performance of the whole assimilation system is dependent on the specification of the model evolution error covariance matrix. In this paper, doing twin experiments, we have been able to prescribe this. However in reality a great deal of work would be required to approximate this matrix. This is an area which is in need of much research, which has largely been neglected by the wider community up to now.

The EWPF has been applied to a high dimensional system and as a result the posterior PDF has been approximated, which is not possible with single-run data assimilation methods like a simple nudging, or variational methods like 4DVar. The computational cost of this method is similar to using an EnKF with a stochastic model. This cost would be comparable to the computational cost

of 4DVAR, where the repeated iterations and running of the slower tangent linear and adjoint models would offset the cost of the ensemble of model runs. An additional issue with 4DVar is that if the ocean and atmosphere are treated as one in the adjoint calculations their different time scales lead to difficulties when specifying the optimal optimization window length. To avoid this problem and to avoid having to generate an adjoint of the full ocean-atmosphere system two separate adjoint are often used. Treating the coupled system together minimizes the effects of initialisation shock that is known to affect methods that treat the ocean and atmosphere separately.

The use of the EWPF, a fully nonlinear data assimilation method, has shown that some marginal PDFs could be multimodal. This is a result of both the dynamics of the model, and the observations which we have used. In other circumstances, the posterior PDF may be Gaussian, in which case using a method such as the EnKF would be advantageous as it is designed specifically for these cases. However if the variables of interest clearly do not follow a Gaussian distribution (for example bounded humidity variables) then the EWPF should be seriously considered as a data assimilation method to be used.

Incorporating a new data assimilation scheme into a GCM could be a daunting task. However the use of the EMPIRE (Browne and Wilson 2015) coupling system allowed us to do so extremely quickly within a particularly small team. This system, along with the open source codes implementing the equivalent weights particle filter, should encourage the community that it is possible to quickly and easily use fully nonlinear data assimilation methods with new models of ever increasing size and complexity.

Acknowledgements

This work was supported by NERC grant NE/J005878/1 and by the National Centre for Earth Observation. This work used the ARCHER UK National Supercomputing Service (`http://www.archer.ac.uk`).

References

Ades, M. and van Leeuwen, P. J. (2013). An exploration of the equivalent weights particle filter. *Quarterly Journal of the Royal Meteorological Society*, 139(672):820–840.

Ades, M. and van Leeuwen, P. J. (2015). The equivalent-weights particle filter in a high dimensional system. *Quarterly Journal of the Royal Meteorological Society*, 141(687):484–503.

Baker, J. E. (1987). Reducing Bias and Inefficiency in the Selection Algorithms. In *Proceedings of the Second International Conference on Genetic Algorithms and their Application*, pages 14–21, Hillsdale, New Jersey, US. Lawrence Erlbaum Associates.

Balmaseda, M. A., Alves, O. J., Arribas, A., Awaji, T., Behringer, D. W., Ferry, N., Fujii, Y., Lee, T., Rienecker, M., Rosati, T., and Stammer, D. (2009). Ocean initialization for seasonal forecasts. *Oceanography*, 22(3):154–159.

Bengtsson, T., Bickel, P., and Li, B. (2008). Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems. *IMS Collections Probability and Statistics: Essays in honor of David A. Freedman*, 2:316–334.

Bocquet, M. (2011). Ensemble Kalman filtering without the intrinsic need for inflation. *Nonlinear Processes in Geophysics*, 18(5):735–750.

Browne, P. and Wilson, S. (2015). A simple method for integrating a complex model into an ensemble data assimilation system using MPI. *Environmental Modelling & Software*, 68:122–128.

Chen, D. and Zebiak, S. (1997). Initialization and Predictability of a Coupled ENSO Forecast Model. *Monthly Weather Revieweview*, 125(5):773–788.

Chorin, A. J. and Tu, X. (2009). Implicit sampling for particle filters. *Proceedings of the National Academy of Sciences of the United States of America*, 106(41):17249–17254.

Cohn, S. (1997). An introduction to estimation theory. *Journal of the Meteorological Society of Japan. Ser. II*, 75(1B):257–288.

Collins, M., Tett, S. F. B., and Cooper, C. (2001). The internal climate variability of HadCM3, a version of the Hadley Centre coupled model without flux adjustments. *Climate Dynamics*, 17:61–81.

Dubinkina, S. and Goosse, H. (2013). An assessment of particle filtering methods and nudging for climate state reconstructions. *Climate of the Past*, 9:1141–1152.

Dubinkina, S., Goosse, H., Sallaz-Damaz, Y., Crespin, E., and Crucifix, M. (2011). Testing a particle filter to reconstruct climate changes over the past centuries. *International Journal of Bifurcation and Chaos*, 21(12):3611–3618.

Goosse, H., Crespin, E., Dubinkina, S., Loutre, M.-F., Mann, M. E., Renssen, H., Sallaz-Damaz, Y., and Shindell, D. (2012a). The role of forcing and internal dynamics in explaining the Medieval Climate Anomaly. *Climate Dynamics*, 39(12):2847–2866.

Goosse, H., Guiot, J., Mann, M. E., Dubinkina, S., and Sallaz-Damaz, Y. (2012b). The medieval climate anomaly in Europe: Comparison of the summer and annual mean signals in two reconstructions and in simulations with data assimilation. *Global and Planetary Change*, 84-85:35–47.

Gordon, C., Cooper, C., Senior, C., Banks, H., Gregoire, L. J., Johns, T., Mitchell, J., and Wood, R. (2000). The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dynamics*, 16:147–168.

Gropp, W., Lusk, E., Doss, N., and Skjellum, A. (1996). A high-performance, portable implementation of the MPI message passing interface standard. *Parallel Computing*, 22(6):789–828.

Hamill, T. M., Whitaker, J. S., and Snyder, C. (2001). Distance-Dependent Filtering of Background Error Covariance Estimates in an Ensemble Kalman Filter. *Monthly Weather Review*, 129(11):2776–2790.

Hogg, J., Reid, J., and Scott, J. (2009). A DAG-based Sparse Cholesky Solver for Multicore Architectures. Technical report, Science and Technology Facilities Council.

Martone, M., Filippone, S., Tucci, S., Paprzycki, M., and Ganzha, M. (2010). Utilizing recursive storage in sparse matrix-vector multiplication - preliminary considerations. In Philips, T., editor, *CATA*, pages 300–305. ISCA.

