Assessment of a Nonlinear Ensemble Transform Filter for High-Dimensional Data Assimilation

JULIAN TÖDTER

Institute for Atmospheric and Environmental Sciences, Goethe University, Frankfurt am Main, Frankfurt, Germany

PAUL KIRCHGESSNER AND LARS NERGER

Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany

BODO AHRENS

Institute for Atmospheric and Environmental Sciences, Goethe University, Frankfurt am Main, Frankfurt, Germany

ABSTRACT

This work assesses the large-scale applicability of the recently proposed nonlinear ensemble transform filter (NETF) in data assimilation experiments with the NEMO ocean general circulation model. The new filter constitutes a second-order exact approximation to fully nonlinear particle filtering. Thus, it relaxes the Gaussian assumption contained in ensemble Kalman filters. The NETF applies an update step similar to the local ensemble transform Kalman filter (LETKF), which allows for efficient and simple implementation. Here, simulated observations are assimilated into a simplified ocean configuration that exhibits globally high-dimensional dynamics with a chaotic mesoscale flow. The model climatology is used to initialize an ensemble of 120 members. The number of observations in each local filter update is of the same order resulting from the use of a realistic oceanic observation scenario. Here, an importance sampling particle filter (PF) would require at least 10^6 members. Despite the relatively small ensemble size, the NETF remains stable and converges to the truth. In this setup, the NETF achieves at least the performance of the LETKF. However, it requires a longer spinup period because the algorithm only relies on the particle weights at the analysis time. These findings show that the NETF can successfully deal with a large-scale assimilation problem in which the local observation dimension is of the same order as the ensemble size. Thus, the second-order exact NETF does not suffer from the PF's curse of dimensionality, even in a deterministic system.

1. Introduction

Data assimilation (DA) refers to the combination of predictions from numerical models with real-world observations. The resulting state estimates are of high relevance, particularly in atmospheric or oceanic modeling (e.g., Bennett 2002). Its application in model initialization represents an essential contribution to forecast quality. However, DA also allows one to consistently reconstruct the system state over extended time periods, and the resulting reanalyses are an extremely valuable data source for climate studies and diagnostics (e.g.,

Bengtsson et al. 2007). Additionally, systematic model or observation errors can be diagnosed from DA output (e.g., Haimberger 2007).

In principle, the DA problem is entirely determined by Bayes's theorem (e.g., Wikle and Berliner 2007). However, a major challenge in geophysical applications (e.g., van Leeuwen 2010) consists of the high dimensionality of the involved state and observation spaces, together with the lack of knowledge about the involved probability density functions (pdfs). This work focuses on sequential, ensemble-based square root filters that iteratively improve the state estimate throughout the assimilation window. They perform an ensemble prediction during the forecast step and update the prior ensemble with the current observations in the analysis step. Compared to variational methods (e.g., 4DVAR; Talagrand and Courtier 1987), which fit a whole trajectory,

Corresponding author address: Julian Tödter, Institute for Atmospheric and Environmental Sciences, Goethe University, Frankfurt am Main, Altenhöferallee 1, 60438 Frankfurt, Germany.
E-mail: julian.toedter@gmail.com

these filters are conceptually attractive and easy to implement since no tangent linear and adjoint models are required (Kalnay et al. 2007). Particle filters (PFs) offer a direct Monte Carlo solution to Bayes's theorem without a parametric assumption on the forecast pdf (Doucet et al. 2001). However, the likelihood weights are computed in a high-dimensional probability space in which the region of significant probability density is extremely narrow. Consequently, many ensemble members obtain insignificant weight, which leads to filter divergence (Snyder et al. 2008). This is known as the curse of dimensionality (Silverman 1986).

The ensemble Kalman filter (EnKF; Evensen 1994; Burgers et al. 1998) avoids this issue by assuming Gaussian distributions, where the required mean and covariance of the state are directly estimated from the forecast ensemble. Over the past two decades, the EnKF has evolved to a robust scheme that is applicable to large-scale systems with small ensemble sizes, such as in numerical weather prediction (e.g., Reich et al. 2011; Miyoshi and Kunii 2012) or oceanography (e.g., Nerger et al. 2007; Losa et al. 2012). Deterministic variants such as the (local) ensemble transform Kalman filter [(L)ETKF, Bishop et al. 2001; Hunt et al. 2007] avoid sampling noise in the analysis step by applying a matrix square root transform. However, the implicit Gaussian assumption leads to a linear update mechanism and renders the analysis suboptimal in nonlinear systems (Lei and Bickel 2011). Consequently, there is broad research activity toward enabling the applicability of non-linear filters in high dimensions (e.g., van Leeuwen 2009). A more recent development is the equivalent weights PF (EWPF; van Leeuwen 2010). It explores proposal densities to ensure that the particles lie in the important region and exhibit small variability of the weights, which allows for applicability to large-scale DA (van Leeuwen and Ades 2013; Ades and van Leeuwen 2015). However, the EWPF only works in models with stochastic model errors and requires an adequate adaption of the forecast step to properly guide the particles (Ades and van Leeuwen 2013), which increases the implementation complexity.

While the EWPF aims at considering the full analysis pdf, approximations to fully nonlinear filtering have been suggested as well (van Leeuwen 2009). This work builds upon Tödter and Ahrens (2015), who introduced the nonlinear ensemble transform filter (NETF). It is based on the nonlinear ensemble adjustment filter (NLEAF; Lei and Bickel 2011), which updates each member with perturbed observations. In contrast, the NETF applies a matrix square root transform, similar to the ETKF, to ensure that the analysis mean and covariance exactly match the Monte Carlo estimates of the Bayesian expectations. To overcome the curse of dimensionality inherent to most PFs, the effective dimensionality is reduced by localizing the analysis as done in the LETKF. The NETF only acts in the analysis step, which allows the use of deterministic models. The empirical findings resulting from assimilation experiments with ensemble sizes not larger than 100 (Tödter and Ahrens 2015) point out that it produces reasonable and stable analyses, even in higher-dimensional and chaotic systems with state dimensions up to 10^3. In these experiments with simplified models, the NETF outperforms the stochastic EnKF and the (L)ETKF in the presence of nonlinearity. Additionally, it improves upon the stochastic NLEAF in larger-dimensional cases because of the deterministic update mechanism. Thus, the NETF exhibits potential applicability to large-scale DA.

Motivated by these results, this work assesses the NETF's performance in an advanced circulation system. This system is characterized by more realistic and higher-dimensional dynamics with a state dimension that is orders of magnitudes larger than studied before. For that purpose, we apply the Nucleus for European Modelling of the Ocean (NEMO v3.3; Madec 2012), a state-of-the-art ocean general circulation model (OGCM), within an identical twin experiment that includes a realistic observation scenario and a challenging filter initialization. The principal objective of this paper is to demonstrate that the generic NETF is applicable to high-dimensional DA, even though it only relies on the Bayesian weights in the analysis step. In particular, no proposal densities, as in the EWPF, or other modifications need to be used. Furthermore, a comparison with the LETKF reveals aspects that influence the filter's performance. This allows us to gain more insight into its behavior and to enable future improvements.

The remainder of this paper is organized as follows. Section 2 contains a brief review of ensemble square root filters and establishes the formal analogy of NETF and ETKF. In section 3, the configuration of the ocean model and its circulation characteristics are presented. Section 4 specifies the observation scenario and filter setup. The results of the assimilation experiments are shown and discussed in section 5. Finally, section 6 draws the principal conclusions and outlines potential continuative research paths.

2. Ensemble square root filtering

In the following work, we exclusively concentrate on the analysis step and, therefore, neglect any time indices. We consider a dynamical system $\mathcal{M}$, and the state vector $\mathbf{x}$ of size d contains the prognostic model variables. We assume that an ensemble of m independent and

identically distributed prior states $\{\mathbf{x}_f^i\}_{i=1,\dots,m}$ is available. Usually, it arises from a preceding forecast step and can be regarded as the best representation of the forecast pdf, $p(\mathbf{x})$ (i.e., before assimilating the new observation). The ensemble vectors are stored in the columns of the $d \times m$ matrix $\mathbf{X}_f$. Subtracting the prior ensemble mean, $\bar{\mathbf{x}}_f$, from each column yields the ensemble perturbation matrix $\mathbf{X}_f'$. The observation, represented by the vector $\mathbf{y}$ of size k, contains all measurements available at that time. The observation operator $\mathcal{H}(\mathbf{x})$ maps any model state into observation space. We define the $k \times m$ matrix $\mathbf{Y}_f$ that contains the mapped ensemble vectors [i.e., $\mathbf{y}_f^i = \mathcal{H}(\mathbf{x}_f^i)$]. The observational uncertainty is represented by the likelihood pdf, $p(\mathbf{y} \,|\, \mathbf{x})$, which may be of arbitrary form in the NETF. In this work, we restrict the likelihood to a Gaussian distribution with covariance $\mathbf{R}$ {i.e., $\mathcal{N}[\mathbf{y}; \mathcal{H}(\mathbf{x}), \mathbf{R}]$}.

As the prior pdf is usually unknown, approximate solutions have been developed, thus leading to a large variety of filters and smoothers. Second-order exact filters aim at an analysis ensemble such that its mean, $\bar{\mathbf{x}}_a$, and covariance, $\mathbf{P}_a = [1/(m-1)]\mathbf{X}_a'\mathbf{X}_a'^{\mathrm{T}}$, exactly match some specified values. Both the ETKF and NETF can be described by the same square root filtering framework (Nerger et al. 2012). The analysis ensemble, $\mathbf{X}_a = \bar{\mathbf{X}}_a + \mathbf{X}_a'$, incorporates the observation by updating the prior mean and perturbations as follows:

$$\bar{\mathbf{x}}_a = \bar{\mathbf{x}}_f + \mathbf{X}_f'\mathbf{w}, \tag{1}$$

$$\mathbf{X}_a' = \mathbf{X}_f'\mathbf{T}\mathbf{\Lambda}. \tag{2}$$

The mean analysis increment is a linear combination of the prior perturbations, and it is determined by a weight vector $\mathbf{w}$, while the $m \times m$ matrix $\mathbf{T}$, typically a matrix square root, transforms $\mathbf{X}_f'$ into analysis perturbations, $\mathbf{X}_a'$. The optional random $m \times m$ matrix $\mathbf{\Lambda}$ constitutes a rotation in the ensemble subspace. The weight vector and transform matrix depend on the prior ensemble and observational properties. Their specific forms in the ETKF and NETF are presented next.

a. Linear filtering and the ETKF

The Kalman filter (KF; Kalman 1960) assumes that the prior as well as the likelihood pdf are Gaussian. Then, according to Bayes's theorem, the analysis pdf remains Gaussian, and its mean and covariance can be computed analytically from the prior moments. Ensemble Kalman filters extract the prior moments from an ensemble integration. This partly accounts for the nonlinear nature of the model (Anderson 2012). Stochastic EnKFs (Burgers et al. 1998; Houtekamer and Mitchell 1998) introduce additional sampling errors (Whitaker and Hamill 2002). In contrast, deterministic

EnKFs (Tippett et al. 2003) transform the prior ensemble such that its first two moments exactly match the theoretical KF values. The ETKF (Bishop et al. 2001) is determined by the update applied with Eqs. (1) and (2) if the weight vector and transform matrix are chosen as follows:

$$\mathbf{w}^{\mathrm{ETKF}} = \frac{1}{m-1}(\mathbf{T}^{\mathrm{ETKF}})(\mathbf{T}^{\mathrm{ETKF}})^{\mathrm{T}}\mathbf{Y}_f'^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y} - \bar{\mathbf{y}}_f), \tag{3}$$

$$\mathbf{T}^{\mathrm{ETKF}} = \sqrt{m-1}[(m-1)\mathbf{I}_k + \mathbf{Y}_f'^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{Y}_f']^{-1/2}. \tag{4}$$

