Quarterly Journal of the
Royal Meteorological Society   RMetS

RESEARCH ARTICLE

Data assimilation for nonlinear systems with a hybrid nonlinear Kalman ensemble transform filter

Lars Nerger

Alfred-Wegener-Institut
Helmholtz-Zentrum für Polar- und
Meeresforschung, Bremerhaven, Germany

Correspondence
L. Nerger, Alfred-Wegener-Institut
Helmholtz-Zentrum für Polar- und
Meeresforschung, Am Handelshafen12,
27570 Bremerhaven, Germany.
Email: lars.nerger@awi.de

Abstract

Ensemble Kalman filters are widely used for data assimilation applications in the geosciences. While they are remarkably stable even with nonlinear systems, it is known that they are not optimal in this case. The alternative particle filters are fully nonlinear, but difficult to apply with high-dimensional models. To combine the strengths of both filter types, a hybrid filter is introduced that combines the local ensemble transform Kalman filter (LETKF) with the nonlinear ensemble transform filter (NETF). Three variants of the hybrid filter are formulated. The hybridization is controlled by a hybrid weight. Different hybrid weights are examined and a new adaptive approach based on the ensemble skewness and kurtosis is introduced. The different hybrid filters and the schemes to compute the hybrid weight are assessed in numerical experiments with the nonlinear Lorenz-63 and Lorenz-96 models at different degrees of nonlinearity. A hybrid variant that first applies the NETF followed by the LETKF yields the best results. For the Lorenz-96 model, error reductions by up to 21.5% compared with the LETKF are obtained for the same ensemble size. Computing the hybrid weight based on skewness and kurtosis combined with the effective sample size yields the lowest estimation errors and the overall highest stability of the hybrid filter. The new hybrid filter applies localization and inflation and is hence also usable with high-dimensional models and can potentially provide a robust way to account for leading nonlinearity with small ensembles in nonlinear data assimilation applications.

KEYWORDS

data assimilation, ensemble, hybrid filter, Kalman filter, nonlinear

1 | INTRODUCTION

Ensemble data assimilation is widely applied in geoscientific applications, for example, to initialize weather, ocean, or sea-ice forecasts (e.g., Sakov *et al.*, 2012; Houtekamer *et al.*, 2014; Martin *et al.*, 2015; Liang *et al.*, 2017), to estimate the ocean biogeochemical state or parameters (e.g., Nerger and Gregg, 2008; Ciavatta *et al.*, 2011; Mattern *et al.*, 2014; Pradhan *et al.*, 2019), to generate re-analyses (e.g., Oke *et al.*, 2015; Ciavatta *et al.*, 2016), to estimate

hydrological parameters (e.g., Baatz *et al.*, 2017), or to simulate the dynamics of the Earth's mantle (Bocher *et al.*, 2018). Most commonly, different variants of ensemble Kalman filters (EnKFs) are used in these applications. While EnKFs assume Gaussian error distributions and hence imply linear models (because a nonlinear model will transform a Gaussian distribution into a non-Gaussian one), the different applications demonstrate that the EnKFs are stable and successful with nonlinear models. Nonetheless, it is known that the EnKFs are suboptimal for nonlinear systems. Different approaches have been developed to improve the performance of EnKFs in the case of nonlinearity. For example, the application of so-called Gaussian anamorphosis, in which the forecast ensemble is transformed into a Gaussian distribution, has improved state estimates in biogeochemical models (Simon and Bertino, 2009; Doron *et al.*, 2011). Further, rank histogram and regression filters (Anderson, 2010; 2019; Metref *et al.*, 2014) have been introduced, which aim to improve the performance of the Kalman filter analysis step by accounting for a non-Gaussian ensemble distribution.

A fully nonlinear alternative to EnKFs are particle filters (PFs: see, e.g., van Leeuwen, 2009). However, these filters cannot be applied to high-dimensional models without particular adaptions, due to the so-called "curse of dimensionality" (see Snyder *et al.*, 2015). Basically, ensembles collapse to a single member if the ensemble size does not increase exponentially with the system dimension. Several variants of PFs (e.g., Ades and van Leeuwen, 2013; Zhu *et al.*, 2016) have been developed that, by construction, avoid negligible particle weights. However, these methods cannot fully avoid the "curse of dimensionality" (see Snyder *et al.*, 2015). In addition, they depend crucially on the application of stochastic model errors for the full model state, which has to be very carefully tuned. These filters also do not converge to the full posterior probability distribution for the large ensemble limit. Alternatively, a localized analysis can reduce the effective system size and hence improve the performance of PFs. While this does not completely avoid the curse of dimensionality, different schemes have been introduced and tested with high-dimensional applications (Poterjoy, 2016; Poterjoy *et al.*, 2019; Potthast *et al.*, 2019). The PFs usually need a very careful tuning, with higher complexity than the localization and inflation approaches used by EnKFs. Further, transportation PFs (e.g., Reich, 2013) transform the ensemble according to some optimal transportation rule, so that no explicit resampling of the analysis ensemble, which would introduce further randomness and sampling errors, is required. However, these filters require quite costly computations and have not yet been implemented in high-dimensional systems. Overall, none of these filter methods is yet fully usable for realistic data assimilation

problems, as also discussed in the recent review on PFs for high-dimensional applications (van Leeuwen *et al.*, 2019).

Related to transport filters are second-order exact transformation PFs like the nonlinear ensemble adjustment filter (Lei and Bickel, 2011) and the nonlinear ensemble transform filter (Tödter and Ahrens, 2015). These filters use the weights computed for the PF to update the ensemble mean, but then transform the ensemble perturbations analogous to EnKFs to fulfill the analysis covariance matrix. While these filters rely on the covariance matrix, they do not imply Gaussian distributions, but only that the covariance matrix still contains representative information. The nonlinear ensemble transform filter (NETF), which is applied here, uses a transformation in ensemble space like the local ensemble transform Kalman filter (LETKF) or local error-subspace transform Kalman filter (LESTKF) methods (Nerger *et al.*, 2012), which is computationally very efficient. The NETF can be applied with the same localization and inflation approaches as these EnKFs, so that the tuning is also analogous. This makes the NETF rather easily usable with high-dimensional models (Tödter *et al.*, 2016; Kirchgessner *et al.*, 2017). For sufficiently large ensembles, the NETF has been shown to yield errors comparable with and sometimes lower than the LETKF or LESTKF.

To benefit from the stability of the EnKFs and the ability of the PF to handle nonlinear systems, hybrid filters have been proposed (Frei and Künsch, 2013; Chustagulprom *et al.*, 2016; Robert *et al.*, 2018). These filters combine the analysis steps of an EnKF with those of a PF and use a hybrid weight to shift the analysis in between these extremes. Robert *et al.* (2018) discussed the successful application of a hybrid filter to a high-dimensional weather forecast model. Some filter variants are proposed to apply the ensemble Kalman filter before the PF (Frei and Künsch, 2013; Robert *et al.*, 2018). Chustagulprom *et al.* (2016) also used this order, but in addition the variant where the PF is applied before the ensemble Kalman filter. They found smaller root-mean-square (RMS) errors when applying the PF first and mention that this might be due to the fact that the prior distribution is more non-Gaussian than the posterior. The algorithmic formulations of these filters differ from pure EnKFs or PFs by requiring either iterative solvers or the solution of a linear programming problem.

Here, a new hybrid filter is introduced, which combines the LETKF with the NETF. The NETF is an attractive choice for this algorithm, because it was already demonstrated that it can be applied to high-dimensional nonlinear models and can outcompete the LETKF for nonlinear cases. It also performs an ensemble transformation so that no resampling step is required. Further, the

same localization and inflation approaches as used in the LETKF can be applied. In addition, one can implement the hybrid filter on the basis on existing implementations of both filters, for example, in the parallel data assimilation framework (PDAF: Nerger *et al.*, 2005; Nerger and Hiller, 2013),[1] without particular iterative solvers or linear programming. Just as the previous hybrid filters were proposed in different variants (applying PF before EnKF or EnKF before PF), here also different variants are introduced and examined with the widely used low-dimensional Lorenz-63 and Lorenz-96 models (Lorenz, 1963; 1996). For the Lorenz-63 model, the dependence of the filter performance on the nonlinearity is studied, while the Lorenz-96 model is used in a strongly nonlinear assimilation configuration to assess the additional effect of small ensembles and localization.

The study is structured as follows. Section 2 reviews the LETKF and NETF algorithms, which are combined into the hybrid filter. The hybrid filter variants are introduced in Section 3. The hybrid filter is assessed with the Lorenz-96 model in Section 4 and the Lorenz-96 model in Section 5. Finally, the results are discussed in Section 6 and conclusions are drawn in Section 7.

2 | ENSEMBLE FILTERS

To represent the state estimate and its uncertainty at some time k, ensemble filter algorithms use an ensemble of N_e state realizations $\mathbf{x}_k^{(i)}$, $i = 1, \ldots N_e$ of dimension N_x. We write the ensemble as a matrix $\mathbf{X}_k = \left[\mathbf{x}_k^{(1)}, \ldots, \mathbf{x}_k^{(N_e)}\right]$. During the forecast phase, all ensemble states are integrated by the model $\mathcal{M}$, which provides the forecast ensemble $\mathbf{X}_k^f = \mathcal{M}(\mathbf{X}_{k-1}^a)$. The superscript "a" denotes the analysis ensemble. At the initial time, the ensemble $\mathbf{X}_0^a$ is initialized using some estimate of the initial state along with an estimate of the uncertainty, which is often prescribed by the temporal variability of the model state.

2.1 | General form of the analysis step

The analysis step, in which the observations for time k are assimilated, transforms the forecast ensemble into the analysis ensemble, which represents the updated state estimate and its uncertainty. We can write the analysis step in a general form as a right-sided product with a weight matrix, which is valid for a wide family of filter methods (see Vetra-Carvalho *et al.*, 2018). We omit the time index

[1]http://pdaf.awi.de

k, as all operations occur at this time. In general, we can write the update separately for the ensemble mean state $\overline{\mathbf{x}}^f = 1/N_e \sum_{\alpha=1}^{N_e} \mathbf{x}^{f(\alpha)}$ and the ensemble perturbation matrix $\mathbf{X}'^f = \mathbf{X}^f - \overline{\mathbf{x}}^f \mathbf{1}^T$, where $\mathbf{1}$ is a vector of length N_e with all elements equal to 1:

$$\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^f + \mathbf{X}'^f \tilde{\mathbf{w}}, \tag{1}$$

$$\mathbf{X}'^a = \mathbf{X}'^f \mathbf{W}. \tag{2}$$

Here, $\tilde{\mathbf{w}}$ is a weight vector of length N_e and $\mathbf{W}$ is a weight matrix of size $N_e \times N_e$. To ensure an unbiased ensemble transformation, the vector $\mathbf{1}$ must be an eigenvector of $\mathbf{W}$. We can combine the update into a single equation as

$$\mathbf{X}^a = \overline{\mathbf{x}}^f \mathbf{1}^T + \mathbf{X}^f \mathbf{T} \left(\tilde{\mathbf{w}} \mathbf{1}^T + \mathbf{W}\right). \tag{3}$$

Here, we introduced the $N_e \times N_e$ matrix

$$\mathbf{T}_{i,j} := \begin{cases} 1 - \frac{1}{N_e} & \text{for } i = j, \\ -\frac{1}{N_e} & \text{for } i \neq j, \end{cases} \tag{4}$$

which subtracts the ensemble mean in the product $\mathbf{X}^f \mathbf{T}$.

The ensemble transform Kalman filter (ETKF) and the NETF both apply this generic form of the analysis step, but they use distinct definitions of $\tilde{\mathbf{w}}$ and $\mathbf{W}$. The computation of the weight vector and matrix utilizes the observation vector $\mathbf{y}$ of size N_y. The observations are related to a state vector through

$$\mathbf{y} = \mathbf{H}\mathbf{x}^f + \epsilon, \tag{5}$$

where ϵ is the observation error with covariance matrix $\mathbf{R}$. The observation operator $\mathbf{H}$ is here written as a linear operator, but a nonlinear operator is also possible as long as it is applied directly to a state vector.

The analysis is computed in a localized form, frequently denoted as domain localization combined with observation localization (e.g., Janjić *et al.*, 2011). With domain localization, $\tilde{\mathbf{w}}$ and $\mathbf{W}$ are computed for a series of a local analysis domains. The local analysis domain can be a single model grid point, but, for example, for ocean models it is often a single vertical column. For the update of a local analysis domain, only observations within a prescribed distance (the localization radius r_{loc}) around it are taken into account. Further, observation localization (Hunt *et al.*, 2007) is applied, in which the observations are weighted according to their distance from the local analysis domain. Practically, the inverse observation-error covariance matrix $\mathbf{R}$ is multiplied elementwise with a

weight matrix, which is constructed using the fifth-order polynomial of Gaspari and Cohn (1999). This weight ranges between 1 for zero distance and 0 for the prescribed radius r_{loc} and mimics a Gaussian function.

2.2 | Ensemble transform Kalman filter

The ETKF was introduced by Bishop *et al.* (2001) and later formulated with localization, then denoted LETKF, by Hunt *et al.* (2007). Here, we follow Nerger (2015).