Meehl, G. a., Goddard, L., Boer, G., Burgman, R., Branstator, G., Cassou, C., Corti, S., Danabasoglu, G., Doblas-Reyes, F., Hawkins, E., Karspeck, A., Kimoto, M., Kumar, A., Matei, D., Mignot, J., Msadek, R., Navarra, A., Pohlmann, H., Rienecker, M., Rosati, T., Schneider, E., Smith, D., Sutton, R., Teng, H., van Oldenborgh, G. J., Vecchi, G., and Yeager, S. (2014). Decadal Climate Prediction: An Update from the Trenches. *Bulletin of the American Meteorological Society*, 95(2):243–267.

Meinhold, R. and Singpurwalla, N. (1983). Understanding the Kalman filter. *The American Statistician*, 37(2):123–127.

Price, J., Weller, R., and Schudlich, R. (1987). Wind-driven ocean currents and Ekman transport. *Science*, 238(4833):1534–1538.

Smith, P. J., Fowler, A. M., and Lawless, A. S. (2015). Exploring strategies for coupled 4D-Var data assimilation using an idealised atmosphere-ocean model by idealised atmosphere-ocean model. *Tellus A*, In Press.

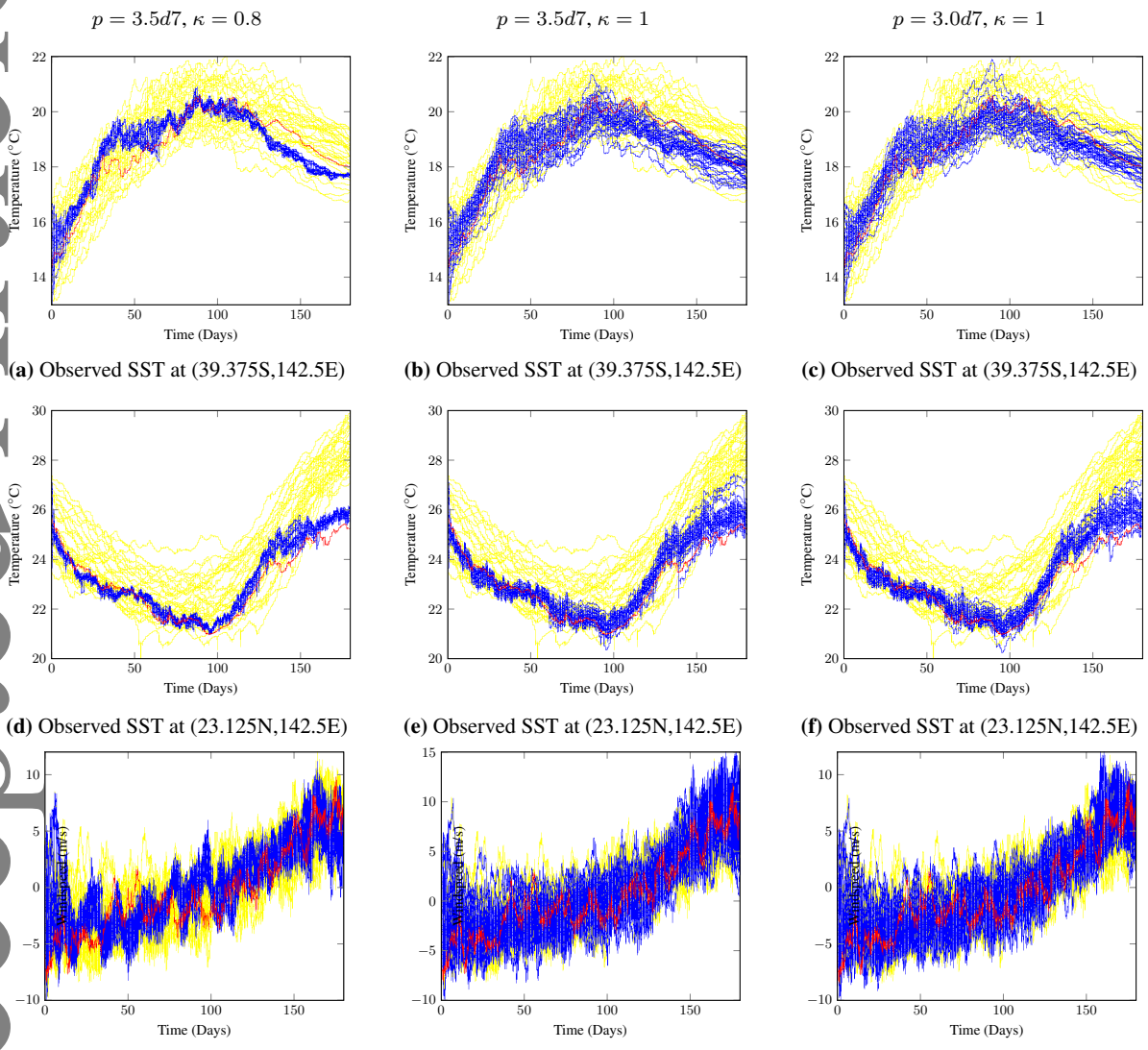
Snyder, C., Bengtsson, T., Bickel, P., and Anderson, J. (2008). Obstacles to High-Dimensional Particle Filtering. *Monthly Weather Review*, 136(12):4629–4640.

Solomon, S. (2007). *Climate change 2007-the physical science basis: Working group I contribution to the fourth assessment report of the IPCC*, volume 4. Cambridge University Press.

Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M. (2013). Climate change 2013: The physical science basis. *Intergovernmental Panel on Climate Change, Working Group I Contribution to the IPCC Fifth Assessment Report (AR5)(Cambridge Univ Press, New York)*.

van Leeuwen, P. J. (2009). Particle Filtering in Geophysical Systems. *Monthly Weather Review*, 137(12):4089–4114.

van Leeuwen, P. J. (2010). Nonlinear data assimilation in geosciences: an extremely efficient particle filter. *Quarterly Journal of the Royal Meteorological Society*, 136(653):1991–1999.