Following Wang et al. (2004), the unique symmetric square root should be chosen to compute the transform matrix in Eq. (4).

b. Nonlinear filtering and the NETF

The Gaussian assumption in the EnKFs leads to a robust solution, but the resulting filters are suboptimal in nonlinear systems. In contrast, PFs (Doucet et al. 2001) offer a nonparametric solution of the assimilation problem. The basic importance sampling PF (Gordon et al. 1993) interprets the prior ensemble as a mixture of delta distributions that approximate the prior pdf [i.e., $p(\mathbf{x}) \approx 1/m\sum_i \delta(\mathbf{x} - \mathbf{x}_f^i)$ for an equally weighted ensemble]. Using this particle representation, Bayes's theorem states that the analysis pdf is approximated as

$$p(\mathbf{x} \,|\, \mathbf{y}) \approx \sum_{i=1}^m w^i \delta(\mathbf{x} - \mathbf{x}_f^i), \quad \text{where} \quad w^i = \frac{p(\mathbf{y} \,|\, \mathbf{x}_f^i)}{\sum_{j=1}^m p(\mathbf{y} \,|\, \mathbf{x}_f^j)}. \tag{5}$$

The analysis weights w^i contain all the information extracted from the observation. For Gaussian observation errors,

$$w^i \propto \mathcal{N}[\mathbf{y}; \mathcal{H}(\mathbf{x}_f^i), \mathbf{R}] \propto \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{y}_f^i)^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y} - \mathbf{y}_f^i)\right]. \tag{6}$$

While the EWPF (van Leeuwen 2010) attempts to sample the full analysis pdf, the NETF (Tödter and Ahrens 2015) represents a generic, yet approximative solution to nonlinear filtering. It applies a deterministic update mechanism in analogy to the ETKF such that the first two moments of the analysis ensemble exactly match the unbiased PF estimates of the Bayesian mean and covariance. The analysis mean and perturbations of the NETF are computed with Eqs. (1) and (2) by using the Bayesian weights defined in Eq. (5) and the following transform matrix:

$$\mathbf{w}^{\mathrm{NETF}} = (w^1, \dots, w^m)^{\mathrm{T}}, \tag{7}$$

$$\mathbf{T}^{\mathrm{NETF}} = \sqrt{m}[\mathbf{W} - \mathbf{w}^{\mathrm{NETF}}(\mathbf{w}^{\mathrm{NETF}})^{\mathrm{T}}]^{1/2}. \tag{8}$$

Here, the matrix $\mathbf{W} \equiv \text{diag}(\mathbf{w}^{\text{NETF}})$ is the diagonal $m \times m$ matrix formed by the weights. As in the ETKF, the unique symmetric square root is used in Eq. (8). For stability reasons, the random rotation with $\mathbf{\Lambda}$ [see Eq. (2)] is mandatory in the NETF. Details on the theoretical background of the NETF can be found in Tödter and Ahrens (2015) and Tödter (2015).

The ETKF and NETF are described by update equations of identical form. They only differ in the explicit equations of the weight vector and the transform matrix, thus allowing simple implementation of the NETF. Both filters perform the update in the ensemble subspace, and their computational expenses are similar for a given ensemble size.

c. Localization and inflation

In large-scale applications, most ensemble filters require modifications because the ensemble size is usually much smaller than the system's dimension (e.g., Anderson 2012). Observation localization (OL; Hunt et al. 2007) suppresses spurious correlations associated with distant locations and increases the rank of the analysis covariance by partitioning the state vector into subsets, the so-called local domains. Typically, a local domain contains all state variables at each grid point or in each vertical column (e.g., Houtekamer and Mitchell 1998; Losa et al. 2012). Each local domain is updated independently using only a part of the global observation vector. Typically, nearby observations are included by choosing an appropriate localization radius. The overlapping observation regions of different local domains ensure that the covariances in the ensemble covariance matrix are taken into account. Mathematically, the observation impact is reduced with distance by multiplying $\mathbf{R}^{-1}$ by an appropriate correlation function through the use of a Schur product (Whitaker and Hamill 2002; Kirchgessner et al. 2014). This term appears in Eqs. (3)–(4) and (6) for the LETKF and NETF, respectively. Therefore, OL can be directly adopted for the NETF. It reduces the effective dimension of the observation space, where the Bayesian weights are computed. This counteracts filter divergence in higher dimensions. A detailed algorithm for the localized NETF is given in Tödter and Ahrens (2015, their section 4f).

Inflation is usually applied to counteract the tendency of the ensemble to underestimate the uncertainty. Multiplicative inflation adjusts the prior ensemble perturbations by $\mathbf{X}'_f \rightarrow \sqrt{\gamma}\mathbf{X}'_f$. This increases the prior covariance by a factor of γ, which is slightly larger than 1 (Anderson and Anderson 1999). The localization radius and inflation factor constitute the main tuning factors for most ensemble filters. They allow the EnKFs to achieve results that are competitive to established variational schemes (e.g., Buehner et al. 2010; Fairbairn et al. 2014), and they can be used similarly in the NETF.

3. Ocean model and its configuration

This section gives an overview of the ocean model applied in our experiment. Additionally, we discuss the characteristic circulation that emerges from our model setup.

a. Model characterization

The NEMO model (Madec 2012, http://www.nemo-ocean.eu) contains an OGCM that numerically solves the primitive equations that determine the ocean's dynamics and thermodynamics. Its prognostic variables are temperature T, salinity S, zonal and meridional fluid velocity (U, V), as well as sea surface height (SSH). Processes that occur on a subgrid scale, such as turbulence or convection, are parameterized. For example, the diffusive fluxes are described by second-order closure schemes that use the gradients of the large-scale fields together with associated eddy coefficients. The NEMO model applies a leapfrog time stepping scheme for all nondiffusive terms in conjunction with a Robert–Asselin time filter. For the lateral diffusive and damping parts of the equations, a forward scheme is used, while in the vertical direction, an implicit scheme is required. The spatial discretization applies second-order central finite differences on a curvilinear Arakawa C-type grid.

Boundary conditions are required to determine the fluxes of momentum, mass, and energy at the ocean interfaces. At the bottom, usually all fluxes are assumed to vanish. In contrast, the fluxes at the upper boundary are derived from atmospheric forcing fields by using bulk formulations or by applying analytical prescriptions. The lateral boundary conditions depend on the specific model setup.

b. Model configuration and setup

The aim of this work is to demonstrate the large-scale applicability of the NETF in principle. Therefore, we chose a simplified configuration of the NEMO model with a wind-driven ocean in a closed basin. It idealizes the ocean circulation that is representative for the midlatitudes (e.g., the Gulf Stream in the North Atlantic). This configuration is an established test bed in oceanography (e.g., Carrier and Robinson 1962; Cosme et al. 2010; Lévy et al. 2010; Yan et al. 2014). The forcing induces a large-scale double-gyre circulation that is complemented by mesoscale eddies.

The model is applied to a closed, rectangular basin in the North Atlantic in an area ranging from 60° to 30°W

and from 24° to 44°N. The horizontal grid employs a distance of 0.25°. This resolution corresponds to an eddy-permitting setup. This discretization results in $n_x = 121$ grid points in the zonal and $n_y = 81$ grid points in the meridional direction. In the vertical, 11 layers are defined with exponentially increasing thickness. The bottom is assumed to be flat and located at a depth of 5054 m. The leapfrog scheme is used with a time step of 15 min and a smoothing parameter of 0.1 for the Robert–Asselin time filter. External gravity waves are damped explicitly in the horizontal momentum equations according to Roullet and Madec (2000). The space variation of the lateral eddy coefficients is constrained to the horizontal, while the vertical ones are fixed. Here, the default values $1.2 \times 10^{-4} \, \mathrm{m^2 \, s^{-1}}$ for momentum and $1.2 \times 10^{-5} \, \mathrm{m^2 \, s^{-1}}$ for temperature are set. The parameterization of lateral mixing is realized by a biharmonic diffusion operator with an eddy coefficient of $-8 \times 10^{10} \, \mathrm{m^4 \, s^{-1}}$ for both momentum and temperature. At the bottom, linear friction with a drag coefficient of $4 \times 10^{-4} \, \mathrm{m \, s^{-1}}$ is prescribed, while the lateral boundaries are assumed to be frictionless.

The model is entirely forced by a prescribed zonal wind stress τ_x, which varies with latitude ϕ, but is constant in time t and longitude λ (Cosme et al. 2010):

$$\tau_x(\phi, \lambda, t) = -\frac{1}{10} \cos\left[\frac{2\pi}{\Delta \phi}(\phi - \phi_1) \right] \quad (\mathrm{N \, s^{-2}}). \quad (9)$$

Here, $\phi_1 = 24°$ is the latitude at the southern boundary and $\Delta \phi = 20°$ is the latitude range of the domain. The forcing is symmetric with respect to $\phi = 34°$ in the domain center, where the zonal wind directs to the east between 29° and 39°. In the northern and southern parts, the wind blows to the west. No freshwater influx is considered, and therefore, salinity does not vary and is disregarded in all following considerations.

c. Model initialization and truth run

We create a reference trajectory for the identical twin experiment, which is referred to as the truth, by initializing the model with an ocean at rest. Salinity remains constant at a prescribed value of $35.5 \, \mathrm{g \, kg^{-1}}$. Each vertical ocean column is initialized by the same temperature profile that corresponds to stratification typically observed in ocean climatologies (Chassignet and Gent 1991):

$$T(x, y, z, t = 0) = 25 + 24.06(e^{-z/800\mathrm{m}} - 1) \quad (°C) \quad (z \text{ in m}). \quad (10)$$

After initialization, the model is integrated forward for 75 years. One year is idealized to 360 days. The first 50 years are considered as the spinup phase toward reaching the model climatology. Afterward, the dynamic equilibrium of the model with respect to the applied forcing is reached (Cosme et al. 2010). The actual DA experiment is performed in year 75, while the years 51–74 allow us to estimate the model climatology.

Before turning to the actual assimilation problem, it is useful to gain an overview of the dynamical structure of the wind-driven ocean. For this purpose, we refer to the true initial state in year 75. Figure 1 shows the surface fields. The large-scale double-gyre circulation, which intensifies at the western boundary, is directly visible in the SSH and T fields. In the west, the inhomogeneous zonal wind forcing leads to boundary current (see Fig. 1d). They support an eastward jet in the center, which is located around $\phi = 34°$ in Fig. 1c. However, the jet is subject to dynamic instabilities. Consequently, chaotic behavior can be observed that leads to a mesoscale flow besides the large-scale double-gyre circulation. It is characterized by eddies, which exhibit notable local differences in velocity, temperature, and SSH, and they may also influence the large-scale flow (Holland 1978). This underlines the model's high-dimensional dynamics, which is important concerning the aim of our study. More quantitatively, the dynamics of a system can be regarded as high dimensional if a considerable number of eigenvectors are necessary to explain most of its variability. The black line in Fig. 2 shows the amount of variability that can be explained for a model run over 4 years. This is derived from a multivariate singular value decomposition (Lermusiaux and Robinson 1999) of states that were sampled each second day. To explain 90% of the variability, about 200 eigenvectors are needed while 500 eigenvectors can explain 99%. These numbers are slightly smaller, but of the same order of magnitude as was found for data from atmosphere or atmosphere–ocean general circulation systems (Achatz and Branstator 1999; Achatz and Opsteegh 2003).

