A novel Gauss-Hermite High-Order Sampling Hybrid ensemble filter for computationally efficient data assimilation in geosciences - Part 1: application to Lorenz-96 in PythonDA v1.2~~.1~~.2

Simone Spada[1], Anna Teruzzi[1], Stefano Maset[2], Stefano Salon[1], Cosimo Solidoro[1], and Gianpiero Cossarini[1]

[1]National Institute of Oceanography and Applied Geophysics - OGS, 34010 Trieste, Italy
[2]University of Trieste, 34127 Trieste, Italy

Correspondence: Simone Spada (sspada@ogs.it)

Abstract. Data assimilation is used in a number of geophysical applications to optimally integrate ~~observations and model knowledge~~information from observations and models. Providing an estimation of both state and uncertainty, ensemble algorithms are among the most successful data assimilation approaches. Since the estimation quality depends on the ensemble, the sampling method is a crucial step in ensemble data assimilation. ~~Among other options to improve the capability of generating an effective ensemble,~~ This work introduces a sampling method featuring a higher polynomial order of approximation~~represents a novelty. Indeed~~, and an ensemble filter, the Gauss-Hermite High-Order Sampling Hybryd filter (GHOSH), which exploits the higher order of the novel sampling method. In contrast, the order of the most ~~widespread ensemble algorithms~~ frequently adopted ensemble algorithms in geosciences is usually equal to or lower than 2. ~~We propose a novel hybrid ensemble algorithm, the Gauss-Hermite High-Order Sampling Hybrid (GHOSH) filter, version 1.0.0.~~ In the directions where the uncertainty is larger, the GHOSH filter's sampling method achieves a higher order of approximation than in other ~~ensemble based~~ ensemble-based filters, without increasing the asymptotic computational complexity that is comparable to ~~the one~~ that of second-order deterministic filters. To evaluate the benefits of the higher approximation order, a set of twin experiments of Lorenz96 simulations has been carried out using the GHOSH filter and a second-order ensemble Kalman filter (SEIK; singular evolutive interpolated Kalman filter). ~~The comparison between the GHOSH and the SEIK filter has been done by varying a number of data assimilation settings.~~ The twin-experiment results show that GHOSH outperforms SEIK in most of the assimilation settings, with up to a ~~70%~~ 56% reduction of the root mean square error on assimilated and non-assimilated variables when best-tuned forgetting factors are adopted for each filter.

1 Introduction

Data assimilation (DA) methodologies provide a conceptual framework to integrate the information ~~contents~~ content embedded in observations and numerical models, and ~~plays~~ play a pivotal role in deriving accurate estimates of the state of ~~earth~~ Earth systems components and reducing the uncertainties of geophysical systems ~~(among the others, see the methods and applications reviewed in C~~

(among others, see the methods and applications reviewed in Carrassi et al., 2018; Houtekamer and Zhang, 2016; Lahoz and Schneider, 20
.

Thanks to the scalability of parallel implementations, the use of ensemble algorithms ~~(see e.g., Vetra-Carvalho et al., 2018; Houtekamer a~~
~~have~~ (see e.g., Vetra-Carvalho et al., 2018; Houtekamer and Zhang, 2016; Bannister, 2017, and references therein) has been proposed to estimate uncertainty and improve assimilation skills in Kalman filters and variational methodologies. On the other hand, some of the strong points of ~~the~~ ensemble and variational methods have been merged in hybrid filters (e.g., Hamill and Snyder, 2000). Moreover, recent developments of EnKF have been conceived to face increasing ~~non-linearities~~ nonlinearities and model complexity that are also related to the expanding availability of observations and computational resources ~~(see e.g., Vetra-Carvalh~~
(as discussed in, e.g., Vetra-Carvalho et al., 2018).

However, the definition of the strategy for ensemble generation is not a trivial task (see e.g., Moore et al., 2019). Straightforward Monte Carlo approaches are not usually a viable option in geoscience applications, because they would require ~~a too large~~ too large a number of ensemble members, and consequently, computational effort. The number of ensemble members can be reduced by adopting deterministic sampling methods ~~(as opposed to stochastic EnKF methods, see e.g. Carrassi et al., 2018).~~
~~Examples~~ (as opposed to stochastic EnKF methods, see e.g. Carrassi et al., 2018; Houtekamer and Zhang, 2016; Nerger et al., 2005; Laho
.

Common examples of deterministic EnKF ~~using second-order sampling methods~~ are SEIK (Pham, 2001) and ETKF (Bishop et al., 2001). Extending the definition of second-order exact methods in Pham (2001), ~~with the term "order" we refer to the polynomial order of approximation~~ we want to introduce here the concept of *polynomial order of approximation* of the filter or of its sampling method ~~. Namely, in the case of~~ or, more concisely, *order*: a h^{th}-order ~~approximation, the~~ ensemble has the property of providing a forecast mean with no error ~~as long as the evolution function~~ if a h^{th}-order polynomial model is used for forecasting~~(i.e. , the model) is a polynomial of order h~~. According to this definition, SEIK and ETKF are second-order methods, while Monte Carlo sampling has order zero. As proven in Appendix A, ~~this~~ h^{th}-order is achieved by sampling an ensemble that matches the first h statistical moments of the uncertainty probability density function (pdf) before evolution (also called *prior* in Bayesian statistics).

Most of the models used in geoscience applications are based on systems of differential equations that cannot be represented by a second-order polynomial and in all of these cases the second-order sampling methods ~~provides~~ provide a non-exact estimation of the forecast mean, which is affected by an error strictly related to the ~~error made in~~ approximation error that would be made if approximating the model by a second-order polynomial. Furthermore, the second-order approximation of the forecast mean is more effective the closer the ensemble members are to each other (i.e., small uncertainty), thus the higher the uncertainty the worse will be the approximation error in the mean computation. Since the state estimation is often affected by a relatively high uncertainty in geosciences data assimilation applications, this approximation error may ~~be not~~ not be negligible.

A potential strategy to reduce this error is the use of a higher order of approximation~~, but~~ . In general, this would require a larger ensemble with respect to the second-order methods and consequently larger computational costs~~, given that~~ . Indeed, a

higher order of approximation implies a larger number of ensemble members to represent the same uncertainty subspace[1]. For instance, second-order methods use an ensemble of $r+1$ members to span an r-dimensional uncertainty subspace (e.g., Pham, 2001), while ~~it can be proven~~ $2r$ ensemble members (see Appendix C, Section C4) ~~that $2r$ ensemble members~~ are needed to achieve order 3 in the same subspace (see e.g., Wan and Van Der Merwe, 2000; Ambadan and Tang, 2009; Luo and Moroz, 2009, for examples of third-order method with $2r+1$ ensemble members).

~~The approximation order can also be increased by using properly weighted ensemble members. A weighted ensemble based on~~ Weighted ensembles can achieve higher order while mitigating the ensemble size increase with respect to non-weighted ensembles. In literature, based on the multi-dimensional Gauss-Hermite quadrature rule (Mastroianni and Milovanovic, 2008, cap. 2)~~could arbitrarily increase the~~ , the use of a weighted ensemble has been applied to achieve a higher order of approximation ~~, but always~~ but still at the cost of a larger ensemble size (Ito and Xiong, 2000) ~~. Weighted~~ with respect to second-order methods. It is worth noting that weighted ensembles are also exploited in particle filter methods (van Leeuwen et al., 2019), where weights are used to evolve both the ensemble members and their probability.

In the present work, we propose a novel weighted ensemble method based on a new high-order sampling ~~, that~~ that identifies subspaces of larger uncertainties and provides ensemble mean estimates of order higher than 2 in those specific subspaces, without increasing the overall number of ensemble members, and therefore without major impacts on the computational cost.

The proposed high-order filter (as does its sampling method) exploits a Gauss-Hermite-like quadrature rule, along with a principal component analysis (PCA), and we named it Gauss-Hermite high-order sampling hybrid (GHOSH) filter. Members and related weights are chosen in such a way as to guarantee accuracy of order higher than 2 for a subset of principal modes, and no less than 2 for the remaining ones. The GHOSH filter exploits a hybrid approach that considers the uncertainty as composed ~~by~~ of a constant and a time-evolving ensemble-based part.

The reliability of the new filter is ~~proven~~ demonstrated in a large set of twin experiments based on an idealized model commonly used to test DA methodologies (Lorenz, 1996). In a companion paper (Spada et al., 2024) the new filter is tested in the much more complex and realistic marine biogeochemical application currently used in ~~Copernicus Marine CMEMS System~~ the Copernicus Marine (CMEMS) system (Salon et al., 2019), featuring assimilation of satellite observations.

Section 2 introduces the novel elements of the high-order sampling and Section 3 presents a synthetic description of the GHOSH algorithm and its localized version. The Lorenz96 numerical experiment is introduced in Section 4 and its results are presented in Section 5. The algorithm and the experimental results are discussed in Section 6. Finally, additional mathematical and algorithmic details along with examples are provided in appendices.

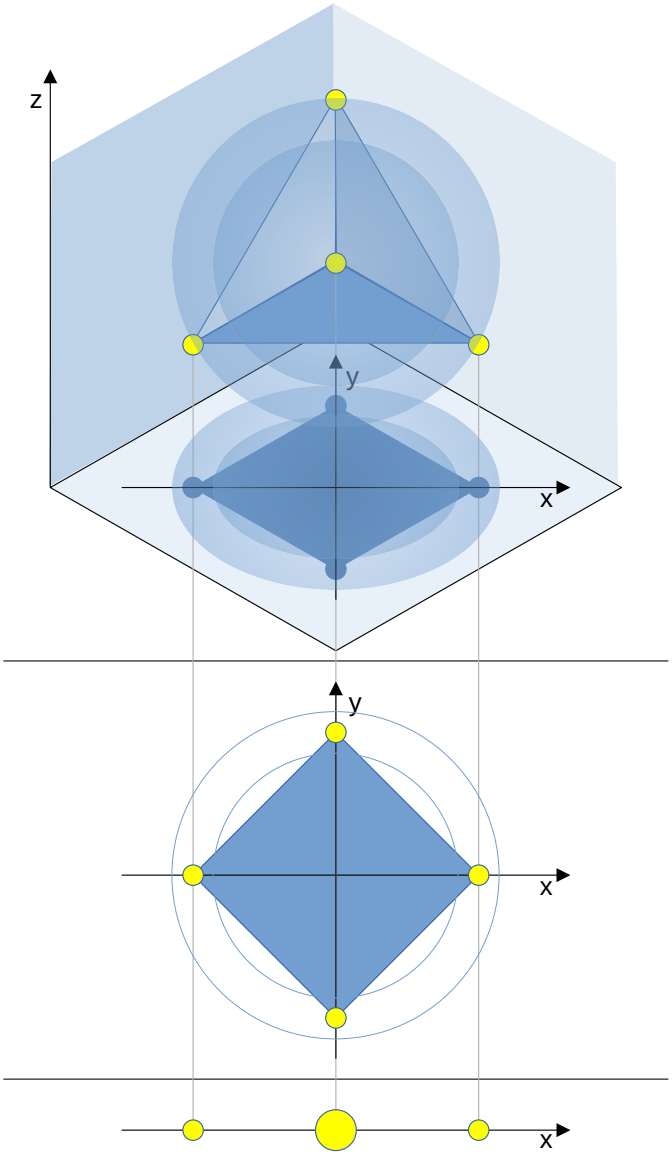

Figure 1. 4 tetrahedron-shaped ensemble members (yellow dots) in a 3-dimensional space (top) represent a second-order approximation of a standard normal distribution (~~in shadow~~ with the spherical isosurfaces shown in shadow). The same 4 members form a square in 2 dimensions (middle), producing a third-order approximation. By projecting further into 1 dimension (bottom) and by assigning proper weights, a fifth-order approximation is obtained.

2 The high-order sampling

In this work we propose a novel method to ~~sample the ensemble members in~~ generate ensembles for data assimilation applications. The preliminary idea is that the more statistical moments are shared by an uncertainty pdf and an ensemble representing it, the lower the error made by the ensemble in the forecast mean (Appendix A). In this sense, an ensemble of order h, that is characterized by the property of matching the uncertainty pdf up to the moment of order h, provides a more accurate forecast mean than an ensemble of order lower than h. Compared to deterministic second-order square-root sampling methods~~(see e.g., Carrassi et al., 2018)~~, whose ensembles match the first 2 statistical moments of the uncertainty pdf, the novel strategy achieves an ensemble with a higher order of approximation (i.e., matching more than 2 statistical moments) without increasing the ensemble size. The strategy is based on the choice, among all the second-order ensembles, of those that feature a higher order of approximation along the principal components of the uncertainty (i.e., those directions where the ensemble spread is larger and the approximation is consequently worse). ~~To better understand the GHOSH approach,~~

Fig. 1 shows an uncertainty represented by a 3-dimensional standard normal distribution ~~(~~in an xyz Cartesian coordinate system (top of Fig. 1 where spherical uncertainty isosurfaces are represented by shadowed areas). ~~Using~~ According to the second-order-exact sampling (Pham, 2001), 4 ~~ensemble members (yellow points)~~ opportunely chosen ensemble members are needed to represent this uncertainty probability density function ~~. In a~~ up to the second order. The ensemble members are not uniquely determined by the second-order exact sampling, which indeed allows for random rotations. In the xyz Cartesian coordinate system ~~, the~~ of Fig. 1, 4 ensemble members ~~are~~ that ensure the second-order exact sampling (yellow points at the vertices of a regular tetrahedron~~, and each~~) are chosen among all the possible second-order sampling ~~corresponds~~ (each corresponding to a rotation of ~~this tetrahedron. With the appropriate rotation, the~~ the tetrahedron). Thanks to their particular orientation, the 4 vertices draw (project) a square (a non-skewed shape) in the xy plane (middle panel of Fig. 1; Appendix C, Section C3.1), achieving a third-order approximation in the xy plane. Indeed, ~~it can be proven that~~ 4 square-shaped ensemble members provide a third-order approximation for a 2-dimensional normal distribution (Appendix C, Section C2.4). Similarly, ~~a further projection on a single axis returns~~ there are other possible choices of second-order vertices that project a square (third-order approximation) in the xy plane, but thanks to this particular orientation, the 4 ~~points or, as in the figure, only~~ members chosen in Fig. 1 can be further projected along the x-axis to obtain the special 3~~points, one of which represents~~ -point distribution (shown at the bottom of Fig. 1), where the central projected point weights as 2 ~~of the original points. In the~~ of the original points. As it is, this 1~~D case, the highest approximation order that can be obtained using 4 members is still limited to~~ -dimensional distribution does not achieve an order higher than 3 ~~, unless proper weights are assigned to each ensemble member. In fact, 4 weighted members (or even only~~ (already achieved in the xy plane with the square projection) along the x axis but 3 properly weighted ensemble members can reach an approximation order as high as 5 on the x axis (not shown in Fig. 1, see Appendix C at Section C1.3)~~are sufficient to reach an approximation order as high as 5 on the x axis~~. On the contrary, if

[1] In the space of the possible state vectors, the uncertainty subspace is the subspace generated by the ensemble members (i.e., the smallest subspace that contains all the ensemble members). The uncertainty subspace was originally called error subspace (see Nerger et al., 2005, 2012, for an introduction to the error subspace concept), but in the context of this work it is more natural to interpret the ensemble as a proxy of the uncertainty, avoiding unnecessary confusion with, for instance, the approximation error.

proper weights are not assigned, the highest approximation order that can be obtained using 4 members in the 1-dimensional case is limited to 3. In summary, Fig. 1 represents ~~a weighted~~ an example of an ensemble with approximation order 2 in the xyz 3-dimensional space, which is opportunely oriented to have approximation order 3 in the xy plane and, by assigning proper weights, can reach approximation order 5 along the x axis (Appendix C, Section C3.2). ~~Meaning that~~Generally, the possibility to increase the approximation order using appropriate rotations and weights opens the way to build an ensemble with a higher order of approximation along dimensions where uncertainties are higher. For instance, in the case of a non-standard normal distribution (differently from the 3-dimensional standard distribution of Fig. 1), if the orientation of the xyz coordinate system was chosen so that the z component of the distribution was less relevant than the x and y components~~(e.g., a non-standard normal distribution with smaller variance on the z coordinate)~~, then the ~~GHOSH~~ ensemble built with the third-order approximation ~~mitigates~~ would mitigate the approximation error in the x and y dimensions where the spread is higher. ~~And~~Similarly, if the coordinate system is oriented so that the variance is larger on x than y, then the achieved fifth-order approximation reduces the error more efficiently where the spread is maximum (x coordinate). As shown later in this work, the high-order sampling mimics this strategy in the ensemble data assimilation context, in order to achieve a higher order in the uncertainty principal components.

The high-order sampling generalizes the idea described above ~~can be generalized~~ considering that any probability distribution (under reasonable hypothesis) can be sampled with an arbitrary high order of approximation with a weighted ensemble that, in the special case of the Gaussian distribution, is represented by the nodes and weights of the Gauss-Hermite quadrature rule (Mastroianni and Milovanovic, 2008, cap. 2). Once the approximation order (h larger than 2) and the ensemble size ($r+1$) are fixed, the h^{th}-order weighted ensemble is computed for a relatively small s-dimensional subspace (s lower than r); then, a new $(r+1)$-sized second-order weighted ensemble is computed such that its projections along the uncertainty principal components coincide with the h^{th}-order ensemble. The resulting ensemble has a second-order approximation in the r-dimensional subspace and a h^{th}-order approximation in the s-dimensional subspace of the principal uncertainty directions. By comparison, a second-order-exact sampling (e.g., SEIK) with $r+1$ ensemble members would provide a second-order approximation for the whole r-dimensional subspace.

It is worth noting that this approach is beneficial independently of the uncertainty pdf. Even in the Gaussian case, where the uncertainty pdf is uniquely determined by its mean and covariance, the ensemble is not uniquely determined in the same way. Thus, building an ensemble matching also higher Gaussian moments helps to better represent the uncertainty pdf, which in turn translates in a smaller error on the forecast mean.

The two phases of the high-order sampling (*initialization* and *sampling*) are described in the following from an algorithmic point of view. The *initialization* includes the steps that must be executed only once, while the *sampling* presents the ensemble generation procedure. ~~Mathematical details,~~ In the sampling description, the system of equations to compute the ensemble perturbations and weights is presented. As the ensemble moments are used in the sampling phase but are calculated only once, their calculation is presented in the initialization section. Mathematical explanations can be found in Appendix A.

2.1 Initialization

~~First, 3~~

2.1.1 Initialization of the hyper-parameters

In the high-order sampling, two sets of hyper-parameters ~~r,~~ need to be defined.

The first set includes: the higher approximation order h ~~and s are chosen in the inizialization: r is~~ reached in the principal components of the uncertainty subspace; the dimension of the uncertainty subspace r, which implies an ensemble size of $r+1$ members; ~~h is the higher approximation order reached in the most relevant directions;~~ and s, that represents the number of principal components that are approximated with the higher order h and that must be smaller than r.

~~Then, the~~ The second set includes the parameterized statistical moments of the typical uncertainty, i.e. the centered moments up to order h of an uncorrelated and normalized pdf in the subspace of dimension s. The statistical moments of the normalized pdf are preliminary to the sampling algorithm, which takes mean and covariance as input and rescales all the moments accordingly (see later equation (6)).