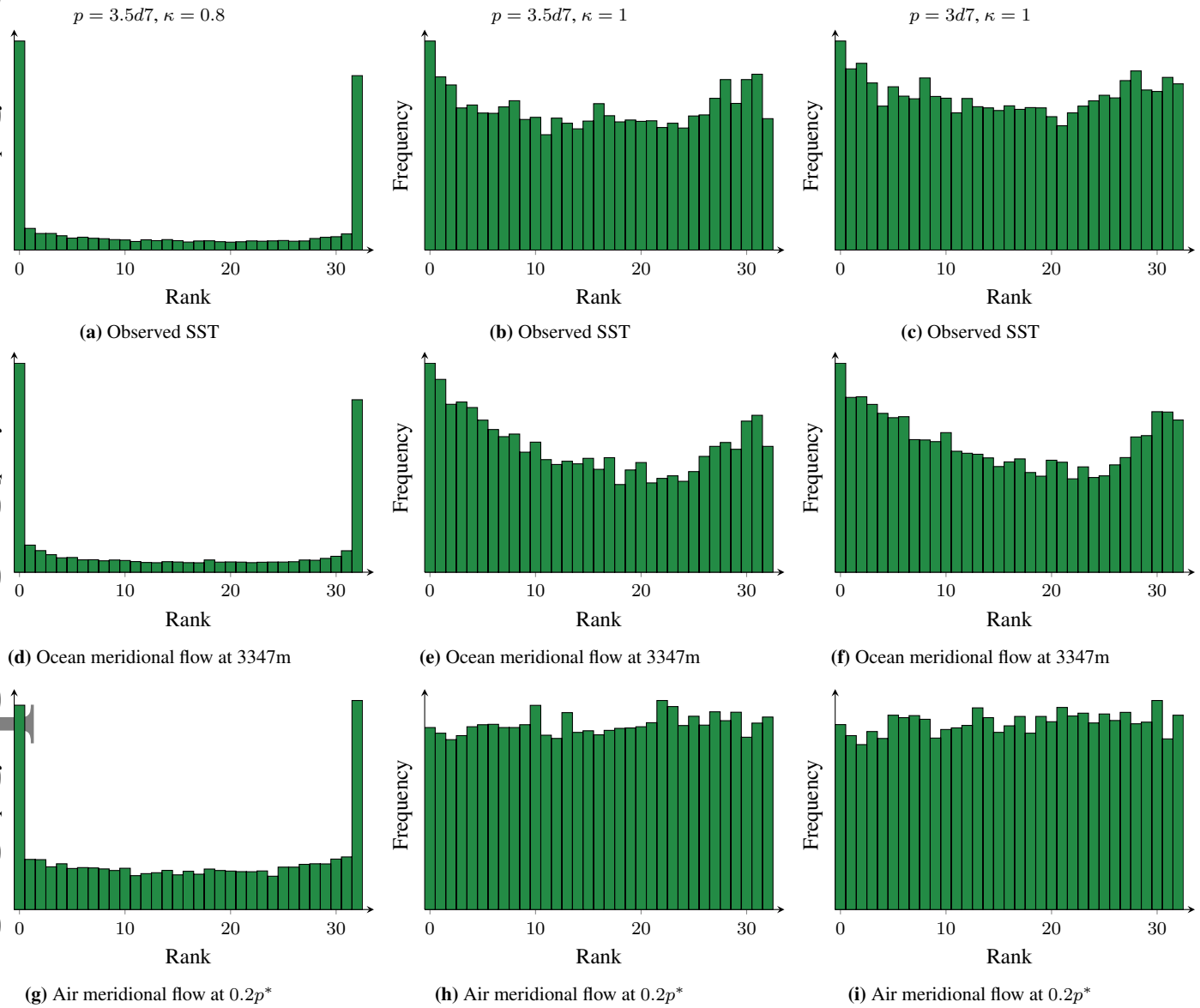


(a) Observed SST at (39.375S,142.5E) **(b)** Observed SST at (39.375S,142.5E) **(c)** Observed SST at (39.375S,142.5E)

(d) Observed SST at (23.125N,142.5E) **(e)** Observed SST at (23.125N,142.5E) **(f)** Observed SST at (23.125N,142.5E)

(g) Unobserved eastward surface wind at (11.25N,69.375E) **(h)** Unobserved eastward surface wind at (11.25N,69.375E) **(i)** Unobserved eastward surface wind at (11.25N,69.375E)

Figure 10. Trajectories of 3 separate state variables when using different parameters in the equivalent weights particle filter. In the background in yellow are the trajectories of the stochastic ensemble. In blue are the assimilated trajectories and in red is the truth