The ETKF uses a transform matrix $\mathbf{A}$ of size $N_e \times N_e$, which is defined by its inverse as

$$\mathbf{A}^{-1} = \rho(N_e - 1)\mathbf{I} + (\mathbf{HX}'^f)^T \mathbf{R}^{-1} \mathbf{HX}'^f . \quad (6)$$

Here ρ, with $0 < \rho \leq 1$, is the so-called "forgetting factor" (Pham *et al.*, 1998), which is used to inflate the forecast error estimate. Matrix $\mathbf{A}$ is inverted using a singular-value decomposition $\mathbf{USV} = \mathbf{A}^{-1}$ such that $\mathbf{A} = \mathbf{US}^{-1}\mathbf{V}$. Further, we utilize the square root $\mathbf{A}^{1/2} = \mathbf{US}^{-1/2}\mathbf{U}^T$. The weight vector and weight matrix for the ensemble transformation are now given by

$$\tilde{\mathbf{w}}_{\text{ETKF}} = \mathbf{A}(\mathbf{HX}'^f)^T \mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}\bar{\mathbf{x}}^f) , \quad (7)$$

$$\mathbf{W}_{\text{ETKF}} = \sqrt{N_e - 1}\,\mathbf{A}^{1/2}\mathbf{\Lambda}, \quad (8)$$

where $\mathbf{\Lambda}$ is the identity or a random matrix that preserves the mean and covariance, which implies that $\mathbf{\Lambda}$ is orthogonal and has the eigenvector $\mathbf{1}$.

2.3 | Nonlinear ensemble transform filter

The NETF was introduced by Tödter and Ahrens (2015). Like particle filters, it uses weights computed from the statistical likelihood of each ensemble state. However, the NETF does not resample the ensemble members, but transforms them from the forecast to the analysis ensemble such that the covariance matrix of the fully nonlinear particle filter is obtained. To this end, it has been classified as a second-order exact particle filter (see Acevedo *et al.*, 2016; van Leeuwen *et al.*, 2019).

The NETF uses a transform matrix $\hat{\mathbf{A}}$ of size $N_e \times N_e$, which is defined by

$$\hat{\mathbf{A}} = N_e \left(\text{diag}(\mathbf{w}) - \mathbf{w}\mathbf{w}^T \right), \quad (9)$$

where $\text{diag}(\mathbf{w})$ denotes the diagonal matrix, the diagonal of which holds the values of the vector $\mathbf{w}$ of size N_e. This

vector holds the likelihood weights of all ensemble states. It is computed according to the assumed distribution of the observation error. For Gaussian errors, the weight for ensemble state i, $i = 1, \ldots, N_e$, is first computed as

$$\hat{w}^{(i)} = \exp\left(-0.5\left(\mathbf{y} - \mathbf{Hx}^{f(i)}\right)^T \mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{Hx}^{f(i)}\right) \right) . \quad (10)$$

The weights are then normalized so that their sum is one:

$$\mathbf{w} = \hat{\mathbf{w}} / \sum_{i=1}^{N_e} \hat{w}^{(i)}. \quad (11)$$

The weight vector and matrix for the ensemble transformation of the NETF are now

$$\tilde{\mathbf{w}}_{\text{NETF}} = \mathbf{w}, \quad (12)$$

$$\mathbf{W}_{\text{NETF}} = \hat{\mathbf{A}}^{1/2}\mathbf{\Lambda}, \quad (13)$$

with the symmetric square root $\hat{\mathbf{A}}^{1/2} = \hat{\mathbf{U}}\hat{\mathbf{S}}^{1/2}\hat{\mathbf{U}}^T$ computed from the singular-value decomposition $\hat{\mathbf{U}}\hat{\mathbf{S}}\hat{\mathbf{V}} = \hat{\mathbf{A}}$. $\mathbf{\Lambda}$ is the identity or a random matrix that preserves the mean and covariance. Using the random variant stabilized the NETF and reduced the risk of filter divergence in numerical experiments (Tödter and Ahrens, 2015; Tödter *et al.*, 2016).

Covariance inflation is applied in the NETF by inflating the forecast ensemble spread directly with a factor $1/\sqrt{\rho}$, where ρ is the forgetting factor; see Section 2.2. An alternative inflation was used by Feng *et al.* (2020) following Poterjoy *et al.* (2019) based on the effective sample size:

$$N_{\text{eff}} = \left(\sum_{i=1}^{N_e} (w^{(i)})^2 \right)^{-1}, \quad (14)$$

where $w^{(i)}$ are the normalized weights (Equation 11). The value of N_{eff} range between N_e if all ensemble states have the same weight and one if one ensemble state gets the weight one, while all others get zero weight. A particle filter will be efficient for values well in between these extremes. For $N_{\text{eff}} = 1$, the ensemble would collapse to a single member, while for $N_{\text{eff}} = N_e$ the ensemble and hence the state estimate in a particle filter would be unchanged. Feng *et al.* (2020) apply an inflation factor β as $\beta\mathbf{R}^{-1}$. β is computed from the condition

$$\frac{N_{\text{eff}}[\beta]}{N_e} \geq \alpha, \quad (15)$$

where $N_{\text{eff}}[\beta]$ denotes the effective sample size that is obtained with weights computed using $\beta\mathbf{R}^{-1}$ and $0 \leq \alpha \leq 1$ is a chosen threshold. The inflation ensures that N_{eff} stays above the prescribed α. This inflation method, denoted

α - inflation below, is analogous to that proposed by Frei and Künsch (2013) as an adaptive choice to regulate a hybrid filter (see Section 3.2). Numerically, we found that the lowest estimation errors are obtained if both the inflations using ρ and β are combined.

3 | HYBRID FILTERING

The theoretical basis for a filter that hybridizes between an EnKF and a PF was described by Frei and Künsch (2013). The weights in a PF with resampling and in the NETF are computed from the likelihood $p(\mathbf{y}|\mathbf{x})$ of the state $\mathbf{x}$ given the observation $\mathbf{y}$. Equation 10 gives the likelihood for Gaussian observation errors. The likelihood can be factorized as

$$p(\mathbf{y}|\mathbf{x}) = p(\mathbf{y}|\mathbf{x})^{\gamma} p(\mathbf{y}|\mathbf{x})^{(1-\gamma)}, \qquad (16)$$

with $0 \le \gamma \le 1$. For Gaussian observation errors, the potentiation with γ is equivalent to dividing the covariance matrix $\mathbf{R}$ by γ (see Equation 10), that is, increasing the observation error. For the hybrid filter, the ensemble analysis steps are computed from the left and the intermediate result will be used as input for the second step, usually denoted as tempering (see van Leeuwen *et al.*, 2019). On this basis, Frei and Künsch (2013) derived a hybrid filter that combined the stochastic EnKF with a PF, while Robert *et al.* (2018) introduced a variant using the LETKF. In both variants, the Kalman filter is applied before the PF. Further, Chustagulprom *et al.* (2016) introduced variants that combine an ensemble Kalman filter with the ensemble transform particle filter by Reich (2013).

An alternative to the tempering above is to combine the analysis increments of two filters linearly in the form

$$\mathbf{X}^{a} = \overline{\mathbf{x}}^{f}\mathbf{1}^{T} + (1-\gamma)\Delta\mathbf{X}_{A} + \gamma\Delta\mathbf{X}_{B}, \qquad (17)$$

where $\Delta\mathbf{X}_{A}$ is the analysis increment computed with filter method A. In contrast to the tempering in Equation 16, γ will shift the increment between the two filters linearly, but will not modify $\mathbf{R}$.

The hybrid filter variants based on Equations 16 or 17 have in common that the limiting cases for $\gamma = 1$ and $\gamma = 0$ are the two original filters.

3.1 | Hybrid Kalman–nonlinear ensemble transform filter

Here, the LETKF is combined with the NETF to define variants of a hybrid filter. To formulate the hybrid Kalman–nonlinear ensemble transform filter (KNETF),

we use as basis the combined Equation 3 for the ensemble update. There are three variants of the hybrid update.

The two-step update schemes motivated above compute in the first step the analysis with one of the two filters. The updated ensemble is then used as the input to the update of the second filter, which yields the final analysis ensemble. The two filter updates use a modified observation-error covariance matrix: the ETKF update uses $\gamma\mathbf{R}^{-1}$, while the NETF update uses $(1-\gamma)\mathbf{R}^{-1}$.

Let the notation $\mathbf{X}_{A}^{a}\left(\mathbf{X}^{f}, \phi\mathbf{R}^{-1}, \mathbf{y}\right)$ denote the analysis ensemble computed with filter method A from the forecast ensemble $\mathbf{X}^{f}$ using the inverse observation covariance matrix $\phi\mathbf{R}^{-1}$. Now we can write the two-step update, in which the NETF is applied first followed by the LETKF, as

$$\tilde{\mathbf{X}}_{HNK} = \mathbf{X}_{NETF}^{a}\left(\mathbf{X}^{f}, (1-\gamma)\mathbf{R}^{-1}, \mathbf{y}\right), \qquad (18)$$

$$\mathbf{X}_{HNK} = \mathbf{X}_{LETKF}^{a}\left(\tilde{\mathbf{X}}_{HNK}, \gamma\mathbf{R}^{-1}, \mathbf{y}\right). \qquad (19)$$

We denote this two-step update scheme as HNK.

The hybrid analysis scheme in which the LETKF is applied before the NETF is denoted HKN. It is given by

$$\tilde{\mathbf{X}}_{HKN} = \mathbf{X}_{LETKF}^{a}\left(\mathbf{X}^{f}, \gamma\mathbf{R}^{-1}, \mathbf{y}\right), \qquad (20)$$

$$\mathbf{X}_{HKN} = \mathbf{X}_{NETF}^{a}\left(\tilde{\mathbf{X}}_{HKN}, (1-\gamma)\mathbf{R}^{-1}, \mathbf{y}\right). \qquad (21)$$

The third variant uses a one-step update scheme (denoted HSync below). Let $\Delta\mathbf{X}$ denote the assimilation increment of a filter, that is, $\Delta\mathbf{X} = \mathbf{X}^{a} - \mathbf{X}^{f}$. Then the hybrid update can be written as

$$\mathbf{X}_{HSync}^{a} = \overline{\mathbf{X}}^{f} + (1-\gamma)\Delta\mathbf{X}_{NETF} + \gamma\Delta\mathbf{X}_{LETKF}. \qquad (22)$$

The hybrid weight γ shifts the filter behavior between the LETKF (for $\gamma = 1$) and the NETF (for $\gamma = 0$).

3.2 | Choosing the hybrid weight

The choice of the hybrid weight γ shifts the filter behavior between the LETKF and the NETF. Its value is expected to be crucial for the performance of the hybrid filters. However, it is not obvious according to which rule the value should be set. Here different approaches will be tested.

A simple approach is to set a constant value of γ. A limitation of a constant γ is that this cannot adapt to the dynamical changes in the ensemble distribution.

3.2.1 | Hybrid weights based on effective sample size

As an alternative to a constant γ, Frei and Künsch (2013) proposed to set γ based on the effective sample size, Equation 14, analogous to the inflation that was later introduced by Feng *et al.* (2020). Frei and Künsch (2013) propose to choose γ so that

$$\frac{N_{\mathrm{eff}}[1 - \gamma_\alpha]}{N_e} \geq \alpha, \qquad (23)$$

where $N_{\mathrm{eff}}[1 - \gamma_\alpha]$ denotes the effective sample size that is obtained with weights computed using $(1 - \gamma_\alpha)\mathbf{R}^{-1}$, where $0 \leq \alpha \leq 1$ is a chosen parameter. γ_α is determined iteratively by varying it from 0 with a pre-defined step size of 0.05 and computing the weight vector $\mathbf{w}$ and N_{eff} for each case. γ_α is found as soon as α is exceeded. This α - method introduces a nonlinear function for γ that is controlled by α.

The α - method finds a value of γ according to N_{eff}. However, the method requires the tuning of α for an optimal performance of the hybrid filter. It would be desirable to obtain an adaptive scheme that does not require tuning. Here, an adaptive scheme based directly on the ratio of N_{eff}/N_e is proposed and examined. Namely, we can compute γ according to

$$\gamma_{\mathrm{lin}} = 1 - N_{\mathrm{eff}}/N_e . \qquad (24)$$

Thus, γ_{lin} will be close to one if N_{eff} is small, while for $N_{\mathrm{eff}} \approx N_e$ one obtains $\gamma_{\mathrm{lin}} \approx 0$. Note that the maximum of γ_{lin} is $(N_e - 1)/N_e$, so that there is always a small contribution of the NETF. In our numerical tests, this leads to better results compared with using $1 - (N_{\mathrm{eff}} - 1)/(N_e - 1)$, unless one introduces a maximum limit. Obviously, the rule for γ_{lin} could be augmented, for example, by multiplying N_{eff}/N_e with a factor ω with $0 < \omega \leq 1$, which would shift the hybridization toward the LETKF and would ensure $\gamma_{\mathrm{lin}} \geq \omega$, but add a tunable parameter.

Note that γ acts on the NETF in the same way as the α - inflation discussed in Section 2.3. Due to this, one cannot apply both methods in combination. However, one could interpret the hybridization as completing the NETF with α - inflation by applying the LETKF to the part of the observational information that is omitted by the α - inflated NETF analysis.