Noted as $\mu_{j_1,\dots,j_\xi}$, a generic moment of order ξ has ξ indices, each of which runs between 1 and s. A common choice for the parameterized pdf is the standard normal distribution $\mathcal{N}(\boldsymbol{z};\mathbf{0},\mathbf{I}_s)$, which leads to moments

$$\mu_{j_1,\dots,j_\xi} = \int_{\mathbb{R}^s} z_{j_1}\cdots z_{j_\xi}\mathcal{N}(\boldsymbol{z};\mathbf{0},\mathbf{I}_s)\,d\boldsymbol{z}. \tag{1}$$

In the Gaussian case, the moments $\mu_{j_1,\dots,j_\xi}$ can easily be computed (i.e., without solving numerically the integral, see Appendix A, equations (A6) and (A7)). Instead, other non-Gaussian probability distributions can be explored considering their specific moments $\mu_{j_1,\dots,j_\xi}$.

However, since the parameterized pdf is uncorrelated and normalized, only centered moments of order higher than 2 are actual hyper-parameters (i.e., $\mu_{j_1,\dots,j_\xi}$ with $\xi > 2$), while moments up to order 2 are always already defined.

2.1.2 Initialization of the ensemble weights and of the sampling matrix

In the *initialization* phase, the ensemble weights and the sampling matrix $\mathbf{\Omega}^h$ are computed by imposing the matching between the statistical moments up to order h of the ensemble and the parameterized uncertainty pdf (equation 2). The ensemble weights and the sampling matrix are constant elements that must be prepared only once for a given set of hyper-parameters.

The real numbers $u_{m,j}$ and ~~v_j~~ v_m, with $j \in \{1,\dots,s\}$ and $m \in \{1,\dots,r+1\}$, represent the entries of $\mathbf{\Omega}^h$ and the square roots of the ensemble weights respectively. The ensemble matrix is built such that the numbers $u_{m,j}/v_m$ represent the ensemble anomalies in a s-dimensional space (see Appendix A), thus they are calculated as a solution of the nonlinear system

$$\begin{cases} \vdots \\ \displaystyle\sum_{m=1}^{r+1} u_{m,j_1}\cdots u_{m,j_\xi} {v_m}^{2-\xi} = \mu_{j_1,\dots,j_\xi} \\ \vdots \end{cases} \tag{2}$$

with one equation for each $\xi \in \{0, \ldots, h\}$ and for each $j_1, \ldots, j_\xi \in \{1, \ldots, s\}$, for a total of $\left(s^{h+1} - 1\right) / (s - 1)$ equations. The system represents the matching between the statistical moments up to order h of the ensemble and the ~~normalized uncertainty pdf~~ parameterized uncertainty pdf (this condition is necessary to achieve h^{th}-order, as proven in Appendix A). In fact, the equation of system (2) for $\xi = 0$, i.e.,

$$\sum_{m=1}^{r+1} {v_m}^2 = \mu = 1,$$

ensures that the weights will sum to 1; the s equations for $\xi = 1$, i.e.,

$$\begin{cases} \sum_{m=1}^{r+1} u_{m,1} v_m = \mu_1 = 0 \\ \vdots \\ \sum_{m=1}^{r+1} u_{m,s} v_m = \mu_s = 0, \end{cases}$$

ensure that the ~~ensemble has~~ anomalies have 0-mean; the s^2 equations for $\xi = 2$, i.e.,

$$\begin{cases} \sum_{m=1}^{r+1} u_{m,1} u_{m,1} = \mu_{1,1} = \delta_{1,1} = 1 \\ \vdots \\ \sum_{m=1}^{r+1} u_{m,1} u_{m,s} = \mu_{1,s} = \delta_{1,s} = 0 \\ \\ \vdots \\ \\ \sum_{m=1}^{r+1} u_{m,s} u_{m,1} = \mu_{s,1} = \delta_{s,1} = 0 \\ \vdots \\ \sum_{m=1}^{r+1} u_{m,s} u_{m,s} = \mu_{s,s} = \delta_{s,s} = 1, \end{cases}$$

force the ensemble covariance matrix to be equal to the identity matrix; and the same holds for higher moments up to h ($2 < \xi \leq h$). The general equation reported in system (2) represents all the moment-matching equations up to order h, where $\mu_{j_1,\ldots,j_\xi}$ are the prescribed statistical moments of a s-dimensional probability distribution that approximates the assumed uncertainty distribution shape. ~~Such probability distribution must be uncorrelated, normalized and have 0 mean, since the ensemble mean and covariance will be later adjusted by the sampling algorithm, which will also rescale the other moments accordingly (see equation (6)). A typical choice is the standard normal distribution $\mathcal{N}(\boldsymbol{z}; 0, \mathbf{I}_s)$, i.e.,~~

$$\mu_{j_1,\ldots,j_\xi} = \int_{\mathbb{R}^s} z_{j_1} \cdots z_{j_\xi} \mathcal{N}(\boldsymbol{z}; 0, \mathbf{I}_s)\, d\boldsymbol{z}.$$

~~In the Gaussian case, the moments $\mu_{j_1,\dots,j_\xi}$, which are other hyper-parameters of the GHOSH algorithm, can easily be computed (~~

The system solution is used to define $\boldsymbol{\Omega}^h$, as the $(r+1)\times s$ matrix with entries $u_{m,j}$, and the ensemble weight vector $\boldsymbol{w}\in\mathbb{R}^{(r+1)}$, as the element-wise square of $\boldsymbol{v}$, i.e., ~~without solving numerically the integral, see Appendix A, equations (A6) and (A7)). Changing the value of the moments $\mu_{j_1,\dots,j_\xi}$ allows to explore other non-Gaussian probability distributions.~~

$$\boldsymbol{w}=(v_1{}^2,\dots,v_{r+1}{}^2). \tag{3}$$

Note that, in order to actually have at least ~~a~~ one solution of the system, the number of independent equations cannot be larger than the number of variables, which implies that given r (and consequently the ensemble size) a larger approximation order h implies a smaller s (i.e., a smaller number of components approximated with order h). ~~The system solution is used to define $\boldsymbol{\Omega}_h$, as the $(r+1)\times s$ matrix with coordinates $u_{m,j}$ (i.e., $u_{m,j}$ is the entry of the matrix $\boldsymbol{\Omega}_h$ at row m and column j), and the ensemble weight vector $\boldsymbol{w}\in\mathbb{R}^{(r+1)}$, as the element-wise square of $\boldsymbol{v}$, i.e. ,~~

$\boldsymbol{w}=(v_1{}^2,\dots,v_{r+1}{}^2)$.

2.2 Sampling

Unlike from non-varying quantities (as those defined in the *initialization* phase), i-indexed quantities appearing in this section are specific of each sampling.

Given an N-dimensional state space, in the ~~sampling~~ *sampling* phase a new ensemble is generated based on the ensemble mean $\boldsymbol{x}_i\in\mathbb{R}^N$ and the covariance matrix $\mathbf{L}_i\mathbf{L}_i^T$. ~~Differently from non-varying quantities (like those defined in the *initialization* phase), i-indexed quantities are specific of each sampling. The columns of~~ Since an explicit covariance matrix might be intractable for large N, the $N\times r$ matrix $\mathbf{L}_i$ ~~adopted to obtain a~~ is adopted as an r-rank factorization of the covariance matrix ~~represent the base~~ and its columns represent the basis of the uncertainty subspace spanned by the ensemble members.
The key idea of this section is to project the high-order ensemble encoded in the sampling matrix $\boldsymbol{\Omega}^h$ onto the principal components of the uncertainty subspace, identified by an eigenvalue decomposition.

~~A~~ In order to avoid inducing a bias in the available unexploited degrees of freedom of the sampling, the matrix $\boldsymbol{\Omega}_i$ is built from $\boldsymbol{\Omega}^h$ by a procedure that includes random symmetries and rotations. These random transformations explore only degrees of freedom that do not compromise the high-order approximation on the principal components of the uncertainty subspace. To achieve this, a random $s\times s$ orthogonal matrix $\boldsymbol{\Omega}_i^{rnd}$ is drawn, then the $(r+1)\times(s+1)$ matrix $\left(\begin{array}{c|c}\boldsymbol{v} & \boldsymbol{\Omega}^h\boldsymbol{\Omega}_i^{rnd}\end{array}\right)$ is completed to an $(r+1)\times(r+1)$ orthogonal matrix by the $(r+1)\times(r-s)$ matrix $\boldsymbol{\Omega}_i^{\perp}$, i.e.,

$$\left(\begin{array}{c|c|c}\boldsymbol{v} & \boldsymbol{\Omega}^h\boldsymbol{\Omega}_i^{rnd} & \boldsymbol{\Omega}_i^{\perp}\end{array}\right)^T\left(\begin{array}{c|c|c}\boldsymbol{v} & \boldsymbol{\Omega}^h\boldsymbol{\Omega}_i^{rnd} & \boldsymbol{\Omega}_i^{\perp}\end{array}\right)=\mathbf{I}_{r+1}, \tag{4}$$

where $\boldsymbol{v}$ is the array of the square roots of the weights (as in equation (3)) and $\mathbf{I}_{r+1}$ is the identity matrix of rank $r+1$. A procedure for randomly building or completing orthogonal matrices is described in Appendix B.

Now, the $(r+1) \times r$ matrix $\mathbf{\Omega}_i$, defined as

$$\mathbf{\Omega}_i = \left(\begin{array}{c|c} \mathbf{\Omega}^h \mathbf{\Omega}_i^{rnd} & \mathbf{\Omega}_i^{\perp} \end{array} \right), \tag{5}$$

can be used to build the $N \times r$ ensemble matrix $\mathbf{X}_i$:

$$\mathbf{X}_i = \boldsymbol{x}_i \mathbb{1}_{1\times(r+1)} + \mathbf{L}_i {\mathbf{C}_i}^T {\mathbf{\Omega}_i}^T \mathbf{W}^{-\frac{1}{2}}, \tag{6}$$

where $\mathbb{1}$ is a matrix (of the subscripted size) filled with ones, $\mathbf{W} = \mathrm{diag}\,(\boldsymbol{w})$ is the $(r+1)\times(r+1)$ diagonal matrix of weights and $\mathbf{C}_i$ is a $r \times r$ orthogonal change-of-basis matrix such that

$${\mathbf{L}_i}^T {\mathbf{\Lambda}_i}^{-1} \mathbf{L}_i = \mathbf{C}_i^T \mathbf{E}_i \mathbf{C}_i. \tag{7}$$

In the last equation, the right-hand side is an eigenvalue decomposition of the left-hand side, with $\mathbf{E}_i$ being an $r \times r$ diagonal matrix of eigenvalues in descending order and $\mathbf{\Lambda}_i$ being an appropriate $N \times N$ matrix. The purpose of $\mathbf{\Lambda}_i$ is to weight the product between $\mathbf{L}_i$ and its ~~transposed~~transpose; therefore, $\mathbf{\Lambda}_i$ defines the criterion that decides the relative importance of different parts of the state vector, affecting the uncertainty principal components. In relatively simple applications, $\mathbf{\Lambda}_i$ can be the identity matrix, but in most scenarios, $\mathbf{\Lambda}_i$ needs to be properly designed (see Section 6 and Spada et al., 2024).

The orthogonal matrix $\mathbf{C}_i$ is fundamental for orienting the ensemble in such a way that the uncertainty principal components align with the directions achieving the higher approximation order (which are encoded in the $\mathbf{\Omega}_i$ matrix). After computing $\mathbf{X}_i$~~computation~~, the ensemble members can be retrieved from the ~~column~~ columns of the matrix, i.e.,

$$\mathbf{X}_i = \left(\boldsymbol{x}_i^1, \ldots, \boldsymbol{x}_i^{(r+1)} \right). \tag{8}$$

3 The GHOSH filter

Similar to other data assimilation ensemble schemes, the GHOSH filter provides an estimate of the state of a system at some ~~pre-fixed~~ discrete times t_i in terms of the state vector and the covariance matrix representing the estimation uncertainty. Hence, at time t_i, a forecast is composed ~~by~~ of the forecast state vector $\boldsymbol{x}_i^f \in \mathbb{R}^N$ (N is the dimension of the state space) and ~~by~~ the forecast uncertainty covariance matrix $\mathbf{P}_i^f$. If an observation vector $\boldsymbol{y}_i$ is available at time t_i, the information is assimilated in the analysis state vector $\boldsymbol{x}_i^a$ with uncertainty covariance matrix $\mathbf{P}_i^a$.

Being an ensemble-based filter scheme, the GHOSH filter represents these quantities by an ensemble of state vectors, e.g.,

$$\mathbf{X}_i^f = \left(\boldsymbol{x}_i^{f,1}, \ldots, \boldsymbol{x}_i^{f,(r+1)} \right) \tag{9}$$

of $r+1$ model state realizations. However, ~~differently from~~ unlike other ensemble filters that uniformly weight the ensemble members, the GHOSH filter assigns to $\mathbf{X}_i^f$ a vector of corresponding weights

$$\boldsymbol{w} = \begin{pmatrix} w_1 \\ \vdots \\ w_{r+1} \end{pmatrix}. \tag{10}$$

The state estimate is given by the weighted mean

$$\boldsymbol{x}_i^f = \sum_{j=1}^{r+1} \boldsymbol{x}_i^{f,j} w_j = \mathbf{X}_i^f \boldsymbol{w}, \tag{11}$$

while the uncertainty covariance is approximated by the ensemble covariance matrix

$$\mathbf{P}_i^f \approx \mathbf{X}_i^f \mathbf{W} \, {\mathbf{X}_i^f}^T - \boldsymbol{x}_i^f \, {\boldsymbol{x}_i^f}^T, \tag{12}$$

with $\mathbf{W} = \mathrm{diag}\,(\boldsymbol{w})$ being the diagonal matrix of weights. As in other ensemble-based filter schemes, the uncertainty covariance never needs to be explicitly computed. However, equation (12) and the like (e.g., 19, 22, 31, 33) are still shown in this form for convenience.

The GHOSH algorithm consists of 5 phases, i.e., *initialization*, *forecast*, *forecast high-order sampling*, *analysis* and *analysis high-order sampling* (Fig. 2). The *initialization* is performed once at the beginning of the filter application, the *forecast* and *forecast high-order sampling* phases are carried out at each time t_i, and *analysis* and *analysis high-order sampling* are executed only when observations are available.

The *analysis* equations are a novel weighted version of the ensemble Kalman filter equations (Evensen, 1994), while the *forecast* phase equations rely on a hybrid approach. In fact, the forecast uncertainty is obtained by combining the ensemble covariance (weighted by a forgetting factor) and a parametric covariance matrix $\mathbf{Q}_i$, which is built from the existing knowledge of the system, as done, for instance, for the background matrix in many variational schemes (Bannister, 2017, 2008).

The method includes two resampling phases, after *analysis* and after *forecast* (Fig. 2), at which the whole ensemble is rebuilt using the high-order sampling method (Section 2). The resampling produces a sample of state vectors that matches the moments of the uncertainty probability density function with better precision than in the other commonly used ensemble DA algorithms, resulting in an order of approximation greater than 2 (see Appendix A) in the principal uncertainty modes. In this way the GHOSH filter takes advantage of the high approximation order of the sampling method twice: before applying the model operator and before applying the observation operator.

3.1 The global GHOSH filter

3.1.1 Initialization

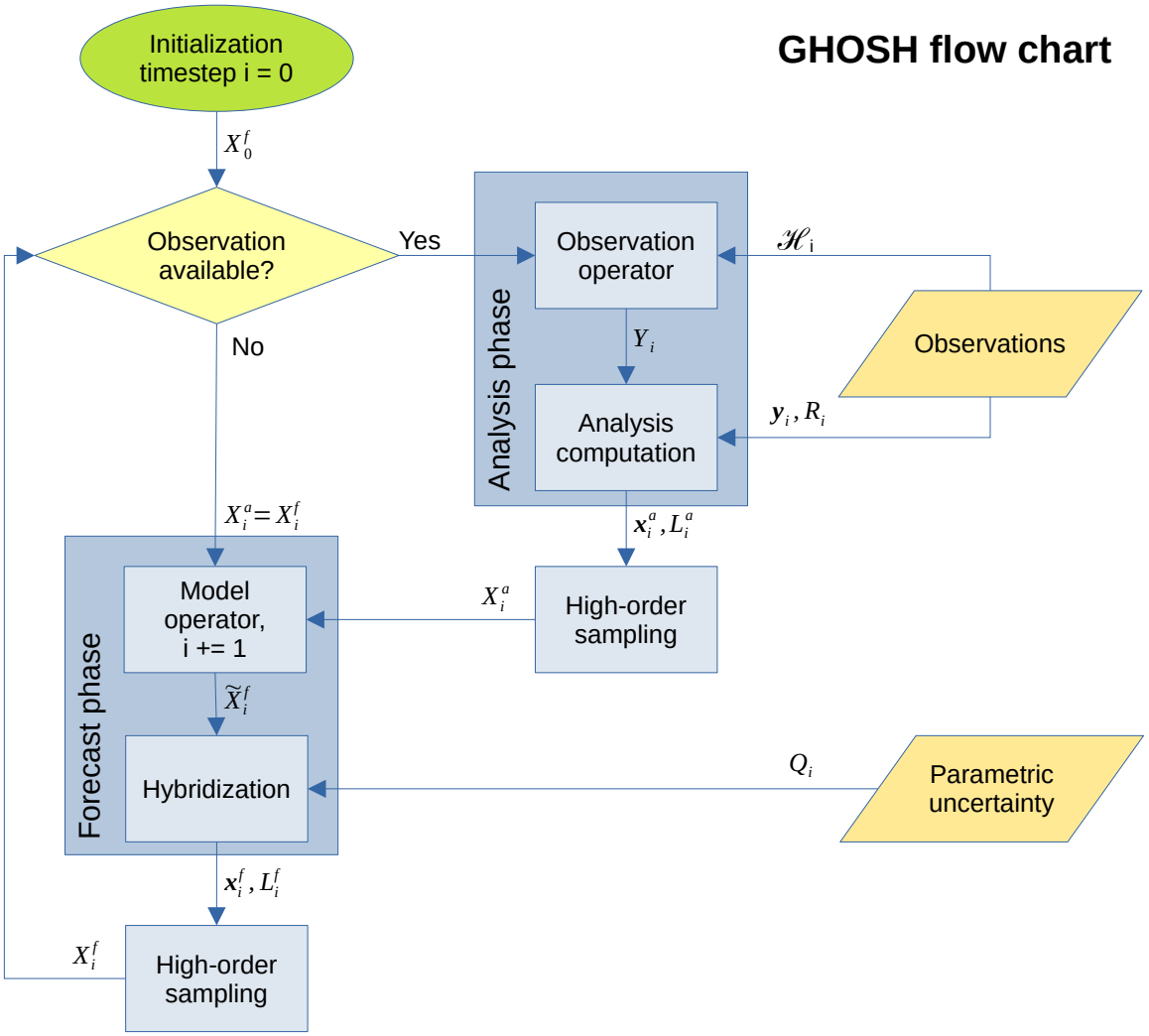

Figure 2. GHOSH filter flow chart. Subscripts represent the time steps of the numerical simulation.