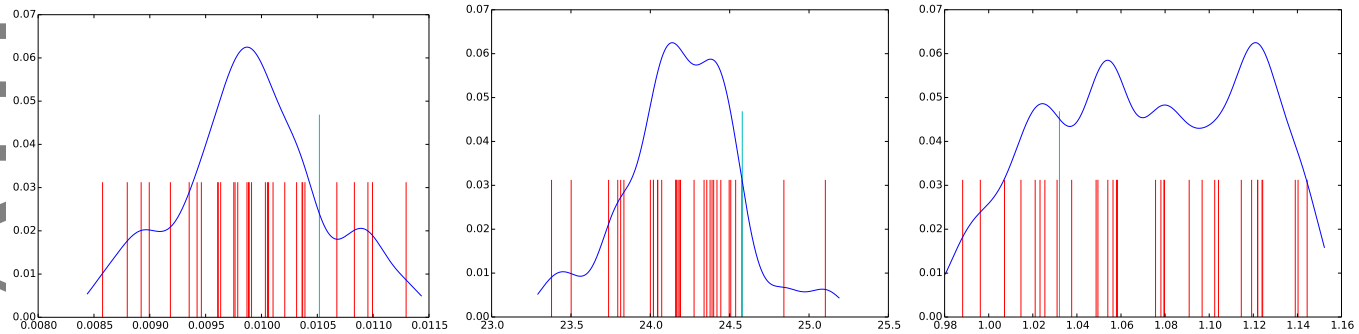


(a) Observed SST **(b)** Observed SST **(c)** Observed SST

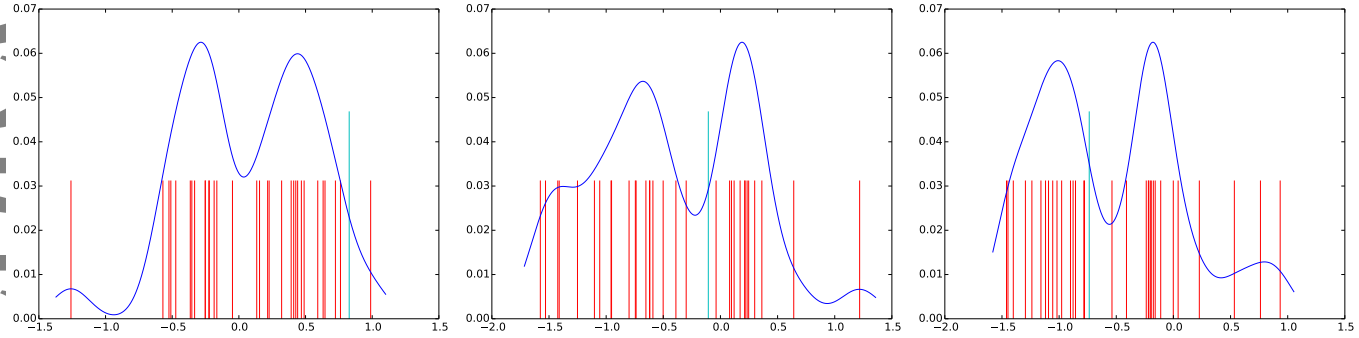
(d) Ocean meridional flow at 3347m **(e)** Ocean meridional flow at 3347m **(f)** Ocean meridional flow at 3347m

(g) Air meridional flow at $0.2p^*$ **(h)** Air meridional flow at $0.2p^*$ **(i)** Air meridional flow at $0.2p^*$

Figure 11. Rank histograms of different model variables at a given model level when using different parameters in the equivalent weights particle filter for the case that $Q = Q_2$.



(a) Surface level specific humidity at (20S,270E) **(b)** Seawater temperature at (1.875S,213.75E) at a depth of 48m **(c)** Seawater temperature at (51.875S,228.75E) at a depth of 3347m



(d) Meridional seawater flow at (15S,100.625E) at 3347m depth on day 75 **(e)** Meridional seawater flow at (15S,100.625E) at 3347m depth on day 161 **(f)** Meridional seawater flow at (15S,100.625E) at 3347m depth on day 177

Figure 12. Delta function representations of the marginal PDFs for different variables at a given instance in time for the experiment $Q = Q_2$, $p = 3.5d7$ and $\kappa = 1$. In red are shown the particles and in cyan we show the truth (raised for clarity). To guide the eye, we have overlaid in blue the results of fitting a Gaussian kernel, where the standard deviation of each Gaussian function is given by the quotient of the range in the marginal PDF and half the number of ensemble members.

A. Results with $Q = Q_1$

In this supplementary section we show the results from using the model error covariance matrix that retains the seasonal cycle.

A.1. Root Mean Squared Errors

Figure 13 shows the root mean square errors for a different variables in the case $Q = Q_1$. In green is shown the RMSE for a stochastically forced ensemble, i.e. simply forecasting from the initial conditions.

Figure 13a shows the root mean squared errors for the observed variables. We can see that sufficient nudging brings the ensemble closer to the truth than the stochastic ensemble. The best performing assimilations are those that have a strong nudging and set the proportion of particles kept in the equivalent weights step, κ to 1.0. In this case we were unable to increase the size of the nudging beyond what is shown here without the model crashing. This shows the sensitive nature of the model to perturbations.

Figure 13c shows the root mean squared errors of the northward sea water velocities at the surface. In this case we can see that only assimilation runs with strong nudging and $\kappa = 1$ follow the stochastic ensemble. In the other case, when $\kappa < 1$ the ensembles lose their spread as the resampling step replaces some particles with copies of other, better performing particles. This has been observed for many other cases of $\kappa < 1$ which we do not include in Figure 13b. With the stochastic error term that we have introduced to HadCM3, this does not provide a sufficient spread in the remaining ensemble to encompass the truth. In this way, the RMSE increases markedly as the truth then lies outside of the ensemble. This is evidenced further by the rank histograms shown later in Section 6.3.

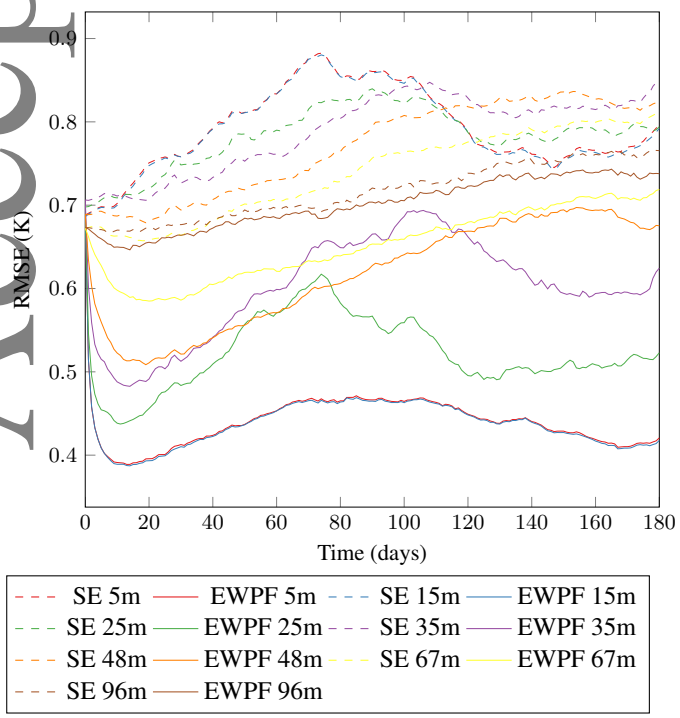


Figure 14. RMSE of sea water temperature at various model levels. Note that at 96m, the RMSE of the assimilation becomes larger than that of the stochastic ensemble (SE).

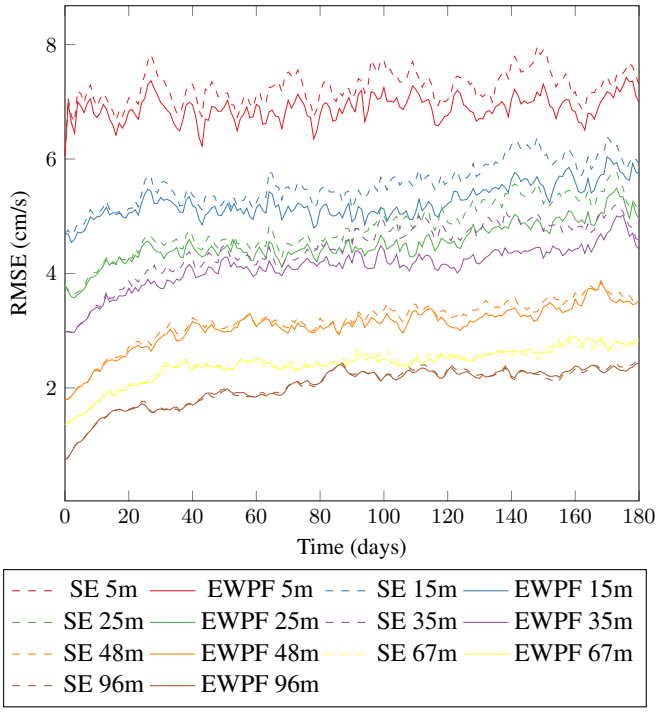


Figure 15. RMSE of sea water meridional flow at various depths. Note that the absolute errors reduce at lower depths as there is less variability in the deeper ocean. Note also that at levels deeper than 67m, the EWPF is performing similarly to the stochastic ensemble.