3.2.2 | Hybrid weights based on non-Gaussianity

The motivation for adaptive schemes using N_{eff} is based on the fact that, for $N_{\mathrm{eff}} \approx 1$, the ensemble of a PF, and

hence the NETF, will collapse. Increasing N_{eff} will lead to more similar weights and hence a more stable filter. However, there is no guarantee that a nonlinear filter will work well for larger N_{eff}. For example, it is known that the NETF suffers from higher sampling errors than the LETKF (see Tödter *et al.*, 2016; Kirchgessner *et al.*, 2017). For this reason, one expects that for small ensembles the LETKF should perform better in situations where the ensemble distribution is close to Gaussian (see also the theoretical considerations by Morzfeld and Hodyss, 2019). Thus, it is not obvious that N_{eff} is the right quantity to base the hybrid weight on.

As an alternative, we propose here to base the computation of the hybrid weight on the nonlinearity of the data assimilation problem, which is represented by the non-Gaussianity of the ensemble distribution. To quantify the non-Gaussianity, we use the skewness *skew* and excess kurtosis *kurt* of the ensemble. For some values $a^{(i)}, i = 1, \ldots, N_e$, they are defined by

$$skew = \frac{\frac{1}{N_e} \sum_{i=1}^{N_e} \left(a^{(i)} - \overline{a}\right)^3}{\left[\frac{1}{(N_e - 1)} \sum_{i=1}^{N_e} \left(a^{(i)} - \overline{a}\right)^2\right]^{3/2}}, \qquad (25)$$

$$kurt = \frac{\frac{1}{N_e} \sum_{i=1}^{N_e} \left(a^{(i)} - \overline{a}\right)^4}{\left[\frac{1}{(N_e)} \sum_{i=1}^{N_e} \left(a^{(i)} - \overline{a}\right)^2\right]^2} - 3 . \qquad (26)$$

For infinite sample size, the skewness is zero for a symmetric distribution and deviates increasingly from zero if the asymmetry grows. The kurtosis specifies the width of the distribution. For a Gausssian distribution it is zero. A narrow distribution with long tails has positive *kurt*, while *kurt* is negative for a wide distribution with short tails. An ensemble filter that accounts for skewness was introduced by Hodyss (2012). This filter is based on the square of the ensemble perturbation and does not involve a nonlinear filter scheme.

To use *skew* and *kurt* as conditions for γ, we compute these separately for each observation in the observed ensemble. Thus, we use $a^{(i)} = [\mathbf{Hx}^{f(i)}]_k$, where k denotes the vector element. This yields $skew_k$ and $kurt_k$. The absolute values are then averaged over all observations to obtain the mean absolute skewness (*mas*) and kurtosis (*mak*):

$$mas = \frac{1}{N_y} \sum_{k=1}^{N_y} |skew_k|, \qquad mak = \frac{1}{N_y} \sum_{k=1}^{N_y} |kurt_k|. \qquad (27)$$

Finally, we normalize *mas* and *mak*, since the maxima of the skewness and kurtosis depend on the ensemble

size. As shown in the Appendix, $mas \leq \sqrt{N_e}$ and $mak \leq N_e$. However, normalizing directly by these limits would decrease the effect of the skewness and kurtosis if N_e is increased, while their values should be approximately constant for a given assimilation case. To this end, we normalize by a parameter κ, which can be chosen so that minimum errors are obtained. This leads to the normalized mean absolute skewness (*nmas*) and kurtosis (*nmak*):

$$nmas = \frac{1}{\sqrt{\kappa}} \, mas, \qquad nmak = \frac{1}{\kappa} \, mak. \qquad (28)$$

Note that these statistics do not measure the non-Gaussianity of the joint distribution of $\mathbf{Hx}^{f(i)}$, but represent an average non-Gaussianity over the observations.

Using *nmas* and *nmak*, we can define rules for the hybrid weight γ as

$$\gamma_{\text{sk},\alpha} = \text{argmax} \left[\text{argmin} \left(1 - |nmak|, 1 - |nmas| \right), \gamma_\alpha \right], \qquad (29)$$

$$\gamma_{\text{sk,lin}} = \text{argmax} \left[\text{argmin} \left(1 - |nmak|, 1 - |nmas| \right), \gamma_{\text{lin}} \right], \qquad (30)$$

where γ_α and γ_{lin} are defined in Equations 23 and 24, respectively. These rules compute the hybrid weight by taking the stronger effect of either the skewness or kurtosis and limiting this value by γ_α or γ_{lin}. We have to stress that taking a linear dependence on *nmas* or *nmak* is not necessarily optimal, but it is used here as a simple choice to study whether these rules are effective in the numerical tests.

4 | FILTER PERFORMANCE WITH THE LORENZ-63 MODEL

4.1 | Configuration

The Lorenz-63 (L63) model (Lorenz, 1963) is defined by a set of three coupled differential equations describing three variables X, Y, Z. The model has been used in several data assimilation studies on nonlinearity (recently, e.g., in Metref *et al.*, 2014; Morzfeld and Hodyss, 2019). Here, it is used with the typical configuration using the parameter values $\sigma = 10$, $\rho = 28$, $\beta = 8/3$. The time integration is performed with the fourth-order Runge–Kutta scheme using a time-step size of 0.05. Following Bocquet (2011), if all three variables are observed, one can classify the nonlinearity for a forecast duration $\Delta t = 0.1$ as mild, $\Delta t = 0.25$ as medium,

and $\Delta t = 0.5$ as strong. Here, we assess the performance of the hybrid filter for the range $0.1 \leq \Delta t \leq 0.7$ to cover the different regimes of nonlinearity.

The truth for the experiments is computed by a model integration over 11,000 time steps. All three state variables are observed. The observations are generated from the truth by adding Gaussian random noise with a variance of 4.0. Ensemble sizes between 15 and 100 are used. The initial ensembles are generated by random draws from the true trajectory. The assimilation experiments are performed over 10,000 time steps, with varying Δt. There is no localization applied. The ensemble inflation (forgetting factor ρ for the ETKF, combined inflation by ρ, and α inflation for the NETF) is varied and the minimum errors are reported. The experiment is run with a random matrix Λ in Equations 8 and 13. The filter performance is assessed based on the time-averaged continuous-ranked probability score (CRPS: Hersbach, 2000) with respect to the truth over all three state variables. The CRPS evaluates the whole ensemble distribution, while the root-mean-square error (RMSE) would only assess the ensemble mean. This makes the CRPS better suited to assess non-Gaussian distributions than the RMSE. For computing the CRPS, the first 200 time steps are omitted to avoid the spin-up of the data assimilation process. For each Δt and choice of inflation, the experiment is repeated 10 times with different random numbers to generate the initial ensemble and the average CRPS is reported.

The L63 model and the filters are implemented using the parallel data assimilation framework (PDAF: Nerger *et al.*, 2005; Nerger and Hiller, 2013). This provides high computational efficiency and the possibility to perform the assimilation with the different toy models, but also realistic models using the same software.

4.2 | Assimilation results

4.2.1 | Dependence on ensemble size

Before comparing the effect of the different rules to compute the hybrid weight, we assess the convergence behavior of the filters with increasing ensemble size. Figure 1 shows the minimum CRPS that was obtained when varying the forgetting factor (for all filters) and the inflation limit α (for NETF). Shown are three cases of forecast lengths $\Delta t = \{0.1, 0.4, 0.7\}$, which represent the different regimes of nonlinearity. The CRPS of the ETKF shows only very small changes when the ensemble size is increased. The NETF shows a higher error than the ETKF for small N_e, but a strong decrease of the CRPS when N_e is increased. When N_e exceeds 30, the CRPS of the NETF is smaller than that of the ETKF for $\Delta t = 0.4$ and 0.7. For the short
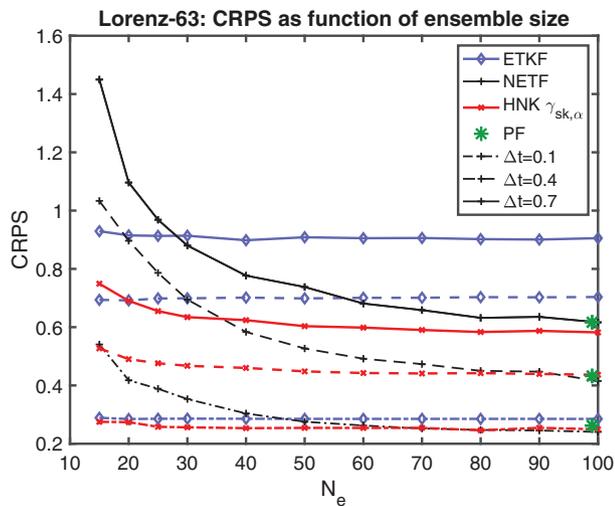
FIGURE 1 Minimum CRPS as a function of the ensemble size N_e. Shown are the ETKF (blue), NETF (black), and hybrid filter variant HNK, using the hybrid rule $\gamma_{sk,\alpha}$ that yielded the smallest errors. The CRPS are shown for $\Delta t = 0.1$ (dash–dotted), $\Delta t = 0.4$ (dashed), $\Delta t = 0.7$ (solid). The effect of the hybrid filter grows with increasing forecast length. The green asterisks show the CRPS of a particle filter at $N_e = 100$ [Colour figure can be viewed at wileyonlinelibrary.com]

forecast $\Delta t = 0.1$, the NETF reaches the same error as the ETKF at $N_e = 50$ and yields significantly smaller CRPS than the ETKF for larger ensembles. The reduction of the CRPS by the NETF also increases with Δt. For $N_e = 100$, the CRPS obtained with the NETF is nearly identical to those obtained with a PF with resampling.

Next to the ETKF and NETF, the CRPS for the hybrid filter variant HNK is shown for the rule $\gamma_{sk,\alpha}$ for $\alpha = 0.25$, which resulted in the overall lowest CRPS for $N_e = 25$ (see Section 4.2.3). For $\Delta t = 0.1$, thus small nonlinearity, the hybrid filter reduces the CRPS by up to 12.5% compared with the ETKF. For the longer forecasts, which result in larger nonlinearity, the CRPS of the HNK filter is up to 38.0% smaller than that of the ETKF. The HNK filter is particularly efficient in reducing the CRPS for small ensembles. Only for $N_e < 30$ does the HNK filter show a significant increase of the CRPS, but the error remains below that of the ETKF. For increasing ensemble size, the CRPS of the NETF approaches that of the HNK filter. The HNK filter yields overall lower CRPS than the NETF except for $N_e = 100$ for $\Delta t = 0.1$ and $\Delta t = 0.4$. For a very large ensemble size of $N_e = 500$, the NETF and PF with resampling still yield nearly the same CRPS. In addition, the $\gamma_{sk,\alpha}$ and γ_α cases can be tuned to yield comparable CRPS. However, setting $\alpha = 0.01$ is required for this very large ensemble so that N_{eff} influences the hybridization only minimally.

## 4.2.2 | Comparison of hybrid filter variants and weight rules

We now focus on the ensemble size $N_e = 25$ and vary the forecast length Δt. Figure 2 shows the CRPS for the hybrid filter variants and the ETKF and NETF. The columns show the CRPS for the three hybrid variants HNK, HKN, and HSync, while the panels of each row show the CRPS obtained with different rules to compute the hybrid weight γ. The NETF exhibits a larger CRPS than the ETKF for this ensemble size for all Δt.

Overall, the hybrid filter variant HNK (left column) results in the smallest CRPS. The HKN filter (center column) is not able to reduce the error through hybridization, but the different rules used to compute the hybrid weight lead to a deterioration of different strength. Thus, combining the ETKF with a subsequent NETF update does not improve the result for the Lorenz-63 model. Lower assimilation impact when an EnKF was applied first was also found with the hybrid filter discussed by Chustagulprom *et al.* (2016). For the HSync filter (right column), the different rules for the hybrid weight decrease the CRPS. However, the effect is smaller than for the HNK filter.

Focusing first on the HNK filter in the left column, Figure 2a shows that, for fixed values of γ, only choices close to one lead to a stable filter behavior. Varying γ, $\gamma = 0.9$ resulted in the smallest CRPS, well below the CRPS of the ETKF and NETF. For smaller γ, the HNK filter shows low CRPS for shorter forecasts, but starts to diverge if the forecast length increases. If γ is increased beyond 0.9, the CRPS will approach that of the ETKF. For $\Delta t > 0.3$, the CRPS of the optimal case varies significantly for the 10 repetitions with different initial ensembles as is shown by the shaded region in Figure 2a, which marks the range of one standard deviation around the case with $\gamma = 0.9$.

Figure 2b shows the CRPS for cases computing γ using N_{eff} with either Equation 23 for γ_α or Equation 24 for γ_{lin}. For γ_α, the hybrid weight is calculated, so that the ratio N_{eff}/N_e does not fall below the threshold α. In this case, high CRPS values close to those of the NETF are obtained for small α, as shown for $\alpha = 0.2$. When increasing α, the CRPS first decreases and finally increases again and approaches that of the ETKF. A minimum CRPS below that of the ETKF is obtained for $\alpha = 0.4$. For γ_{lin}, no tuning is necessary. In this case, the CRPS lies below the minimum error obtained with γ_α for all Δt. Compared with the ETKF, the CRPS is reduced by 8% for $\Delta t = 0.1$, and by 25–31% for $\Delta t \geq 0.3$.