~~First,~~ Based on the same hyper-parameters as the high-order sampling ~~*initialization* phase (Section 2.1.2), the three GHOSH hyper-parameters~~ (Section 2.1.1), namely r (the dimension of the uncertainty subspace), h (the higher approximation order reached in the most relevant directions) and s (the number of principal components that are approximated with order h)~~are used to compute~~, GHOSH *initialization* computes the sampling matrix $\boldsymbol{\Omega}^h$ and the weights vector $\boldsymbol{w}$ ~~.~~

~~A~~ as in Section 2.1.2. Furthermore, a starting weighted ensemble must be provided, with the ~~coordinates~~ entries of $\boldsymbol{w}$ as weights. Such an ensemble can come from any previous forecast/assimilation using the GHOSH filter, or it can be built from scratch with an ensemble generation method (e.g., a PCA on historical data followed by the GHOSH sampling method

described in Section 2.2). The ensemble members are stored in the columns of ensemble matrix $\mathbf{X}_0^f$, i.e.,

$$\mathbf{X}_0^f = \left(\boldsymbol{x}_0^{f,1}, \ldots, \boldsymbol{x}_0^{f,(r+1)}\right), \tag{13}$$

with the subscripted index 0 representing the first time step.
The mean $\mathbf{x}_0^f$ ~~,~~can be computed as

$$\mathbf{x}_0^f = \mathbf{X}_0^f \mathbf{w}. \tag{14}$$

3.1.2 Analysis phase

When observations are available at time step i (Fig. 2), $\boldsymbol{y}_i \in \mathbb{R}^n$ represents the observation array and $\mathcal{H}_i$ is the observation operator, which incorporates all operations needed to obtain the observed quantities from the state vector. This operator is used on each ensemble member $\boldsymbol{x}_i^{f,k}$ to build the $n \times (r+1)$ matrix $\mathbf{Y}_i$, i.e.,

$$\boldsymbol{y}_i^k = \mathcal{H}_i\left(\boldsymbol{x}_i^{f,k}\right) \tag{15}$$

$$\mathbf{Y}_i = \left(\boldsymbol{y}_i^1, \ldots, \boldsymbol{y}_i^{(r+1)}\right), \tag{16}$$

with $k \in \{1, \ldots, r+1\}$.
The base $\tilde{\mathbf{L}}_i^a$ of the uncertainty subspace spanned by the ensemble is computed by

$$\tilde{\mathbf{L}}_i^a = \mathbf{X}_i^f \mathbf{T}, \tag{17}$$

where $\mathbf{T}$ is an $(r+1) \times r$ full-rank matrix with zero column sums. A matrix fulfilling these requirements, that also makes the product in equation (17) computationally fast, is

$$\mathbf{T} = \begin{pmatrix} & \mathbf{I}_r & \\ \hline 0 & \cdots & 0 \end{pmatrix} - \boldsymbol{w}\mathbb{1}_{1\times r}, \tag{18}$$

with $\mathbf{I}_r$ being the identity matrix of size r and $\mathbb{1}$ ~~is~~ being a matrix (of the subscripted size) filled with ones. Equation (18) implies the use of the ensemble anomalies as basis members, and it is a common choice in SEIK implementations (e.g., Triantafyllou et al., 2003), but other $\mathbf{T}$s can be explored without affecting the main structure of the algorithm (see e.g., Nerger et al., 2012).
The analysis uncertainty covariance matrix $\mathbf{P}_i^a$ is obtained by

$$\mathbf{P}_i^a = \tilde{\mathbf{L}}_i^a \mathbf{A}_i^a \, \tilde{\mathbf{L}}_i^{a\,T}, \tag{19}$$

where

$$\mathbf{A}_i^{a\,-1} = \mathbf{T}^T W^{-1} \mathbf{T} + \mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{Z}_i, \tag{20}$$

and

$$\mathbf{Z}_i = \mathbf{Y}_i \mathbf{T}, \tag{21}$$

$W = \mathrm{diag}(\boldsymbol{w})$ being the $(r+1) \times (r+1)$ diagonal matrix of weights and $\mathbf{R}_i$ the $n \times n$ covariance matrix representing the uncertainty in the observation y_i.

An opportune change in basis is used to obtain the simpler decomposition

$$\mathbf{P}_i^a = \mathbf{L}_i^a \mathbf{L}_i^{aT}, \tag{22}$$

where

$$\mathbf{L}_i^a = \tilde{\mathbf{L}}_i^a \mathbf{S}_i^a \tag{23}$$

and $\mathbf{S}_i^a$ is the symmetric square root of $\mathbf{A}_i^a$, i.e.,

$$\mathbf{S}_i^a \mathbf{S}_i^a = \mathbf{A}_i^a. \tag{24}$$

Finally, the analysis state $\boldsymbol{x}_i^a$ is estimated by

$$\boldsymbol{x}_i^a = \boldsymbol{x}_i^f + \mathbf{L}_i^a \mathbf{S}_i^a \mathbf{Z}_i^T \mathbf{R}_i^{-1} \left(\boldsymbol{y}_i - \mathbf{Y}_i \boldsymbol{w} \right). \tag{25}$$

Equation (25) differs from the usual ensemble Kalman filter equations in the use of $\mathbf{Y}_i \boldsymbol{w}$ in the last term instead of $\mathcal{H}_i \left(\boldsymbol{x}_i^f \right)$. The two are the same if the observation operator $\mathcal{H}_i$ is linear. However, in the general case, $\mathbf{Y}_i \boldsymbol{w}$, which relies on the whole ensemble instead of the ensemble mean only, is a better estimator of the expected value of the observed quantities. Moreover, ensembles generated with GHOSH's sampling method lead to a further advantage due to the higher order of approximation.

3.1.3 Analysis high-order sampling

The $N \times (r+1)$ ensemble matrix $\mathbf{X}_i^a = \left(\boldsymbol{x}_i^{a,1}, \ldots, \boldsymbol{x}_i^{a,(r+1)} \right)$, whose columns are the ensemble members, is obtained by

$$\mathbf{X}_i^a = \boldsymbol{x}_i^a \mathbb{1}_{1 \times (r+1)} + \mathbf{L}_i^a \mathbf{C}_i^{aT} \boldsymbol{\Omega}_i^{aT} \mathbf{W}^{-\frac{1}{2}}. \tag{26}$$

The sampling procedure is described in Section 2.2; therefore, $\mathbf{C}_i^a$ and $\boldsymbol{\Omega}_i^a$ are computed as $\mathbf{C}_i$ and $\boldsymbol{\Omega}_i$ used in equation (6).

3.1.4 Forecast phase

In this phase, the time step index i is increased by 1 and the ensemble is evolved through the model operator $\mathcal{M}_i$ (which usually represents the numerical integration of a system of differential equations), i.e., for $k \in \{1, \ldots, r+1\}$,

$$\tilde{\boldsymbol{x}}_i^{f,k} = \mathcal{M}_i \left(\boldsymbol{x}_{i-1}^{a,k} \right), \tag{27}$$

$$\tilde{\mathbf{X}}_i^f = \left(\tilde{\boldsymbol{x}}_i^{f,1}, \ldots, \tilde{\boldsymbol{x}}_i^{f,(r+1)} \right). \tag{28}$$

The forecast state $\boldsymbol{x}_i^f$ is estimated as the weighted mean of the ensemble by

$$\boldsymbol{x}_i^f = \tilde{\mathbf{X}}_i^f \boldsymbol{w}. \tag{29}$$

To track the uncertainty of this estimation, the basis of the uncertainty subspace is obtained by

$$\tilde{\mathbf{L}}_i^f = \tilde{\mathbf{X}}_i^f \mathbf{T}. \tag{30}$$

The covariance matrix $\mathbf{P}_i^f$ is approximated by

$$\mathbf{P}_i^f = \tilde{\mathbf{L}}_i^f \left(\mathbf{T}^T \mathbf{W}^{-1} \mathbf{T}\right)^{-1} \tilde{\mathbf{L}}_i^{f\,T} + \mathbf{Q}_i \approx \tilde{\mathbf{L}}_i^f \mathbf{A}_i^f \tilde{\mathbf{L}}_i^{f\,T}, \tag{31}$$

where $\mathbf{Q}_i$ is the $N \times N$ covariance matrix of an additive unbiased noise parameterizing some of the sources of uncertainty in the forecast estimation not already accounted for by the ensemble, while $\mathbf{A}_i^f$ is the $r \times r$ matrix approximating the uncertainty covariance in the reduced basis expressed by $\tilde{\mathbf{L}}_i^f$. $\mathbf{A}_i^f$ can be computed in many ways, depending on the form of $\mathbf{Q}_i$ and on the chosen hybridization strategy. Here we consider the case of $\mathbf{Q}_i$ being a full rank matrix (e.g., diagonal), thus we suggest

$$\mathbf{A}_i^f = \left(\rho \mathbf{T}^T \mathbf{W}^{-1} \mathbf{T}\right)^{-1} + \left(\tilde{\mathbf{L}}_i^{f\,T} \mathbf{Q}_i^{-1} \tilde{\mathbf{L}}_i^f\right)^{-1}, \tag{32}$$

where ρ is the forgetting factor, which is added to the equation to introduce inflation ~~(see e.g., Pham et al., 1998; Carrassi et al., 2018)~~ (see e.g., Pham et al., 1998; Anderson, 2007). The last term in equation (32) is one of the novel elements of the present work. It represents the projection of the parametric uncertainty matrix $\mathbf{Q}_i$ on the uncertainty subspace defined by the ensemble. While the use of $\mathbf{Q}_i$ in equation (31) follows Pham (2001), equation (32) projects $\mathbf{Q}_i$ in a novel non-orthogonal way that is induced by the scalar product defined by $\mathbf{Q}_i^{-1}$. This approach can be interpreted as a form of hybridization that aims at focusing on the effects of the parametric uncertainty in the ensemble uncertainty subspace.

$\mathbf{Q}_i$ should be built ~~accordingly~~ according to some knowledge of the system, as done, for example, for the background covariance matrix in variational methods (Bannister, 2017, 2008). Since $\mathbf{Q}_i$ can have a very large size, it should be sparse or managed in a decomposed form.

Finally, equation (31) is conveniently rewritten as

$$\mathbf{P}_i^f \approx \mathbf{L}_i^f \mathbf{L}_i^{f\,T}, \tag{33}$$

providing a covariance decomposition consistent with the high-order sampling formulation. Here, $\mathbf{L}_i^f$ is the new basis of the uncertainty subspace and it is obtained by

$$\mathbf{L}_i^f = \tilde{\mathbf{L}}_i^f \mathbf{S}_i^f, \tag{34}$$

where $\mathbf{S}_i^f$ is the symmetric square root of $\mathbf{A}_i^f$, i.e.,

$$\mathbf{S}_i^f \mathbf{S}_i^f = \mathbf{A}_i^f, \tag{35}$$

which can be computed by an eigenvalue decomposition of $\mathbf{A}_i^f$.

3.1.5 Forecast high-order sampling

The $N \times (r+1)$ ensemble matrix $\mathbf{X}_i^f = \left(\boldsymbol{x}_i^{f,1}, \ldots, \boldsymbol{x}_i^{f,(r+1)}\right)$, whose columns are the ensemble members, is obtained by

$$\mathbf{X}_i^f = \boldsymbol{x}_i^f \mathbb{1}_{1\times(r+1)} + \mathbf{L}_i^f \mathbf{C}_i^{f^T} \mathbf{\Omega}_i^{f^T} \mathbf{W}^{-\frac{1}{2}}. \tag{36}$$

The sampling procedure is similar to ~~the one~~ that of Section 3.1.3; therefore, $\mathbf{C}_i^f$ and $\mathbf{\Omega}_i^f$ are computed as $\mathbf{C}_i$ and $\mathbf{\Omega}_i$ used in equation (6).

Compared to data assimilation schemes with resampling only after analysis, the GHOSH filter uses this sampling phase to produce a better representation of the uncertainty.

In fact, it takes into account the ~~$\mathbf{Q}_i$ effects~~ effects of $\mathbf{Q}_i$ and ρ in equations (31) and (32), which~~would be otherwise neglected. If~~, by augmenting the uncertainty after the forecast, modify the second-order moments of the uncertainty pdf. Without this resampling, the forecast ensemble would no longer be either a high-order or a second-order sampling, since the ensemble does not match the changed covariance anymore. However, if $\mathbf{Q}_i$ is supposed to be not significant (i.e., $\mathbf{Q}_i = 0$) then the covariance is only modified by the forgetting factor ρ and the *forecast high-order sampling* phase can be widely simplified~~and~~ : in this specific case, indeed, the ensemble members can be ~~obtained by~~ safely obtained by inflating the ensemble anomalies, i.e.,

$$\boldsymbol{x}_i^{f,k} = \boldsymbol{x}_i^f + \frac{1}{\sqrt{\rho}} \left(\tilde{\boldsymbol{x}}_i^{f,k} - \boldsymbol{x}_i^f\right), \tag{37}$$

for $k \in \{1, \ldots, r+1\}$.

3.1.6 Asymptotic computational complexity

~~Here,~~ In order to assess the scaling capability of the proposed algorithm, this section presents the asymptotic computational complexity of the GHOSH filter. The estimate does not include the cost of applying the model operator $\mathcal{M}_i$ and the observation operator $\mathcal{H}_i$. These operators can ~~change dramatically~~ dramatically change the total computational cost, in particular in realistic applications, where the model operator can be very computationally expensive.

Further, the *initialization* cost is not taken into account, since it is done only once. This includes the computational cost of solving system (2), which can substantially vary depending on the adopted method. For instance, in the provided python code (see the *Code availability* section), an analytical solution for system (2) is built with a constructive method in $\mathcal{O}\left(r^2\right)$ steps.

Finally, the covariance matrices $\mathbf{Q}_i^{-1}$ and $\mathbf{R}_i^{-1}$ and the scaling matrix $\mathbf{\Lambda}_i^{-1}$ are considered sparse or low rank, such that, for instance, the computational complexity of $\mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{Z}_i$ is the same as $\mathbf{Z}_i^T \mathbf{Z}_i$.

Under these assumptions, the most expensive operations in the *analysis* phase are: the computation of $\mathbf{A}_i^{a^{-1}}$ (which is $\mathcal{O}\left(nr^2\right)$) and of its inverse ($\mathcal{O}\left(r^3\right)$), and the change of basis of equation (23), which is a $\mathcal{O}\left(Nr^2\right)$ operation. Excluding the cost of the observation operator $\mathcal{H}_i$, the analysis computational complexity sums to $\mathcal{O}\left(nr^2 + r^3 + Nr^2\right)$.

During the *forecast* phase, the computation of $\mathbf{A}_i^f$ and the following change of ~~base~~ basis (equation (34)) lead to $\mathcal{O}\left(r^3 + Nr^2\right)$, without taking into account the cost of the model operator $\mathcal{M}_i$.

Finally, each sampling phase adds $\mathcal{O}\left(r^3 + Nr^2\right)$ operations given the eigenvalue decomposition and the matrix products in

equation (6).

All together, the asymptotic computational complexity of the GHOSH filter is $\mathcal{O}\left(nr^2 + r^3 + Nr^2\right)$. This is similar to SEIK and ETKF (see, e.g., Nerger et al., 2012; Tippett et al., 2003), meaning that the filters scale in the same way. However, even if not changing the asymptotic complexity, the GHOSH filter's eigenvalue decomposition and its re-sampling after *forecast* are GHOSH-specific operations that add computational time with respect to other ensemble filters that do not execute such operations. On the other hand, in the vast majority of realistic geoscience applications of ensemble data assimilation, the most demanding computational cost is represented by the model integration scaled by the ensemble size (see e.g., Vetra-Carvalho et al., 2018), making the cost of the other data assimilation operations almost negligible.

3.2 The local GHOSH filter

Localization is a widely used procedure that aims to avoid spurious correlations induced by an overly small ensemble while reducing the degrees of freedom of the system (see e.g., Janjic et al., 2011).

A localized version of the GHOSH algorithm is obtained by modifying some of the equations in Sections 3.1.2 and 3.1.4. Three operators $\mathcal{L}_p$, $\mathcal{L}_p^{\mathcal{H}}$ and $\mathcal{D}$ are used to improve readability. The localization operator $\mathcal{L}_p$ takes a global (array or) matrix with N rows and returns its localized counterpart with l rows with respect to the domain point p. The simplest and most common example of localization operator is taking just the variable values of the cells in a small radius around p. $\mathcal{L}_p^{\mathcal{H}}$ works in the same way in the observation space, reducing the number of rows from n to l'. The delocalization operator $\mathcal{D}$ takes a set of local matrices (or arrays), ideally one for each point of the domain, and returns the global counterpart. Building the global matrix by taking the values at the central points of each local matrix is a common example of delocalization operator.

3.2.1 Local forecast

Instead of equations (31-35), the localized version of the forecast step of the algorithm computes, for each point p of the domain, the local covariance matrix $\mathbf{A}_i^{p,f}$ by

$$\mathbf{A}_i^{p,f} = \left(\rho \mathbf{T}^T \mathbf{W}^{-1} \mathbf{T}\right)^{-1} + \left(\mathcal{L}_p\left(\tilde{\mathbf{L}}_i^f\right)^T \left(\mathbf{Q}_i^p\right)^{-1} \mathcal{L}_p\left(\tilde{\mathbf{L}}_i^f\right)\right)^{-1}, \tag{38}$$

where $\mathbf{Q}_i^p$ is the $l \times l$ covariance matrix at the point p of the parametric uncertainty non-dependent on the ensemble.

The global basis is built by changing the basis and delocalizing, i.e.,

$$\mathbf{L}_i^f = \mathcal{D}\left(\left\{\mathcal{L}_p\left(\tilde{\mathbf{L}}_i^f\right)\mathbf{S}_i^{p,f}\right\}_p\right), \tag{39}$$

where $\mathbf{S}_i^{p,f}$ is the symmetric square root of $\mathbf{A}_i^{p,f}$. Note that the change in basis induced by $\mathbf{S}_i^{p,f}$ is continuous (see, e.g., Nerger et al., 2012); thus, the delocalization operator does not need to include averaging or other smoothing procedures to avoid discontinuities.

3.2.2 Local analysis

The localized version of the analysis step substitutes equations (19) and (20) by calculating, for each point p of the domain, the local covariance matrix $\mathbf{A}_i^{p,a}$, i.e.,

$$\mathbf{A}_i^{p,a\,-1} = \mathbf{T}^T \mathbf{W}^{-1} \mathbf{T} + \mathcal{L}_p^{\mathcal{H}} \left(\mathbf{Z}_i\right)^T \left(\mathbf{R}_i^p\right)^{-1} \mathcal{L}_p^{\mathcal{H}} \left(\mathbf{Z}_i\right), \tag{40}$$

where $\mathbf{R}_i^p$ is the $l' \times l'$ observation uncertainty covariance matrix at the point p. This matrix can be extracted ~~by~~ from $\mathbf{R}_i^{-1}$, but it is convenient to increase the uncertainty at the points far from p (as in Nerger and Gregg, 2007). ~~A sound choice~~ One widely used option is to rescale by ~~a function that is equal to 1 in p and goes smoothly to zero out of the localization radius, e.g.,~~

$$f\left(d\right) = \begin{cases} \left(\left(\frac{d}{d_l}\right)^2 - 1\right)^2 & \text{if } d < d_l \\ 0 & \text{otherwise,} \end{cases}$$

~~with d being the distance from p and d_l the localization radius. Equation (??) is the simplest (i.e., lowest order) polynomial function with the required properties, but other options are viable (e.g., the fifth-order piecewise rational function of Gaspari and Cohn, 199~~ applying the fifth-order piecewise rational function of Gaspari and Cohn (1999).

Similarly to equation (39) for local forecast, equation (23) becomes

$$\mathbf{L}_i^a = \mathcal{D}\left(\left\{\mathcal{L}_p\left(\tilde{\mathbf{L}}_i^a\right)\mathbf{S}_i^{p,a}\right\}_p\right), \tag{41}$$

where $\mathbf{S}_i^{p,a}$ is the symmetric square root of $\mathbf{A}_i^{p,a}$.