4. Experimental setup

Having shown the model and its circulation characteristics, the remaining task is to define an assimilation problem by simulating observations. Furthermore, we now describe and discuss the filter setup.

a. Observation scenario

Even though our experiment employs a simplified ocean configuration with simulated observations, we create a situation that resembles a typical ocean assimilation problem. Cosme et al. (2010) and Yan et al. (2014) applied the NEMO configuration described above to investigate different DA algorithms. We closely follow
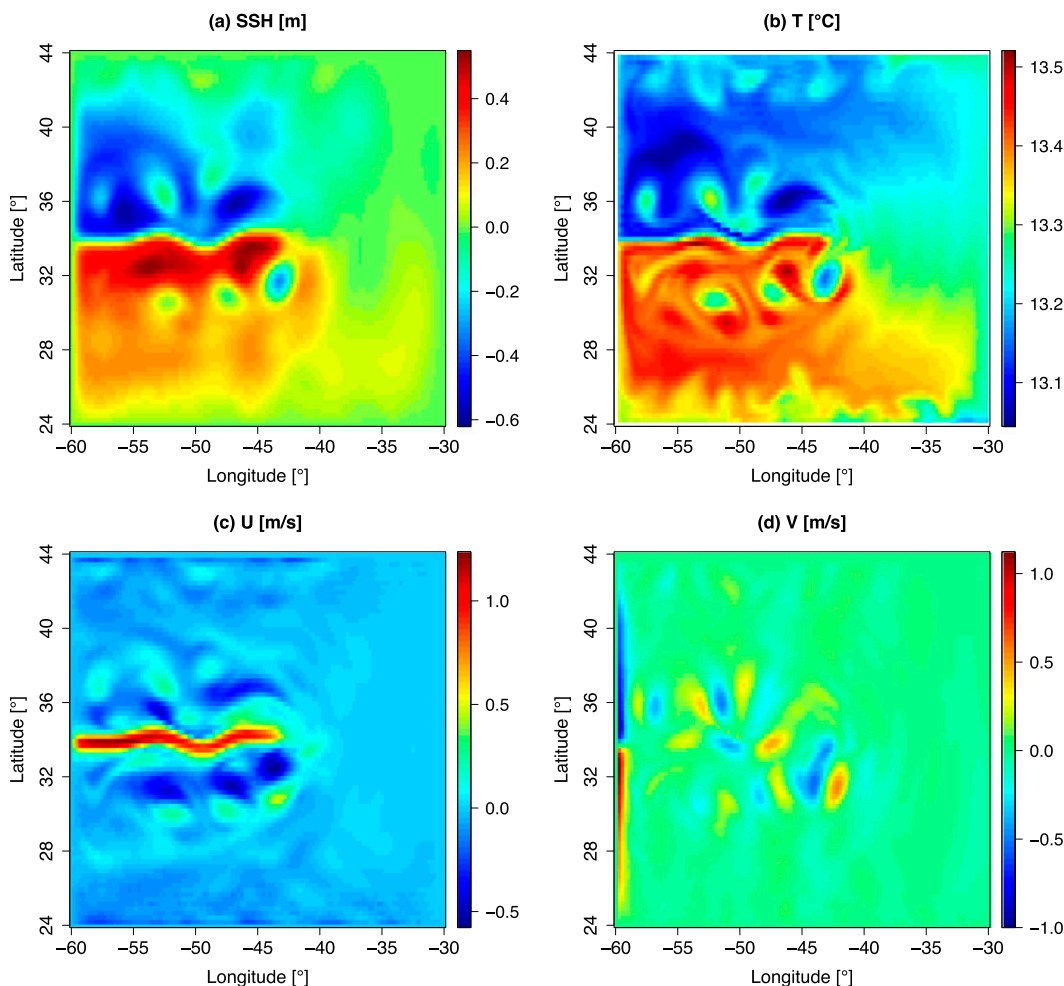
FIG. 1. Snapshots of the model fields at the beginning of year 75 at the (a) surface (SSH) or in the (b)–(d) first layer (*T, U, V*), respectively. They visualize the double-gyre circulation with a central jet, meridional currents at the western boundary, and mesoscale eddies.

their setup and simulate two types of observations with distinct spatial structures and densities.

First, *Environmental Satellite* (*Envisat*) SSH observations (e.g., Durrant et al. 2009) are simulated on the flight tracks falling into our model domain during the year 2009. This yields on average 150 measurements at each analysis time. As the observation operator, a bilinear interpolation from the model grid to the observation grid is applied. The observation error, which summarizes the measurement and representivity errors, is modeled as uncorrelated Gaussian noise with a standard deviation of 0.06 m. Second, temperature observations mimic the Argo network (Carval et al. 2013) on a horizontal $3° \times 3°$ grid that is shifted at each analysis time to reflect the nonstationarity of the profilers. Vertically, all Argo levels located below the first

layer's center are considered for each profile. The standard deviation of the observation error variance is 0.3°C, and the observation operator is given by trilinear interpolation.

The observations are assimilated every second day (i.e., 192 model time steps define one analysis cycle). Since the truth run covers one idealized year, 180 analysis steps are performed in total. Figure 3 visualizes the observation characteristics on day 8 (i.e., at the fourth analysis step). Figure 3a is a horizontal snapshot that shows both the *Envisat* locations and the Argo network available at this time. The SSH observations along the satellite tracks are indicated by their color. Figure 3b gives an example of the temperature profiles along the $\lambda = -50°$ line and their location in the vertical NEMO grid.
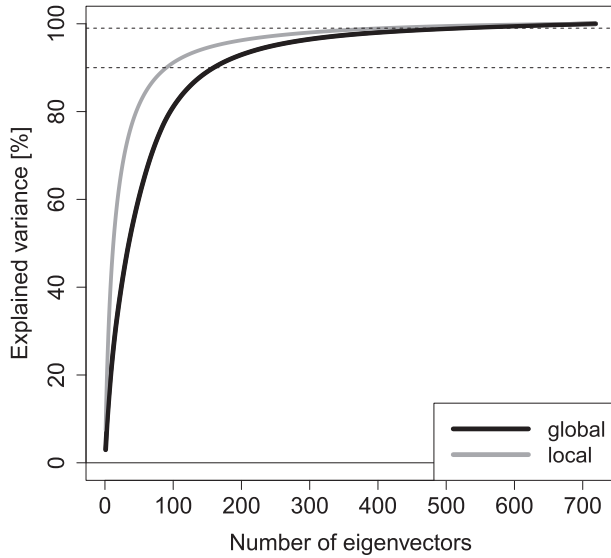
FIG. 2. Number of eigenvectors needed to explain a certain percentage of the variability of the system, derived from a model run over four years (sampled every other day). The black line refers to the global state, while the gray line is based on the average singular value spectrum of the state vector in each local region, as defined by the localization radius of 2.5°. The dashed lines mark the 90% and 99% thresholds.

b. Technical implementation

For the assimilation experiments, the NEMO model was coupled with the parallel DA framework (PDAF; Nerger and Hiller 2013, available online at http://pdaf.awi.de). In PDAF, numerous EnKF variants including the LETKF are already implemented and used for different applications (e.g., Losa et al. 2012; Fournier et al.

2013). Because of the similarity of the LETKF and NETF equations (see section 2), almost the same implementation is used. Only minor changes were needed to account for the different transform matrices and weight vectors. At an analysis time, each ensemble state vector $\mathbf{x}_f^i$ ($i = 1, \ldots, m$) is constructed from the model fields (T, U, V, SSH) and mapped into observation space by interpolation (see section 4a), thus yielding $\mathbf{y}_f^i = \mathcal{H}(\mathbf{x}_f^i)$. Localization is implemented as described in Nerger et al. (2006) by performing a local analysis for each vertical column. For each column, nearby measurements (see section 4c) are selected to create the local observation vector $\mathbf{y}$ and the local part of $\mathbf{y}_f^i$. Then, a local analysis is computed as explained in section 2. Finally, all local analyses are accumulated and each analysis state vector $\mathbf{x}_a^i$ is distributed into the model fields to initialize the next forecast phase.

Technically, NEMO has been extended to call PDAF directly from the model code, and it has been compiled into a single program that contains the full assimilation system (Tödter 2015, see chapter 6.4.1). Therefore, all parallel features that are already present in the PDAF library are directly usable. In addition, the same model or observation specific routines can be applied directly to both filters and the twin experiments can be conveniently performed.

c. Filter setup

For our experiment, we apply the NETF algorithm as presented in section 2 without any model-specific modifications. A fixed multiplicative inflation as shown in section 2c is applied. We found that a factor of $\gamma = 1.02$ yields the best performance for the NETF (see section 5).
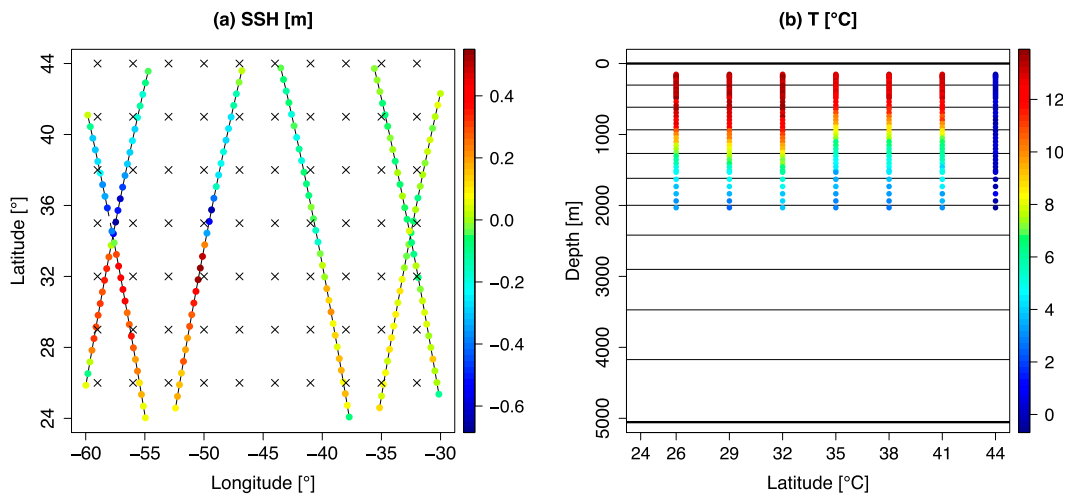

FIG. 3. Observation characteristics on day 8. (a) The horizontal domain is shown, together with the Argo profiler locations (crisscrosses) and the synthetic SSH observations (colored) on the *Envisat* tracks (thin lines). (b) The vertical grid of 11 layers is visualized, and embedded are the artificial Argo temperature profiles (46 values each) along the $\lambda = -50°$ longitude line. At $\phi = 44°$, the true temperature field is zero because of the lateral boundary conditions.

TABLE 1. Statistics on the observations and on localization. All values are temporal averages over the 180 analysis steps. The total number of local domains is 9801 (all vertical columns) and this table refers to the localization radius of 2.5°.

Statistics	Only SSH	Only T	Combined
Global observation dimension	145	3128	3273
Max local observation dimension	16	184	199
Avg local observation dimension	7	90	94
No. of local domains with observations	6262	9654	9742
No. of local domains without observations	3539	147	59

To localize the analysis, the $n_x \times n_y = 9801$ vertical ocean columns are updated independently. A horizontal localization radius of 2.5° is used in conjunction with a fifth-order polynomial correlation function [Gaspari and Cohn 1999, their Eq. (4.10)] to reduce the observational influence with distance. This radius, which roughly corresponds to 250 km, arises from experiences in prior LETKF experiments in this setting and is in agreement with similar experiments and the statistical properties of the system (Cosme et al. 2010; Yan et al. 2014). In tuning experiments, we found that this radius also allows the NETF to achieve its best performance. However, in other applications, the optimal radius might differ for both filters. Table 1 provides some statistics for the observation dimensions from a global and local point of view. On average, at each analysis time, 3273 independent observations (mostly temperature data) are assimilated. The average local observation dimension is nearly 100, but it may reach up to 200. Thus, on average, the likelihood weights are evaluated in a 100-dimensional subspace.