The CRPS for the cases that compute γ based on the ensemble skewness and kurtosis is shown for the HNK filter in Figure 2c. Here, the results for the value of κ that leads to the minimum CRPS are shown. The dependence on κ is discussed in Section 4.2.3. For $\gamma_{sk,\alpha}$ (Equation 29),
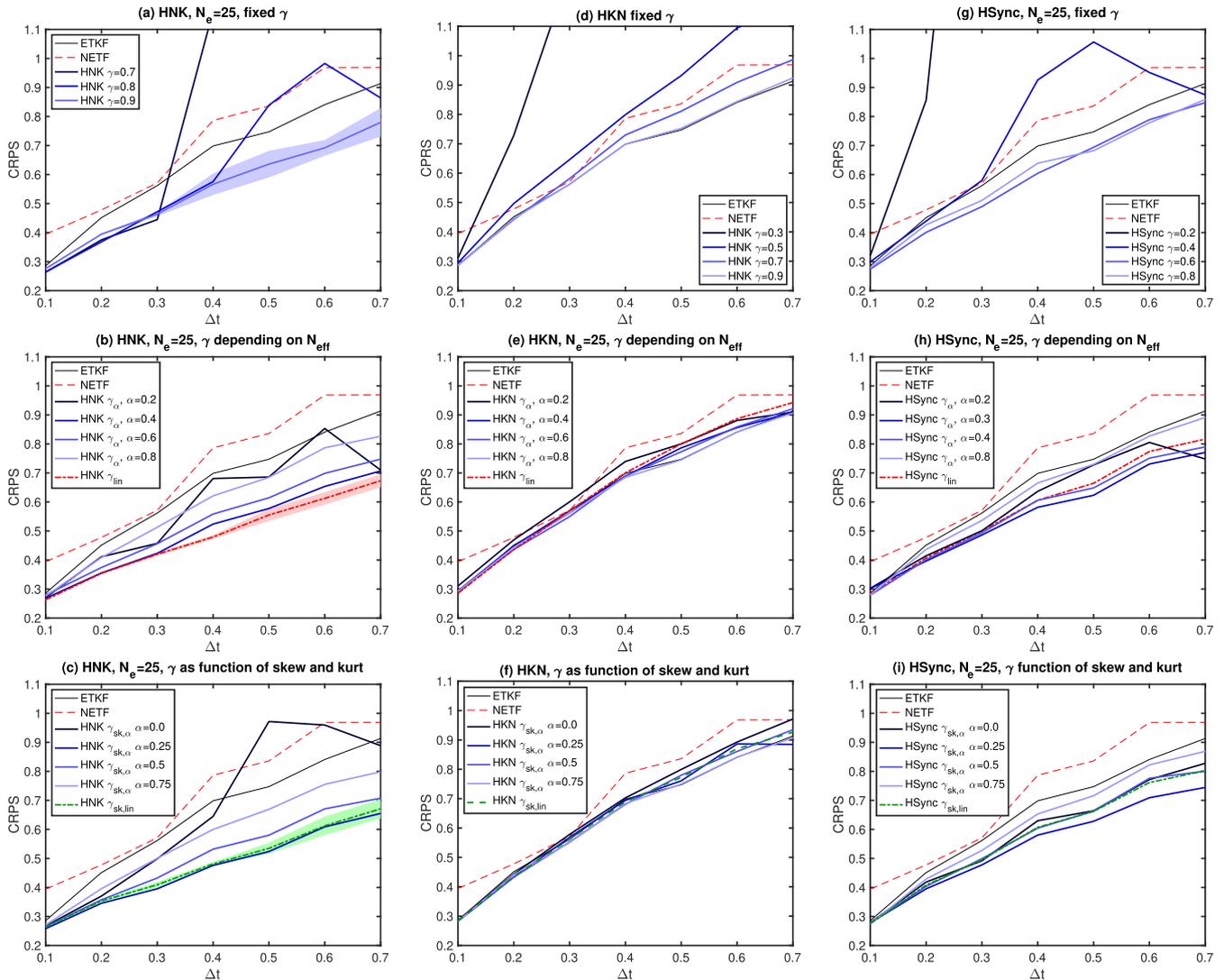
FIGURE 2 Minimum CRPS as a function of the forecast length Δt. Shown are the ETKF (thin solid), NETF (thin dashed), and hybrid filter (thick lines; different colors). The columns show the filter variants HNK (left), HKN (center), and HSync (right). The rows show (top) fixed values of γ, (middle) γ computed using N_{eff}, and (bottom) γ computed using the ensemble skewness and kurtosis. In the panels in the left column, the shaded areas around the line for one of the hybrid rules show the range of one standard deviation over each ten experiments. Only a selection of choices for the fixed γ or α is shown to demonstrate the effect of the parameters [Colour figure can be viewed at wileyonlinelibrary.com]

the hybrid weight varies with the choice of α. For $\alpha = 0.0$, γ is computed from the skewness and kurtosis without accounting for N_{eff}. In this case, the filter shows small CRPS for $\Delta t \leq 0.2$, but significantly larger errors occur for longer forecasts. The smallest overall CRPS is obtained for $\alpha = 0.25$, while for larger α the CRPS increases and approaches the CRPS of the ETKF. For $\gamma_{sk,lin}$ (Equation 30), the hybrid weight is computed with only κ as tuning parameter. In this case, the CRPS is statistically not distinguishable from the smallest RMSE obtained for $\gamma_{sk,\alpha}$. For $\Delta t \geq 0.3$, the reduction of the CRPS relative to the ETKF is between 29 and 32%. The shaded areas in Figure 2b,c show that the variation within the ten repetitions of each set of

parameters is slightly larger for $\gamma_{sk,lin}$ compared with γ_{lin}. Both show a smaller variation than the fixed γ in Figure 2a. Thus, the hybridization stabilizes the filter process.

While the HSync filter variant in the right column of Figure 2 yields larger CRPS than the HNK filter, the hybridization is also able to reduce the CRPS. The adaptive rules γ_{lin} and $\gamma_{sk,lin}$ lead to similarly reduced CRPS. For γ_α, the filter is less sensitive to the choice of α than in the HNK filter. The smallest CRPS values are obtained for $\alpha = 0.3$. For HSync, the CRPS values for γ_α are lower than for γ_{lin} for α between about 0.2 and 0.5. For a fixed choice of γ, values between 0.6 and 0.8 lead to the smallest CRPS. For the rule $\gamma_{sk,\alpha}$, the smallest CRPS is obtained for $\alpha = 0.25$.

However, the CRPSs for $\alpha = 0.0$ is also smaller than that of the ETKF. Thus, for HSync the hybrid weight can also be computed just based on skewness and kurtosis, without the additional constraint specified by $\alpha > 0$.

4.2.3 | Effects of different hybridization rules

For a further assessment of the hybridization, we focus on the HNK filter variant and the longest forecast time of $\Delta t = 0.7$. To get more insight into the effect of accounting for skewness and kurtosis, we first assess in Figure 3 how the CRPS for the hybrid rules $\gamma_\alpha, \gamma_{sk,lin}$, and $\gamma_{sk,\alpha}$ depends on the parameters α and κ. Only $\gamma_{sk,\alpha}$ depends on both parameters, while γ_α only depends on α and $\gamma_{sk,lin}$ only depends on κ. The central panel of Figure 3 shows the CRPS for $\gamma_{sk,\alpha}$. Small CRPS is generally obtained for κ between 5 and 30 and $0.1 \le \alpha < 0.4$. The smallest CRPS is obtained for $\alpha = 0.1$ and $\kappa = 10$. This optimal κ is smaller than the ensemble size $N_e = 25$. The CRPS varies more strongly with α than with κ. If α is reduced below 0.1, the CRPS increases strongly. This increase is largest for $\kappa \le 30$, thus when both α and κ allow for a small value of γ. For $\alpha = 0.0$, the smallest CRPS is obtained for $\kappa = 95$. With this large value, the dependence rule using *nmas* and *nmak* counters the missing limitation by the effective sample size. For $\alpha > 0.5$, the effect of varying κ diminishes and the CRPS increases. Here, the limit according to $\alpha = N_{eff}/N_e$ dominates the hybrid rule and the possible effect of the skewness and kurtosis is not taken into account.

The left column in Figure 3 shows the CRPS for γ_α. The result is nearly identical to the case of $\gamma_{sk,\alpha}$ with $\kappa = 1$. In this case, the criterion using *nmas* and *nmak* has very little effect and the hybrid rule is dominated by the limit set by α. The rule $\gamma_{sk,lin}$ also profits from small values of κ, as is visible in the bottom row of Figure 3. Here, the minimum CRPS is obtained for $\kappa = 5$. The CRPS values obtained with $\gamma_{sk,lin}$ are comparable with those of $\gamma_{sk,\alpha}$ obtained with $\alpha = 0.35$.

To assess how the adaptive hybrid rules act differently during the DA process, Table 1 summarizes different statistics of selected cases for $\Delta t = 0.7$. The left four columns show cases with small CRPS. Here, $\gamma_{sk,\alpha=0.1}$ results in the smallest CRPS, followed by $\gamma_{sk,lin}$ and γ_{lin} with nearly identical CRPS, and finally $\gamma_{\alpha=0.4}$. This order also holds for the root-mean-square error (RMSE). Comparable values of CRPS are obtained for significantly different mean values $\overline{\gamma}$. In particular, γ_{lin} and $\gamma_{sk,lin}$ have nearly identical CRPS, but the mean value $\overline{\gamma}$ of 0.628 for γ_{lin} is significantly smaller than $\overline{\gamma} = 0.756$ for $\gamma_{sk,lin}$. The case $\gamma_{\alpha=0.4}$ uses a much smaller $\overline{\gamma}$ of 0.399. For $\gamma_{sk,\alpha=0.1}$, the mean value $\overline{\gamma}$ of 0.717 lies in between the cases $\gamma_{sk,lin}$ and γ_{lin}. For $\gamma_{sk,\alpha=0.1}$, the average N_{eff}/N_e is also larger than for the other cases,
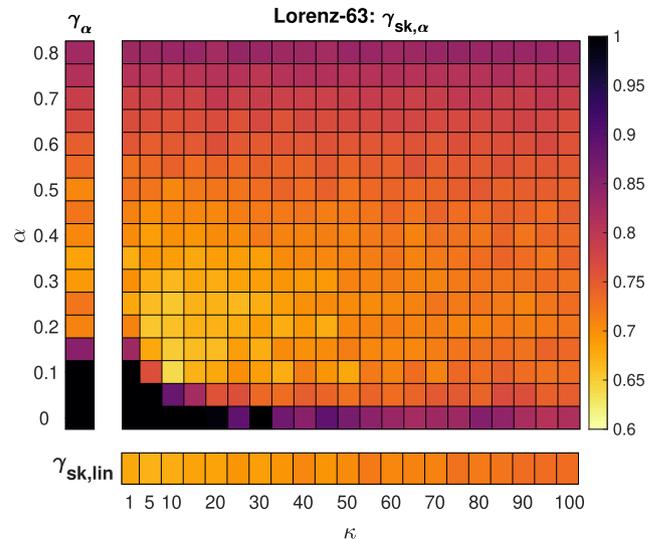

FIGURE 3 Minimum CRPS over all choices of α and κ for the HNK filter and $N_e = 25$. Shown are (center) the CRPS for $\gamma_{sk,\alpha}$, (left) γ_α, and (bottom) $\gamma_{sk,lin}$ [Colour figure can be viewed at wileyonlinelibrary.com]

which might indicate a more stable DA process. In contrast, the statistics *mas* and *mak* show no obvious relation to the CRPS. However, *nmas* and *nmak* are smaller for $\gamma_{sk,\alpha=0.1}$ than for $\gamma_{sk,lin}$. This is related to the different optimal choices of κ. The smaller values of *nmas* and *nmak*, however, do not translate directly into smaller values of $\overline{\gamma}$.

For the hybrid filter, it is overall relevant that during the DA process a "suitable" value of γ is chosen, that is, one that yields an analysis ensemble with low errors and sufficiently large N_{eff}. Thus, the mean value $\overline{\gamma}$ does not relate directly to the CRPS. The different hybrid rules lead to distinct solutions, as is visible in Figure 4, which shows the cumulative distribution of γ over all analysis steps. For the case $\gamma_{\alpha=0.4}$, a value $\gamma = 0$ is used in about 40% of the analysis steps. In these cases, the limit of $\alpha \ge 0.4$ is fulfilled without inflating the observation errors by the hybridization. For the remaining analysis steps, the distribution of γ is rather uniform in the range 0.4–1.0, with some increased occurrence of values above 0.8. This distribution leads to the overall low value of $\overline{\gamma}$ in Table 1. Adding the condition on skewness and kurtosis in the case $\gamma_{sk,\alpha=0.1}$ changes the distribution of γ drastically and leads to values of γ that are mainly above 0.4. This is despite the lower threshold of $\alpha = 0.1$. A change of the distribution of γ is also visible when comparing the cases γ_{lin} and $\gamma_{sk,lin}$. The case γ_{lin} shows a rather uniform distribution of γ in the range 0.2–1, but small probability of $\gamma < 0.2$. Accounting in addition for skewness and kurtosis in $\gamma_{sk,lin}$ leads to a similar distribution as for $\gamma_{sk,\alpha=0.1}$, with values of γ mainly above 0.4. However, $\gamma_{sk,\alpha=0.1}$ shows about three times higher probability of $\gamma < 0.4$ than the case $\gamma_{sk,lin}$. If N_{eff} is large enough,

TABLE 1   Mean errors and parameter values over experiments with L63 for $\Delta t = 0.7$