Finally, the analysis state $\boldsymbol{x}_i^a$ in equation (25) is instead estimated by

$$\boldsymbol{x}_i^a = \boldsymbol{x}_i^f + \mathcal{D}\left(\left\{\mathcal{L}_p\left(\tilde{\mathbf{L}}_i^a\right)\mathbf{A}_i^{p,a}\mathcal{L}_p^{\mathcal{H}}\left(\mathbf{Z}_i\right)^T\left(\mathbf{R}_i^p\right)^{-1}\mathcal{L}_p^{\mathcal{H}}\left(\boldsymbol{y}_i - \mathbf{Y}_i\boldsymbol{w}\right)\right\}_p\right). \tag{42}$$

4 Experimental setup

The GHOSH filter has been tested in a very large set of twin experiments based on the Lorenz96 model (Lorenz, 1996) to evaluate the performance of GHOSH compared with a second-order filter (SEIK). The Lorenz96 model, which has a chaotic behaviour that is comparable to ~~the one~~ that of fluid dynamics equations at a ~~largely~~ much lower computational cost, is a standard choice for testing new data assimilation schemes (e.g., Brajard et al., 2020; Fertig et al., 2007; Gharamti, 2018; Grooms and Robinson, 2021; Nerger, 2022).

In each experiment, the *truth* trajectory is provided by integrating the model for a certain time interval. Then, at regular time intervals, observations of a subset of the system variables have been extracted from the *truth*, adding a random error to each of them to represent the observation uncertainty. The same set of observations is then used for two assimilation experiments, one using the SEIK filter (as a second-order approximation filter Nerger et al., 2005) , and one using the GHOSH filter with ~~fifth approximation order~~ fifth-order approximation (i.e., $h = 5$, see Section 3). The SEIK and GHOSH assimilation experiments

are ~~initialised with the the~~ initialized with the same initial condition, chosen randomly around the *truth* initial condition. After an initial ~~spin up~~ spin-up of 10 time units, the skill of each filter is evaluated by computing, during the whole simulation, the root mean square error (RMSE) between the filter forecast and the *truth* before the assimilation step. ~~The~~ Representing both how often and how far the filters deviate from the truth, RMSE is a common proxy for filter skill. In the experiments, the RMSEs on assimilated and non-assimilated variables have been evaluated for each filter. Separately assessing filter skills on non-assimilated variables provides indications on the capability of transferring information gathered with observations to the whole state of the system, exploiting correlations.

In order to produce reliable statistics, the same procedure has been repeated 90000 times varying the *truth* trajectory, the observations, the filters' initial conditions and some filters hyper-parameters such as the ensemble size and the forgetting factor.

4.1 The Lorenz96 model

The Lorenz96 model (Lorenz, 1996) is a dynamical system commonly used as a testing framework in data assimilation. The state vector $\boldsymbol{x} = (x_1, \dots, x_N) \in \mathbb{R}^N$ is evolved according to the system of differential equations given by, for $1 \leq j \leq N$,

$$\frac{dx_j}{dt} = (x_{j+1} - x_{j-2})\, x_{j-1} - x_j + F, \tag{43}$$

with periodic conditions, i.e., $x_{-1} = x_{N-1}$, $x_0 = x_N$ and $x_{N+1} = x_1$. In our implementation, the size of the state vector is $N = 62$ and the forcing term is $F = 8$, which is a common value used to generate chaotic behaviour.

The equations have been implemented in python and numerically solved using SciPy's *solve_ivp* routine with its default solver method (i.e., *RK45*).

At time $t = 0$, the state vector has been initialized by adding a small perturbation ($x_N = 0.01$) to the null solution $\boldsymbol{x} = 0$. The model has been integrated for 20 time units as spin-up, then, the following 1000 time units have been used as historical data to compute the climatological mean and covariance (performing a PCA on 10000 snapshots taken every 0.1 time units). The model has been further integrated for other 20 time units of spin-up, followed by 80 time units, divided ~~in~~ into 4 blocks of 20 time units. The model trajectory of each one of these 4 blocks has been taken as reference in a set of twin experiments and referred to as *truth*. In the same way, a longer block of 150 time units has been used as *truth* in a set of twin experiments aimed ~~to test~~ at testing long-run performances.

4.2 Observations

Observations are generated for each *truth* trajectory by extracting from the state vector the values of the ~~even indexed~~ even-indexed variables (i.e., ~~$x_2, x_4, \dots, x_{62}$~~ $x_2, x_4, \dots, x_{62}$) every Δt time units and adding to each observed variable a random number sampled from a standard normal distribution (i.e., with variance equal to 1). Tests have been ~~made~~ carried out for $\Delta t \in \{0.1, 0.15, 0.2, 0.25, 0.3\}$.

4.3 Filters setup

The same settings are used in both SEIK and GHOSH filters. The ensemble size is chosen among $7, 15, 31$ and 63 members.

The initial ~~conditions~~ condition of the ensemble mean is chosen randomly around a *truth* initial condition, adding a random Gaussian noise with 0-mean and same covariance matrix as the climatological covariance of the Lorenz96 model (Section 4.1). The initial ensemble is sampled with each filter's own sampling method (second-order-exact sampling for SEIK, high-order sampling with $h = 5$ for GHOSH), setting the prior covariance matrix accordingly to the PCA approximation of the ~~climatolgical~~ climatological covariance (i.e., r principal components are taken into account if the ensemble size is $r + 1$).

Both filters use the same forgetting factor (equation (32)), chosen among $0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95$ and 1 (the last value describes a condition without inflation). The GHOSH filter is run without hybridization ($\mathbf{Q}_i = \mathbf{0}$) to keep GHOSH and SEIK as similar as possible.

The order of the GHOSH filter was fixed at $h = 5$, with s (i.e., the number of principal components approximated with order $h = 5$) as ~~big~~ large as possible, depending on the ensemble size. Hence, the values adopted for s are $2, 3, 4$ and 5, respectively, for $7, 15, 31$ and 63 ensemble members.

Localization has not been applied, since the number of variables N is relatively small and comparable with the ensemble size.

4.4 Experimental design

The twin experiments have been performed in two phases, the first one for the 20~~-time-units-long~~ -time-unit-long *truth* trajectories and the second one for the 150~~-time-units-long trajectory.~~ -time-unit-long trajectory. The relatively short time series in the first experiment phase are motivated by the aim to focus on comparing DA convergence after the assimilation and not on long-term convergence. On the other hand, the 150-time-unit experiments have been carried out to verify the robustness of the filter on longer time scales.

In the first phase, for each of the 4 20~~-time-units-long~~ -time-unit-long *truth* trajectories, 100 twin experiments have been launched for each of the 5 observation frequencies, 4 ensemble sizes and 11 forgetting factors (for a total of 88000 twin experiments). Each of the 100 twin experiments has its own set of random observations and initial conditions.

The RMSE score of the first phase experiments has been used to identify the best forgetting factor for each filter in each configuration of observation frequency and ensemble size. Then, in the second phase, the 150~~-time-units-long~~ -time-unit-long *truth* trajectory has been employed in a set of 100 twin experiments for each of the 5 observation frequencies and 4 ensemble sizes (for a total of 2000 longer twin experiments), using for each filter its best forgetting factor in each configuration.

All the code was written in python and executed ~~in~~ on a Linux computer equipped with an Intel(R) Core(TM) i7-11800H @2.30GHz.

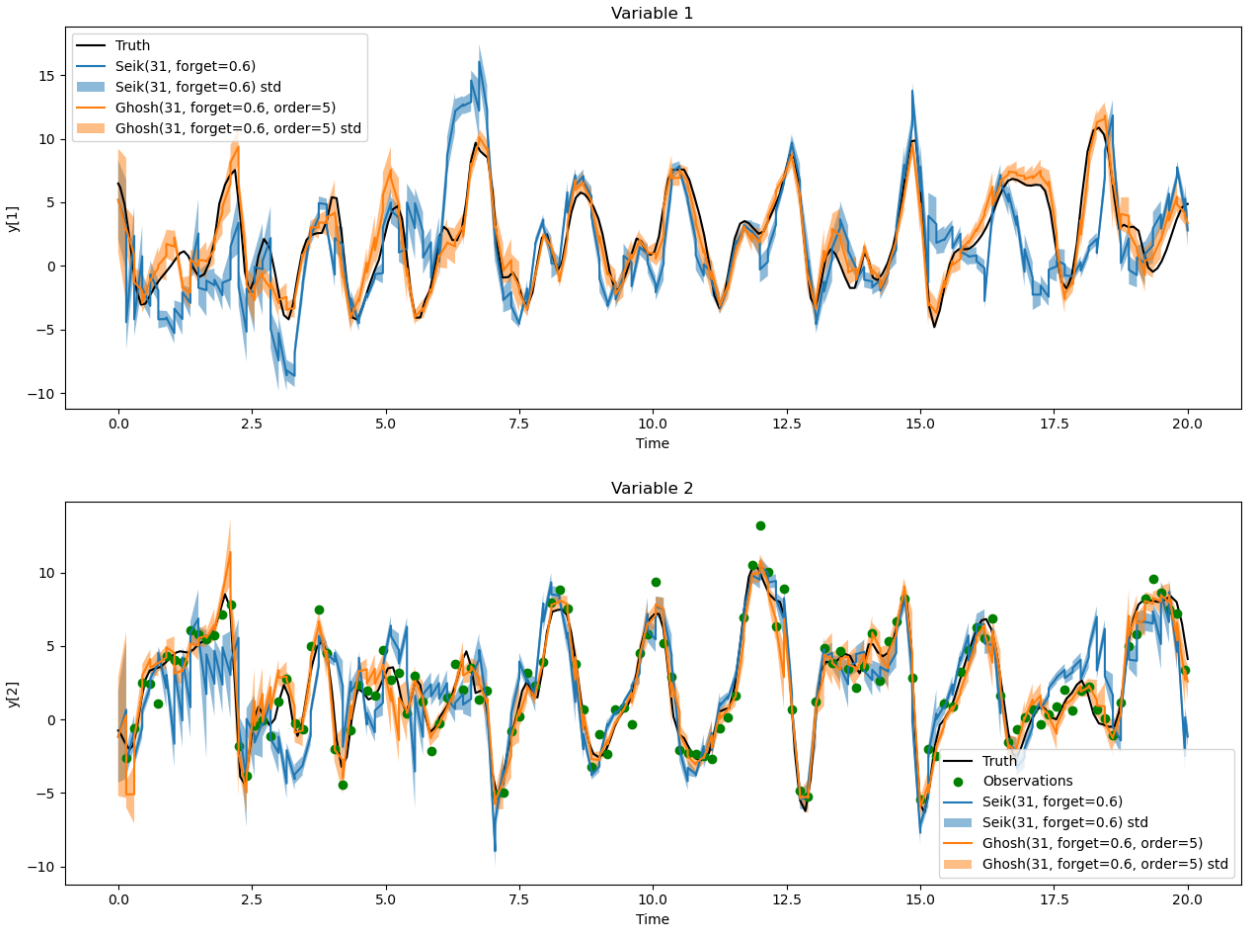

Figure 3. Results from one example twin experiment~~example~~: time between ~~observation~~ observations (green dots) is 0.15, forgetting factor is 0.6, ensemble size is 31. Shady areas around blue (SEIK) and orange (GHOSH) lines represent the ensemble standard deviation, *truth* is the black line. Top panel: time evolution of the first variable (non-assimilated); bottom panel: time evolution of the second variable (assimilated).

5 Results

5.1 20-~~time-units-long~~ -time-unit-long experiments

Figure 3 reports an example of the evolution of the first two variables (one assimilated and one non-assimilated) of the Lorenz96 model for the two filters (GHOSH and SEIK). The selected simulation out of 88000 experiments shows that GHOSH filter evolution is closer to truth evolution (black line in Fig. 3) than SEIK. As expected, the uncertainty of variable 2 (shaded areas) of both filters is lower than in variable 1 because of the assimilation of observations in even variables.

In order to be statistically reliable, the RMSE of the 88000 experiments under different conditions ~~have~~ has been computed and reported in aggregated form in Fig. 4. Each small square reports the RMSE value computed over a set of 400 experiments[2], which includes, for each of the four available truths, 100 experiments with different random observations and initial conditions. Aggregating 400 experiments together reduces the outcome variability by a factor 20 (i.e., $\sqrt{400}$) with respect to a single twin experiment, removing ~~stocasticity~~ stochasticity concerns in the analysis of the twin experiment results.

The first two rows of colour-maps refer to ~~the SEIK filter~~ assimilated and non-assimilated variables for the SEIK filter. In each colour-map, the time interval between assimilated observations increases from left to right, while the different forgetting factor values decrease from top to bottom. The first line in each colour-map, labelled "*best*", represents the best result, in terms of lowest RMSE, obtained among the set of tested forgetting factors (i.e., the skill of the filters when optimally tuned). The color scale is capped to an RMSE of 4, the climatological standard deviation of the model, that represents the threshold under which the filter performs a successful assimilation.

The GHOSH experiments are summarized in the same way in the 3rd and 4th rows of Fig. 4, while the last two rows show the ratio between GHOSH and SEIK RMSE values, with red color indicating RMSE reduction of GHOSH with respect to SEIK.

The column corresponding to the smallest ensemble size (i.e., 7 members) is not shown, since in this case SEIK and GHOSH behave very similarly, with very poor performances, due to lack of capability of describing the system complexity with an overly small ensemble size.

The yellow 15~~-members~~ -member columns also show poor performances in all configurations but, looking at the last two rows of Fig. 4, the GHOSH RMSE is significantly better than SEIK RMSE (red squares), specially for low values of forgetting factor. In fact, while the GHOSH filter keeps the RMSE around the climatological values, the SEIK filter is more prone to numerical divergence ~~landing in trajectory~~ ending up on trajectories far from the *truth* (see Fig. 5 for an example of diverging behaviour).

Considering the cases with 31 and 63 ensemble members, the GHOSH filter ~~produce~~ produces a successful assimilation (i.e., achieving an RMSE smaller than the climatological standard deviation, which is represented in yellow in Fig. 4) in almost all configurations (except two yellow squares representing experiments without inflation in the case of observation interval smaller than 0.15 time units). The SEIK filter instead shows a more limited capacity ~~of producing~~ to produce a successful assimilation, as proved by a larger number of yellow squares in Fig 4. Remarkably, the GHOSH filter with 31 ensemble members improves the state estimation error compared to the climatological standard deviation of the model for every observation interval. The same is not true for the SEIK filter, which achieves convergence (i.e., RMSE lower than the climatological standard deviation, at least in its *best* forgetting factor configuration) only for observation intervals shorter than 0.2 time units.

Looking at the last two rows of Fig. 4, the GHOSH filter outperforms the SEIK filter in most conditions, up to a 3~~times~~ -times RMSE reduction. Furthermore~~GHOSH proves to be~~ , GHOSH is always at least as good as SEIK. In the range of explored forgetting factors, the largest improvements (dark red, RMSE ratio 0.30) mainly occur when: i) inflation is low (forgetting factor ~~near~~ close to 1)~~; ii)~~ , ii) a large amount of information is injected into the assimilation scheme (i.e., high frequency

[2] Computing the RMSE over the whole set of 400 experiments is equivalent to computing the aggregated RMSE with the formula $\sqrt{\frac{1}{400}\sum_{i=1}^{400} \mathrm{MSE}_i}$, where MSE_i is the mean square error of the i^{th} experiment.

observations, left side of Fig. 4)~~;~~ , and iii) the filter can take into account a high dimensional uncertainty subspace (i.e., the ensemble size is large).

On the other hand, the comparison of each filter tuned with its best forgetting factor (the "*best*" top line in each colour-map) clearly shows that the GHOSH filter has considerably better ~~performances~~ performance than the SEIK filter (up to ~~a~~ an RMSE ratio of 0.44) with a moderate number of ensemble members (31). In case of maximum ensemble size the improvement is still significant but less intense (up to 0.90 RMSE ratio).

The first four lines of Fig. 4 ~~makes~~ make evident that the GHOSH filter, compared to SEIK, is less dependent on inflation, reaching its best performance with a forgetting factor closer to 1.

The skill of both ~~the~~ filters is closely related with the observations frequency: the shorter the time between observations, the higher the accuracy.

Finally, as expected, for both ~~the~~ filters the skill is better on assimilated variables than non-assimilated ones but, quite interestingly, the RMSE ratio shows that the GHOSH advantage over SEIK is slightly larger in all the non-assimilated variables compared to the assimilated ones with the same settings (up to 0.08 RMSE ratio improvement in the *best* forgetting factor configuration, from 0.69 for assimilated to 0.61 for non-assimilated variables).

5.2 150~~-time-units-long~~ -time-unit-long experiments

Similarly to the case of the 20~~-time-units-long~~ -time-unit-long twin experiments, the results of the 150~~-time-units-long~~ -time-unit-long tests are summarized in Fig. 6. Each small square reports the RMSE value computed over a set of 100 experiments with different random observations and initial conditions. In the top panel, the 5 colour-maps correspond to different time intervals between observations, increasing from left to right. Each colour-map shows, for assimilated and non-assimilated variables and for different ensemble sizes, the RMSE of the SEIK filter. The RMSE of the GHOSH filter is represented in the same way in the middle panel, while the ratio of the GHOSH RMSE over SEIK RMSE is shown in the bottom panel.

The longer experiments ~~confirms~~ confirm the result obtained by the shorter ones, with GHOSH performing better ~~the~~ than SEIK in every configuration. In particular, the improvement is maximum (0.51 RMSE ratio) in the case of 31 ensemble members. Both filters show poor performance at 15 or ~~less~~ fewer ensemble members, and both of them achieve a good convergence at the maximum ensemble size. In the case of 31 ensemble members, the behaviour of GHOSH and SEIK differs when the observation interval is 0.25 time units or longer, with GHOSH achieving a successful assimilation (blue-green color) while SEIK performs worse than the climatological error (yellow color).

The long-run experiments also confirm that the improvement is slightly larger for non-assimilated variables than for assimilated ones, scoring a better RMSE ratio (maximum ratio improvement of 0.05) in almost all configurations.

5.3 Computational cost

From the computational point of view, the whole experiment set ~~needed~~ required around 9 computational hours. SEIK and GHOSH schemes used 12% and 16% of the total time respectively, while the rest was ~~committed to model integration. The GHOSH filter executes more operations than SEIK (e.g., an eigenvalue decomposition) resulting in more computational time~~

~~even if the asymptotic computational complexity of the two methods is the same (Section 3.1.6). However, even in the case of a relatively simple model like the Lorenz96, the time to solution is dominated by the model integrationand the difference between GHOSH and SEIK only accounts for 4% of the total time.~~ devoted to model integration. Unexpectedly, the model integration time (averaged every 100 twin experiments) when applying the SEIK filter ~~sometimes is longer than~~ is sometimes longer than in the GHOSH filter case, up to twice ~~the time~~as long, depending on some settings and random factors. This difference in computational time is more evident and occurs more often when the forgetting factor is lower (i.e., when the inflation is more pronounced). The reason can be understood by looking at Fig. 5: in this experiment with forgetting factor equal to 0.6, the SEIK filter presents an unstable and diverging behaviour. When it happens, the SciPy's *solve_ivp* integrating routine reduces the time step size ~~for granting~~ to ensure the required accuracy, which comes with longer integration time.