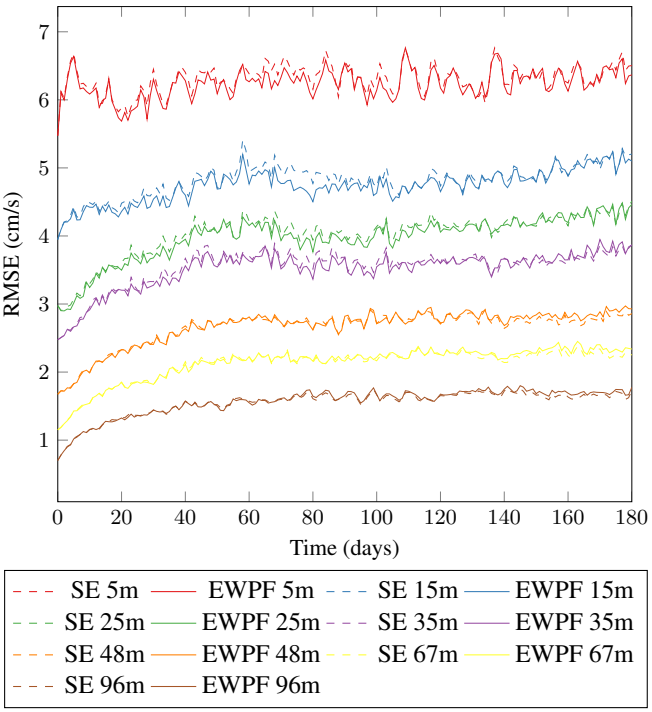


Figure 16. RMSE of sea water zonal flow at various depths.

A.2. Trajectories of individual state variables

Figure 19 shows the trajectories of the same points as shown in Figure 10 but for $Q = Q_1$. Figure 19a shows that with $\kappa < 1$ we have lost the appropriate spread in the ensemble. After around 130 days, the truth diverges from the ensemble and the nudging is not strong enough to bring the ensemble back to the truth. Figure 19d shows similar behaviour, but note that in this variable, in the final 30 days, the truth is far away from the stochastic ensemble. In this case, although the spread is still poor, the ensemble is significantly closer to the truth than the stochastically forced

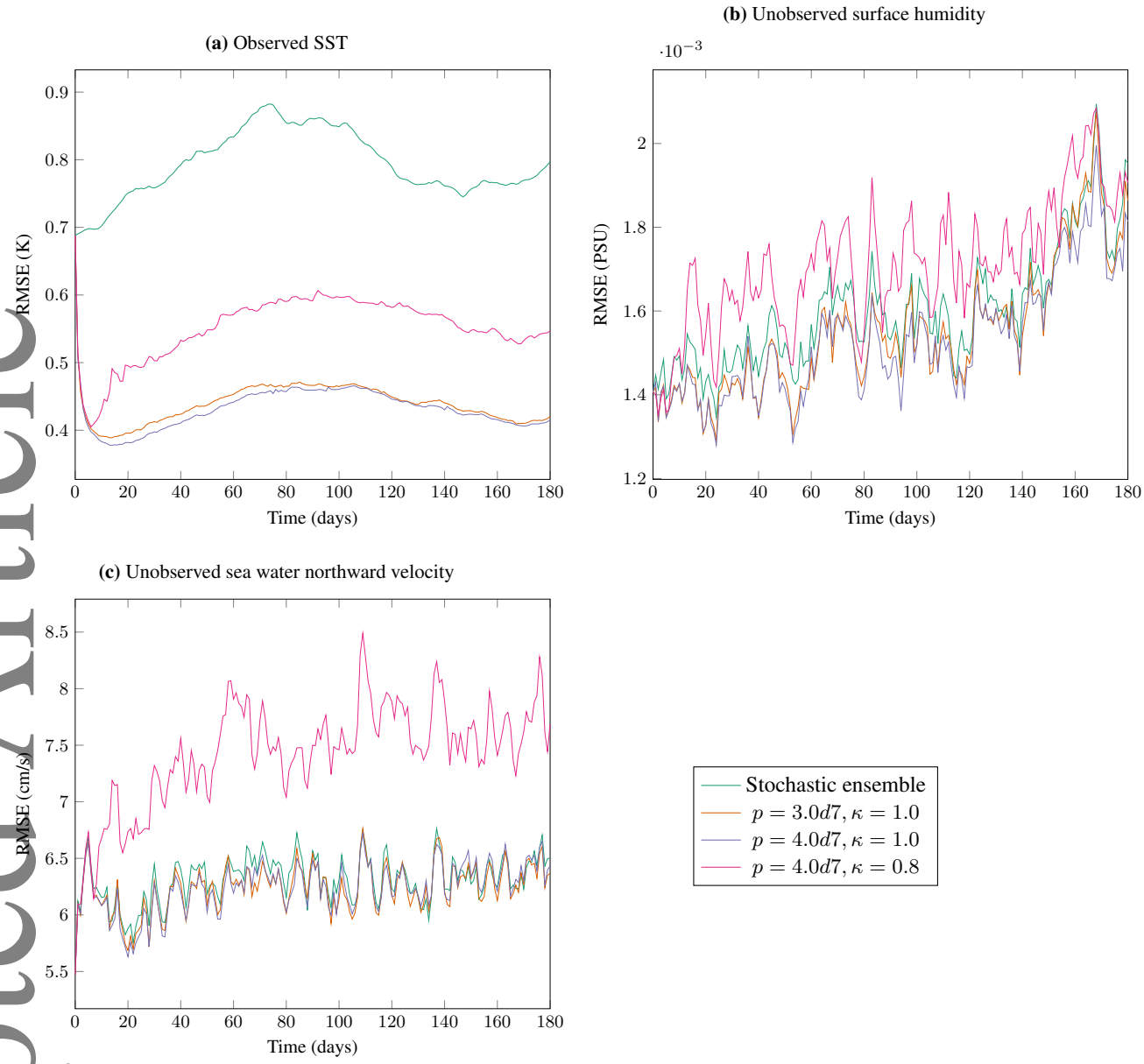


Figure 13. Root mean squared errors through time for different types of variable in the coupled climate model for different parameters, with the model evolution error covariance matrix Q_1 – the model evolution error covariance matrix containing the seasonal cycle.

ensemble. Figure 19g is behaving broadly similarly to Figure 10g, in that the truth is not well constrained and the spread in the ensemble has been reduced.