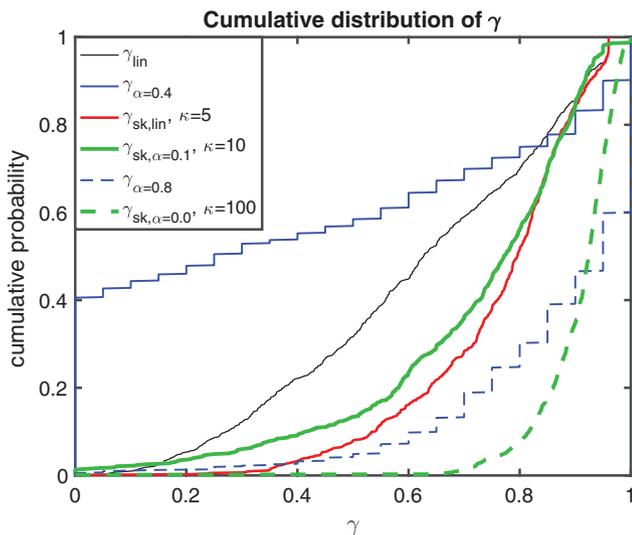| Filter:	γ_{lin}	$\gamma_{\alpha=0.4}$	$\gamma_{\text{sk,lin}}$	$\gamma_{\text{sk},\alpha=0.1}$	$\gamma_{\alpha=0.8}$	$\gamma_{\text{sk},\alpha=0.0}$
κ	–	–	5	10	–	100
CRPS	0.673	0.707	0.671	0.639	0.826	0.873
RMSE	1.125	1.178	1.113	1.105	1.372	1.575
$\overline{\gamma}$	0.628	0.400	0.756	0.717	0.863	0.862
$N_{\text{eff}}/N_{\text{e}}$	0.372	0.355	0.354	0.404	0.289	0.289
mas	0.935	0.934	0.948	0.906	0.964	0.964
mak	2.026	1.972	2.073	1.935	2.126	2.126
nmas	–	–	0.424	0.286	–	0.096
nmak	–	–	0.415	0.194	–	0.021

FIGURE 4   Cumulative distribution of γ for six cases of the Lorenz-63 model over all analysis steps of an experiment. The cases correspond to those for which mean statistics are shown in Table 1 [Colour figure can be viewed at wileyonlinelibrary.com]

a small value of γ can be used without causing an ensemble collapse. However, the cases $\gamma_{\text{sk,lin}}$ and $\gamma_{\text{sk},\alpha}$ show that such a low value is not necessary to obtain a low CRPS if the skewness and kurtosis are small enough.

Two suboptimal cases of higher CRPS are shown in the rightmost columns in Table 1. The higher errors are caused by episodes of increased CRPS, in which the filter diverges for a few analysis cycles during the DA process. The case $\gamma_{\alpha=0.8}$ uses too large a limit compared with $\gamma_{\alpha=0.4}$. With the condition $\alpha = 0.8$, $\overline{\gamma}$ more than doubles compared with $\gamma_{\alpha=0.4}$. Figure 4 shows that a value of $\gamma = 0$ is almost never reached for $\gamma_{\alpha=0.8}$, while in 40% of the analysis steps the maximum value $\gamma = 1.0$ is used. The second suboptimal case is $\gamma_{\text{sk},\alpha=0.0}$, for which a high value $\kappa = 100$ leads to the lowest CRPS. In this case, the choice of κ counters the missing limit regarding α. However, this alone is not

sufficient to obtain a low CRPS. The distribution of γ shows increased values compared with $\gamma_{\text{sk},\alpha=0.1}$. The two suboptimal cases use nearly identical mean values of γ, but these are obtained by different distributions. However, both methods have in common that particularly high values of γ are used, so that too much weight is given to the ETKF.

Table 1 also provides information on the non-Gaussianity of the DA problem. The average *mas* and *mak* are about 0.93 and 2.0, respectively. However, in the different experiments, maximum values of about *mas* = 4.0 and *mak* = 16.0 occur at single analysis steps. Given the maximum limits of 5 for *mas* and 25 for *mak* (see Appendix), these distributions are strongly non-Gaussian, with both significant skewness and kurtosis.

Overall, the analysis shows that there is no unique solution for γ. Small values can sometimes be used without deteriorating the analysis. In fact, the hybridization is sometimes insensitive within some range of γ. It is important that the filter has a "suitable" value of γ at a given analysis step. This aspect points to an obvious weakness of using a constant hybrid weight. In this case, the hybrid filter cannot adjust to the current ensemble distribution. This leads to the higher CRPS visible in the upper row of Figure 2.

5 | HYBRID FILTERING WITH THE LORENZ-96 MODEL

5.1 | Configuration

While the L63 model is a chaotic model that poses significant challenges to the ensemble DA methods, there is the obvious limitation of a very small state dimension. Thus, experiments with ensembles that are smaller than the state dimension are not possible and applying localization is not meaningful. For a higher-dimensional case including

localization, we perform assimilation experiments with the chaotic Lorenz-96 (L96) model (Lorenz, 1996), which was used in many studies to assess data assimilation methods, for example, by Tödter and Ahrens (2015) and Kirchgessner *et al.* (2017) for studying the NETF filter and smoother variants. The configuration uses a state dimension of 40 grid points, forcing parameter $F = 8.0$, and time-stepping with the fourth-order Runge–Kutta scheme using a dimensionless time-step size of $\Delta t = 0.05$. The true state was generated by a model integration over 11,000 time steps. The observations are available at each second grid point. They are generated by adding Gaussian random noise of variance 1.0 to the true states.

The assimilation experiment starts at time step 2,000 of the true trajectory to avoid the model spin-up period. The observations are assimilated at each eighth time step ($\Delta t = 4.0$) over a period of 5,000 time steps. The long interval between successive analysis steps leads to a non-Gaussian forecast ensemble (Lei and Bickel, 2011). Together with the incomplete observations, this configuration leads to significantly higher assimilation errors than the frequently used case of full observations and an analysis update at each time step.

Two cases with different ensemble sizes are used: $N_e = 15$ represents a particular small-ensemble case, while $N_e = 40$ has lower sampling errors. The initial ensembles are generated by random draws from the true trajectory starting at time-step 2001. The experiment is run with a random matrix Λ in Equations 8 and 13.

In the experiments, the assimilation performance in terms of the CRPS averaged over the last 3,000 time steps of each experiment is analyzed varying the forgetting factor and the localization radius, as well as the parameters controlling the hybridization. For each configuration, the experiment is repeated ten times with different random numbers for the ensemble initialization and the CRPS is averaged over these experiments. As for the L63 model, the L96 model is implemented using PDAF (Nerger *et al.*, 2005; Nerger and Hiller, 2013).

Following the discussion in Section 4.2.3, which clarified that a fixed choice of γ cannot react in the time-varying ensemble distribution and hence needs to be suboptimal, we focus here on the adaptive rules to set γ.

5.2 | Assimilation results

5.2.1 | Influence of inflation and localization

To first assess the dependence of the filter results on the inflation and the localization radius r_{loc}, Figure 5 shows the time-mean CRPS for the LETKF, LNETF, and hybrid filter variants with rule $\gamma_{sk,lin}$. The white fields indicate

cases with CRPS above 1.2, which we consider here as the convergence limit, because values above 1.2 are often due to filter divergence in at least one of the 10 repetitions. For the LNETF, two cases are shown. The NETF that only applies inflation by ρ (Figure 5b) is the limiting case of the hybrid filters for $\gamma = 0$. In contrast, the LNETF with combined inflation using ρ and α (Figure 5c) represents the optimal case when using the LNETF without hybridization.

The LETKF (Figure 5a) converges for all tested combinations of forgetting factor ρ and radius r_{loc}. The minimum CRPS of 0.756 is obtained for $r_{loc} = 6$ grid points. In contrast, the LNETF using only inflation by ρ (Figure 5b) converges only for $r_{loc} < 4$. The minimum CRPS of 0.862 is obtained here for $\rho = 0.85$ and $r_{loc} = 2$. When α - inflation is used in addition to ρ (Figure 5c), the LNETF converges for a wider range of localization radii. However, r_{loc} still needs to be much smaller than for the LETKF. The convergence also depends on ρ. The figure shows the smallest errors, obtained with $\rho = 0.85$, as is the case when only ρ is used. The CRPS is smallest for $r_{loc} = 2$ and $\alpha = 0.2$ with a value of 0.848; compared with using only inflation by ρ, the CRPS is reduced by only 1.6%. However, if the ensemble is increased to 40 members, the effect of the α - inflation grows to 7.0% and the minimum CRPS is 0.667, which is only 0.5% larger than the CRPS of the LETKF. Further, the radius for the minimum CRPS in the LNETF is larger with $r_{loc} = 5$.

The lower row of Figure 5 shows the CRPS for the three variants of the hybrid filter using $\gamma_{sk,lin}$. All hybrid filter variants reduce the CRPS compared with the LETKF and LNETF, albeit to a different extent. The filter variants show an overall dependence of ρ and r_{loc} that is also typical for the LETKF. In particular, the optimal r_{loc} increases when ρ is decreased. As for the LETKF, the minimum CRPS is found for some intermediate value of ρ and r_{loc}. For high values of r_{loc} and ρ (i.e., small inflation), all hybrid filter variants show increased CRPS or divergence. This effect is stronger in the HNK and HSync variants (Figure 5d,f) than in the filter variant HKN (Figure 5e) and the LETKF. The similar dependence of the hybrid filters on r_{loc} and ρ indicates that the configuration of the LETKF should be usable as a baseline for tuning the hybrid filter.

5.2.2 | Comparison of hybrid filter variants and weight rules

Here, the performance of the hybrid filter variants is compared and the effect of the different hybrid rules based either on the effective sample size using α (Section 3.2.1) or, in addition, on the ensemble skewness and kurtosis (Section 3.2.2), is assessed.
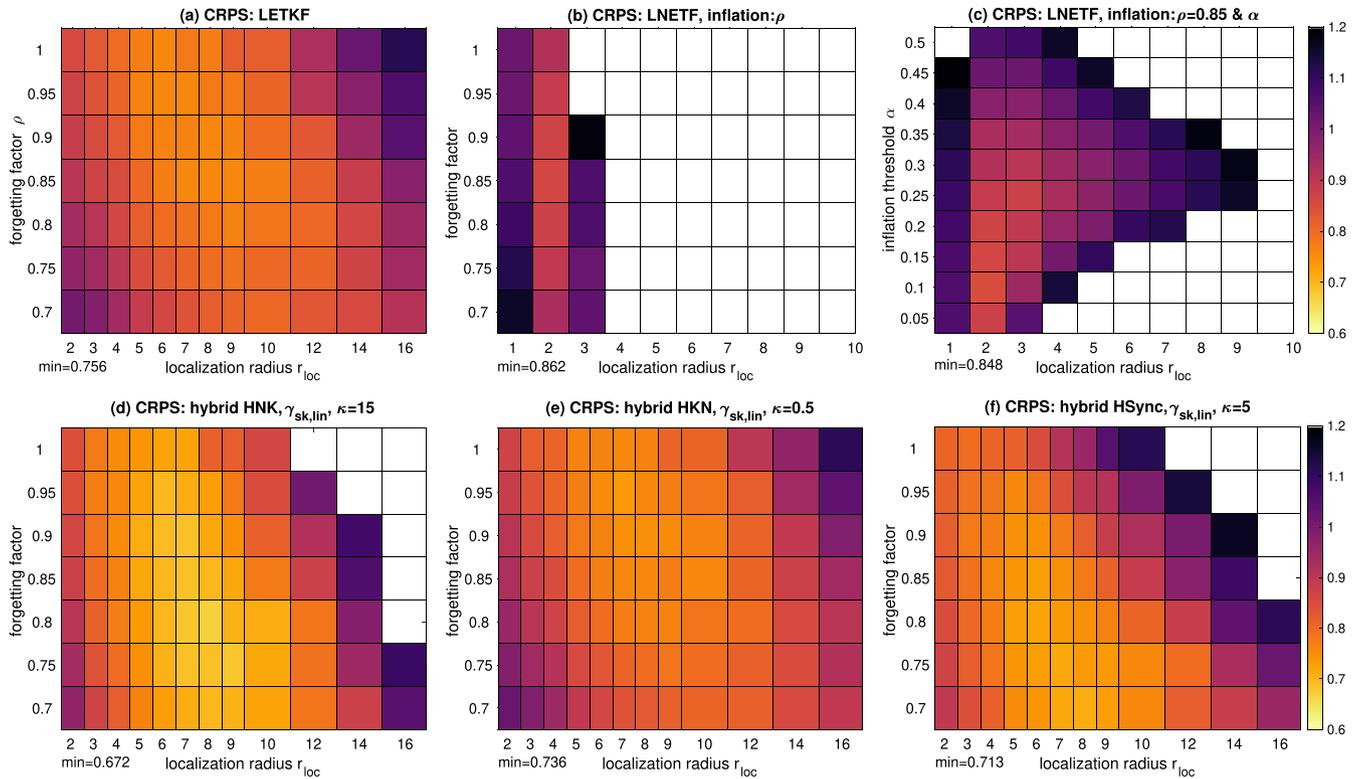
FIGURE 5 CRPS for varying inflation (ρ and α) and localization radius r_{loc} for $N_e = 15$. The upper row shows (a) the LETKF for varying forgetting factor ρ, (b) the NETF applying only ρ, and (c) the NETF with combined use of ρ and α-inflation for $\rho = 0.85$ and varying limit α. The lower row shows three hybrid filter variants for the hybrid rule $\gamma_{\text{sk,lin}}$. The white fields indicate CRPS values above 1.2. The number in the bottom left corner of each panel shows the overall minimum CRPS for each filter [Colour figure can be viewed at wileyonlinelibrary.com]

Table 2 summarizes the reduction of the CPRS relative to the LETKF obtained for the different filter variants and hybrid rules for the two ensemble sizes 15 (left) and 40 (right). As for the L63 model, the HNK filter yields the largest reductions of the CRPS, with up to 11.2% for $N_e = 15$. With the larger ensemble size $N_e = 40$, the reduction of the CRPS is almost twice as large, with 21.5%. The effect of the HSync filter is only about half as large as that of the HNK filter. The HKN filter is also able to reduce the CRPS compared with the LETKF, but the reduction is the smallest of the three hybrid filter variants and reaches only 4.9% for $N_e = 40$. This is different from the L63 model, where the HKN filter was not able to reduce the CRPS.