The filter is initialized from model climatology. To allow an investigation of its influence on the filter performance, we prepared both a long- and short-term climatology by extracting a model state from the truth run every other month within the years 51–74 and 65–74, respectively. Table 2 summarizes the basic statistics of these climatologies. They exhibit a similar deviation from the true initial state, but the long-term climatology has significantly more spread, particularly in the temperature field. The ensemble of 120 members is constructed from the corresponding climatology via second-order exact sampling (Pham 2001). This guarantees that the ensemble contains the dominant error directions, but it could only explain at best 80% of the system's variability (see Fig. 2). Additionally, following Table 2, we increased its spread by 25% to ensure that the truth lies within the ensemble range. The initial members contain mesoscale

TABLE 2. Basic properties of the climatological samples that are used to generate the initial ensemble (see section 4c). The statistical measures, defined in section 5b, are evaluated for each model variable. The RMSE is computed for the climatological mean state with respect to the true initial state in year 75.

Statistics	Climatology	T (°C)	SSH (m)	U (m s^{-1})	V (m s^{-1})
RMSE	Years 65–74	0.3336	0.0775	0.0641	0.0546
	Years 51–74	0.2931	0.0810	0.0627	0.0544
Spread	Years 65–74	0.1151	0.0662	0.0527	0.0524
	Years 51–74	0.2056	0.0772	0.0583	0.0571

features, however, their mean only reflects the large-scale double-gyre circulation and the central jet, as visualized in Fig. 4 for the long-term climatology.

d. Complexity of the assimilation problem

The twin experiment employs a state-of-the-art, complex circulation system, as used in ocean or atmospheric modeling, including advanced physical parameterizations. The global state vector consists of all prognostic variables of NEMO except salinity [i.e., T, U, V (three-dimensional fields), and SSH (two-dimensional)]. Therefore, the global state dimension is $3 \times (121 \times 81 \times 11) + 121 \times 81 = 333\,234 \approx 3.3 \times 10^5$. On the global scale, the system exhibits high-dimensional dynamics, as reflected by the black line in Fig. 2 and discussed in section 3c. The dynamics are characterized by the chaotic mesoscale flow. Each *local* region, as defined by the localization radius, contains nearly 300 water columns with about 1×10^4 state variables. The gray line in Fig. 2 shows the average amount of eigenvectors needed to explain the system's variance in a local region. It is smaller than for the global case, but the difference is not very pronounced. Thus, because of the eddies, the dynamics also exhibit considerably high variability on the local scale.

The number of SSH and T observations at each analysis step is about 3300, which is one order of magnitude larger than the ensemble size of 120. Yet, the average *local* observation space dimension is of the same order as the ensemble size. This is a typical situation in large-scale ocean problems (e.g., Yan et al. 2014). For the NETF as a particle-based technique, it represents a major challenge. Ordinary PFs (with or without resampling) would require at least 10^6 (10^{11}) members to assimilate 100 (200) independent observations (Snyder et al. 2008). However, in atmospheric DA applications, the local observation dimension may strongly exceed the ensemble size (e.g., Buehner et al. 2010). This remains a limitation of our study.

Finally, the initial ensemble chosen here does not contain information about the actual mesoscale flow.
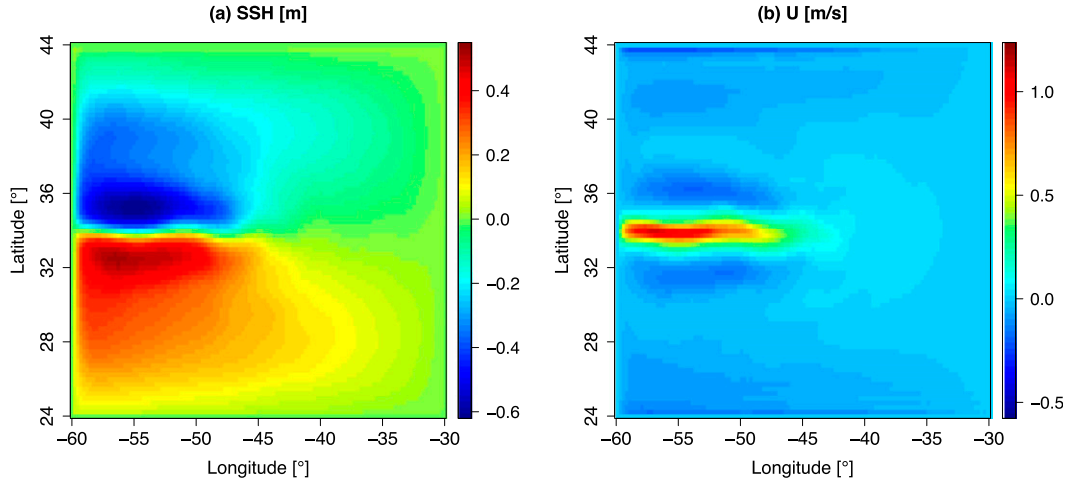
FIG. 4. Ensemble mean state at the initial time for (a) SSH and (b) U (first layer). The mesoscale patterns average out over the initial ensemble. For comparison, the true initial fields are shown in Fig. 1 (with identical color legend).

Hence, we not only assess the capability of the NETF to track the truth, but also to converge to it at first. In summary, our experiment considers a full three-dimensional ocean including its thermodynamics, together with a realistic observation scenario and a climatological initialization. For comparison, Ades and van Leeuwen (2015) initialized the EWPF around the truth in a highly nonlinear single-layer primitive equation system with a state dimension of 6×10^4 using only 24 particles. We conclude that the experiment conducted here is suited to assess the NETF's general applicability for nonlinear, high-dimensional DA.

5. Results and discussion

This section presents and discusses the results of the experiment described above. For comparative reasons, we also integrated the initial ensemble throughout the time window but without assimilating any observations. This free ensemble serves as a reference to show the impact of the observations on the model evolution.

a. Qualitative evaluation

For now, we apply the long-term climatology to create the initial ensemble, as presented in section 4. Thus, the free ensemble can only deliver climatological information of the system during the forecast phase (i.e., the large-scale double-gyre circulation created by the wind forcing). However, it is not able to resolve the mesoscale patterns, which are of a chaotic nature and average out over the free ensemble. This is supported by snapshots of the SSH fields on day 260 in Fig. 5. In principle, the mean of the free ensemble, Fig. 5b, has not changed compared to the initial time, as revealed by a comparison with

Fig. 4a. In contrast, the NETF analysis at that time, Fig. 5c, yields an estimate that closely resembles the true SSH field including the mesoscale features, as shown in Fig. 5a. The field difference, plotted in Fig. 5d, shows only small deviations of at most 10%. They mainly concern the magnitude of the mesoscale perturbations in the dynamically very active region in the domain center, which is highlighted by the box in Fig. 5d. These qualitative findings are similarly observed for all other variables as well, including the unobserved velocity fields. These results demonstrate that the NETF successfully assimilates the observations in this high-dimensional problem.

b. Quantitative evaluation

1) EVALUATION MEASURES

The main evaluation criterion used for filter performance is the root-mean-square error (RMSE) of the ensemble mean at each analysis time level $j \in \{0, 1, 2, \ldots, 180\}$. It is based on the spatial average of the squared deviations, and it is computed separately for each model variable according to

$$\text{RMSE}_j(X) = \sqrt{\frac{\sum_{\text{gp}} \left(\overline{X_j} - X_{j,\text{true}} \right)^2}{\dim(X)}}. \qquad (11)$$

Here, X stands for each of the four different model fields (T, U, V, SSH) and $\overline{X_j} = 1/m \sum_{i=1}^{m} X_j^i$ represents the ensemble mean field. The term $\dim(X)$ is the dimension of the field (i.e., 107 811 for T, U, and V and 9801 for SSH), and $\sum_{\text{gp}}$ indicates summation over all model grid points. As an overall measure, the RMSE may not capture all performance details, but a low RMSE represents a necessary condition for a successful analysis. Ensemble
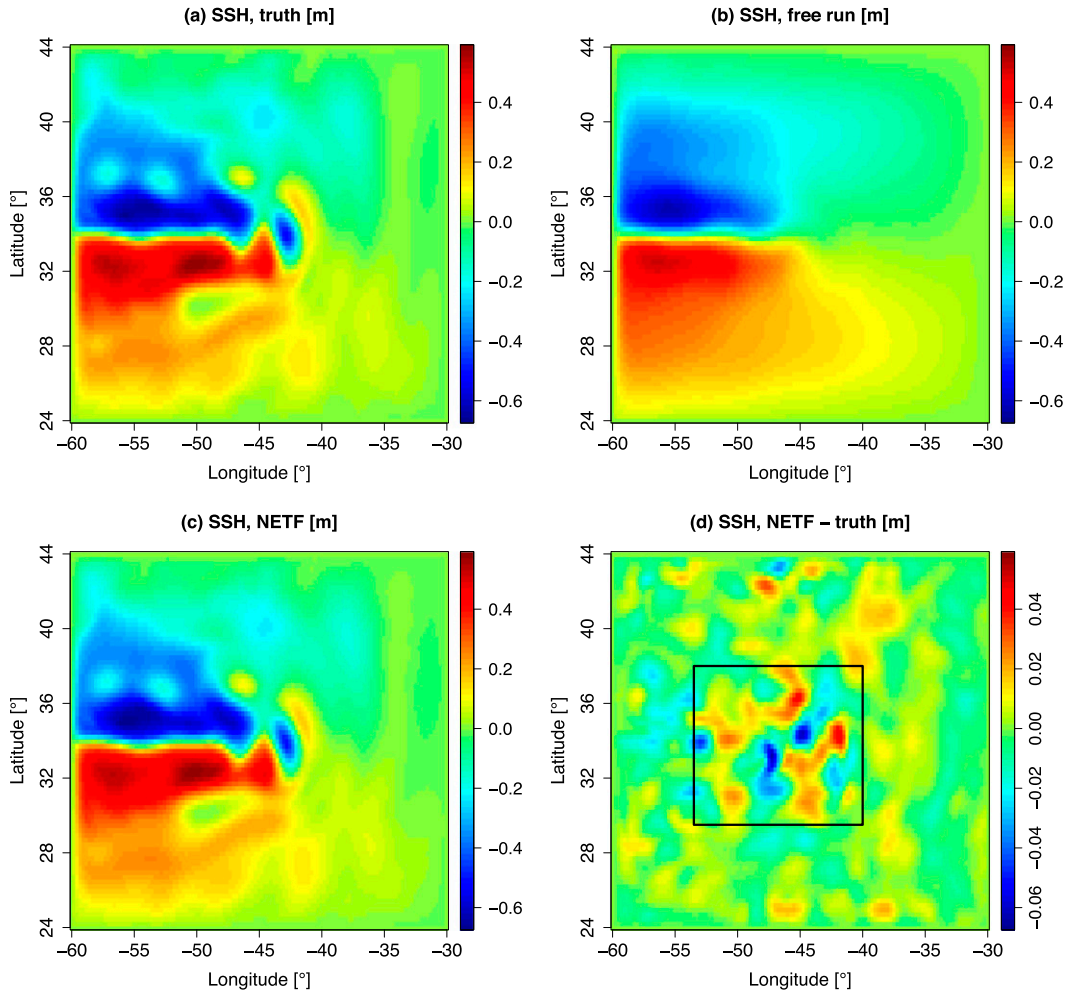
FIG. 5. Results for SSH at day 260. Shown are the fields of (a) truth, (b) free ensemble mean, and (c) NETF ensemble mean. (d) The field is the difference between the NETF analysis and the truth. Note that in (d), the color scale is enlarged by a factor of 10, as otherwise both fields can be barely distinguished. The box marks the region of relatively more pronounced deviations.

filters allow one to calculate an *estimated* RMSE by the field-averaged ensemble spread:

$$\text{SPREAD}_j(X) = \sqrt{\frac{\sum_{\text{gp}} \left[\frac{1}{m-1} \sum_{i=1}^{m} (X_j^i - \overline{X_j})^2 \right]}{\dim(X)}}. \quad (12)$$

This measure can be computed independently from the truth. In a well-calibrated DA system, it should be of similar magnitude as the RMSE, since then the ensemble distribution will allow for reliable diagnostics of the filter uncertainty (e.g., Palmer et al. 2005; Hopson 2014).