As in Figure 10, Figures 19h and 19i show how the unobserved wind variables are poorly constrained by the SST observations

In Figure 19c a single particle in the ensemble can be seen to drop substantially below the rest of the ensemble. Even with $\kappa = 1$ we still perform Universal Importance Resampling after each equivalent weights step. This provides a small chance that the particle will be replaced by a copy of another from the ensemble. This can be seen happening at around the 112 day mark where the trajectory of the low particle is brought up back into the main body of the ensemble. This resampling was found to happen less than 0.1% of the possible times it could occur.

A.3. Rank histograms

In Figure 20, the columns correspond to the same columns as in Figure 19, i.e. the different parameter values in the equivalent weights particle filter but with the different model evolution error covariance matrix $Q = Q_1$. Figures 20a to 20c are rank histograms of the observed variables, Figures 20d to 20f rank histograms of eastward sea water flow in the deep ocean and Figures 20g to 20i rank histograms of northward wind in the high atmosphere.

Considering Figures 20a 20d and 20g, where $\kappa = 0.8$ we can see that these are marginally more underdispersive than the corresponding plots in Figure 11.

Figures 20b and 20c, rank histograms of the observed variables, show signs of being underdispersive, unlike the corresponding plots 11b and 11c. The remaining Figures 20e, 20f, 20h and 20f appear to be qualitatively similar to their counterparts in Figure 20.

A.4. Delta function representation of marginal PDFs

Figure 21 shows the delta function representation of marginal PDFs arising from the assimilation run with $p = 4d7$ and $\kappa = 1$

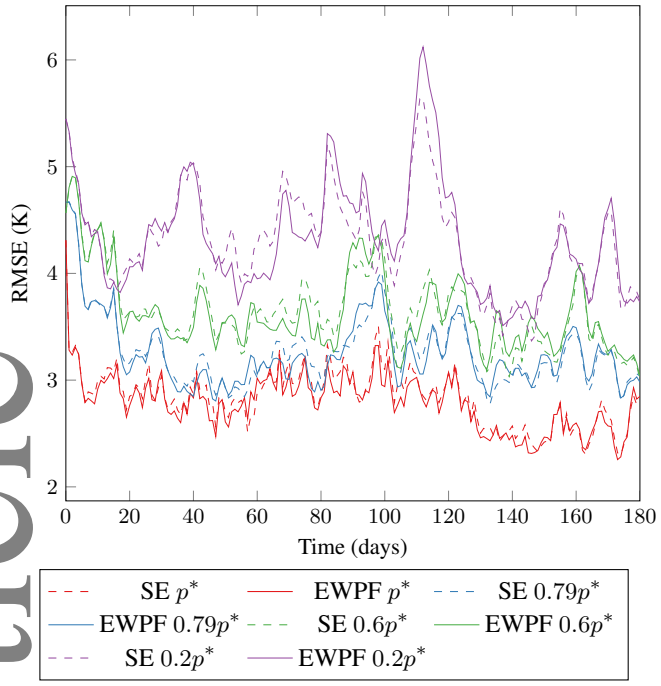


Figure 17. RMSE of atmosphere temperature at various heights.