Among the different hybrid rules, $\gamma_{\text{sk,lin}}$ has the largest effect for $N_e = 15$ for the HNK filter, while the effect of $\gamma_{\text{sk},\alpha}$ is slightly smaller. This changes for the larger ensemble, where $\gamma_{\text{sk},\alpha}$ shows the largest reduction of the CRPS. The rule γ_{lin}, which does not require tuning, is also efficient in reducing the CRPS. However, its effect is smaller than the other rules (except for $N_e = 15$ for the HNK filter, where γ_α is less efficient). For the HKN and HSync filters, the rule $\gamma_{\text{sk},\alpha}$ shows the largest effect, followed by γ_α.

To analyze the influence of the parameters α and κ further, we focus on the HNK filter. Figure 6 shows the CRPS

of the hybrid filter relative to the CRPS of the LETKF as a function of α and κ for the hybrid rules γ_α, $\gamma_{\text{sk},\alpha}$, and $\gamma_{\text{sk,lin}}$ for both ensemble sizes. Here, the white fields indicate that the hybrid filter results in a higher CRPS than the LETKF. For both ensemble sizes, the rule γ_α leads to values of the CRPS very close to $\gamma_{\text{sk},\alpha}$ with $\kappa = 1$. These values are larger than the optimal CRPS obtained with $\gamma_{\text{sk,lin}}$ or $\gamma_{\text{sk},\alpha}$. For small values of α, the rule γ_α leads to increased CRPS. For $N_e = 15$, this is the case for $\alpha < 0.4$, which corresponds to a minimum N_{eff} of 6 according to Equation 23. For $N_e = 40$, the hybrid filters show lower CRPS than the LETKF down to $\alpha = 0.1$. This corresponds to a minimum N_{eff} of 4, thus lower than for $N_e = 15$. Furthermore, the optimal value of α shifts from 0.6 for $N_e = 15$ to 0.4 for $N_e = 40$. Keeping in mind that α is the limit for N_{eff}/N_e and N_e is increased by a factor of 2.7, one sees that for $N_e = 40$ the optimal filter results are obtained at larger N_{eff} than for $N_e = 15$, despite the decreased value of α. However, even for the same N_{eff} ($\alpha \approx 0.2$ for $N_e = 40$), smaller CRPSs are obtained.

For $\gamma_{\text{sk},\alpha}$, combinations of small values of α and κ also result in an increase of the CRPS relative to the LETKF. For the larger ensemble $N_e = 40$, this only happens for $\alpha = 0.0$, $\kappa = 1$, while the range is larger for $N_e = 15$. For the smaller ensemble, the minimum CRPS for $\gamma_{\text{sk},\alpha}$ is obtained

TABLE 2 Maximum reduction of the CRPS in per cent obtained for the different hybrid weight rules for the Lorenz-96 model experiments

Filter	γ_α	γ_{lin}	$\gamma_{\text{sk},\alpha}$	$\gamma_{\text{sk,lin}}$	γ_α	γ_{lin}	$\gamma_{\text{sk},\alpha}$	$\gamma_{\text{sk,lin}}$
	$N_e = 15$				$N_e = 40$			
HNK	8.6	9.3	10.9	***11.2***	19.6	17.6	***21.5***	17.6
HKN	2.7	2.0	*3.2*	2.7	4.7	2.0	*4.9*	1.4
HSync	6.5	4.9	*6.6*	6.0	8.4	6.0	*10.6*	6.4

Note: An italic font indicates the largest reduction of the CRPS over each 4 hybrid rules, while the bold italic number indicates the overall largest reduction.
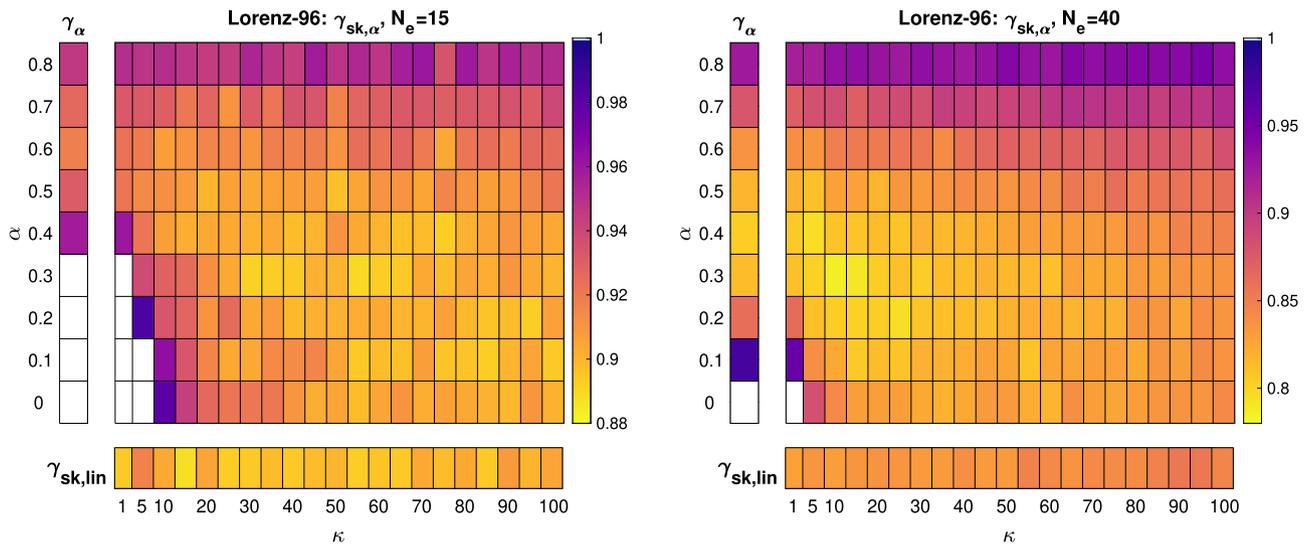
FIGURE 6 Minimum CRPS relative to the minimum CRPS of the LETKF over all choices of α and κ for (left) $N_e = 15$ and (right) $N_e = 40$. Shown are (center) the CRPS for $\gamma_{\text{sk},\alpha}$, (left vertical column) γ_α, and (bottom row) $\gamma_{\text{sk,lin}}$. White fields indicate relative CRPS values above one. Note the different color scale of both panels [Colour figure can be viewed at wileyonlinelibrary.com]

for $\kappa \geq 30$ and $\alpha \leq 0.4$. There is no clear optimal choice of κ and α, but a wider range of these parameters results in low CRPS. For decreased α, there is a tendency for the optimal choice of κ to increase. When the ensemble size is increased to 40 members, a clear minimum of the CRPS is obtained for $\kappa = 10$, $\alpha = 0.3$. As for $N_e = 15$, decreasing α requires us to increase κ. Thus, the lesser constraint by α is to some extent compensated by κ. The well-identifiable minimum CRPS for $N_e = 40$ indicates that the sampling errors are small enough that accounting for skewness and kurtosis does not just stabilize the filter, but the hybrid filter successfully utilizes N_{eff} in α and the skewness and kurtosis to reduce its error. In contrast, the larger sampling errors for $N_e = 15$ result in a wide range of parameters for which a small CRPS is obtained. This further leads to the fact that the minimum CRPS is obtained for larger values of κ compared with $N_e = 40$. The larger κ reduces the influence of skewness and kurtosis and low γ is only used for particularly high non-Gaussianity.

Increasing α beyond 0.5 leads to growing CRPS. Further, the dependence on κ diminishes, since the limitation

by α dominates. While the CRPS increases for $\gamma_{\text{sk},\alpha}$ when α approaches zero, this increase is smaller than in the L63 model. This different behavior can be related to the localization. In the case of the L63 model, the whole model state is treated at once, while for the L96 model a sequence of local analyses is computed. In this case, only single local analysis updates might have $N_{\text{eff}} = 1$, while $N_{\text{eff}} > 1$ for the others. Thus, only part of the model state is affected when no constraint regarding N_{eff} is active. This effect stabilizes the analysis and leads to a smaller increase of the CRPS. As for the L63 model, larger values of κ need to be applied for $\alpha = 0$. In this case, the minimum CRPS is obtained for $N_e = 15$ for $\kappa = 100$, while $\kappa = 60$ is optimal for $N_e = 40$. Utilizing the skewness and kurtosis of the ensemble in addition to N_{eff} allows the filter to obtain small CRPS over the full range of α. However, for very small α, the choice of κ compensates for the missing limitation of α.

The rule $\gamma_{\text{sk,lin}}$ at the bottom of the panels in Figure 6 shows small CRPS over the full range of κ with a tendency of increased CRPS for very small and large κ. Thus, consistent with Table 2, the hybridization is mainly determined

TABLE 3   Mean errors and parameter values over experiments with L96 for $N_e = 40$

Filter	γ_{lin}	$\gamma_{\alpha=0.4}$	$\gamma_{sk,lin}$	$\gamma_{sk,\alpha=0.3}$	$\gamma_{\alpha=0.1}$
κ	–	–	2	10	–
r_{loc}	14	12	12	14	5
CRPS	0.548	0.534	0.551	0.522	0.649
RMSE	1.092	1.063	1.090	1.034	1.348
$\bar{\gamma}$	0.896	0.713	0.877	0.835	0.060
$\bar{\gamma}_{min}$	0.766	0.393	0.723	0.782	0.0
$\bar{\gamma}_{max}$	0.962	0.899	0.959	0.893	0.414
N_{eff}/N_e	0.104	0.121	0.125	0.102	0.303
mas	0.581	0.589	0.580	0.589	0.598
mak	0.913	0.937	0.917	0.932	0.960
nmas	–	–	0.410	0.186	–
nmak	–	–	0.459	0.093	–

by the linear dependence on N_{eff}, with a smaller effect of κ. For $N_e = 15$, the CRPS is comparable with the lowest values obtained for $\gamma_{sk,\alpha}$, while the CRPS for $\gamma_{sk,lin}$ is larger in the case of $N_e = 40$. Here, the CRPS is comparable with that obtained for $\kappa = 0.5$.

5.2.3 | Effect of different hybrid rules on γ

Finally we assess how the different hybrid rules lead to different values of γ for the HNK filter. Table 3 shows several statistics for four experiments with small CRPS and one suboptimal case for $N_e = 40$. These are averages over space and time. The table can be compared with Table 1, which shows the statistics for the L63 model experiments. In contrast to the L63 model, the DA experiments with the L96 model apply localization and the values of γ vary for each local analysis domain. To take this into account, Table 3 also shows, next to the mean value $\bar{\gamma}$, the time mean of the minimum ($\bar{\gamma}_{min}$) and maximum value ($\bar{\gamma}_{max}$) of γ for each analysis step.

On average, the nonlinearity of the L96 experiments is smaller than for L63, as is visible from the lower values of *mas* and *mak*. However, the values vary strongly over the local analysis domains. The time average of the minimum *mas* is 0.2, while the maximum *mas* is 1.42 for the low-CRPS cases in the left columns of the table. For *mak*, the range is between 0.24 and 4.0. Consistent with the lower nonlinearity, the time-averaged values $\bar{\gamma}$ are larger here compared with the L63 model.

Among the four low-CRPS cases in the left columns, the case γ_{α} shows particular small values of both $\bar{\gamma}$ and $\bar{\gamma}_{min}$. This smaller value is consistent with the results from the L63 model. Thus, again this hybrid rule results in

similar CRPS to the other cases, by using the LNETF more strongly. The other cases use rather similar average values of γ. This is also the case for $\gamma_{sk,lin}$ and $\gamma_{sk,\alpha=0.3}$, where $\gamma_{sk,lin}$ uses the smaller value $\kappa = 2$ compared with $\kappa = 10$ in $\gamma_{sk,\alpha=0.3}$. This leads to larger values of *nmas* and *nmak* and hence a stronger potential influence of skewness and kurtosis for $\gamma_{sk,lin}$. Overall, $\gamma_{sk,lin}$ uses a larger spread of γ compared with $\gamma_{sk,\alpha=0.3}$, as is visible from $\bar{\gamma}_{min}$ and $\bar{\gamma}_{max}$.

The rightmost column of Table 3 shows the suboptimal case $\gamma_{\alpha=0.1}$, which results in a CRPS comparable with that of the LETKF. Here, the minimum CRPS is obtained for the much smaller radius $r_{loc} = 5$. This results in an overall higher likelihood and hence a higher value of N_{eff}/N_e. The lower limiting condition $\alpha = 0.1$ results in small values of γ. The mean value of $\bar{\gamma} = 0.060$ indicates that the hybrid filter analysis is strongly shifted to the LNETF. However, the mean maximum value $\bar{\gamma}_{max} = 0.414$ shows that, for some of the local analysis domains, the LNETF is still applied with a non-negligible fraction. This appears to stabilize the filter overall.