Due to the increase in integration time when the filter is not converging, the percentage of computational time used by the filters with respect to the total time greatly varies from case to case, depending on the particular experiment parameters and on filter convergence. On average, in our experiments on the Lorenz96 model, the time to solution is dominated by the model integration and the difference between GHOSH and SEIK only accounts for 4% of the total time. This difference between the two filters is explained by the fact that the GHOSH filter executes more operations than SEIK (mainly an eigenvalue decomposition) resulting in more computational time even if the asymptotic computational complexity of the two methods is the same (Section 3.1.6).

6 Discussion

The name "Gauss-Hermite high-order sampling hybrid filter" was chosen to indicate the two main features of this novel filter. First, it is a hybrid filter, in the sense that the uncertainty covariance matrix is obtained by averaging the ensemble covariance and a parametric covariance matrix non-depending ~~by~~ on the ensemble, as described in, e.g., Carrassi et al. (2018). Second, it exploits a novel sampling method based on a Gauss-Hermite-like quadrature rule to reach an ~~arbitrary~~ arbitrarily high (depending on the ensemble size) polynomial order of approximation.

The latter feature introduces, to the best of our knowledge, a completely new class of DA algorithms with an order of approximation higher than 2. The advantage of moving to a higher order of approximation has previously been documented in the literature, in particular Nerger et al. (2005) compares order 1 algorithms (e.g., SEEK Pham et al., 1998) and order 2 algorithms (e.g., SEIK Pham, 2001). ~~Results~~ The results of Section 5 confirm the expected benefits of the high-order sampling adopted in the GHOSH filter. In particular, in the Lorenz96 idealized and controlled conditions, the novel method outperforms (up to ~~70%~~ 56% lower RMSE) a typical second-order method like SEIK and features higher stability and better ~~performances~~performance.

Thanks to the large number of settings tested in the Lorenz96 twin experiment, it is possible to suggest the conditions where the GHOSH filter considerably increases the assimilation skill and reliability. Concerning the ensemble size, the larger improvements occur when the dimension of the uncertainty subspace spanned by the ensemble is ~~big~~ large enough to be representative. If the number of ensemble members is too small, the assimilation simply fails independently ~~from~~ of the tested assimilation algorithm, but even in that case, the GHOSH filter error is less detrimental than the SEIK error. Moreover, it has been shown

that the GHOSH increased accuracy reduces instabilities and numerical divergence in our Lorenz96 implementation (Fig. 5). This aspect is consistent with the fact that a higher order of approximation helps to manage ~~non-linearities~~ nonlinearities (Nerger et al., 2005). GHOSH filter showed a lower need for inflation with respect to SEIK, removing one of the instability causes observed in the twin experiment (Section 5). Indeed, in the twin experiment, a larger number of instability cases were observed at lower inflation values both for SEIK and GHOSH (one example with diverging SEIK is shown in Fig. 5). The lower need ~~of~~ for inflation can be also seen as a GHOSH's improved capacity to take into account ~~non-linearity~~nonlinearity. Indeed, ~~non-linearity~~ nonlinearity typically introduces errors that are not considered by ensemble filters, which ~~needs~~ need inflation to compensate for this uncertainty covariance underestimation (see e.g., Raanes et al., 2019; Bocquet et al., 2015; Rainwater and Hunt, 2013). Finally, the GHOSH filter could remarkably perform a successful assimilation even when the SEIK filter is failing due to observation sparsity.

Even if a large number of settings have been tested in the twin experiments, other Lorenz96 literature applications (see e.g., Brajard et al., 2020; Fertig et al., 2007; Gharamti, 2018; Grooms and Robinson, 2021; Grooms, 2022; Kirchgessner et al., 2014; Lei and Bickel, 2011; Nerger, 2022; Scheffler et al., 2022) tested ensemble DA methods applying different conditions and strategies. In particular, longer time series, different observation operators or different state vector dimensions have been considered. The shorter time series in the first experiment phase in the present work (Section 4.4) are motivated by the aim to focus on comparing DA convergence in the phase after the assimilation and not on long-term convergence. ~~Thank~~ Thanks to this approach, our results provide information for possible ~~GHOSH applications in realistic applications~~ realistic applications of GHOSH in geoscientific operational simulations. In this framework, the shown initial fast convergence of the GHOSH filter is strongly beneficial to improve the simulation accuracy on temporal scales of interest.

In addition to a higher polynomial order, the GHOSH filter features a resampling taking place twice per step. Ensemble Kalman-like filters usually adopt an interpolation strategy by applying the observation operator directly to the $\tilde{\mathbf{X}}_i^f$ ensemble of the states after the model evolution (see equation (28)). However, this strategy might lead to inaccurate estimations because the covariance matrix $\mathbf{Q}_i$ of equation (31) is not taken into account in the interpolation. For this reason, if ~~Q_i~~ $\mathbf{Q}_i$ is not zero, GHOSH applies the observation operator to a new ensemble, resampled between the forecast and analysis ~~phase~~ phases (Fig. 2). In this way, $\mathbf{Q}_i$ is taken into account and, especially in the case of a nonlinear observation operator, the high order of the sampling method is exploited to improve the accuracy of the projection into the observation space (as in Part 2 Spada et al., 2024).

The statistical moments μ in equation (2) are key hyper-parameters that describe the uncertainty pdf moments for orders higher than 2. In our experiments we used the moments of a Gaussian distribution, but it is worth noting that the GHOSH algorithm does not enforce this choice and other pdfs could be used. However, Kalman filter ~~'s~~ analysis equations somehow prescribe a Gaussian approximation, thus the GHOSH filter keeps a link to Gaussianity, even if the GHOSH sampling does not. Interestingly, other filters, like Hodyss (2011), try to overcome this limitation and could represent good candidates to study the effects of a non-Gaussian GHOSH sampling.

The scaling matrix $\mathbf{\Lambda}_i$ used in the sampling procedure plays an important role in the filter, since $\mathbf{\Lambda}_i$ defines how variables are compared to each ~~others~~ other in the computation of the most relevant components of the state uncertainty (Equation (7)).

In the Lorenz96 case, it is straightforward to set $\mathbf{\Lambda}_i$ equal to the identity matrix, since each variable is equivalent to any other. In more complex applications, $\mathbf{\Lambda}_i$ should be carefully designed to focus on the variables relevant for the processes of interest (see also Spada et al., 2024, for a non-identity $\mathbf{\Lambda}_i$ and further discussion).

In the present implementation, the GHOSH filter derives the uncertainty subspace basis (equation 17) using a $\mathbf{T}$ matrix (equation 18) inherited ~~by~~ from the SEIK filter. This is an advantage from the computational point of view, but ensemble members after the resampling may be very different from the original ensemble. For this reason, in some applications issues can arise, for instance, in the case of the physical balance constraints. As a solution, Nerger et al. (2012) propose a filter (the error subspace transform Kalman filter, ESTKF) that~~minimizes the differences after the resampling procedure~~, during the analysis step, minimizes the unnecessary changes to the forecast ensemble, and such approach can also be applied to the GHOSH filter algorithm by choosing an ~~opportunely~~ appropriately designed $\mathbf{T}$ matrix. ~~On the other hand, a GHOSH filter built with the ESTKF approach would present an additional complexity layer, since~~ However, differently from the ESTKF case, the $\mathbf{T}$ matrix should change at every forecast ~~step. Given~~ to take into account GHOSH projections onto the principal components of the uncertainty subspace, introducing an additional complexity layer to the newly presented GHOSH filter. Thus, given the additional complexity of the ESTKF approach and considering that: i) its effects are less important if the state vector does not strongly need to preserve specific ~~non-linear~~ nonlinear constrains; ii) Nerger et al. (2012) showed that, in Lorenz96, SEIK with random rotation (adopted also in GHOSH) ~~results to be~~ is as good as ESTKF; in the present work we limited to ~~adopt~~ adopting a GHOSH filter that relies on a SEIK-like ~~T matrix~~approach.

Compared to other ensemble filters, the key feature of the GHOSH filter is the higher polynomial order of its sampling method. The advantage of a higher order is not limited to a better estimation of the mean state (Appendix A) but it extends to the covariance matrix of the uncertainty probability distribution. In fact, the polynomial order in the estimation of the covariance matrix is half the order of the mean (not shown). Since the covariance matrix is involved in the state estimation, the more accurate GHOSH results shown in Section 5 can be partly related to an improvement in the approximation error affecting the uncertainty covariance matrix. Moreover, it is worth noting that a higher order of the sampling positively impacts the approximation of the nonlinearity of the model independently of its Gaussianity. The examples in Appendix C show how moments of order higher than two can affect a sampling of a Gaussian pdf. In fact, even if its first two moments completely characterize the distribution, the sampling needs to also match the higher moments to produce an ensemble capable of achieving an order of approximation higher than two.

Similarly to other ensemble filters, a weighted ensemble is used in the GHOSH filter. However, the GHOSH filter differs substantially from the particle filter (e.g., van Leeuwen et al., 2019) and from other types of weighted ensemble filters, e.g., Hoteit et al. (2008) and Stordal et al. (2011). In the listed filters, weights (representing likelihood) change over time, while particles remain the same (and strategies are applied to resample the particles when weights "collapse", e.g., van Leeuwen et al., 2019). In the case of GHOSH, the ensemble is used to estimate specific moments and then resampled to keep ~~constant the weights~~ the weights constant in time, because those weights have specific desired properties that help reduce the error of the mean estimation. In this, GHOSH has similarities with the nonlinear ensemble transform filter (Tödter and Ahrens, 2015),

which does a resampling to keep the weights constant but, differently from GHOSH, with uniform weights and a second-order sampling procedure.

Finally, the asymptotic computational complexity of the GHOSH filter (Section 3.1.6) is comparable to second-order deterministic filters (e.g., SEIK, ETKF). ~~The computational cost, as in most ensemble methods, mainly~~ In our experiments (Section 5.3), GHOSH showed an increased computational time with respect to SEIK (on average, 4% of the total computational time), mainly due to the GHOSH filter's eigenvalue decomposition. However, it is worth noting that the cost related to the eigenvalue decomposition only depends on the ~~ensemble size, which multiplies the most demanding model integration cost (e.g., Nerger et al., 2005). Hence, the computational cost of the GHOSH filter is no more a concern than in other second-order data assimilation schemes,~~ number of ensemble members, affecting the ~~total computational time only by a 4% even in the case of the simple~~ $\mathcal{O}(r^3)$ part of the full GHOSH complexity $\mathcal{O}(r^3 + (N+n)r^2)$. As such, the eigenvalue decomposition is an important term if $N+n$ (i.e., the sum of the dimensions of model and observations) is small (in our experiments on the Lorenz96 ~~toy model implementation (Section 5.3). The~~ model it is smaller than 100), but becomes orders of magnitude less relevant in a realistic geoscience application, where degrees of freedom can be of the order of the millions and above. Moreover, while the model integration time and the filter computational time are still comparable in our experiments (12% and 16% for SEIK and GHOSH respectively), in a realistic application this will not be the case. Indeed, the filters computational complexity shows that filter computational time grows linearly with model dimension and the same holds for the Lorenz96 model that, however, is orders of magnitude less complex and computationally expensive than a realistic geoscience model. This is consistent with results obtained in a companion paper (Spada et al., 2024), where the GHOSH filter computational time in a realistic application accounts for less than 1% of the total computational cost.

In summary, the benefits of the higher order offered by the GHOSH filter rely on a mathematical advantage with limited impacts on computational effort. In fact, ~~the~~ GHOSH allows the exploitation of the degrees of freedom in the randomness of the sampling process, choosing ensembles that have a better consistency with the uncertainty distribution.

7 Conclusions

~~This work presented a novel ensemble~~ This work introduces two novel ensemble strategies: a sampling method (the high-order sampling) and an ensemble hybrid DA filter (GHOSH). ~~The~~ Exploiting the high-order sampling, the GHOSH filter features a higher order of approximation compared to other ensemble filters, which resulted in better data assimilation ~~performances~~performance. ~~The~~ Based on the results of the present work, the order of an ensemble method ~~is~~ can be one of the ~~most relevant proxies of a filter skill. In fact, the order 2 schemes (e.g., SEIK, ETKF, ESTKF) are preferred nowadays to older order 1 methods (e.g., SEEK) . This study represents a natural further step in this direction, opening to a new class of data assimilation schemes with~~ proxies for filter skill, as shown by comparing SEIK (order ~~greater~~ 2) with GHOSH (order higher than 2).

While increasing the order usually involves increasing the number of ensemble members (and consequently the computational cost), the GHOSH approach keeps the same computational complexity and number of ensemble members as SEIK or ETKF and exploits the advantage of a higher order only where the approximation errors are larger.

The feasibility of the GHOSH filter in a realistic geophysical application has also been assessed in a companion paper (Spada et al., 2024) by testing a 3D parallel implementation of a Mediterranean physical-biogeochemical model.

The ~~outstanding~~ improvements of the GHOSH filter with respect to the SEIK filter are demonstrated in an extensive set of twin experiments based on the Lorenz96 model. The GHOSH filter showed an improved capacity to take into account ~~non-linearities by a lesser need of~~ nonlinearities through a lower need for inflation with respect to SEIK, along with the ~~remarkable~~ capability of performing a successful assimilation even when observation sparsity was ~~producing a SEIK failure~~causing SEIK to fail. In every configuration, the GHOSH filter proved to be as good as SEIK or better~~, reaching up to a 70% RMSE reduction with respect to SEIK when tuned with the same forgetting factor.~~. Considering each filter tuned with its optimal forgetting factor, the GHOSH filter significantly improved the assimilation skill with respect to SEIK, achieving its best RMSE reduction (~~54%~~56%) in the setting with moderate ensemble size (31 members).

Code availability. A GHOSH python implementation is available from the GitHub page: https://github.com/Sword-Code/PythonDA under the licence GNU GPLv3. The exact version used to produce the twin experiment results and the related plots (Section 5) is archived on Zenodo (https://doi.org/10.5281/zenodo.18931130).

Appendix A: High-order sampling

Sampling a limited number of ensemble members that effectively represent the uncertainty of the state estimation is a crucial part of any ensemble algorithm. The GHOSH filter uses multi-dimensional Gauss-Hermite-like quadrature rules to choose the ensemble members, achieving a high polynomial order of convergence.

The core idea behind the GHOSH sampling can be proved by taking two independent random variables with the same moments up to a certain order h. Thus, they have the same mean after applying any polynomial operator of order h. In fact, given a probability density function (pdf) $p : \mathbb{R}^s \to \mathbb{R}$, if $\mathcal{F} : \mathbb{R}^s \longrightarrow \mathbb{R}$ is a polynomial operator such that

$$\mathcal{F}(\boldsymbol{x}) = \sum_{\xi=0}^{h} \sum_{j_1,\dots,j_\xi=1}^{s} f_\xi^{j_1,\dots,j_\xi} x_{j_1} \cdots x_{j_\xi}, \tag{A1}$$

where s is the space dimension, $f_\xi^{j_1,\dots,j_\xi}$ are the polynomial coefficients and $x_1, \dots, x_s$ are the ~~coordinates~~ entries of $\boldsymbol{x}$, then the mean $\bar{\mathcal{F}}$,

$$\bar{\mathcal{F}} = \int_{\mathbb{R}^s} \mathcal{F}(\boldsymbol{x})\, p(\boldsymbol{x})\, dx = \sum_{\xi=0}^{h} \sum_{j_1=1}^{s} \dots \sum_{j_\xi=1}^{s} f_\xi^{j_1,\dots,j_\xi} \int_{\mathbb{R}^s} x_{j_1} \cdots x_{j_\xi} p(\boldsymbol{x})\, d\boldsymbol{x}, \tag{A2}$$

depends only on the first h moments of the pdf p, which are represented by the last integral in equation (A2).

This proves that the mean of an operator can be computed exactly up to a certain order h by substituting the sampled probability distribution with any discrete finite probability distribution (such as a weighted ensemble) as long as their moments match (up to order h).

In order to build an ensemble with this property, the moment-matching equations are summarized by the nonlinear system

$$\begin{cases} \vdots \\ \sum_{m=1}^{r+1} x_{j_1,m} \cdots x_{j_\xi,m} w_m = \mu_{j_1,\ldots,j_\xi}, \\ \vdots \end{cases} \tag{A3}$$

with one equation for each $\xi \in \{0,\ldots,h\}$ and for each $j_1,\ldots,j_\xi \in \{1,\ldots,s\}$.

In system (A3), $x_{j,m}$ is the j^{th} ~~coordinate~~ entry of the m^{th} ensemble member, $r+1$ is the ensemble size, $w_m \geq 0$ are the weights and $\mu_{j_1,\ldots,j_\xi}$ are the statistical moments of p, i.e.,

$$\mu_{j_1,\ldots,j_\xi} = \int_{\mathbb{R}^s} x_{j_1} \cdots x_{j_\xi} p(\boldsymbol{x})\, d\boldsymbol{x}. \tag{A4}$$

The system is solved considering $x_{j,m}$ and w_m as the unknowns, while p can be taken uncorrelated, normalized and with ~~0~~ $\mathbf{0}$ mean without loss of generality. In fact, the ensemble matrix $\tilde{\mathbf{X}} = (x_{j,m})$ (which contains in its columns the ensemble members with ~~0~~ $\mathbf{0}$ mean and covariance matrix equal to the identity matrix $\mathbf{I}_s$) can be transformed ~~in~~ into the ensemble matrix ~~X~~ $\mathbf{X}$ with arbitrary mean $\bar{\boldsymbol{x}}$ and covariance matrix LL^T through the projection

$$\mathbf{X} = \bar{\boldsymbol{x}} \mathbb{1}_{1\times(r+1)} + \mathbf{L}\tilde{\mathbf{X}}, \tag{A5}$$

where $\mathbb{1}_{1\times(r+1)}$ is a matrix (of the subscripted size) filled with ones.

In the special case of the standard normal distribution, i.e., ~~$p(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x},0,\mathbf{I}_s)$~~ $p(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x};\mathbf{0},\mathbf{I}_s)$, the moments $\mu_{j_1,\ldots,j_\xi}$ are given by

$$\int_{\mathbb{R}^s} {x_1}^{k_1} \cdots {x_s}^{k_s} \mathcal{N}\left(\boldsymbol{x};\underline{0}\mathbf{0},\mathbf{I}_s\right) d\boldsymbol{x} = 0 \tag{A6}$$

if any exponent k_j is an odd number, or

$$\int_{\mathbb{R}^s} {x_1}^{k_1} \cdots {x_s}^{k_s} \mathcal{N}\left(\boldsymbol{x};\underline{0}\mathbf{0},\mathbf{I}_s\right) dx d\boldsymbol{x} = \frac{k_1!}{\sqrt{2}^{k_1}\left(\frac{k_1}{2}!\right)} \cdots \frac{k_s!}{\sqrt{2}^{k_s}\left(\frac{k_s}{2}!\right)} \tag{A7}$$

otherwise. Further, the solutions $x_{j,m}$ and w_m are the nodes and weights of the Gauss-Hermite quadrature rule.