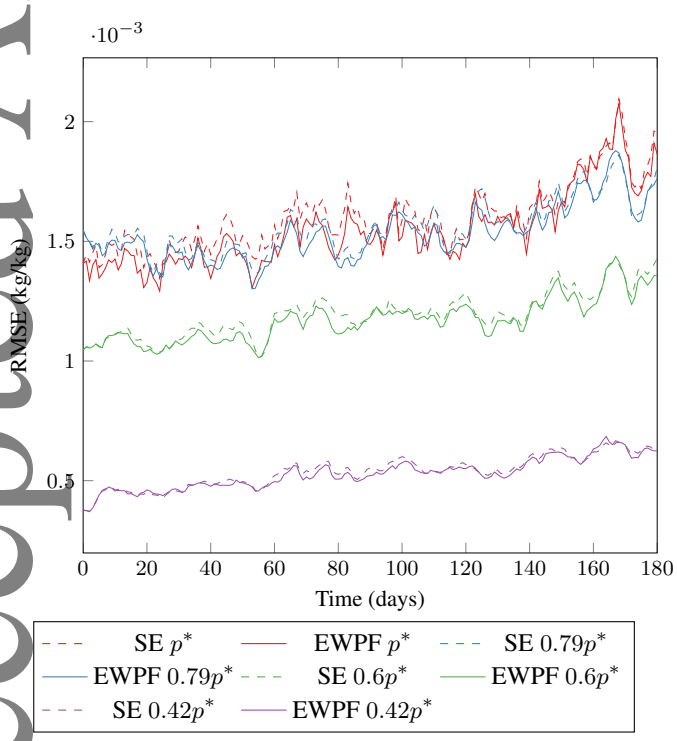
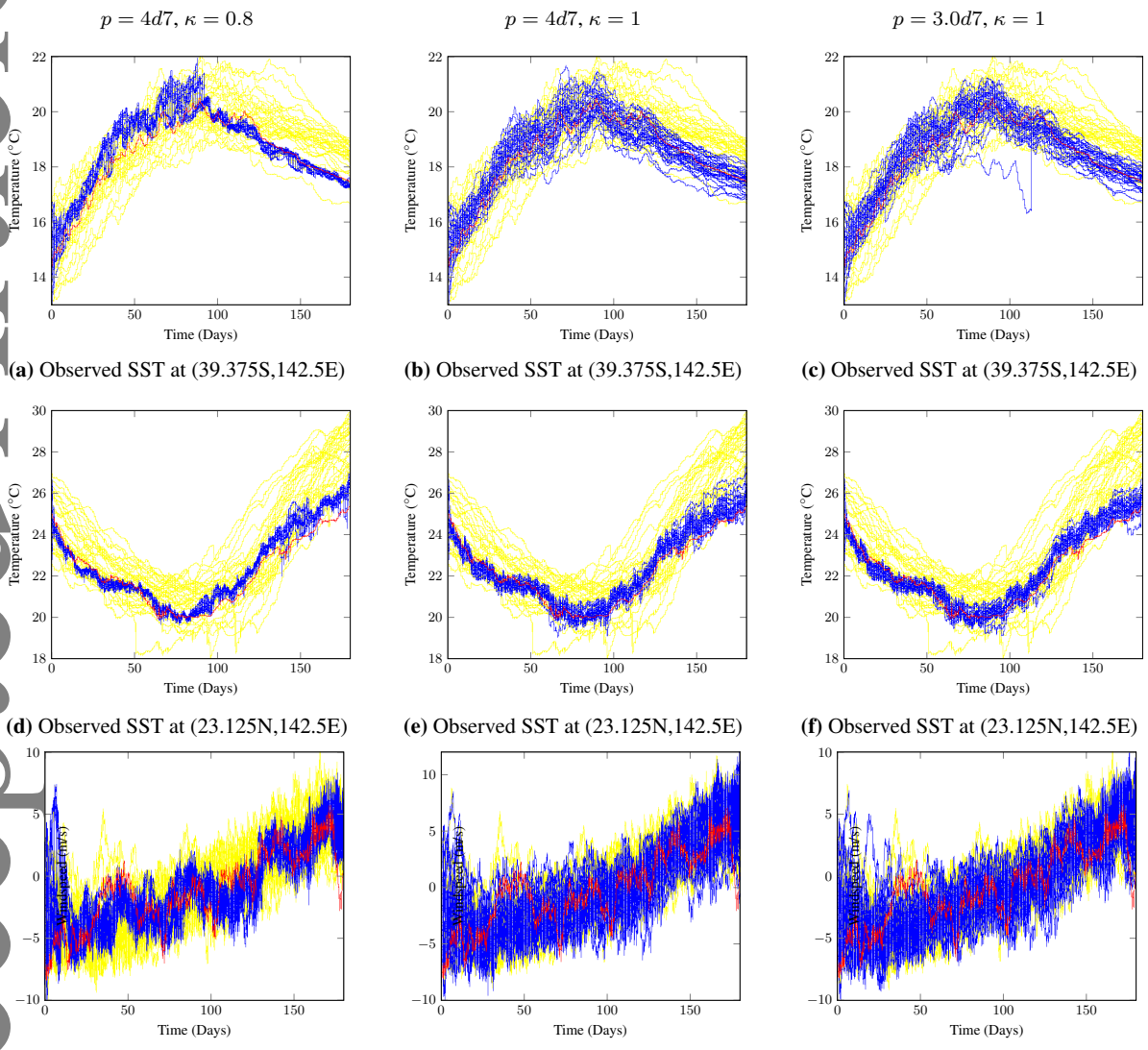


Figure 18. RMSE of atmosphere humidity at various heights.

for the experiment with $Q = Q_1$. The variables are the same as shown in Figure 12. If we compare these figures to Figure 12, we can see that the spread of the sea water temperature variables close to the surface (Figure 21b) is increased. As in the case that $Q = Q_2$, the variables shown in Figures 21e and 21f appear to be bimodal, however in this case with $Q = Q_1$, this bimodality occurs later as it is not seen in Figure 21d.