6 | DISCUSSION

6.1 | The hybrid filter algorithm

The hybrid filter schemes discussed in this study are based on the factorization of the likelihood (for HNK and HKN) or on linear interpolation (for HSync). To the author's knowledge, there is so far no analytical proof that such types of hybrid filter should lead to a better assimilation performance compared with applying the LETKF and LNETF alone. The analysis update of the KF is linear in the

observations. This implies that if one applies the analysis update without localization and with a linear observation operator, for example, twice with doubled observation errors, the final assimilation result, that is, the state estimate and covariance matrix, is unchanged. This also holds for the ETKF, which results in the same analysis ensemble for single-step or multi-step updates if Λ in Equation 8 is the identity. This is true even if the ensemble distribution is non-Gaussian, thus when a nonlinear model is used. The latter is due to the fact that only the mean and covariance of the ensemble are taken into account. Note that these properties can be easily checked numerically, but it is unclear whether all of them can be proofed analytically due to the matrix algebra involved. The properties will likely change when localization is applied, as is, for example, known for localized filters with serial observation processing (see Nerger, 2015). It is also known that KFs and different PF schemes do not converge to the correct estimate of the unbiased analysis probability distribution in the limit of infinite ensemble size (hence the methods are denoted as "biased"). Further, nonlinear filters perform poorly with small ensembles, due to the curse of dimensionality and sampling errors that are larger than in EnKFs. In this study, we are particularly interested in this case of small ensembles, since, for realistic data assimilation applications with high-dimensional models, only small ensembles—of $\mathcal{O}(10)$ to $\mathcal{O}(100)$ and hence much smaller than the state dimension—are feasible. To this end, it is not obvious that combining two schemes that have different approximations and sampling errors should lead to improved estimates.

Intuitively, we consider the two-step variants as iterative solutions. Consider that the forecast probability distribution is non-Gaussian, while the assimilation of observations with Gaussian errors results in an analysis distribution that is closer to Gaussian (see, e.g., the theoretical considerations by Morzfeld and Hodyss, 2019). We now have three cases. In the first case (HNK), the nonlinear filter is applied first and acts on the non-Gaussian distribution. The resulting intermediate analysis distribution is closer to Gaussian. It is then provided to the KF, the equations of which assume Gaussianity. Increasing the observation errors using γ should stabilize the nonlinear filter due to less variance of the weights, while the KF completes the overall analysis so that the full observational information is utilized. In the second case (HKN), the KF is applied first. Thus, the KF acts on the non-Gaussian distribution. The analysis will be suboptimal for this case, but usually very stable. The intermediate analysis ensemble is closer to Gaussian and will, in general, deviate less from the observations. Accordingly, in the following analysis with the nonlinear filter, the likelihoods of the different ensemble states will be higher and overall more

uniform. This should lead to a better analysis result. On the other hand, the intermediate distribution will be less non-Gaussian, so that the nonlinear filter can make less use of its ability to handle non-Gaussian distributions. The third case (HSync) uses the linear combination. Here, the nonlinear filter likely gives most weight to a single ensemble state. This can lead to a rather accurate state estimate, perhaps even overfitting, but leads to a low effective sample size of the analysis ensemble. In contrast, the KF can lead to too little correction, because it only uses the covariance matrix. Combining both analyses could potentially lead to a better state estimate and, with sufficient tuning of, for example, the inflation, to better error estimates. Using the hybrid weight γ in the different hybrid filter variants ensures that the combined filter analysis is consistent in the sense that, for Gaussian distributions and without sampling error, all cases should yield the same result independent of γ.

6.2 | Accounting for skewness and kurtosis

The hybrid rules $\gamma_{sk,lin}$ and $\gamma_{sk,\alpha}$ utilize the mean absolute skewness (*mas*) and kurtosis (*mak*) to determine γ. This approach is motivated by the fact that the LETKF can only be optimal for Gaussian distributions, while the LNETF does not assume Gaussianity. Thus, the LNETF should perform better in non-Gaussian cases, which are caused by nonlinear models. On the other hand, the LNETF has larger sampling errors, so that the LETKF should outperform it in the case of Gaussianity. Computing the hybrid weight based on the skewness and kurtosis allows the hybrid filter to stay close to the LETKF for nearly Gaussian cases, but to shift closer to the LNETF for increasing non-Gaussianity.

For small N_e, both *mas* and *mak* will have significant sampling errors, so that even for Gaussian distributions these quantities deviate from zero. Likewise, the normalized quantities *nmas* and *nmak* will deviate from zero. Here, averaging over N_y observations will reduce, but not eliminate, the deviation. We point out that the purpose for the hybrid filter is not to detect Gaussianity, but rather to detect whether a distribution is significantly non-Gaussian. This appears possible even for rather small ensembles. Consider the example of $N_e = 25$ and $\kappa = 25$ as used for the L63 model. A rough estimate for the standard error of skewness is $\sqrt{6/N_e}$, which for $N_e = 25$ yields 0.49. The maximum skewness as used for the normalization in *nmas* is approximated by $\sqrt{\kappa} = 5$. Thus, for a Gaussian distribution sampled by $N_e = 25$ states, we obtain *nmas* ≈ 0.1, corresponding to a sampling error of 10%. This error will be further reduced by the averaging over multiple

observations used for a local analysis update. For L63, this averaging was only done over three observations, while for the L96 model between 9 and 15 observations were assimilated locally.

Relying on only skewness and kurtosis to account for non-Gaussianity is a simplification. However, with small ensembles it is not possible to determine the full probability distribution, so that relying on these statistics is a practical approach.

6.3 | Numerical results

The experiments with the L63 and L96 models show a clear improvement of the assimilation result by using the hybrid filter compared with using either the linear LETKF or the nonlinear LNETF. The hybrid variant HNK shows the overall best performance. Utilizing the ensemble with non-Gaussian distribution first in the nonlinear filter, followed by the linear Kalman filter, can apparently make the best use of the non-Gaussianity. Even though applying the (L)ETKF first in the HKN filter would not completely eliminate non-Gaussianity, it at least leads to an intermediate ensemble that is closer to Gaussian. The effect of bringing the ensemble closer to the observation by the stable LETKF appears to have a lesser influence than the non-Gaussianity. The results of the HSync filter show that the linear combination of the analysis ensembles of LETKF and LNETF can also improve the analysis. Here, the filters appear to be truly biased, with one filter acting too strong and the other too weak, so that the linear combination of their analysis ensembles yields an improved ensemble.

For the L63 model, it was shown that, of the different approaches to compute the hybrid weight, using a fixed value of γ resulted in larger CRPS than the adaptive rules. This shows that the adaptive choices can indeed utilize the information from the ensemble statistics (N_{eff}, skewness, kurtosis) to account for the current ensemble situation at each analysis update. For the L96 model, the behavior is similar (not shown here). Utilizing information from the skewness and kurtosis improved the stability, but also the estimates of the hybrid filter variants. This happened despite the sampling errors in the skewness and kurtosis. However, for the smaller ensemble in the L96 case, the influence of the skewness and kurtosis had to be reduced compared with the larger ensemble. This reduced the influence of the sampling errors and let the hybridization weight focus on the ensemble distributions with sufficiently large non-Gaussianity, which were detectable despite the sampling errors.

Localization was applied for the assimilation in the L96 model. The localization stabilizes the hybrid filtering. Very low N_{eff} will only occur for some local analysis domains, while other domains show larger N_{eff}. This averaging effect avoids the case in which the ensemble for the full state degenerates.

The hybrid filters tested here are particularly efficient for small ensembles. For larger ensembles, as, for example, tested for the L63 model, the advantage over the full use of the LNETF or even a PF, if applicable, diminishes. This is mainly due to the reduced sampling errors. However, a large ensemble can also represent more than just the leading statistics (mean, covariances, skewness, and kurtosis). This additional information can be utilized by the LNETF or PF, but not in the hybrid rules using skewness and kurtosis. Small ensembles lead to higher sampling errors, which deteriorate the nonlinear filters in particular. However, they also limit the mathematical analysis, so that this study relied on numerical experiments. An analysis of the asymptotic behavior for large ensembles seems to be of limited value, since this does not translate into a filter performance for the small-ensemble regime. Nonetheless, the hybrid rules γ_α and $\gamma_{\text{sk},\alpha}$ can also be tuned to work well for large ensembles. In contrast, the linear relationship on N_{eff} using γ_{lin} and $\gamma_{\text{sk,lin}}$ appears to be too limiting for large ensembles.

The hybrid rules using skewness and kurtosis only used both higher-order moments in combination. The values for *nmas* and *nmak* in Tables 1 and 3 show that, in particular for small κ, both quantities can have a similar magnitude. With the current definition of the hybrid rules, it is not possible to distinguish the influence of skewness from that of kurtosis. However, separating them would require the introduction of a further parameter and hence a larger tuning effort. The lowest CRPS was obtained for $\kappa = 10$ for both the L63 experiment and the L96 model with $N_e = 40$. However, this numerical evidence is not sufficient to consider this value as a standard value. Choosing $\kappa = N_e$ and $\alpha \approx 0.3$ resulted in the different experiments in low, but not optimal CRPS. This might indicate that these choices are a suitable starting value for tuning.

The models used in this study are highly idealized. Obvious limitations of the models are that they have a low dimension, no physical balances, and uniform scales. Further, the L96 model is univariate and has spatially homogeneous dynamics. The L63 model does not allow for the case in which the ensemble size is smaller than the model dimension, which is common in all high-dimensional DA cases. However, a general conclusion from the experiments is that the non-Gaussianity caused by nonlinear model dynamics can be utilized by the hybrid filter to generate improved state estimates. For the small ensembles used here, this approach significantly improves the analysis ensemble. The hybrid filter variants allow us to apply inflation and localization in the same way

as the LETKF and LNETF. Further, the hybrid filters do not rely on any particular properties of the models. Thus, the hybrid filters can be applied to realistic high-dimensional DA cases and there is the expectation that they also yield improved estimates if the nonlinearity is sufficiently large. The magnitude of the improvement will depend on the model nonlinearity. A critical aspect that was not tested here is the multivariate character of a realistic application. Skewness and kurtosis are univariate quantities. They are computed here as mean absolute skewness and kurtosis for the locally assimilated observations. Accordingly, with observations of different quantities, it could be advantageous to assimilate each observation type separately to be able to account for its specific non-Gaussianity. A successful multivariate correction, such that assimilating observations of one variable also improves other variables, can be expected from the fact that both the LETKF and LNETF can successfully perform multivariate state updates. This feature is preserved by the hybrid filter.

An essential open question is what amount of nonlinearity is required so that the hybrid filter can yield improved estimates. The non-Gaussianity in the ensemble has to be detectable with the ensemble size feasible for the DA application. The experiments showed that this was possible in the L63 model for $N_e = 25$ and in the L96 model for $N_e = 40$. Initial experiments with an idealized configuration of an ocean model simulating a wind-driven double gyre at 0.25° resolution (the setup used by Tödter *et al.*, 2016 and Kirchgessner *et al.*, 2017 to assess the performance of the NETF) were performed. At ensemble size 120, they showed only 2–3% improvement compared with the LETKF when simulated sea-surface height observations were assimilated with the HNK filter variant and the hybrid rules utilizing the skewness and kurtosis. Only for small localization radii, for which the LETKF showed increased errors, was the effect of the hybridization larger. These results might indicate that the nonlinearity of this data assimilation problem is not large enough.

7 | CONCLUSION

This study introduced variants of a hybrid ensemble filter combining the local ensemble transform Kalman filter (LEKTF) with the localized nonlinear ensemble transform filter (LNETF). A hybrid weight γ shifts the behavior of the filter in between the LETKF and LNETF solutions. Three variants of this hybrid local Kalman–nonlinear ensemble transform filter (LKNETF) have been proposed. Two variants can be considered as iterative solutions. The HNK variant applies the nonlinear LNETF first to generate an intermediate ensemble, which is then used in the subsequent LETKF update. The HKN variant switches the order

and applies the LETKF before the LNETF. In these filter variants, γ is used to inflate the observation-error covariance matrix to distribute the observation information over both filter updates. The third hybrid filter variant, HSync, applies a linear interpolation of the analysis ensembles obtained with the LNETF and LETKF. Here, γ determines the weight for the linear interpolation. As the hybrid filter variants combine the LETKF and LNETF directly, they can easily be implemented based on existing implementations of the LETKF and LNETF. The LETKF and LNETF compute their update in the ensemble space and only require the eigenvalue decomposition of a matrix with dimension of the ensemble size squared, but no iterative solvers or particular solver library as in the particle filters used in the hybrid filters of Robert *et al.* (2018) and Chustagulprom *et al.* (2016). A particular feature of the LKNETF filter is that it combines two transform filters and does not apply resampling. Both filters can be localized in the same way, so that the hybrid LKNETF can also be applied with high-dimensional models.

The filters have been assessed in numerical experiments using the chaotic Lorenz-63 and Lorenz-96 models. In these experiments, the HNK variant resulted in the largest effect of the hybridization. The estimation errors (quantified as CRPS) were reduced by up to 38% compared with the error of the LETKF in the case of the Lorenz-63 model and up to 21.5% for the Lorenz-96 model. The HKN variant exhibited the smallest effect, while the HSync variant performed in between the HKN and HNK variants.