In general, by applying to system (A3) the substitution

$$\begin{cases} w_m = {v_m}^2, \\ x_{j,m} = \dfrac{u_{m,j}}{v_m}, \end{cases} \tag{A8}$$

the weights are naturally forced to be positive, and the system can be conveniently rewritten as

$$\begin{cases} \vdots \\ \sum_{m=1}^{r+1} u_{m,j_1} \cdots u_{m,j_\xi} {v_m}^{2-\xi} = \mu_{j_1,\ldots,j_\xi}, \\ \vdots \end{cases} \tag{A9}$$

which, in the case of order $h = 2$, can be expressed in matrix form as

$$\left(\begin{array}{c|cc} \boldsymbol{v} & \boldsymbol{\Omega}^h & \end{array}\right)^T \left(\begin{array}{c|cc} \boldsymbol{v} & \boldsymbol{\Omega}^h & \end{array}\right) = \mathbf{I}_s, \tag{A10}$$

where $\boldsymbol{v} \in \mathbb{R}^{r+1}$ is the column vector with ~~coordinates~~ entries v_m and $\boldsymbol{\Omega}^h$ is an $(r+1) \times s$ matrix with ~~coordinates~~ entries $u_{m,j}$.

Note that equation (A10) can be solved for any unit vector $\boldsymbol{v}$ and for any value of $s \leq r$. If constant weights are chosen and $s = r$, then this method is equivalent to the second-order exact sampling of the SEIK filter (Pham, 2001), which leads to, in the formalism of equation (A5),

$$\mathbf{X} = \bar{\boldsymbol{x}} \mathbb{1}_{1\times(r+1)} + \mathbf{L} {\boldsymbol{\Omega}^h}^T \mathbf{W}^{-\frac{1}{2}}, \tag{A11}$$

where $\mathbf{W} = \mathrm{diag}\,(\boldsymbol{w})$ is the $(r+1) \times (r+1)$ diagonal matrix of weights.

In general, if $h > 2$, s must be lower than r to obtain a solution to system (A9). The obtained s-dimensional ensemble can be extended to an r-dimensional one: its projection on the s-dimensional subspace defined by the first s ~~coordinates~~ entries will retain order h, while the remaining components are sampled with order 2. This is achieved by starting from $\boldsymbol{v}$ and $\boldsymbol{\Omega}^h$, given by a particular solution of system (A9). Then, exploiting symmetry and rotational invariance, randomness is added to avoid biases by multiplying $\boldsymbol{\Omega}^h$ by any random $s \times s$ orthogonal matrix $\boldsymbol{\Omega}^{rnd}$. Finally, the $(r+1) \times (r-s)$ matrix $\boldsymbol{\Omega}^\perp$ is chosen to fulfil

$$\left(\begin{array}{c|cc|c} \boldsymbol{v} & \boldsymbol{\Omega}^h\boldsymbol{\Omega}^{rnd} & & \boldsymbol{\Omega}^\perp \end{array}\right)^T \left(\begin{array}{c|cc|c} \boldsymbol{v} & \boldsymbol{\Omega}^h\boldsymbol{\Omega}^{rnd} & & \boldsymbol{\Omega}^\perp \end{array}\right) = \mathbf{I}_{r+1}. \tag{A12}$$

A procedure for randomly building or completing orthogonal matrices is described in Appendix B.

The sampling matrix $\boldsymbol{\Omega}$ is defined as

$$\boldsymbol{\Omega} = \left(\begin{array}{cc|c} & \boldsymbol{\Omega}^h\boldsymbol{\Omega}^{rnd} & \boldsymbol{\Omega}^\perp \end{array}\right). \tag{A13}$$

Now the columns of $\boldsymbol{\Omega}^T \mathbf{W}^{-\frac{1}{2}}$ represent an r-dimensional ensemble of order h in the first s ~~coordinates~~ dimensions and order 2 in the last $r - s$ ~~coordinates~~dimensions. Equation (A11) needs to be modified to properly project the higher order subspace ~~in~~ onto the principal components of the covariance matrix $\mathbf{L}\mathbf{L}^T$. Such components, ordered by relevance, are the ~~column~~ columns of the matrix $\mathbf{L}\mathbf{C}^T$, where $\mathbf{C}$ is a $r \times r$ orthogonal change-of-basis matrix such that

$$\mathbf{L}^T \boldsymbol{\Lambda}^{-1} \mathbf{L} = \mathbf{C}^T \mathbf{E} \mathbf{C}. \tag{A14}$$

In the last equation, the right-hand side is an eigenvalue decomposition of the left-hand side, with $\mathbf{E}$ being an $r \times r$ diagonal matrix of eigenvalues in descending order and $\mathbf{\Lambda}$ being an appropriate $N \times N$ matrix. The purpose of $\mathbf{\Lambda}$ has already been discussed in Section 2.2.

The eigenvalue decomposition is performed to produce a PCA of the uncertainty represented by the ensemble. In fact, the $\mathbf{C}$ matrix induces a change of basis such that the principal ~~component~~ components are on the left side of the matrix product $\mathbf{L}\mathbf{C}^T$. This is compliant with the ensemble stored in $\mathbf{\Omega}^T\mathbf{W}^{-\frac{1}{2}}$, and the sampling is finally obtained by

$$\mathbf{X} = \bar{\boldsymbol{x}}\mathbb{1}_{1\times(r+1)} + \mathbf{L}\mathbf{C}^T{\mathbf{\Omega}^h}^T\mathbf{W}^{-\frac{1}{2}}. \tag{A15}$$

Appendix B: Orthogonal matrices

The GHOSH sampling algorithm requires ~~to randomly generate~~ the random generation of orthogonal matrices, as in the case of $\mathbf{\Omega}^{rnd}$, or to complete a set of orthonormal vectors to form an orthogonal matrix, as in the case of $\mathbf{\Omega}^{\perp}$ (equations (4) and (A12)). Algorithms to ~~extract~~ sample such matrices have been discussed in literature (e.g., Pham, 2001; Nerger et al., 2012). Here we propose the algorithm that we have implemented in the provided python code.

Given a $m \times (m-s)$ matrix $\mathbf{\Omega}^{=}$, the columns of ~~witch~~ which are a set of $(m-s)$ orthonormal vectors in $\mathbb{R}^m$, we want to randomly sample a $m \times s$ matrix ~~$\mathbf{\Omega}^{\perp}$~~ $\mathbf{\Omega}^{\perp}$, such that the $m \times m$ matrix $\mathbf{\Omega}$ defined as

$$\mathbf{\Omega} := \left(\begin{array}{c|c} & \\ \mathbf{\Omega}^{=} & \mathbf{\Omega}^{\perp} \\ & \end{array} \right) \tag{B1}$$

is an orthogonal matrix.

To do so we proceed in an iterative way, building ~~$\mathbf{\Omega}^{\perp}$~~ $\mathbf{\Omega}^{\perp}$ column by column. Thus, the problem reduces to producing a random vector $\boldsymbol{w} \in \mathbb{R}^m$ orthonormal to the columns of a given $\mathbf{\Omega}^{=}$. This can be done through the following steps:

1. extract a random unitary vector $\boldsymbol{u} \in \mathbb{R}^m$, i.e.:
   (a) generate a column vector $\boldsymbol{v} \in \mathbb{R}^m$ by extracting m random numbers from a standard normal distribution (i.e., 0-mean and unitary variance), one for each ~~coordinate~~ entry of $\boldsymbol{v}$;
   (b) normalize $\boldsymbol{v}$ to get $\boldsymbol{u} = \dfrac{\boldsymbol{v}}{\sqrt{\boldsymbol{v}^T\boldsymbol{v}}}$
2. decompose $\boldsymbol{u} = \boldsymbol{u}^{=} + \boldsymbol{u}^{\perp}$, with the first term ~~laying~~ lying in the subspace generated by the columns of $\mathbf{\Omega}^{=}$ and the last term in its orthogonal subspace, i.e.:
   $$\boldsymbol{u}^{=} = \mathbf{\Omega}^{=}\left(\mathbf{\Omega}^{=}\right)^T\boldsymbol{u},$$
   $$\boldsymbol{u}^{\perp} = \boldsymbol{u} - \boldsymbol{u}^{=};$$
3. normalize ~~$\boldsymbol{u}^{\perp}$~~ $\boldsymbol{u}^{\perp}$ to obtain $\boldsymbol{w} = \dfrac{\boldsymbol{u}^{\perp}}{\sqrt{\left(\boldsymbol{u}^{\perp}\right)^T\boldsymbol{u}^{\perp}}}$.

The vector ~~$\boldsymbol{w}$~~ $\boldsymbol{u}$ has unitary norm and it is orthogonal to $\boldsymbol{\Omega}^{=}$ as $\boldsymbol{u}^{\perp}$. The next columns of $\boldsymbol{\Omega}^{\perp}$ are computed in the same manner, after adding $\boldsymbol{w}$ as a new column of $\boldsymbol{\Omega}^{=}$.

This algorithm also covers how to generate a random orthogonal matrix from scratch. In fact, if $s = m$, then $\boldsymbol{\Omega} = \boldsymbol{\Omega}^{\perp}$ and the first computed column of $\boldsymbol{\Omega}^{\perp}$ is $\boldsymbol{w} = \boldsymbol{u}^{\perp} = \boldsymbol{u}$.

Appendix C: Examples

The aim of this appendix is to help the reader ~~understanding~~ understand the concept of h^{th}-order of approximation by presenting some simple examples of ensembles with a certain order of approximation h. The examples are organized by the space dimension r. In all the examples the ensemble members, noted as $\boldsymbol{P}_1, \boldsymbol{P}_2, \ldots$, are points of $\mathbb{R}^r$, and the target pdf, the moments of which are matched up to order h, is the standard normal distribution.

C1 Dimension 1 ($r = 1$)

The target pdf is the one-dimensional standard normal distribution $\mathcal{N}(x; 0, 1)$. According to equations (A6) and (A7), the moments characterizing the pdf are:

- the first moment (mean): $\mu_x = 0$,
- the second moment (variance): $\mu_{x^2} = 1$,
- the third moment (skewness): $\mu_{x^3} = 0$,
- the fourth moment (kurtosis): $\mu_{x^4} = 3$,
- the fifth moment: $\mu_{x^5} = 0$, ...

C1.1 Second-order ensemble ($h = 2$) with 2 members

$$\begin{aligned} &\boldsymbol{P}_1: \quad x_1 = 1; \\ &\boldsymbol{P}_2: \quad x_2 = -1. \end{aligned} \tag{C1}$$

The ensemble first moment (mean) is:

$$\frac{1}{2}(x_1 + x_2) = \frac{1}{2}(1 + (-1)) = 0. \tag{C2}$$

The ensemble second moment (variance) is:

$$\frac{1}{2}\left({x_1}^2 + {x_2}^2\right) = \frac{1}{2}(1 + 1) = 1. \tag{C3}$$

Thus, the ensemble $\boldsymbol{P}_1, \boldsymbol{P}_2$ matches the moments of the target pdf up to order 2.

C1.2 Third-order ensemble ($h = 3$) with 2 members

The present example uses ensemble (C1) of the previous example.

In one dimension, the ensemble of order 2 is also an ensemble of order 3. The same does not happen for higher dimensions.

The ensemble third moment (skewness) is:

$$\frac{1}{2}\left({x_1}^3 + {x_2}^3\right) = \frac{1}{2}\left(1 + (-1)\right) = 0. \tag{C4}$$

Thus, the ensemble $\boldsymbol{P}_1, \boldsymbol{P}_2$ matches the moments of the target pdf up to order 3.

C1.3 Fifth-order weighted ensemble ($h = 5$) with 3 members

$$\begin{aligned} &\boldsymbol{P}_1: \quad x_1 = \sqrt{3}, \qquad w_1 = \frac{1}{6}; \\ &\boldsymbol{P}_2: \quad x_2 = -\sqrt{3}, \quad w_2 = \frac{1}{6}; \\ &\boldsymbol{P}_3: \quad x_3 = 0, \qquad w_3 = \frac{2}{3}. \end{aligned} \tag{C5}$$

The ensemble first moment (mean) is:

$$w_1 x_1 + w_2 x_2 + w_3 x_3 = \frac{1}{6}\cdot\sqrt{3} + \frac{1}{6}\cdot\left(-\sqrt{3}\right) + \frac{2}{3}\cdot 0 = 0. \tag{C6}$$

The ensemble second moment (variance) is:

$$w_1 {x_1}^2 + w_2 {x_2}^2 + w_3 {x_3}^2 = \frac{1}{6}\cdot 3 + \frac{1}{6}\cdot 3 + \frac{2}{3}\cdot 0 = 1. \tag{C7}$$

The ensemble third moment (skewness) is:

$$w_1 {x_1}^3 + w_2 {x_2}^3 + w_3 {x_3}^3 = \frac{1}{6}\cdot\sqrt{3}^3 + \frac{1}{6}\cdot\left(-\sqrt{3}\right)^3 + \frac{2}{3}\cdot 0 = 0. \tag{C8}$$

The ensemble fourth moment (kurtosis) is:

$$w_1 {x_1}^4 + w_2 {x_2}^4 + w_3 {x_3}^4 = \frac{1}{6}\cdot 9 + \frac{1}{6}\cdot 9 + \frac{2}{3}\cdot 0 = 3. \tag{C9}$$

The ensemble fifth moment is:

$$w_1 {x_1}^5 + w_2 {x_2}^5 + w_3 {x_3}^5 = \frac{1}{6}\cdot\sqrt{3}^5 + \frac{1}{6}\cdot\left(-\sqrt{3}\right)^5 + \frac{2}{3}\cdot 0 = 0. \tag{C10}$$

Thus, the ensemble $\boldsymbol{P}_1, \boldsymbol{P}_2, \boldsymbol{P}_3$ matches the moments of the target pdf up to order 5.

C2 Dimension 2 ($r = 2$)

The target pdf is the two-dimensional standard normal distribution $\mathcal{N}(\boldsymbol{x}; \mathbf{0}, \mathbf{I}_2)$. According to equations (A6) and (A7), the moments characterizing the pdf are:

- ~~first moments:~~
- ~~first moment~~ the first moments associated to x (i.e., the mean along x) ~~is:~~ $\mu_x = 0$ ~~,~~
- ~~first moment associated~~ and to y (i.e., the mean along y) ~~is:~~ $\mu_y = 0$,
- ~~second moments:~~
- ~~second moment~~ the second moments associated to x^2 (i.e., the variance along x) ~~is:~~ $\mu_{x^2} = 1$,
- ~~second moment associated~~ to y^2 (i.e., the variance along y) ~~is:~~ $\mu_{y^2} = 1$,
- ~~second moment associated~~ and to xy (i.e., the covariance between x and y) ~~is:~~ $\mu_{xy} = 0$,
- ~~third moments:~~
- ~~third moment~~ the third moments associated to x^3 ~~is:~~ $\mu_{x^3} = 0$,
- ~~third moment associated~~ to x^2y ~~is:~~ $\mu_{x^2y} = 0$,
- ~~third moment associated~~ to xy^2 ~~is:~~ $\mu_{xy^2} = 0$,
- ~~third moment associated~~ and to y^3 ~~is:~~ $\mu_{y^3} = 0$,
- ~~fourth moments:~~
- ~~fourth moment~~ the fourth moments associated to x^4 ~~is:~~ $\mu_{x^4} = 3$,
- ~~fourth moment associated~~ to x^3y ~~is:~~ $\mu_{x^3y} = 0$,
- ~~fourth moment associated~~ to x^2y^2 ~~is:~~ $\mu_{x^2y^2} = 1$,
- ~~fourth moment associated~~ to xy^3 ~~is:~~ $\mu_{xy^3} = 0$,
- ~~fourth moment associated~~ and to y^4 ~~is:~~ $\mu_{y^4} = 3$,
- ~~fifth moments:~~
- ~~fifth moment~~ the fifth moments associated to x^5 ~~is:~~ $\mu_{x^5} = 0$,
- ~~fifth moment associated~~ to x^4y ~~is:~~ $\mu_{x^4y} = 0$,
- ~~fifth moment associated~~ to x^3y^2 ~~is:~~ $\mu_{x^3y^2} = 0$,
- ~~fifth moment associated~~ to x^2y^3 ~~is:~~ $\mu_{x^2y^3} = 0$,
- ~~fifth moment associated~~ to xy^4 ~~is:~~ $\mu_{xy^4} = 0$,
- ~~fifth moment associated~~ and to y^5 ~~is:~~ $\mu_{y^5} = 0$, ...

C2.1 Second-order ensemble ($h = 2$) with 3 members

$$\begin{aligned}
\boldsymbol{P}_1: &\quad (x_1, y_1) = \left(\sqrt{\frac{3}{2}}, -\frac{\sqrt{2}}{2}\right); \\
\boldsymbol{P}_2: &\quad (x_2, y_2) = \left(-\sqrt{\frac{3}{2}}, -\frac{\sqrt{2}}{2}\right); \\
\boldsymbol{P}_3: &\quad (x_3, y_3) = \left(0, \sqrt{2}\right).
\end{aligned} \tag{C11}$$

This ensemble has the shape of an equilateral triangle. It is one of the most common second-order ensembles sampled by deterministic sampling methods like SEIK's second-order-exact sampling (Pham, 2001).

The ensemble first moment associated to x (i.e., the mean along x) is:

$$\frac{1}{3}(x_1 + x_2 + x_3) = \frac{1}{3}\left(\sqrt{\frac{3}{2}} + \left(-\sqrt{\frac{3}{2}}\right) + 0\right) = 0. \tag{C12}$$

The ensemble first moment associated to y (i.e., the mean along y) is:

$$\frac{1}{3}(y_1 + y_2 + y_3) = \frac{1}{3}\left(-\frac{\sqrt{2}}{2} + \left(-\frac{\sqrt{2}}{2}\right) + \sqrt{2}\right) = 0. \tag{C13}$$

The ensemble second moment associated to x^2 (i.e., the variance along x) is:

$$\frac{1}{3}\left({x_1}^2 + {x_2}^2 + {x_3}^2\right) = \frac{1}{3}\left(\frac{3}{2} + \frac{3}{2} + 0\right) = 1. \tag{C14}$$

The ensemble second moment associated to y^2 (i.e., the variance along y) is:

$$\frac{1}{3}\left({y_1}^2 + {y_2}^2 + {y_3}^2\right) = \frac{1}{3}\left(\frac{1}{2} + \frac{1}{2} + 2\right) = 1. \tag{C15}$$

The ensemble second moment associated to xy (i.e., the covariance between x and y) is:

$$\frac{1}{3}(x_1 y_1 + x_2 y_2 + x_3 y_3) = \frac{1}{3}\left(\sqrt{\frac{3}{2}} \cdot \left(-\frac{\sqrt{2}}{2}\right) + \left(-\sqrt{\frac{3}{2}}\right) \cdot \left(-\frac{\sqrt{2}}{2}\right) + 0 \cdot \sqrt{2}\right) = 0. \tag{C16}$$

Thus, the ensemble $\boldsymbol{P}_1, \boldsymbol{P}_2, \boldsymbol{P}_3$ matches the moments of the target pdf up to order 2.

Observe that any symmetry or rotation applied to this ensemble (i.e., applying an orthogonal transformation to the members) preserves its order.

C2.2 High-order sampling ensemble with 3 members, $h = 3$ along the first dimension

The present example uses ensemble (C11) of the previous example.

This particular second-order ensemble has order 3 along the x-axis. In general, this is no longer true if you apply any rotation (or symmetry) to the ensemble.

The ensemble third moment associated to x^3 is:

$$\frac{1}{3}\left({x_1}^3 + {x_2}^3 + {x_3}^3\right) = \frac{1}{3}\left(\left(\sqrt{\frac{3}{2}}\right)^3 + \left(-\sqrt{\frac{3}{2}}\right)^3 + 0\right) = 0. \tag{C17}$$

Thus, the ensemble $\boldsymbol{P}_1, \boldsymbol{P}_2, \boldsymbol{P}_3$ matches the moments of the target pdf up to order 3 along the x-axis, and up to order 2 elsewhere.