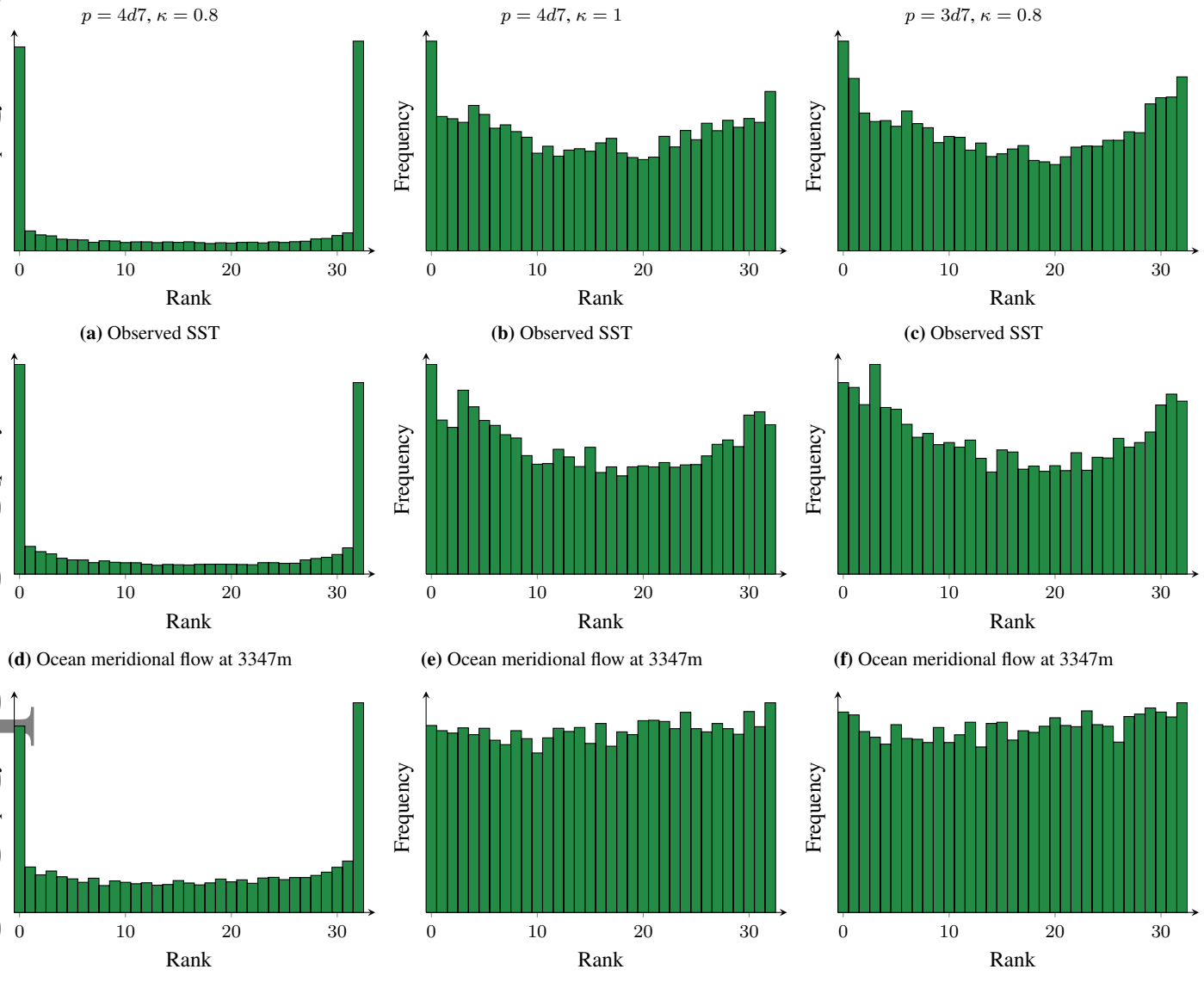


(a) Observed SST at (39.375S,142.5E) **(b)** Observed SST at (39.375S,142.5E) **(c)** Observed SST at (39.375S,142.5E)

(d) Observed SST at (23.125N,142.5E) **(e)** Observed SST at (23.125N,142.5E) **(f)** Observed SST at (23.125N,142.5E)

(g) Unobserved eastward surface wind at (11.25N,69.375E) **(h)** Unobserved eastward surface wind at (11.25N,69.375E) **(i)** Unobserved eastward surface wind at (11.25N,69.375E)

Figure 19. $Q = Q_1$: Trajectories of 3 separate state variables when using different parameters in the equivalent weights particle filter. In the background in yellow are the trajectories of the stochastic ensemble. In blue are the assimilated trajectories and in red is the truth

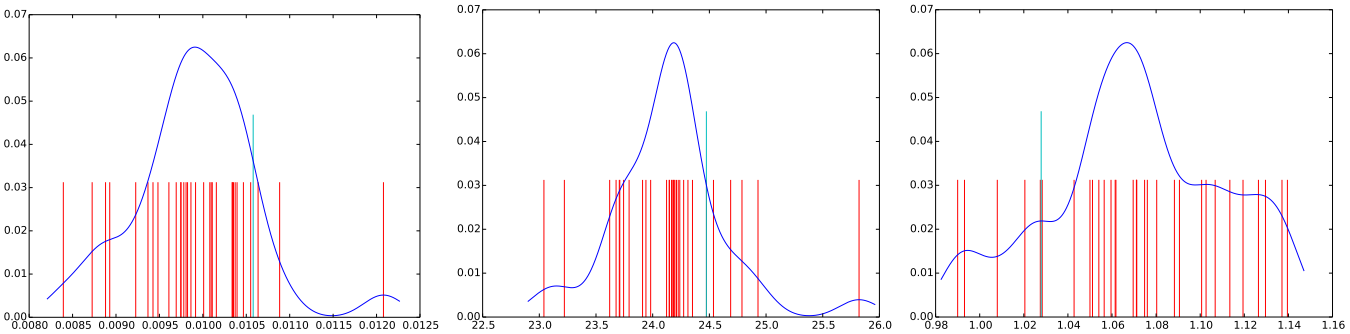


(a) Observed SST **(b)** Observed SST **(c)** Observed SST

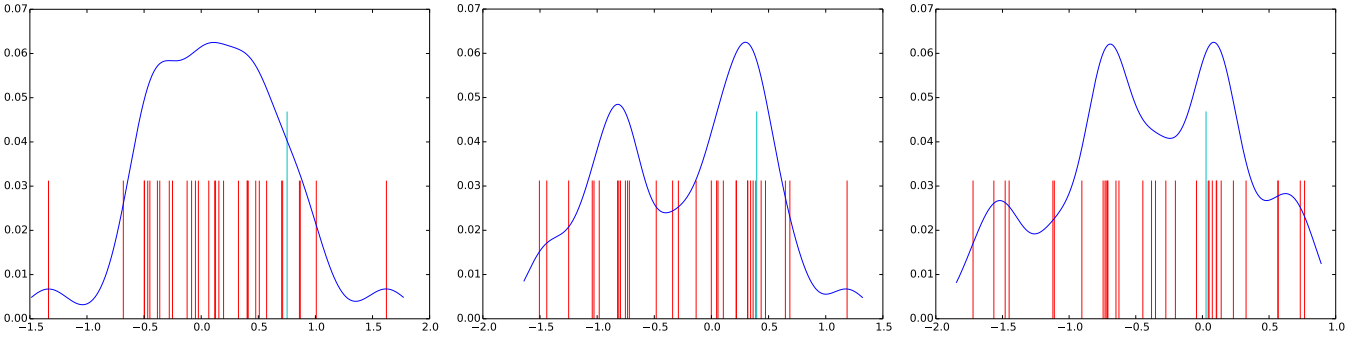
(d) Ocean meridional flow at 3347m **(e)** Ocean meridional flow at 3347m **(f)** Ocean meridional flow at 3347m

(g) Air meridional flow at $0.2p^*$ **(h)** Air meridional flow at $0.2p^*$ **(i)** Air meridional flow at $0.2p^*$

Figure 20. Rank histograms of different model variables at a given model level when using different parameters in the equivalent weights particle filter for the case that $Q = Q_1$.



(a) Surface level specific humidity at (20S,270E) **(b)** Seawater temperature at (1.875S,213.75E) at a depth of 48m **(c)** Seawater temperature at (51.875S,228.75E) at a depth of 3347m



(d) Meridional seawater flow at (15S,100.625E) at 3347m depth on day 75 **(e)** Meridional seawater flow at (15S,100.625E) at 3347m depth on day 161 **(f)** Meridional seawater flow at (15S,100.625E) at 3347m depth on day 177

Figure 21. Delta function representations of the marginal PDFs for different variables at a given instance in time for the experiment $Q = Q_1$, $p = 4d7$ and $\kappa = 1$. In red are shown the particles and in cyan we show the truth (raised for clarity). To guide the eye, we have overlaid in blue the results of fitting a Gaussian kernel, where the standard deviation of each Gaussian function is given by the quotient of the range in the marginal PDF and half the number of ensemble members.