Different rules to specify the hybrid coefficient γ have been studied in the experiments. Here, a new approach was introduced that computes γ based on the absolute mean of the skewness and kurtosis of the observed ensemble in each local analysis update. A linear dependence on a normalized skewness and kurtosis was assumed to compute γ. To avoid the effective sample size N_{eff} becoming too low, this hybrid approach was combined with rules that either set γ so that N_{eff} remains above a prescribed threshold or use a linear dependence on N_{eff}/N_e. With sufficient tuning, the former approach resulted in the lowest estimation errors, but needed an additional parameter. Taking the skewness and kurtosis into account improved the assimilation result compared with only accounting for N_{eff} and stabilized the filter further. In particular, it allowed tuning of the hybrid filter so that it always reduced the CRPS compared with the LETKF, which was not possible when only the threshold condition on N_{eff} was used. Thus, the information from the skewness and kurtosis was utilized successfully.

Overall, the hybrid filter resulted in significant reductions of the estimation error compared with both the LNETF and LETKF, in particular for the variant

LKNETF-HNK. These error levels can otherwise only be reached by the LNETF with a significantly larger ensemble size and hence larger computing cost.

The data assimilation results with small chaotic models are promising. However, one would need to test the hybrid filters with realistic models. Which data assimilation problems have sufficient nonlinearity that the hybrid filter can improve the assimilation estimates beyond those of the LETKF is an open question. Also, the computation of the hybrid weight should be refined further. Utilizing skewness and kurtosis resulted in promising results and they have a theoretical basis in quantifying the non-Gaussianity of the ensemble distribution. However, the linear dependence on the skewness and kurtosis was chosen here as a simple direct approach. Likely, better functions can be found to describe the dependence of the hybrid weight on the non-Gaussianity.

AUTHOR CONTRIBUTIONS

Lars Nerger: conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; project administration; resources; validation; visualization; writing – original draft; writing – review and editing.

ACKNOWLEDGEMENTS

ORCID

Lars Nerger https://orcid.org/0000-0002-1908-1010

REFERENCES

Acevedo, W., de Wiljes, J. and Reich, S. (2016) Second-order accurate ensemble transform particle filters. *SIAM Journal on Scientific Computing*, 39, A1834–A1850. ArXiv:1608.08179.

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

Anderson, J.L. (2010) A non-Gaussian ensemble filter update for data assimilation. *Monthly Weather Review*, 138, 4186–4198.

Anderson, J.L. (2019) A nonlinear rank regression method for ensemble Kalman filter data assimilation. *Monthly Weather Review*, 147, 2847–2860.

Baatz, D., Kurtz, W., Franssen, H.J.H., Vereecken, H. and Kollet, S.J. (2017) Catchment tomography – an approach for spatial parameter estimation. *Advances in Water Resources*, 107, 147–159.

Bishop, C.H., Etherton, B.J. and Majumdar, S.J. (2001) Adaptive sampling with the Ensemble Transform Kalman Filter. Part I: theoretical aspects. *Monthly Weather Review*, 129, 420–436.

Bocher, M., Fournier, A. and Coltice, N. (2018) Ensemble Kalman filter for the reconstruction of the Earth's mantle circulation. *Nonlinear Processes in Geophysics*, 25, 99–123.

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

Chustagulprom, N., Reich, S. and Reinhardt, M. (2016) A hybrid ensemble transform filter for nonlinear and spatially extended dynamical systems. *SIAM/ASA Journal on Uncertainty Quantification*, 4, 552–591.

Ciavatta, S., Kay, S., Saux-Picard, S., Butenschön, M. and Allen, J.I. (2016) Decadal reanalysis of biogeochemical indicators and fluxes in the North West European shelf-sea ecosystem. *Journal of Geophysical Research: Oceans*, 121, 1824–1845.

Ciavatta, S., Torres, R., Suax-Picart, S. and Allen, J.I. (2011) Can ocean color assimilation improve biogeochemical hindcasts in shelf seas? *Journal of Geophysical Research*, 116, C12043.

Doron, M., Brasseur, P. and Brankart, J.-M. (2011) Stochastic estimation of biogeochemical parameters of a 3D ocean coupled physical–biogeochemical model: twin experiments. *Journal of Marine Systems*, 87, 194–207.

Feng, J., Wang, X. and Poterjoy, J. (2020) A comparison of two local moment-matching nonlinear filters: local particle filter (LPF) and local nonlinear ensemble transform filter (LNETF). *Monthly Weather Review*, 148, 4377–4395.

Frei, M. and Künsch, H.R. (2013) Briding the ensemble Kalman filter and particle filters. *Biometrika*, 100, 781–800.

Gaspari, G. and Cohn, S.E. (1999) Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125, 723–757.

Hersbach, H. (2000) Decomposition of the Continuous Ranked Probability Score for ensemble prediction systems. *Weather and Forecasting*, 15, 559–570.

Hodyss, D. (2012) Accounting for skewness in ensemble data assimilation. *Monthly Weather Review*, 140, 2346–2358.

Houtekamer, P.L., He, B. and Mitchell, H.L. (2014) Parallel implementation of an ensemble Kalman filter. *Monthly Weather Review*, 142, 1163–1182.

Hunt, B.R., Kostelich, E.J. and Szunyogh, I. (2007) Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter. *Physica D*, 230, 112–126.

Janjić, T., Nerger, L., Albertella, A., Schröter, J. and Skachko, S. (2011) On domain localization in ensemble based Kalman filter algorithms. *Monthly Weather Review*, 139, 2046–2060.

Kirchgessner, P., Toedter, J., Ahrens, B. and Nerger, L. (2017) The smoother extension of the nonlinear ensemble transform filter. *Tellus A*, 69, 1327766.

van Leeuwen, P.J. (2009) Particle filtering in geophysical systems. *Monthly Weather Review*, 137, 4089–4114.

van Leeuwen, P.J., Künsch, H.R., Nerger, L., Potthast, R. and Reich, S. (2019) Particle filters for high-dimensional geoscience applications: a review. *Quarterly Journal of the Royal Meteorological Society*, 145, 2335–2365.

Lei, J. and Bickel, P. (2011) A moment matching ensemble filter for nonlinear non-Gaussian data assimilation. *Monthly Weather Review*, 139, 3964–3973.

Liang, X., Yang, Q., Nerger, L., Losa, S.N., Zhao, B., Zheng, F., Zhang, L. and Wu, L. (2017) Assimilating Copernicus SST data into a pan-Arctic ice–ocean coupled model with a local SEIK filter. *Journal of Atmospheric and Oceanic Technology*, 34, 1985–1999.

Lorenz, E.N. (1963) Deterministic nonperiodic flow. *J. Atmos. Sci.*, 20, 131–141.

Lorenz, E.N. (1996). Predictability – a problem partly solved. In: *Proceedings Seminar on Predictability*, pp. 1–18. Reading: ECMWF.

Martin, M., Balmaseda, M., Bertino, L., Brasseur, P., Brassington, G., Cummings, J., Fujii, Y., Lea, D., Lellouche, J.-M., Mogensen, K., Oke, P., Smith, G., Testut, C.-E., Waagbø, G., Waters, J. and Weaver, A. (2015) Status and future of data assimilation in operational oceanography. *Journal of Operational Oceanography*, 8S1, 28–48.

Mattern, J.P., Fennel, K. and Dowd, M. (2014) Periodic time-dependent parameters improving forecasting abilities of biological ocean models. *Geophysical Research Letters*, 41, 6848–6854.

Metref, S., Cosme, E., Snyder, C. and Brasseur, P. (2014) A non-Gaussian analysis scheme using rank histograms for ensemble data assimilation. *Nonlinear Processes in Geophysics*, 21, 869–885.

Morzfeld, M. and Hodyss, D. (2019) Gaussian approximations in filters and smoother for data assimilation. *Tellus A*, 71, 1600344.

Nerger, L. (2015) On serial observation processing on localized ensemble Kalman filters. *Monthly Weather Review*, 143, 1554–1567.

Nerger, L. and Gregg, W.W. (2008) Improving assimilation of SeaW-iFS data by the application of bias correction with a local SEIK filter. *Journal of Marine Systems*, 73, 87–102.

Nerger, L. and Hiller, W. (2013) Software for ensemble-based data assimilation systems – implementation strategies and scalability. *Computers & Geosciences*, 55, 110–118.

Nerger, L., Hiller, W. and Schröter, J. (2005) A comparison of error subspace Kalman filters. *Tellus*, 57A, 715–735.

Nerger, L., Janjić, T., Schröter, J. and Hiller, W. (2012) A unification of ensemble square root Kalman filters. *Monthly Weather Review*, 140, 2335–2345.

Oke, P.R., Larnicol, G., Fujii, Y., Smith, G.C., Lea, D.J., Guinehut, S., Remy, E., Balmaseda, M.A., Rykova, T., Surcel-Colan, D., Martin, M.J., Sellar, A.A., Mulet, S. and Turpin, V. (2015) Assessing the impact of observations on ocean forecasts and reanalyses: part 1, global studies. *Journal of Operational Oceanography*, 8S1, 49–62.

Pham, D.T., Verron, J. and Roubaud, M.C. (1998) A singular evolutive extended Kalman filter for data assimilation in oceanography. *Journal of Marine Systems*, 16, 323–340.

Poterjoy, J. (2016) A localized particle filter for high-dimensional nonlinear systems. *Monthly Weather Review*, 144, 59–76.

Poterjoy, J., Wicker, L. and Buehner, M. (2019) Progress toward the application of a localized particle filter for numerical weather prediction. *Monthly Weather Review*, 147, 1107–1126.

Potthast, R., Walter, A. and Rhodin, A. (2019) A localized adaptive particle filter within and operational NWP framework. *Monthly Weather Review*, 147, 345–362.

Pradhan, H.K., Voelker, C., Losa, S.N., Bracher, A. and Nerger, L. (2019) Assimilation of global total chlorophyll OC-CCI data and its impact on individual phytoplankton fields. *Journal of Geophysical Research: Oceans*, 124, 470–490.

Reich, S. (2013) A non-parametric ensemble transform method for Bayesian inference. *SIAM Journal on Scientific Computing*, 35, A2013.

Robert, S., Leuenberger, D. and Künsch, H.R. (2018) A local ensemble transform Kalman particle filter for convective scale data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 144, 1279–1296.

Sakov, P., Counillon, F., Bertino, L., Lisaeter, K.A., Oke, P.R. and Korablev, A. (2012) TOPAZ4: an ocean–sea ice data assimilation system for the North Atlantic and Arctic. *Ocean Science*, 8, 633–656.

Simon, E. and Bertino, L. (2009) Application of the Gaussian anamorphosis to assimilation in a 3-D coupled physical-ecosystem model of the North Atlantic with the EnKF: a twin experiment. *Ocean Science*, 5, 495–510.

Snyder, C., Bengtsson, T. and Morzfeld, M. (2015) Performance bounds for particle filters using the optimal proposal. *Monthly Weather Review*, 143, 4750–4761.

Tödter, J. and Ahrens, B. (2015) A second-order exact ensemble square root filter for nonlinear data assimilation. *Monthly Weather Review*, 143, 1347–1367.

Tödter, J., Kirchgessner, P., Nerger, L. and Ahrens, B. (2016) Assessment of a nonlinear ensemble transform filter for high-dimensional data assimilation. *Monthly Weather Review*, 144, 409–427.

Vetra-Carvalho, S., van Leeuwen, P.J., Nerger, L., Barth, A., Altaf, M.U., Brasseur, P., Kirchgessner, P. and Beckers, J.-M. (2018) State-of-the-art stochastic data assimilation methods for high-dimensional non-Gaussian problems. *Tellus A*, 70, 1445364.

Zhu, M., van Leeuwen, P.J. and Amezcua, J. (2016) Implicit equal-weights particle filter. *Quarterly Journal of the Royal Meteorological Society*, 142, 1904–1919.

APPENDIX A. LIMITING CASES FOR SKEWNESS AND KURTOSIS

Equations 28 define normalized values of skewness and kurtosis. Both statistical moments have bounds when they are computed from finite samples. This can be illustrated by considering limiting distributions, considering two values $d \neq 0$ and a.

- Maximum skewness (max. skew): a maximum amplitude of the skewness is obtained if one member has the value $a + d$ while all other members have the same value a.

- Symmetric distribution with maximum kurtosis (max. kurt): $N_e - 2$ ensemble states have the value a, one member has value $a - d$, one member $a + d$.
- Symmetric distribution with minimum kurtosis (min. kurt): there are two groups of size $N_e/2$ with values $a - d$ and $a + d$, respectively.

Table A1 shows the values of the distributions and the limiting values of the skewness and kurtosis. Numerical tests show that, for $N_e = 100$, the deviation is 5–6%, except for *kurt* in the case of maximum kurtosis, where the value -2 is exact.

TABLE A1 Limiting values for skewness (*skew*) and kurtosis (*kurt*). a and b are arbitrary values with $d \neq 0$

Case	Values	*skew* limit	*kurt* limit
max. skew	$x^{(1)} = a - d, \quad x^{(i)} = a, \ i = 2, \dots, N_e$	$\sqrt{N_e}$	N_e
max. kurt	$x^{(1)} = a - d, \quad x^{(2)} = a + d, \quad x^{(i)} = a, \ i = 3, \dots, N_e$	0	-2
min. kurt	$x^{(i)} = a - d, \ i = 1, \dots N_e/2; \quad x^{(j)} = a + d, \ j = N_e/2 + 1, \dots, N_e$	0	$N_e/2$