C2.3 High-order sampling ensemble with 3 weighted members, $h = 5$ along the first dimension

$$\begin{aligned}
\boldsymbol{P}_1: &\quad (x_1, y_1) = \left(\sqrt{3}, -\sqrt{2}\right), &\quad w_1 = \frac{1}{6};\\
\boldsymbol{P}_2: &\quad (x_2, y_2) = \left(-\sqrt{3}, -\sqrt{2}\right), &\quad w_2 = \frac{1}{6};\\
\boldsymbol{P}_3: &\quad (x_3, y_3) = \left(0, \frac{\sqrt{2}}{2}\right), &\quad w_3 = \frac{2}{3}.
\end{aligned} \tag{C18}$$

Compared to the previous example, this ensemble uses weighted members to achieve a higher order of approximation (5 ~~insted~~ instead of 3) along the x-axis. In fact, looking at the x-coordinate of the ensemble members (i.e., projecting on the x-axis), it results in the same ensemble as (C5). Thus, it only remains to verify the moments involving also the y direction.

The ensemble first moment associated to y (i.e., the mean along y) is:

$$w_1 y_1 + w_2 y_2 + w_3 y_3 = \frac{1}{6} \cdot \left(-\sqrt{2}\right) + \frac{1}{6} \cdot \left(-\sqrt{2}\right) + \frac{2}{3} \cdot \frac{\sqrt{2}}{2} = 0. \tag{C19}$$

The ensemble second moment associated to y^2 (i.e., the variance along y) is:

$$w_1 {y_1}^2 + w_2 {y_2}^2 + w_3 {y_3}^2 = \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 2 + \frac{2}{3} \cdot \frac{1}{2} = 1. \tag{C20}$$

The ensemble second moment associated to xy (i.e., the covariance between x and y) is:

$$w_1 x_1 y_1 + w_2 x_2 y_2 + w_3 x_3 y_3 = \frac{1}{6} \cdot \sqrt{3} \cdot \left(-\sqrt{2}\right) + \frac{1}{6} \cdot \left(-\sqrt{3}\right) \cdot \left(-\sqrt{2}\right) + \frac{2}{3} \cdot 0 \cdot \frac{\sqrt{2}}{2} = 0. \tag{C21}$$

Thus, the ensemble $\boldsymbol{P}_1, \boldsymbol{P}_2, \boldsymbol{P}_3$ matches the moments of the target pdf up to order 5 along the x-axis, and up to order 2 elsewhere.

C2.4 Third-order ensemble ($h = 3$) with 4 members

$$\begin{aligned}
\boldsymbol{P}_1: &\quad (x_1, y_1) = \left(\sqrt{2}, 0\right);\\
\boldsymbol{P}_2: &\quad (x_2, y_2) = \left(-\sqrt{2}, 0\right);\\
\boldsymbol{P}_3: &\quad (x_3, y_3) = \left(0, \sqrt{2}\right);\\
\boldsymbol{P}_4: &\quad (x_4, y_4) = \left(0, -\sqrt{2}\right).
\end{aligned} \tag{C22}$$

This square-shaped ensemble uses $2r$ members to achieve the third-order.

The ensemble first moment associated to x (i.e., the mean along x) is:

$$\frac{1}{4}\left(x_1 + x_2 + x_3 + x_4\right) = \frac{1}{4}\left(\sqrt{2} + \left(-\sqrt{2}\right) + 0 + 0\right) = 0. \tag{C23}$$

The ensemble first moment associated to y (i.e., the mean along y) is:

$$\frac{1}{4}\left(y_1+y_2+y_3+y_4\right)=\frac{1}{4}\left(0+0+\sqrt{2}+\left(-\sqrt{2}\right)\right)=0. \tag{C24}$$

The ensemble second moment associated to x^2 (i.e., the variance along x) is:

$$\frac{1}{4}\left({x_1}^2+{x_2}^2+{x_3}^2+{x_4}^2\right)=\frac{1}{4}\left(2+2+0+0\right)=1. \tag{C25}$$

The ensemble second moment associated to y^2 (i.e., the variance along y) is:

$$\frac{1}{4}\left({y_1}^2+{y_2}^2+{y_3}^2+{y_4}^2\right)=\frac{1}{4}\left(0+0+2+2\right)=1. \tag{C26}$$

The ensemble second moment associated to xy (i.e., the covariance between x and y) is:

$$\frac{1}{4}\left(x_1y_1+x_2y_2+x_3y_3+x_4y_4\right)=\frac{1}{4}\left(\sqrt{2}\cdot 0+\left(-\sqrt{2}\right)\cdot 0+0\cdot\sqrt{2}+0\cdot\left(-\sqrt{2}\right)\right)=0. \tag{C27}$$

The ensemble third moment associated to x^3 is:

$$\frac{1}{4}\left({x_1}^3+{x_2}^3+{x_3}^3+{x_4}^3\right)=\frac{1}{4}\left(\left(\sqrt{2}\right)^3+\left(-\sqrt{2}\right)^3+0+0\right)=0. \tag{C28}$$

The ensemble third moment associated to x^2y is:

$$\frac{1}{4}\left({x_1}^2y_1+{x_2}^2y_2+{x_3}^2y_3+{x_4}^2y_4\right)=\frac{1}{4}\left(2\cdot 0+2\cdot 0+0\cdot\sqrt{2}+0\cdot\left(-\sqrt{2}\right)\right)=0. \tag{C29}$$

The ensemble third moment associated to xy^2 is:

$$\frac{1}{4}\left(x_1{y_1}^2+x_2{y_2}^2+x_3{y_3}^2+x_4{y_4}^2\right)=\frac{1}{4}\left(\sqrt{2}\cdot 0+\left(-\sqrt{2}\right)\cdot 0+0\cdot 2+0\cdot 2\right)=0. \tag{C30}$$

The ensemble third moment associated to y^3 is:

$$\frac{1}{4}\left({y_1}^3+{y_2}^3+{y_3}^3+{y_4}^3\right)=\frac{1}{4}\left(0+0+\left(\sqrt{2}\right)^3+\left(-\sqrt{2}\right)^3\right)=0. \tag{C31}$$

Thus, the ensemble $\boldsymbol{P}_1,\boldsymbol{P}_2,\boldsymbol{P}_3,\boldsymbol{P}_4$ matches the moments of the target pdf up to order 3.

C2.5 Third-order weighted ensemble with 4 members, fifth-order along the first dimension

$$\begin{aligned}
&\boldsymbol{P}_1: \quad (x_1,y_1)=\left(\sqrt{3},0\right), && w_1=\frac{1}{6};\\
&\boldsymbol{P}_2: \quad (x_2,y_2)=\left(-\sqrt{3},0\right), && w_2=\frac{1}{6};\\
&\boldsymbol{P}_3: \quad (x_3,y_3)=\left(0,\sqrt{\frac{3}{2}}\right), && w_3=\frac{1}{3};\\
&\boldsymbol{P}_4: \quad (x_4,y_4)=\left(0,-\sqrt{\frac{3}{2}}\right), && w_4=\frac{1}{3}.
\end{aligned} \tag{C32}$$

This ensemble extends ensemble (C5) to two dimensions. In fact, looking at the x-coordinates, it results in the same ensemble as (C5) after summing the weights of $\boldsymbol{P}_3$ and $\boldsymbol{P}_4$ collapsing in the same point on the x-axis. Differently from ensemble (C18), which also ~~extend~~ extends ensemble (C5), it achieves order 3 by using 4 members instead of 3. This is not considered a high-order sampling ensemble, since high-order sampling produces ensembles with $r+1$ members, as in the case of ensemble (C18). Since the moments along the x-axis are already checked for ensemble (C5), it only remains to verify the moments involving the y direction.

The ensemble first moment associated to y (i.e., the mean along y) is:

$$w_1 y_1 + w_2 y_2 + w_3 y_3 + w_4 y_4 = \frac{1}{6}\cdot 0 + \frac{1}{6}\cdot 0 + \frac{1}{3}\cdot\sqrt{\frac{3}{2}} + \frac{1}{3}\cdot\left(-\sqrt{\frac{3}{2}}\right) = 0. \tag{C33}$$

The ensemble second moment associated to y^2 (i.e., the variance along y) is:

$$w_1 {y_1}^2 + w_2 {y_2}^2 + w_3 {y_3}^2 + w_4 {y_4}^2 = \frac{1}{6}\cdot 0 + \frac{1}{6}\cdot 0 + \frac{1}{3}\cdot\frac{3}{2} + \frac{1}{3}\cdot\frac{3}{2} = 1. \tag{C34}$$

The ensemble second moment associated to xy (i.e., the covariance between x and y) is:

$$w_1 x_1 y_1 + w_2 x_2 y_2 + w_3 x_3 y_3 + w_4 x_4 y_4 = \frac{1}{6}\cdot\sqrt{3}\cdot 0 + \frac{1}{6}\cdot\left(-\sqrt{3}\right)\cdot 0 + \frac{1}{3}\cdot 0\cdot\sqrt{\frac{3}{2}} + \frac{1}{3}\cdot 0\cdot\left(-\sqrt{\frac{3}{2}}\right) = 0. \tag{C35}$$

The ensemble third moment associated to $x^2 y$ is:

$$w_1 {x_1}^2 y_1 + w_2 {x_2}^2 y_2 + w_3 {x_3}^2 y_3 + w_4 {x_4}^2 y_4 = \frac{1}{6}\cdot 3\cdot 0 + \frac{1}{6}\cdot 3\cdot 0 + \frac{1}{3}\cdot 0\cdot\sqrt{\frac{3}{2}} + \frac{1}{3}\cdot 0\cdot\left(-\sqrt{\frac{3}{2}}\right) = 0. \tag{C36}$$

The ensemble third moment associated to xy^2 is:

$$w_1 x_1 {y_1}^2 + w_2 x_2 {y_2}^2 + w_3 x_3 {y_3}^2 + w_4 x_4 {y_4}^2 = \frac{1}{6}\cdot\sqrt{3}\cdot 0 + \frac{1}{6}\cdot\left(-\sqrt{3}\right)\cdot 0 + \frac{1}{3}\cdot 0\cdot\frac{3}{2} + \frac{1}{3}\cdot 0\cdot\frac{3}{2} = 0. \tag{C37}$$

The ensemble third moment associated to y^3 is:

$$w_1 {y_1}^3 + w_2 {y_2}^3 + w_3 {y_3}^3 + w_4 {y_4}^3 = \frac{1}{6}\cdot 0 + \frac{1}{6}\cdot 0 + \frac{1}{3}\cdot\left(\sqrt{\frac{3}{2}}\right)^3 + \frac{1}{3}\cdot\left(-\sqrt{\frac{3}{2}}\right)^3 = 0. \tag{C38}$$

Thus, the ensemble $\boldsymbol{P}_1, \boldsymbol{P}_2, \boldsymbol{P}_3, \boldsymbol{P}_4$ matches the moments of the target pdf up to order 5 along the x-axis, and up to order 3 elsewhere.

C2.6 Fifth-order weighted ensemble ($h = 5$) with 7 members

$$
\begin{aligned}
&\boldsymbol{P}_1: \quad (x_1, y_1) = \left(\sqrt{3}, 1\right), && w_1 = \frac{1}{12};\\
&\boldsymbol{P}_2: \quad (x_2, y_2) = \left(\sqrt{3}, -1\right), && w_2 = \frac{1}{12};\\
&\boldsymbol{P}_3: \quad (x_3, y_3) = \left(-\sqrt{3}, 1\right), && w_3 = \frac{1}{12};\\
&\boldsymbol{P}_4: \quad (x_4, y_4) = \left(-\sqrt{3}, -1\right), && w_4 = \frac{1}{12};\\
&\boldsymbol{P}_5: \quad (x_5, y_5) = (0, 2), && w_5 = \frac{1}{12};\\
&\boldsymbol{P}_6: \quad (x_6, y_6) = (0, -2), && w_6 = \frac{1}{12};\\
&\boldsymbol{P}_7: \quad (x_7, y_7) = (0, 0), && w_7 = \frac{1}{2}.
\end{aligned}
\tag{C39}
$$

This hexagon-shaped ensemble uses its 7 weighted members to achieve the fifth-order. Looking at the x-coordinate of the ensemble members (i.e., projecting on the x-axis), it results in the same ensemble as (C5) by summing the weights of points with the same projection.

The ensemble first moment associated to x (i.e., the mean along x) is:

$$
w_1 x_1 + w_2 x_2 + \ldots + w_7 x_7 = \frac{1}{12} \cdot \sqrt{3} + \frac{1}{12} \cdot \sqrt{3} + \frac{1}{12} \cdot \left(-\sqrt{3}\right) + \frac{1}{12} \cdot \left(-\sqrt{3}\right) + \frac{1}{12} \cdot 0 + \frac{1}{12} \cdot 0 + \frac{1}{2} \cdot 0 = 0. \tag{C40}
$$

The ensemble first moment associated to y (i.e., the mean along y) is:

$$
w_1 y_1 + w_2 y_2 + \ldots + w_7 y_7 = \frac{1}{12} \cdot 1 + \frac{1}{12} \cdot (-1) + \frac{1}{12} \cdot 1 + \frac{1}{12} \cdot (-1) + \frac{1}{12} \cdot 2 + \frac{1}{12} \cdot (-2) + \frac{1}{2} \cdot 0 = 0. \tag{C41}
$$

The ensemble second moment associated to x^2 (i.e., the variance along x) is:

$$
w_1 {x_1}^2 + w_2 {x_2}^2 + \ldots + w_7 {x_7}^2 = \frac{1}{12} \cdot 3 + \frac{1}{12} \cdot 3 + \frac{1}{12} \cdot 3 + \frac{1}{12} \cdot 3 + \frac{1}{12} \cdot 0 + \frac{1}{12} \cdot 0 + \frac{1}{2} \cdot 0 = 1. \tag{C42}
$$

The ensemble second moment associated to y^2 (i.e., the variance along y) is:

$$
w_1 {y_1}^2 + w_2 {y_2}^2 + \ldots + w_7 {y_7}^2 = \frac{1}{12} \cdot 1 + \frac{1}{12} \cdot 1 + \frac{1}{12} \cdot 1 + \frac{1}{12} \cdot 1 + \frac{1}{12} \cdot 4 + \frac{1}{12} \cdot 4 + \frac{1}{2} \cdot 0 = 1. \tag{C43}
$$

The ensemble second moment associated to xy (i.e., the covariance between x and y) is:

$$
\begin{aligned}
&w_1 x_1 y_1 + w_2 x_2 y_2 + \ldots + w_7 x_7 y_7 =\\
&\quad = \frac{1}{12} \cdot \sqrt{3} \cdot 1 + \frac{1}{12} \cdot \sqrt{3} \cdot (-1) + \frac{1}{12} \cdot \left(-\sqrt{3}\right) \cdot 1 + \frac{1}{12} \cdot \left(-\sqrt{3}\right) \cdot (-1) + \frac{1}{12} \cdot 0 \cdot 2 + \frac{1}{12} \cdot 0 \cdot (-2) + \frac{1}{2} \cdot 0 \cdot 0 = 0.
\end{aligned}
\tag{C44}
$$

Given the comparably high number of moments to be matched, it could be useful here to introduce a result that relieves from checking some moments. In fact, if a zero-mean ensemble is symmetric (i.e., if $\boldsymbol{P}$ is a non-zero ensemble member then also $-\boldsymbol{P}$ is an ensemble member with the same weight), then all the odd moments are zero. It can be ~~proven easily~~ easily shown by noting that any couple of symmetric members adds zero to the moment calculation, since both of the members produce the

same quantity but with opposite sign.

The ensemble (C39) is symmetric, then it remains only to prove that fourth-order moments ~~matches~~ match the moments of the pdf.

The ensemble fourth moment associated to x^4 is:

$$w_1 {x_1}^4 + w_2 {x_2}^4 + \ldots + w_7 {x_7}^4 = \frac{1}{12}\cdot 9 + \frac{1}{12}\cdot 9 + \frac{1}{12}\cdot 9 + \frac{1}{12}\cdot 9 + \frac{1}{12}\cdot 0 + \frac{1}{12}\cdot 0 + \frac{1}{2}\cdot 0 = 3. \tag{C45}$$

The ensemble fourth moment associated to $x^3 y$ is:

$$\begin{aligned} &w_1 {x_1}^3 y_1 + w_2 {x_2}^3 y_2 + \ldots + w_7 {x_7}^3 y_7 = \\ &= \frac{1}{12}\cdot\left(\sqrt{3}\right)^3\cdot 1 + \frac{1}{12}\cdot\left(\sqrt{3}\right)^3\cdot(-1) + \frac{1}{12}\cdot\left(-\sqrt{3}\right)^3\cdot 1 + \frac{1}{12}\cdot\left(-\sqrt{3}\right)^3\cdot(-1) + \frac{1}{12}\cdot 0\cdot 2 + \frac{1}{12}\cdot 0\cdot(-2) + \frac{1}{2}\cdot 0\cdot 0 = 0. \end{aligned} \tag{C46}$$

The ensemble fourth moment associated to $x^2 y^2$ is:

$$w_1 {x_1}^2 {y_1}^2 + w_2 {x_2}^2 {y_2}^2 + \ldots + w_7 {x_7}^2 {y_7}^2 = \frac{1}{12}\cdot 3\cdot 1 + \frac{1}{12}\cdot 3\cdot 1 + \frac{1}{12}\cdot 3\cdot 1 + \frac{1}{12}\cdot 3\cdot 1 + \frac{1}{12}\cdot 0\cdot 4 + \frac{1}{12}\cdot 0\cdot 4 + \frac{1}{2}\cdot 0\cdot 0 = 1. \tag{C47}$$

The ensemble fourth moment associated to xy^3 is:

$$\begin{aligned} &w_1 x_1 {y_1}^3 + w_2 x_2 {y_2}^3 + \ldots + w_7 x_7 {y_7}^3 = \\ &\quad = \frac{1}{12}\cdot\sqrt{3}\cdot 1 + \frac{1}{12}\cdot\sqrt{3}\cdot(-1) + \frac{1}{12}\cdot\left(-\sqrt{3}\right)\cdot 1 + \frac{1}{12}\cdot\left(-\sqrt{3}\right)\cdot(-1) + \frac{1}{12}\cdot 0\cdot 8 + \frac{1}{12}\cdot 0\cdot(-8) + \frac{1}{2}\cdot 0\cdot 0 = 0. \end{aligned} \tag{C48}$$

The ensemble fourth moment associated to y^4 is:

$$w_1 {y_1}^4 + w_2 {y_2}^4 + \ldots + w_7 {y_7}^4 = \frac{1}{12}\cdot 1 + \frac{1}{12}\cdot 1 + \frac{1}{12}\cdot 1 + \frac{1}{12}\cdot 1 + \frac{1}{12}\cdot 16 + \frac{1}{12}\cdot 16 + \frac{1}{2}\cdot 0 = 3. \tag{C49}$$

Thus, the ensemble $\boldsymbol{P}_1, \boldsymbol{P}_2, \boldsymbol{P}_3, \boldsymbol{P}_4$ matches the moments of the target pdf up to order 5.