The RMSE enables a general judgment of the analysis quality. However, it only evaluates the ensemble mean and neglects higher-order moments. To assess the quality of the ensemble distribution as well, we also apply the probabilistic continuous ranked probability score (CRPS, e.g., Gneiting et al. 2007; Tödter and Ahrens 2012) for the comparison with the LETKF. The CRPS accounts for the distribution of probability mass around the truth for any state vector component x. Specifically, let $f^j(x)$ be the empirical ensemble pdf. The CRPS quantifies the difference between the cumulative density functions (cdfs) of the ensemble and the verifying truth (a Heaviside step function, Θ) in probability space by using a quadratic norm:

$$\text{CRPS}(x_j) = \int_{-\infty}^{\infty} [F^j(x) - \Theta(x - x_{\text{true},j})]^2 \, dx. \quad (13)$$

The empirical ensemble cdf, $F^j(x) = \int f^j(x) \, dx = 1/m \sum_{i=1}^{m} \Theta(x - x_j^i)$, is a piece-wise constant function,
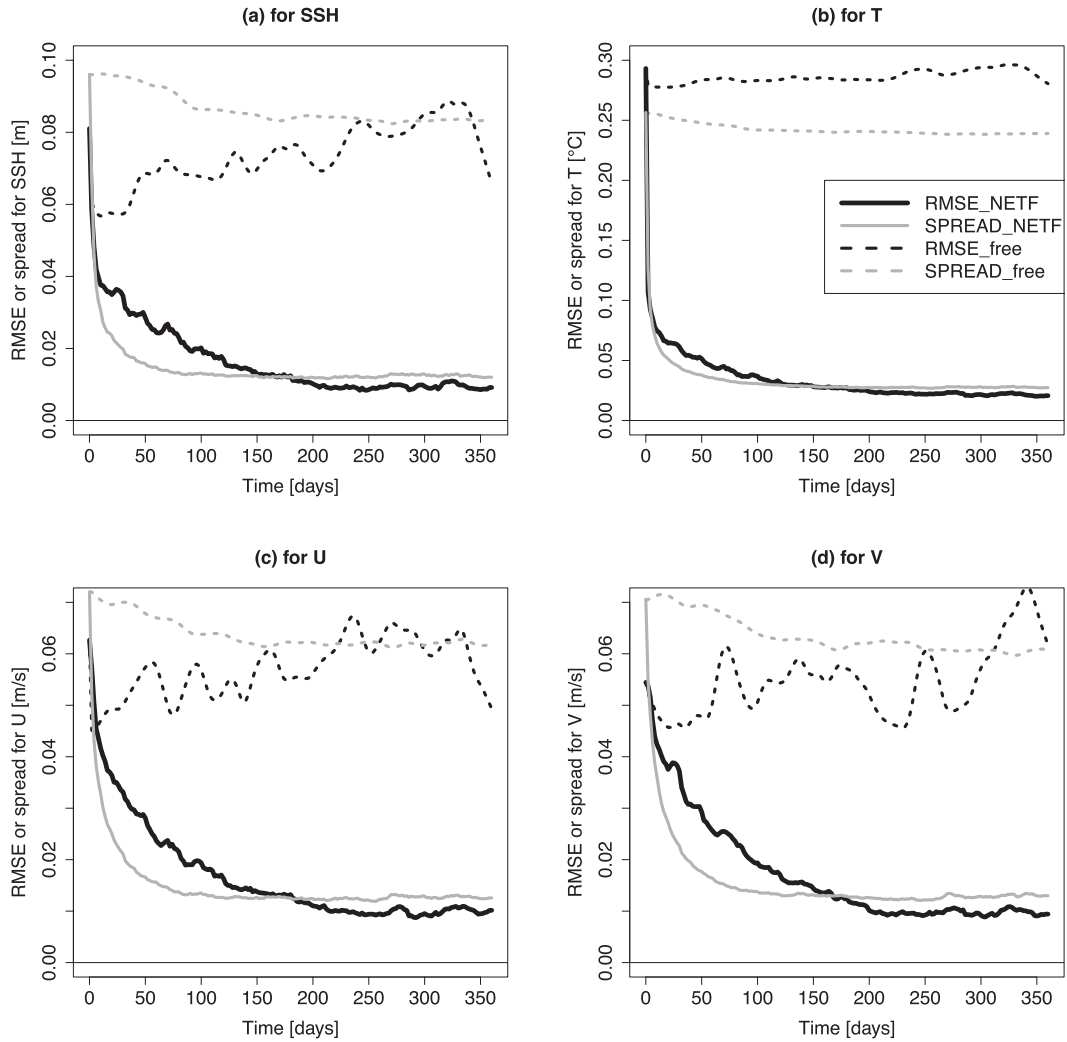
FIG. 6. Temporal evolution of the evaluation measures for the NETF (solid lines) and free ensemble (dashed lines) for all state variables: (a) SSH, (b) *T*, (c) *U*, and (d) *V*. The RMSE is drawn in black and the spread in gray. Both measures refer to the average over the full spatial field, including all vertical layers for *T*, *U*, and *V*. The legend in (b) is valid for all panels.

which allows for the computation of Eq. (13) according to the algorithm suggested by Hersbach (2000). Similar to the RMSE, we evaluate the CRPS at each grid point and then average over the whole field.

2) ERROR REDUCTION

Figure 6 shows the temporal evolution of the analysis RMSE for each control variable, together with the average ensemble spread. The evaluation measures represent the average over the whole field (two-dimensional for SSH, three-dimensional for *T*, *U*, and *V*). The corresponding results for the free ensemble are also included. As expected, the free ensemble does not improve upon the initial time. Its error and spread remain of similar magnitude, and their high variability reflects

the chaotic nature of the model evolution. The RMSEs of the NETF indicate that it performs successfully, as its analyses only exhibit small errors. This reflects the insight that the NETF is able to resolve the mesoscale patterns of the true flow. For all variables, the initial error is reduced nearly monotonically until convergence is reached. This spinup phase is explained by the fact that the initial ensemble does not contain information about the mesoscale circulation (Yang et al. 2012), and this will be discussed in more detail in section 5c. After the spinup, the errors remain approximately constant, which shows that the filter ensemble is in a quasi-stationary balance constrained by the observations. The comparison with the free ensemble demonstrates the error reduction in absolute terms. In addition, the

small variability of the filter errors compared to the free ensemble shows that the NETF tracks the truth continuously. These considerations are further confirmed by the ensemble spread, which shows that the inflation factor is well tuned. After the spinup phase, the spread is nearly constant and typically slightly larger than the RMSE. Therefore, the ensemble distributions are consistent with the truth in a statistical sense.

Next, for a better assessment of the performance, the RMSEs are normalized by their values at the initial time (day 0) at which the NETF and free ensemble are equivalent. Figure 7 shows the temporal evolution of these relative errors. The strongest error reduction is observed for the temperature T. This is easily explained by the fact that at least the upper temperature field is constrained by the majority of the observations. A minimal relative error of about 7.5% can be achieved. SSH, also an observed variable, achieves the second-best minimal relative error, with about 10% beyond the spinup. Even though the velocity fields are not observed, they can also be estimated with fairly high accuracy, and minimal relative errors of about 15% are found. Here U exhibits slightly smaller errors, presumably because it is, at least close to the surface, constrained by the fixed zonal wind forcing. Figure 8 provides a more detailed evaluation of the overall performance by displaying the error evolution in all vertical layers for the three-dimensional variables (T, U, and V). It shows that not only the densely observed surface layer but also the subsurface layers are improved by the filter. As is typical in ocean data assimilation, the spinup time increases with depth because the deeper ocean layers are less constrained by observations and exhibit a slower variability (Zhang and Rosati 2010).

3) COMPARISON TO THE LETKF

As demonstrated, the NETF produces consistent and reasonable analyses, which proves the main objective of
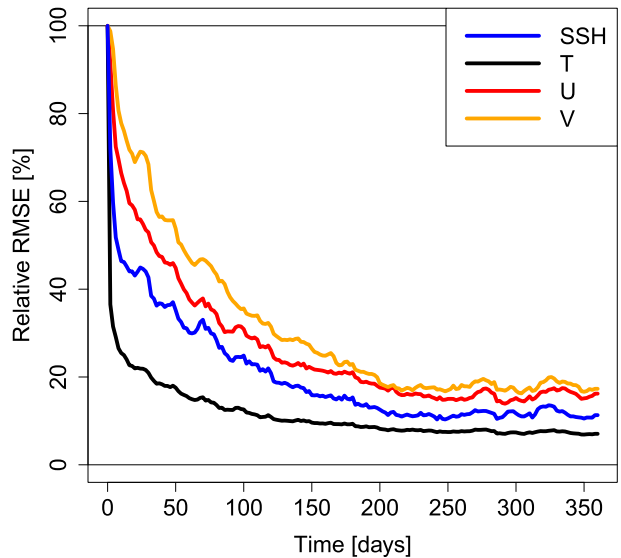

FIG. 7. The lines represent the relative error (RMSE normalized by its initial value) of the NETF analysis for SSH (blue), T (black), U (red), and V (orange).

this work. To further assess its performance, we also applied the LETKF using the same initial ensemble and setup. We found that an inflation factor of $\gamma = 1.01$ results in the best performance for the LETKF. Figure 9 compares the relative errors of both filters in terms of the relative RMSE (black and gray lines) and CRPS (red and orange lines) for all four variables. The LETKF requires only slightly more than 100 days to reach convergence, while the NETF spinup phase takes about 200 days. This difference will be discussed in section 5c. However, from day 200 on, the NETF and LETKF perform almost identical in terms of the RMSE. In principle, the same holds for the CRPS. The spinup time indicated by the CRPS is slightly longer than that for the RMSE. This behavior is independent of the filter. The reason is that the RMSE only measures the
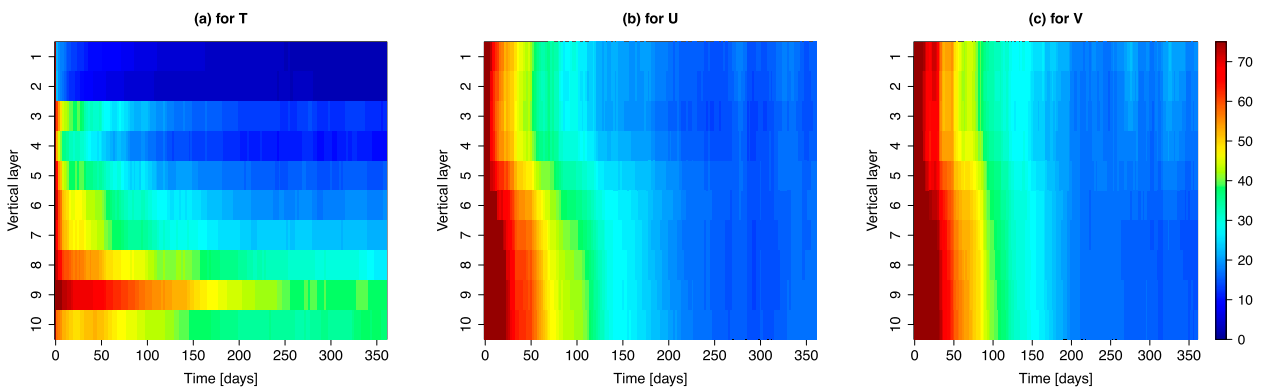

FIG. 8. Temporal evolution of the relative RMSE (in %) in each vertical layer for (a) T, (b) U, and (c) V. The legend bar in (c) is valid for all panels.
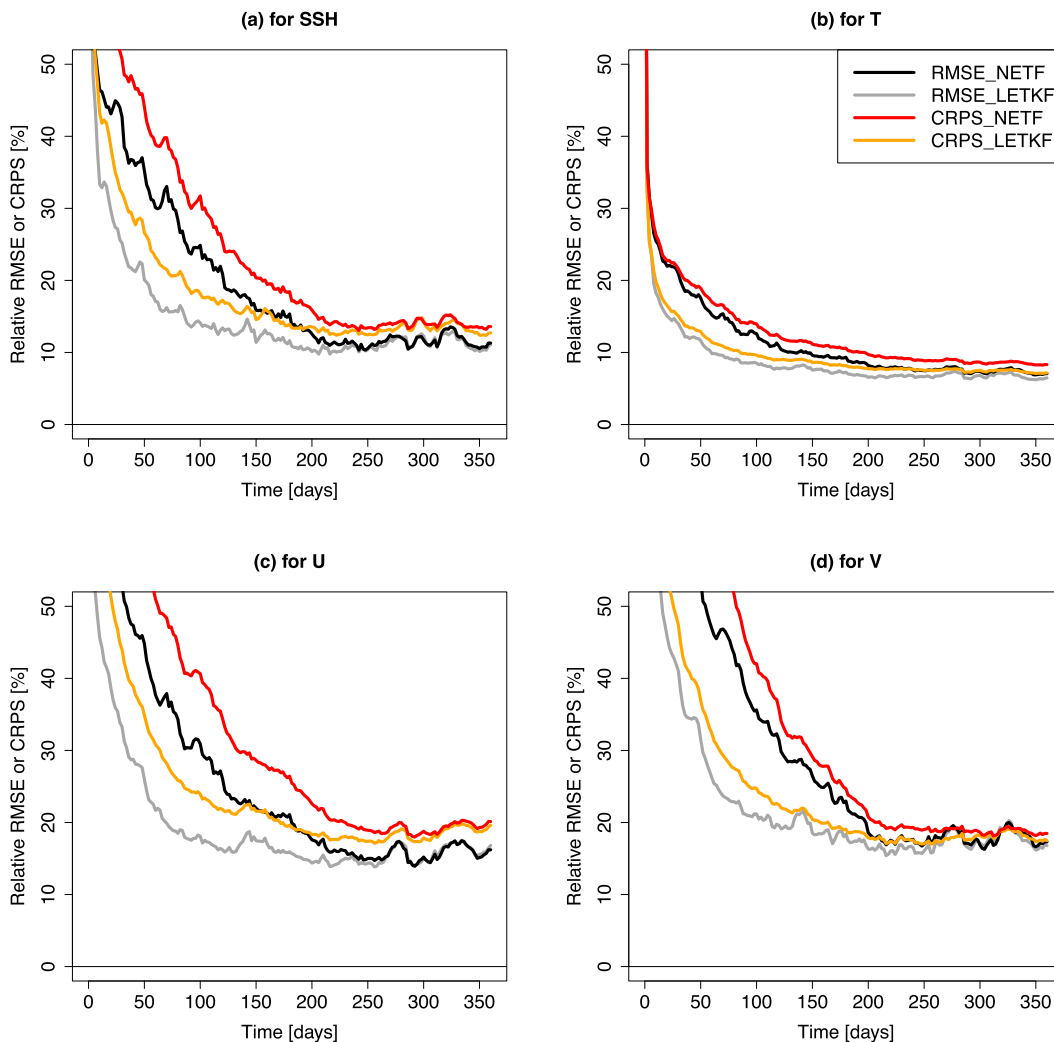
FIG. 9. Comparison of the NETF and LETKF in terms of the RMSE (black/gray) and CRPS (red/orange). Following the definition in section 5b, each line represents the field-averaged relative RMSE and CRPS, respectively, for all prognostic variables: (a) SSH, (b) *T*, (c) *U*, and (d) *V*. The legend in (b) is valid for all panels.

convergence of the ensemble mean, while the CRPS takes the whole ensemble distribution into account. After the spinup, the CRPS and RMSE achieve a similar relative value for each variable. This confirms that the ensemble distribution, which quantifies the filter uncertainty, is consistent with the error of the ensemble mean.

To assess the statistical significance of the similarity of the results, we repeated the assimilation experiment 10 times with different random numbers for the generation of the initial ensemble and the observations (see section 4) as well as for the random rotations [see Eq. (2)]. The truth (see section 3c) was kept fixed to ensure a reproducible reference trajectory of balanced model states. The thick lines in Fig. 10 show the relative RMSEs averaged over all model variables and all repetitive runs. After the spinup phase, they are very similar for the NETF (in black) and LETKF (in gray), respectively. This confirms that, on average, the NETF achieves the performance of the LETKF after the spinup. The minimal and maximal scores of the LETKF are close to the mean, which underlines the high robustness of the LETKF. The scores of the NETF exhibit more variability in the repetition runs, particularly during its spinup phase. For the CRPS, the corresponding results are very similar as in Fig. 10 (not shown).

c. Initial ensemble and spinup phase

The results have confirmed that the NETF yields a reasonable analysis with low errors, but it requires a longer spinup time than the LETKF. During this period, the results are also more sensitive to random variations.
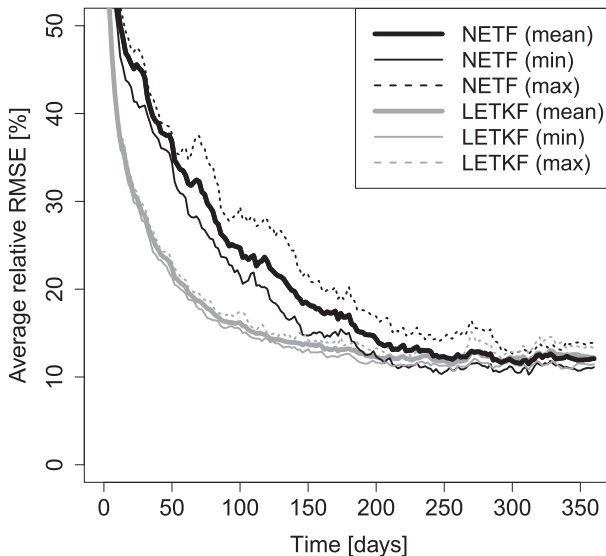
FIG. 10. Results from multiple runs of the experiment with different random realizations of the initial ensemble, random rotations, and observations for the NETF (in black) and LETKF (in gray). All lines show the temporal evolution of the relative RMSE (as in Figs. 9 and 7), averaged over all variables (T, U, V, SSH). The thick line refers to the mean score of the 10 runs, while the thin solid (dashed) line refers to the minimal (maximal) average relative RMSE.
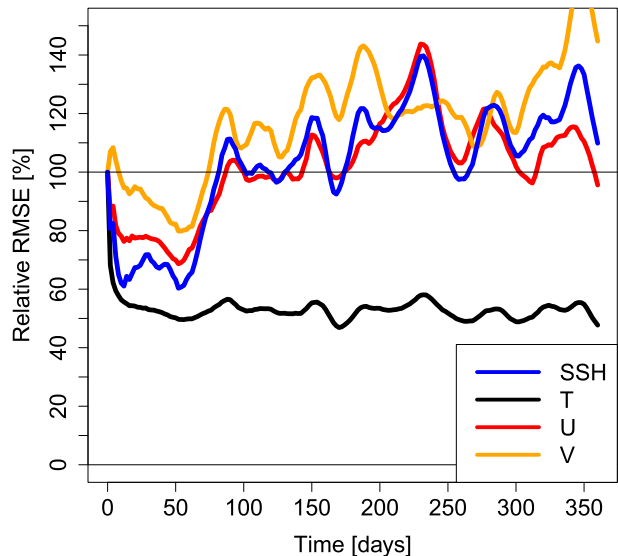

FIG. 11. Filter divergence with an inconsistent initial ensemble. Here, the initial NETF ensemble was constructed from the short-term climatology (years 65–74). Shown are the relative RMSEs for all variables, as in Fig. 7.

In general, the spinup phase of an ensemble filter is related to its initialization (Kalnay and Yang 2010). Hence, the NETF is more sensitive to the specification of the initial ensemble than the LETKF. This property might also be valid for nonlinear filtering in general, and this section provides a more detailed exploration of this issue.

1) SENSITIVITY TO THE INITIAL ENSEMBLE

As demonstrated in section 5b, the long-term climatology (spanning years 51–74) results in a successful NETF run. Initially, we had used the short-term climatology (spanning years 64–74, see also section 4c), for which the LETKF works as well. In contrast, the NETF diverges in this case independently of the tuning parameters. The ensemble spread decreases, but there is no error reduction (Fig. 11). These earlier failures emphasize that the consistent specification of the initial ensemble is a key criterion for the NETF to avoid filter divergence. It should appropriately span the largest sources of variability (Zhang et al. 2004). This claim is supported by the statistical properties of the model climatologies, which are shown in Table 2. The deviations from the true initial state, as measured by the RMSE, are of similar magnitude for both climatologies. However, the spread in T is very small for the short-term climatology (i.e., only about one-third of

the RMSE). Hence, the initial ensemble does not properly cover the truth. For the other variables in Table 2, the situation is less problematic, but the weights are computed in the observation space, which is mainly spanned by temperature measurements. Therefore, in the beginning, most ensemble members receive insignificant relative weights, and this leads to filter divergence. In contrast, the long-term climatology exhibits more consistent statistics. Even though the spread is still too small, its deviation from the RMSE is less pronounced.

2) SPINUP TIME AND INFLATION

The initial ensemble based on the long-term climatology enables the NETF to work successfully in our experiment, but the relatively long spinup time represents a practical restriction for future NETF applications. A first hint toward understanding this behavior can be found by revisiting Fig. 6. During the spinup phase, the spread decreases faster than the RMSE for all variables, particularly for U, V, and SSH. EnKF-like methods do not sufficiently account for the observations if the ensemble spread is too small. This can increase the spinup time or even lead to filter divergence (e.g., Evensen 2009; Yang et al. 2012). Usually, it is counteracted by inflation procedures. Therefore, one might consider that a larger inflation factor might reduce the spinup time of the NETF. However, after the spinup phase, it would be detrimental to the filter and cause the RMSE to increase again. To explore this
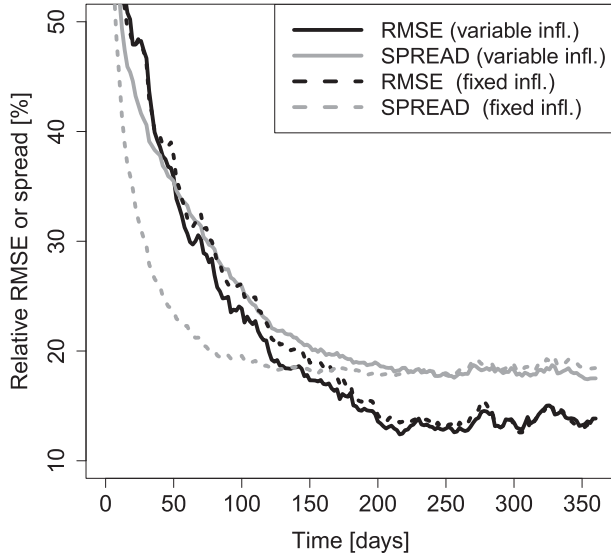
FIG. 12. Spread–skill relation in the experiment with inflation tuning. Shown is the average (over all variables, i.e., T, U, V, SSH) of the relative RMSE (black), together with the average relative spread (gray). The latter has been normalized by the RMSE at initial time. The full lines result from a run with variable inflation during the first 100 days (see section 5c). The dashed lines refer to the standard setup with fixed inflation, which yields underdispersive ensembles in the spinup phase.
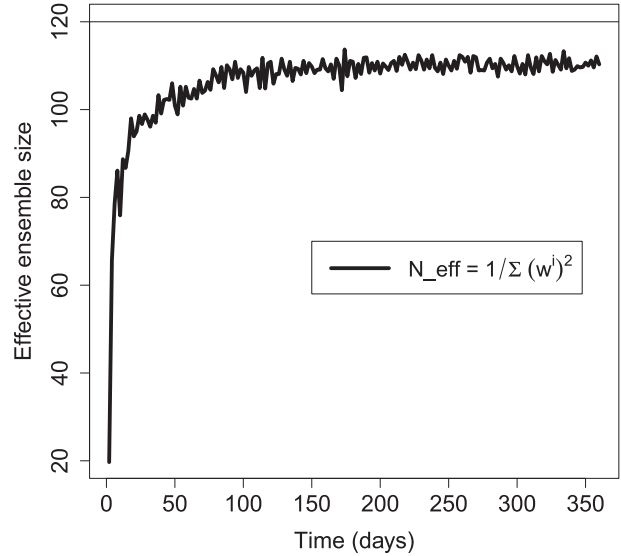

FIG. 13. Temporal evolution of the effective ensemble size for the NETF, using the default setup as described in section 4. The maximum of $m = 120$ can be achieved only if all weights are equal ($w^i \equiv 1/m \, \forall i$).

question, we tested a linearly decreasing inflation factor (1.11–1.02) during the initial phase of the assimilation experiment. Empirically, we found that applying this tuning during the first 100 days results in a much better agreement between the RMSE and spread over the entire assimilation window, as shown in Fig. 12. However, the spinup time is not shortened much by this modified inflation procedure and the RMSE remains almost as before. This indicates that the NETF's slower convergence is not simply a matter of inflation tuning, even though a more objective adaptive inflation technique (e.g., Anderson 2009) might still improve the results.

3) EXPLANATION

The NETF's higher sensitivity toward the initial ensemble in comparison with the LETKF can be explained by a key difference in their update mechanisms that was not explored in Tödter and Ahrens (2015). In EnKFs, the mean increment is *proportional* to the mean innovation (i.e., $\overline{\mathbf{d}} = \mathbf{y} - \overline{\mathbf{y}}_f$) (Hodyss 2012), as visible in Eq. (3). The analysis covariance in the LETKF is determined by the transform matrix in Eq. (4). It only depends on the prior ensemble covariance and the observation error covariance matrix $\mathbf{R}$, but not on the actual observation $\mathbf{y}$ (see Posselt et al. 2014). Therefore, the EnKFs are stable if the initial ensemble exhibits sufficient spread and the observation error is specified

consistently. In contrast, according to Eqs. (7) and (8), both the NETF's analysis mean and covariance are determined directly by the likelihood weights [Eq. (5)]. For Gaussian observation errors, the weights, as given by Eq. (6), decay *exponentially* with the innovations, $\mathbf{d}^i = \mathbf{y} - \mathbf{y}_f^i$. Thus, the relative contribution of members with a large distance to the observation is less pronounced in the NETF compared to the EnKFs. If many weights are insignificant, the analysis mean and covariance are effectively estimated by few members only, which degrades their quality. This property becomes particularly apparent for an ensemble initialized from a model climatology that is characterized by large innovations $\mathbf{d}^i$. In this case, the ensemble might collapse (see above), but at least the marginal contribution of most members results in the longer spinup period. This property also explains the increased sensitivity toward random variations in the observations (see section 5b).

To verify this hypothesis, the average effective ensemble size (Doucet et al. 2001) is computed by $[\sum_i (w^i)^2]^{-1}$ for the experiment in section 5b. It equals m if all members have equal weight and is smaller otherwise. Thus, it serves as a descriptive measure of the variance of the weights (Stordal et al. 2011). As shown in Fig. 13, the effective ensemble size is only about 20 during the first analysis steps and increases to over 100 later on. These numbers support the theoretical expectation stated above that most NETF members contribute only a little information in the beginning.

6. Conclusions and outlook

This work explores the large-scale applicability of the recently proposed NETF (Tödter and Ahrens 2015), which is a model-independent filter that creates a new analysis ensemble whose mean and covariance exactly match the Monte Carlo estimates of the corresponding Bayesian expectations. Thereby, it avoids biases that arise because of the Gaussian assumption inherent in EnKFs. The NETF performs an ensemble transformation like the LETKF, but with a distinct transform matrix and weight vector. It is supplemented by a moment-preserving random rotation. In contrast to most PFs, this transformation increases the stability of the NETF in larger-dimensional settings if the analysis is localized. Compared to Tödter and Ahrens (2015), this paper applies the new filter to a more realistic assimilation problem in an OGCM with high-dimensional dynamics and about 3.3×10^5 state variables. The assimilation of temperature profiles and satellite sea surface heights mimics realistic ocean observation networks. The system is characterized by its wind-driven large-scale dynamics, a double-gyre circulation as occurs in the North Atlantic, and mesoscale eddies that express its chaotic nature.

The NETF remains stable in this setting and shows a reasonable performance. It converges to the truth and keeps track of it, including the mesoscale features. Therefore, the NETF is applicable to this nonlinear, high-dimensional problem even though it only relies on the Bayesian weights. Thus, it is able to overcome to curse of dimensionality with a computationally feasible ensemble size. Furthermore, its implementation in realistic circulation systems requires only little additional effort if an LETKF system is available. To the knowledge of the authors, so far, a similar successful large-scale application could only be achieved with the more complex EWPF, which additionally relies on stochastic models. The circulation system considered here exhibits globally high-dimensional dynamics. A limitation of this study is that the local observation space, where the analysis is computed, exhibits a dimension of the same order of magnitude as the ensemble size. The observation scenario constructed here is realistic and dense for ocean applications, and already very challenging for a PF-based technique. However, in meteorological applications, the observation density can become considerably higher.

The experiments demonstrate that the NETF reduces the analysis errors to between 7.5% and 15% for the different variables, compared to the initial time. Thus, it generates reasonable analysis increments based on the local likelihood weights. The ensembles are

statistically consistent, which shows that the observations are used efficiently. After the spinup period, it can at least match the performance of the LETKF, as determined by both the RMSE and CRPS. However, it is unknown whether more improvement over the LETKF is actually possible in this setup. Because of its nonlinear design, the NETF may offer potential benefits, but this has to be elaborated upon in additional nonlinear, large-scale applications.

The results further emphasize that the NETF is sensitive to its initialization, which becomes particularly relevant in higher dimensions. An inconsistent initial ensemble (e.g., one with too little spread) results in many members with insignificant weights in the first analysis steps and might cause filter collapse. Furthermore, this issue increases the spinup phase in comparison to the LETKF, particularly if the initial ensemble only contains little information about the true flow. This problem is likely of concern to the EWPF as well. In the large-scale EWPF experiments published so far (van Leeuwen and Ades 2013; Ades and van Leeuwen 2015), the initial ensembles were generated by perturbing the true initial state, which prevents a spinup phase. Nevertheless, even with a climatological initialization, the NETF is still able to reconstruct the truth. In future applications, one should make sure to specify consistent initial ensembles that contain as much knowledge about the true flow as possible (e.g., from previous analyses). As shown for EnKFs, this can strongly reduce the spinup time (Yang et al. 2012). The most pragmatic solution consists of simply using an EnKF for the spinup phase and switching to the nonlinear filter afterward. Alternatively, more advanced adaptive inflation methods (e.g., Anderson 2009; Miyoshi 2011) could be adopted for the NETF or the spinup could be reduced by smoothing the estimates with future observations (Cosme et al. 2010). Such a nonlinear smoother based on the NETF can be derived in analogy to the ensemble transform Kalman smoother (Kalnay and Yang 2010; Nerger et al. 2014) and will be presented in an upcoming paper.

The successful application of the NETF in an OGCM offers numerous possibilities for future research. First, we anticipate that the experiences presented here will stimulate further research to reduce the spinup time. Second, in the experiment conducted here, the filter performance was similar to the LETKF at least after the spinup phase. Therefore, the NETF should be applied to other large-scale systems to accentuate potential benefits of the nonlinear approach. This could concern setups where the chaotic dynamics are able to modify the large-scale flow, since the latter is constrained by the analytical wind forcing in our setup. Additionally, it

would be important to assess whether the NETF is able to deal with a local observation space whose dimension is much higher than the ensemble size. Such a situation is characteristic of atmospheric data assimilation, and it might represent a limitation for the filter.

Finally, the NETF should be compared to the EWPF to assess the relevance of higher-order moments in nonlinear, large-scale DA. This necessarily requires the employment of a stochastic model. The NETF can be combined with a nudged forecast step as well (Tödter 2015, chapter 4.9), which allows for a fair comparison of the analysis algorithms in such a scenario. For that purpose, the NETF algorithm does not need to be modified, except that the prior weights resulting from nudging have to be considered in Eq. (7). In this context, it could also be investigated whether the EWPF exhibits a similar sensitivity to the initial ensemble.

REFERENCES

Achatz, U., and G. Branstator, 1999: A two-layer model with empirical linear corrections and reduced order for studies of internal climate variability. *J. Atmos. Sci.*, **56**, 3140–3160, doi:10.1175/1520-0469(1999)056<3140:ATLMWE>2.0.CO;2.

——, and J. D. Opsteegh, 2003: Primitive-equation-based low-order models with seasonal cycle. Part I: Model construction. *J. Atmos. Sci.*, **60**, 465–477, doi:10.1175/1520-0469(2003)060<0465: PEBLOM>2.0.CO;2.

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

——, and ——, 2015: The effect of the equivalent-weights particle filter on dynamical balance in a primitive equation model. *Mon. Wea. Rev.*, **143**, 581–596, doi:10.1175/MWR-D-14-00050.1.

Anderson, J. L., 2009: Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus*, **61A**, 72–83, doi:10.1111/j.1600-0870.2008.00361.x.

——, 2012: Localization and sampling error correction in ensemble Kalman filter data assimilation. *Mon. Wea. Rev.*, **140**, 2359– 2371, doi:10.1175/MWR-D-11-00013.1.

——, and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.*, **127**, 2741–2758, doi:10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2.

Bengtsson, L., and Coauthors, 2007: The need for a dynamical climate reanalysis. *Bull. Amer. Meteor. Soc.*, **88**, 495–501, doi:10.1175/BAMS-88-4-495.

Bennett, A. F., 2002: *Inverse Modeling of the Ocean and Atmosphere.* Cambridge University Press, 256 pp.

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

Buehner, M., P. L. Houtekamer, C. Charette, H. L. Mitchell, and B. He, 2010: Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I: Description and single-observation experiments. *Mon. Wea. Rev.*, **138**, 1550–1566, doi:10.1175/2009MWR3157.1.

Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719–1724, doi:10.1175/1520-0493(1998)126<1719: ASITEK>2.0.CO;2.

Carrier, G. F., and A. R. Robinson, 1962: On the theory of the wind-driven ocean circulation. *J. Fluid Mech.*, **12**, 49–80, doi:10.1017/S0022112062000038.

Carval, T., and Coauthors, 2013: ARGO user's manual, version 3.03. ARGO data management Rep. ar-um-02-01, 90 pp., doi:10.13155/26387. [Available online at http://archimer.ifremer.fr/ doc/00153/26387/24482.pdf.]

Chassignet, E. P., and P. R. Gent, 1991: The influence of boundary conditions on midlatitude jet separation in ocean numerical models. *J. Phys. Oceanogr.*, **21**, 1290–1299, doi:10.1175/ 1520-0485(1991)021<1290:TIOBCO>2.0.CO;2.

Cosme, E., J.-M. Brankart, J. Verron, P. Brasseur, and M. Krysta, 2010: Implementation of a reduced rank square-root smoother for high resolution ocean data assimilation. *Ocean Modell.*, **33**, 87–100, doi:10.1016/j.ocemod.2009.12.004.

Doucet, A., N. de Freitas, and N. Gordon, 2001: *Sequential Monte-Carlo Methods in Practice.* Springer-Verlag, 581 pp.

Durrant, T. H., D. J. M. Greenslade, and I. Simmonds, 2009: Validation of *Jason-1* and *Envisat* remotely sensed wave heights. *J. Atmos. Oceanic Technol.*, **26**, 123–134, doi:10.1175/ 2008JTECHO598.1.

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

——, 2009: *Data Assimilation: The Ensemble Kalman Filter.* Springer, 279 pp.

Fairbairn, D., S. R. Pring, A. C. Lorenc, and I. Roulstone, 2014: A comparison of 4DVar with ensemble data assimilation methods. *Quart. J. Roy. Meteor. Soc.*, **140**, 281–294, doi:10.1002/qj.2135.

Fournier, A., L. Nerger, and J. Aubert, 2013: An ensemble Kalman filter for the time-dependent analysis of the geomagnetic field. *Geochem. Geophys. Geosyst.*, **14**, 4035–4043, doi:10.1002/ ggge.20252.

Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757, doi:10.1002/qj.49712555417.

Gneiting, T., F. Balabdaoui, and A. E. Raftery, 2007: Probabilistic forecasts, calibration and sharpness. *J. Roy. Stat. Soc.*, **69**, 243–268, doi:10.1111/j.1467-9868.2007.00587.x.

Gordon, N., D. J. Salmond, and A. F. M. Smith, 1993: Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc., F, Radar Signal Process.*, **140**, 107–113, doi:10.1049/ip-f-2.1993.0015.

Haimberger, L., 2007: Homogenization of radiosonde temperature time series using innovation statistics. *J. Climate*, **20**, 1377–1403, doi:10.1175/JCLI4050.1.

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

Hodyss, D., 2012: Accounting for skewness in ensemble data assimilation. *Mon. Wea. Rev.*, **140**, 2346–2358, doi:10.1175/MWR-D-11-00198.1.

Holland, W. R., 1978: The role of mesoscale eddies in the general circulation of the ocean numerical experiments using a wind-driven quasi-geostrophic model. *J. Phys. Oceanogr.*, **8**, 363–392, doi:10.1175/1520-0485(1978)008<0363:TROMEI>2.0.CO;2.

Hopson, T. M., 2014: Assessing the ensemble spread–error relationship. *Mon. Wea. Rev.*, **142**, 1125–1142, doi:10.1175/MWR-D-12-00111.1.

Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811, doi:10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2.

Hunt, B. R., E. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D*, **230**, 112–126, doi:10.1016/j.physd.2006.11.008.

Kalman, R., 1960: A new approach to linear filtering and prediction problems. *J. Basic Eng.*, **82**, 35–45, doi:10.1115/1.3662552.

Kalnay, E., and S.-C. Yang, 2010: Accelerating the spin-up of ensemble Kalman filtering. *Quart. J. Roy. Meteor. Soc.*, **136**, 1644–1651, doi:10.1002/qj.652.

——, H. Li, T. Miyoshi, S.-C. Yang, and J. Ballabrera-Poy, 2007: 4D-Var or ensemble Kalman filter? *Tellus*, **59A**, 748–773.

Kirchgessner, P., L. Nerger, and A. Bunse-Gerstner, 2014: On the choice of an optimal localization radius in ensemble Kalman filter methods. *Mon. Wea. Rev.*, **142**, 2165–2175, doi:10.1175/MWR-D-13-00246.1.

Lei, J., and P. Bickel, 2011: A moment matching ensemble filter for nonlinear non-Gaussian data assimilation. *Mon. Wea. Rev.*, **139**, 3964–3973, doi:10.1175/2011MWR3553.1.

Lermusiaux, P. F. J., and A. R. Robinson, 1999: Data assimilation via error subspace statistical estimation. Part I: Theory and schemes. *Mon. Wea. Rev.*, **127**, 1385–1407, doi:10.1175/1520-0493(1999)127<1385:DAVESS>2.0.CO;2.

Lévy, M., P. Klein, A.-M. Tréguier, D. Iovino, G. Madec, S. Masson, and K. Takahashi, 2010: Modifications of gyre circulation by sub-mesoscale physics. *Ocean Modell.*, **34**, 1–15, doi:10.1016/j.ocemod.2010.04.001.

Losa, S. N., S. Danilov, J. Schröter, L. Nerger, S. Massmann, and F. Janssen, 2012: Assimilating NOAA SST data into the BSH operational circulation model for the North and Baltic Seas: Inference about the data. *J. Mar. Syst.*, **105–108**, 152–162, doi:10.1016/j.jmarsys.2012.07.008.

Madec, G., 2012: NEMO ocean engine. Tech. Rep. 27, Note du Pôle de modélisation, Institut Pierre-Simon Laplace (IPSL), 357 pp.

Miyoshi, T., 2011: The Gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter. *Mon. Wea. Rev.*, **139**, 1519–1535, doi:10.1175/2010MWR3570.1.

——, and M. Kunii, 2012: The local ensemble transform Kalman filter with the Weather Research and Forecasting model: Experiments with real observations. *Pure Appl. Geophys.*, **169**, 321–333, doi:10.1007/s00024-011-0373-4.

Nerger, L., and W. Hiller, 2013: Software for ensemble-based data assimilation systems—Implementation strategies and scalability. *Comput. Geosci.*, **55**, 110–118, doi:10.1016/j.cageo.2012.03.026.

——, S. Danilov, W. Hiller, and J. Schröter, 2006: Using sea-level data to constrain a finite-element primitive-equation ocean model with a local SEIK filter. *Ocean Dyn.*, **56**, 634–649, doi:10.1007/s10236-006-0083-0.

——, ——, G. Kivman, W. Hiller, and J. Schröter, 2007: Data assimilation with the ensemble Kalman filter and the SEIK filter applied to a finite element model of the North Atlantic. *J. Mar. Syst.*, **65**, 288–298, doi:10.1016/j.jmarsys.2005.06.009.

——, T. Janjić, J. Schröter, and W. Hiller, 2012: A unification of ensemble square root Kalman filters. *Mon. Wea. Rev.*, **140**, 2335–2345, doi:10.1175/MWR-D-11-00102.1.

——, S. Schulte, and A. Bunse-Gerstner, 2014: On the influence of model nonlinearity and localization on ensemble Kalman smoothing. *Quart. J. Roy. Meteor. Soc.*, **140**, 2249–2259, doi:10.1002/qj.2293.

Palmer, T., R. Buizza, R. Hagedorn, A. Lawrence, M. Leutbecher, and L. Smith, 2005: Ensemble prediction: A pedagogical perspective. *ECMWF Newsletter*, No. 106, ECMWF, Reading, United Kingdom, 10–17.

Pham, D. T., 2001: Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Mon. Wea. Rev.*, **129**, 1194–1207, doi:10.1175/1520-0493(2001)129<1194:SMFSDA>2.0.CO;2.

Posselt, D. J., D. Hodyss, and C. H. Bishop, 2014: Errors in ensemble Kalman smoother estimates of cloud microphysical parameter. *Mon. Wea. Rev.*, **142**, 1631–1654, doi:10.1175/MWR-D-13-00290.1.

Reich, H., A. Rhodin, and C. Schraff, 2011: LETKF for the non-hydrostatic regional model COSMO-DE. *COSMO Newsl.*, **11**, 27–31.

Roullet, G., and G. Madec, 2000: Salt conservation, free surface, and varying levels: A new formulation for ocean general circulation models. *J. Geophys. Res.*, **105**, 23 927–23 942, doi:10.1029/2000JC900089.

Silverman, B. W., 1986: *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, 175 pp.

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

Stordal, A. S., H. A. Karlsen, G. Nævdal, H. J. Skaug, and B. Vallès, 2011: Bridging the ensemble Kalman filter and particle filters: The adaptive Gaussian mixture filter. *Comput. Geosci.*, **15**, 293–305, doi:10.1007/s10596-010-9207-1.

Talagrand, O., and P. Courtier, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory. *Quart. J. Roy. Meteor. Soc.*, **113**, 1311–1328, doi:10.1002/qj.49711347812.

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.*, **131**, 1485–1490, doi:10.1175/1520-0493(2003)131<1485:ESRF>2.0.CO;2.

Tödter, J., 2015: Derivation and characterization of a new filter for nonlinear high-dimensional data assimilation. Ph.D. thesis, Goethe University, Frankfurt/Main, 187 pp. [Available online at http://publikationen.ub.uni-frankfurt.de/frontdoor/index/index/docId/37795.]

——, and B. Ahrens, 2012: Generalization of the ignorance score: Continuous ranked version and its decomposition. *Mon. Wea. Rev.*, **140**, 2005–2017, doi:10.1175/MWR-D-11-00266.1.

——, and ——, 2015: A second-order exact ensemble square root filter for nonlinear data assimilation. *Mon. Wea. Rev.*, **143**, 1347–1367, doi:10.1175/MWR-D-14-00108.1.

van Leeuwen, P. J., 2009: Particle filtering in geophysical systems. *Mon. Wea. Rev.*, **137**, 4089–4114, doi:10.1175/2009MWR2835.1.

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

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

Wang, X., C. H. Bishop, and S. J. Julier, 2004: Which is better, an ensemble of positive–negative pairs or a centered spherical simplex ensemble? *Mon. Wea. Rev.*, **132**, 1590–1605, doi:10.1175/1520-0493(2004)132<1590:WIBAEO>2.0.CO;2.

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

Wikle, C. K., and L. M. Berliner, 2007: A Bayesian tutorial for data assimilation. *Physica D*, **230**, 1–16, doi:10.1016/j.physd.2006.09.017.

Yan, Y., A. Barth, and J. M. Beckers, 2014: Comparison of different assimilation schemes in a sequential Kalman filter assimilation system. *Ocean Modell.*, **73**, 123–137, doi:10.1016/j.ocemod.2013.11.002.

Yang, S.-C., E. Kalnay, and T. Miyoshi, 2012: Accelerating the EnKF spinup for typhoon assimilation and prediction. *Wea. Forecasting*, **27**, 878–897, doi:10.1175/WAF-D-11-00153.1.

Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, **132**, 1238–1253, doi:10.1175/1520-0493(2004)132<1238:IOIEAO>2.0.CO;2.

Zhang, S., and A. Rosati, 2010: An inflated ensemble filter for ocean data assimilation with a biased coupled GCM. *Mon. Wea. Rev.*, **138**, 3905–3931, doi:10.1175/2010MWR3326.1.