C3 Dimension 3 ($r = 3$)

The target pdf is the three-dimensional standard normal distribution $\mathcal{N}(\boldsymbol{x}; \mathbf{0}, \mathbf{I}_3)$. According to equations (A6) and (A7) and similarly to previous cases, the moments characterizing the pdf are:

- the first moments:
- ~~first moment associated to x (i.e., the mean along x) is:~~ $\mu_x = 0$,
- ~~first moment associated to y (i.e., the mean along y) is:~~ $\mu_y = 0$,
- ~~first moment associated to z (i.e., the mean along z) is:~~ and $\mu_z = 0$,
- ~~second moments:~~

- ~~second moment associated to x^2 (i.e., the variance along x) is:~~ the second moments $\mu_{x^2} = 1$,
- ~~second moment associated to y^2 (i.e., the variance along y) is:~~ $\mu_{y^2} = 1$,
- ~~second moment associated to z^2 (i.e., the variance along z) is:~~ $\mu_{z^2} = 1$,
- ~~second moment associated to xy (i.e., the covariance between x and y) is:~~ $\mu_{xy} = 0$,
- ~~second moment associated to xz (i.e., the covariance between x and z) is:~~ $\mu_{xz} = 0$,
- ~~second moment associated to yz (i.e., the covariance between y and z) is:~~ and $\mu_{yz} = 0$, ...

C3.1 High-order sampling ensemble with 4 members, $h = 3$ along the xy-plane

$$\begin{aligned}
&\boldsymbol{P}_1: \quad (x_1, y_1, z_1) = \left(\sqrt{2}, 0, -\frac{1}{2}\right);\\
&\boldsymbol{P}_2: \quad (x_2, y_2, z_2) = \left(-\sqrt{2}, 0, -\frac{1}{2}\right);\\
&\boldsymbol{P}_3: \quad (x_3, y_3, z_3) = \left(0, \sqrt{2}, \frac{1}{2}\right);\\
&\boldsymbol{P}_4: \quad (x_4, y_4, z_4) = \left(0, -\sqrt{2}, \frac{1}{2}\right).
\end{aligned} \tag{C50}$$

This tetrahedron-shaped ensemble is a second-order ensemble. It is one of the many possible outcomes of the SEIK's sampling (Pham, 2001) but, differently from other second-order ensembles, the specific orientation of this ensemble ~~produce~~ produces a square-shaped projection of its members in the xy-plane. In fact, this ensemble is an extension to three dimensions of the third-order ensemble (C22), which has members with the same x and y coordinates.

The moments in the xy-plane are already checked for ensemble (C22). Here it remains to check moments involving z.

The ensemble first moment associated to z (i.e., the mean along z) is:

$$\frac{1}{4}(z_1 + z_2 + z_3 + z_4) = \frac{1}{4}\left(\left(-\frac{1}{2}\right) + \left(-\frac{1}{2}\right) + \frac{1}{2} + \frac{1}{2}\right) = 0. \tag{C51}$$

The ensemble second moment associated to z^2 (i.e., the variance along z) is:

$$\frac{1}{4}\left({z_1}^2 + {z_2}^2 + {z_3}^2 + {z_4}^2\right) = \frac{1}{4}\left(\frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4}\right) = 1. \tag{C52}$$

The ensemble second moment associated to xz (i.e., the covariance between x and z) is:

$$\frac{1}{4}(x_1 z_1 + x_2 z_2 + x_3 z_3 + x_4 z_4) = \frac{1}{4}\left(\sqrt{2}\cdot\left(-\frac{1}{2}\right) + \left(-\sqrt{2}\right)\cdot\left(-\frac{1}{2}\right) + 0\cdot\frac{1}{2} + 0\cdot\frac{1}{2}\right) = 0. \tag{C53}$$

The ensemble second moment associated to yz (i.e., the covariance between y and z) is:

$$\frac{1}{4}(y_1 z_1 + y_2 z_2 + y_3 z_3 + y_4 z_4) = \frac{1}{4}\left(0\cdot\left(-\frac{1}{2}\right) + 0\cdot\left(-\frac{1}{2}\right) + \sqrt{2}\cdot\frac{1}{2} + \left(-\sqrt{2}\right)\cdot\frac{1}{2}\right) = 0. \tag{C54}$$

Thus, the ensemble $\boldsymbol{P}_1, \boldsymbol{P}_2, \boldsymbol{P}_3, \boldsymbol{P}_4$ matches the moments of the target pdf up to order 3 in the xy-plane, and up to order 2 elsewhere.

The present ensemble ~~,~~ is represented in the first two panels of Fig. 2.

C3.2 High-order sampling ensemble with 4 weighted members, $h = 5$ along the x dimension and $h = 3$ along the xy-plane

$$\begin{aligned}
\boldsymbol{P}_1: &\quad (x_1, y_1, z_1) = \left(\sqrt{3}, 0, -\sqrt{2}\right), \qquad w_1 = \frac{1}{6};\\
\boldsymbol{P}_2: &\quad (x_2, y_2, z_2) = \left(-\sqrt{3}, 0, -\sqrt{2}\right), \qquad w_2 = \frac{1}{6};\\
\boldsymbol{P}_3: &\quad (x_3, y_3, z_3) = \left(0, \sqrt{\frac{3}{2}}, \frac{\sqrt{2}}{2}\right), \qquad w_3 = \frac{1}{3};\\
\boldsymbol{P}_4: &\quad (x_4, y_4, z_4) = \left(0, -\sqrt{\frac{3}{2}}, \frac{\sqrt{2}}{2}\right), \qquad w_4 = \frac{1}{3}.
\end{aligned} \tag{C55}$$

This ensemble extends ensemble (C32) to three dimensions. In this way, it keeps the third-order in the xy-plane and the fifth-order along the x-axis. It only remains to prove the matching of moments involving z.

The ensemble first moment associated to z (i.e., the mean along z) is:

$$w_1 z_1 + w_2 z_2 + w_3 z_3 + w_4 z_4 = \frac{1}{6}\cdot\left(-\sqrt{2}\right) + \frac{1}{6}\cdot\left(-\sqrt{2}\right) + \frac{1}{3}\cdot\frac{\sqrt{2}}{2} + \frac{1}{3}\cdot\frac{\sqrt{2}}{2} = 0. \tag{C56}$$

The ensemble second moment associated to z^2 (i.e., the variance along z) is:

$$w_1 {z_1}^2 + w_2 {z_2}^2 + w_3 {z_3}^2 + w_4 {z_4}^2 = \frac{1}{6}\cdot 2 + \frac{1}{6}\cdot 2 + \frac{1}{3}\cdot\frac{1}{2} + \frac{1}{3}\cdot\frac{1}{2} = 1. \tag{C57}$$

The ensemble second moment associated to xz (i.e., the covariance between x and z) is:

$$w_1 x_1 z_1 + w_2 x_2 z_2 + w_3 x_3 z_3 + w_4 x_4 z_4 = \frac{1}{6}\cdot\sqrt{3}\cdot\left(-\sqrt{2}\right) + \frac{1}{6}\cdot\left(-\sqrt{3}\right)\cdot\left(-\sqrt{2}\right) + \frac{1}{3}\cdot 0\cdot\frac{\sqrt{2}}{2} + \frac{1}{3}\cdot 0\cdot\frac{\sqrt{2}}{2} = 0. \tag{C58}$$

The ensemble second moment associated to yz (i.e., the covariance between y and z) is:

$$w_1 y_1 z_1 + w_2 y_2 z_2 + w_3 y_3 z_3 + w_4 y_4 z_4 = \frac{1}{6}\cdot 0\cdot\left(-\sqrt{2}\right) + \frac{1}{6}\cdot 0\cdot\left(-\sqrt{2}\right) + \frac{1}{3}\cdot\sqrt{\frac{3}{2}}\cdot\frac{\sqrt{2}}{2} + \frac{1}{3}\cdot\left(-\sqrt{\frac{3}{2}}\right)\cdot\frac{\sqrt{2}}{2} = 0. \tag{C59}$$

Thus, the ensemble $\boldsymbol{P}_1, \boldsymbol{P}_2, \boldsymbol{P}_3, \boldsymbol{P}_4$ matches the moments of the target pdf up to order 5 along the x-axis, up to order 3 in the xy-plane, and up to order 2 elsewhere.

The present ensemble ~~,~~is represented (not in scale) in Fig. 2.

C4 Arbitrary number r of dimensions: third-order ensemble ($h=3$) with $2r$ members

$$
\begin{array}{llllllllllll}
\boldsymbol{P}_1: & (x_{1,1}, & x_{2,1}, & \dots, & x_{r,1}) & = & (& \sqrt{r}, & 0, & 0, & \dots, & 0); \\
\boldsymbol{P}_{-1}: & (x_{1,-1}, & x_{2,-1}, & \dots, & x_{r,-1}) & = & (& -\sqrt{r}, & 0, & 0, & \dots, & 0); \\
\boldsymbol{P}_2: & (x_{1,2}, & x_{2,2}, & \dots, & x_{r,2}) & = & (& 0, & \sqrt{r}, & 0, & \dots, & 0); \\
\boldsymbol{P}_{-2}: & (x_{1,-2}, & x_{2,-2}, & \dots, & x_{r,-2}) & = & (& 0, & -\sqrt{r}, & 0, & \dots, & 0); \\
\vdots & & & & \vdots & & & & & & & \\
\boldsymbol{P}_r: & (x_{1,r}, & x_{2,r}, & \dots, & x_{r,r}) & = & (& 0, & 0, & \dots, & 0, & \sqrt{r}); \\
\boldsymbol{P}_{-r}: & (x_{1,-r}, & x_{2,-r}, & \dots, & x_{r,-r}) & = & (& 0, & 0, & \dots, & 0, & -\sqrt{r}).
\end{array}
\tag{C60}
$$

This symmetric ensemble has $2r$ members. Thanks to the result presented in Section C2.6, the odd moments are zero. Also the covariances between different variables must be zero, since every ensemble member has only one non-zero ~~coordinate~~entry. Finally, it only remains to prove that variances are equal to 1.

The ensemble second moment associated to the i^{th} ~~coordinate~~ dimension (i.e., the variance along x_i) is:

$$
\frac{1}{2r}\left({x_{i,1}}^2+{x_{i,-1}}^2+{x_{i,2}}^2+{x_{i,-2}}^2+\ldots+{x_{i,i}}^2+{x_{i,-i}}^2+\ldots+{x_{i,r}}^2+{x_{i,-r}}^2\right)=\frac{1}{2r}\left(0+r+r+0\right)=1. \tag{C61}
$$

Thus, the ensemble $\boldsymbol{P}_1, \boldsymbol{P}_{-1}, \boldsymbol{P}_2, \boldsymbol{P}_{-2}, \ldots, \boldsymbol{P}_r, \boldsymbol{P}_{-r}$ matches the moments of the target pdf (i.e., the r-dimensional standard normal distribution $\mathcal{N}(\boldsymbol{x}; \mathbf{0}, \mathbf{I}_r)$) up to order 3.

Author contributions. GC, S Salon and S Spada designed the study. S Spada developed the algorithms, supervised by SM. S Spada, AT and GC designed the experiments. S Spada implemented the code and performed the simulations. S Spada wrote the manuscript, with the contribution of AT, GC and CS. All the authors participated in acquiring the funding for the project.

Competing interests. The authors declare no competing interests.

Acknowledgements. We acknowledge the CINECA award under the ISCRA initiative and the HPC-TRES program, for the availability of computing resources.
The research reported in this work was supported by OGS and by the SEAMLESS project (https://www.seamlessproject.org/).
This work was partly funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004032 (project SEAMLESS).

References

Ambadan, J. T. and Tang, Y.: Sigma-Point Kalman Filter Data Assimilation Methods for Strongly Nonlinear Systems, Journal of the Atmospheric Sciences, 66, 261 – 285, https://doi.org/10.1175/2008JAS2681.1, 2009.

Anderson, J. L.: An adaptive covariance inflation error correction algorithm for ensemble filters, Tellus A: Dynamic Meteorology and Oceanography, https://doi.org/10.1111/j.1600-0870.2006.00216.x, 2007.

Bannister, R.: A review of operational methods of variational and ensemble-variational data assimilation, Quarterly Journal of the Royal Meteorological Society, 143, 607–633, https://doi.org/10.1002/qj.2982, 2017.

Bannister, R. N.: A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance statistics, Quarterly Journal of the Royal Meteorological Society, 134, 1971–1996, https://doi.org/10.1002/qj.340, _eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.340, 2008.

Bishop, C. H., Etherton, B. J., and Majumdar, S. J.: Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects, Monthly Weather Review, 129, 420 – 436, https://doi.org/https://doi.org/10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2, 2001.

Bocquet, M., Raanes, P. N., and Hannart, A.: Expanding the validity of the ensemble Kalman filter without the intrinsic need for inflation, Nonlinear Processes in Geophysics, 22, 645–662, https://doi.org/10.5194/npg-22-645-2015, 2015.

Brajard, J., Carrassi, A., Bocquet, M., and Bertino, L.: Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations: A case study with the Lorenz 96 model, Journal of Computational Science, 44, 101 171, https://doi.org/https://doi.org/10.1016/j.jocs.2020.101171, 2020.

Carrassi, A., Bocquet, M., Bertino, L., and Evensen, G.: Data assimilation in the geosciences: An overview of methods, issues, and perspectives, Wires climate change, 9, 2018.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, Journal of Geophysical Research: Oceans, 99, 10 143–10 162, https://doi.org/10.1029/94JC00572, 1994.

Fertig, E. J., Harlim, J., and Hunt, B. R.: A comparative study of 4D-VAR and a 4D Ensemble Kalman Filter: perfect model simulations with Lorenz-96, Tellus A: Dynamic Meteorology and Oceanography, https://doi.org/10.1111/j.1600-0870.2006.00205.x, 2007.

Gaspari, G. and Cohn, S. E.: Construction of correlation functions in two and three dimensions, Quarterly Journal of the Royal Meteorological Society, 125, 723–757, https://doi.org/https://doi.org/10.1002/qj.49712555417, 1999.

Gharamti, M. E.: Enhanced Adaptive Inflation Algorithm for Ensemble Filters, Monthly Weather Review, 146, 623 – 640, https://doi.org/https://doi.org/10.1175/MWR-D-17-0187.1, 2018.

Grooms, I.: A comparison of nonlinear extensions to the ensemble Kalman filter, Computational Geosciences, 26, 1–18, https://doi.org/10.1007/s10596-022-10141-x, 2022.

Grooms, I. and Robinson, G.: A hybrid particle-ensemble Kalman filter for problems with medium nonlinearity, PLOS ONE, 16, 1–20, https://doi.org/10.1371/journal.pone.0248266, 2021.

Hamill, T. M. and Snyder, C.: A Hybrid Ensemble Kalman Filter–3D Variational Analysis Scheme, Monthly Weather Review, 128, 2905 – 2919, https://doi.org/10.1175/1520-0493(2000)128<2905:AHEKFV>2.0.CO;2, 2000.

Hodyss, D.: Ensemble State Estimation for Nonlinear Systems Using Polynomial Expansions in the Innovation, Monthly Weather Review, 139, 3571 – 3588, https://doi.org/10.1175/2011MWR3558.1, 2011.

Hoteit, I., Pham, D.-T., Triantafyllou, G., and Korres, G.: Particle Kalman Filtering for Data Assimilation in Meteorology and Oceanography, in: 3rd WCRP International Conference on Reanalysis, pp. 1–6, Tokyo, Japan, https://hal.science/hal-00853919, 2008.

Houtekamer, P. L. and Zhang, F.: Review of the Ensemble Kalman Filter for Atmospheric Data Assimilation, Monthly Weather Review, 144, 4489 – 4532, https://doi.org/https://doi.org/10.1175/MWR-D-15-0440.1, 2016.

Ito, K. and Xiong, K.: Gaussian filters for nonlinear filtering problems, IEEE Transactions on Automatic Control, 45, 910–927, https://doi.org/10.1109/9.855552, 2000.

Janjic, T., Nerger, L., Albertella, A., Schroter, J., and Skachko, S.: On Domain Localization in Ensemble-Based Kalman Filter Algorithms, Monthly Weather Review, 139, 2046–2060, 2011.

Kirchgessner, P., Nerger, L., and Bunse-Gerstner, A.: On the Choice of an Optimal Localization Radius in Ensemble Kalman Filter Methods, Monthly Weather Review, 142, 2165 – 2175, https://doi.org/10.1175/MWR-D-13-00246.1, 2014.

Lahoz, W. A. and Schneider, P.: Data assimilation: making sense of Earth Observation, Frontiers in Environmental Science, 2, https://doi.org/10.3389/fenvs.2014.00016, 2014.

Lei, J. and Bickel, P.: A Moment Matching Ensemble Filter for Nonlinear Non-Gaussian Data Assimilation, Monthly Weather Review, 139, 3964 – 3973, https://doi.org/10.1175/2011MWR3553.1, 2011.

Lorenz, E. N.: Predictability: A problem partly solved, in: Proc. Seminar on predictability, vol. 1, Reading, 1996.

Luo, X. and Moroz, I.: Ensemble Kalman filter with the unscented transform, Physica D: Nonlinear Phenomena, 238, 549–562, https://doi.org/https://doi.org/10.1016/j.physd.2008.12.003, 2009.

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 A.T., W.: Status and future of data assimilation in operational oceanography, Journal of Operational Oceanography, 8, s28–s48, https://doi.org/10.1080/1755876X.2015.1022055, 2015.

Mastroianni, G. and Milovanovic, G. V.: Interpolation Processes: Basic Theory and Applications, Springer Publishing Company, Incorporated, 1 edn., 2008.

Moore, A. M., Martin, M. J., Akella, S., Arango, H. G., Balmaseda, M., Bertino, L., Ciavatta, S., Cornuelle, B., Cummings, J., Frolov, S., Lermusiaux, P., Oddo, P., Oke, P. R., Storto, A., Teruzzi, A., Vidard, A., and Weaver, A. T.: Synthesis of Ocean Observations Using Data Assimilation for Operational, Real-Time and Reanalysis Systems: A More Complete Picture of the State of the Ocean, Frontiers in Marine Science, 6, https://doi.org/10.3389/fmars.2019.00090, 2019.

Nerger, L.: Data assimilation for nonlinear systems with a hybrid nonlinear Kalman ensemble transform filter, Quarterly Journal of the Royal Meteorological Society, 148, 620–640, https://doi.org/https://doi.org/10.1002/qj.4221, 2022.

Nerger, L. and Gregg, W.: Assimilation of SeaWiFS data into a global ocean-biogeochemical model using a local SEIK filter, Journal of Marine Systems, 68, 237–254, https://doi.org/10.1016/j.jmarsys.2006.11.009, 2007.

Nerger, L., Hiller, W., and Schröter, J.: A comparison of error subspace Kalman filters, Tellus A, 57, 715–735, https://doi.org/https://doi.org/10.1111/j.1600-0870.2005.00141.x, 2005.

Nerger, L., Janjic Pfander, T., Schröter, J., and Hiller, W.: A Unification of Ensemble Square Root Kalman Filters, Monthly Weather Review, 140, 2335–2345, https://doi.org/10.1175/MWR-D-11-00102.1, 2012.

Pham, D. T.: Stochastic Methods for Sequential Data Assimilation in Strongly Nonlinear Systems, Monthly Weather Review, 129, 1194 – 1207, https://doi.org/10.1175/1520-0493(2001)129<1194:SMFSDA>2.0.CO;2, 2001.

Pham, D. T., Verron, J., and Roubaud, M. C.: A Singular Evolutive Extended Kalman filter for data assimilation in Oceanography, Journal of Marine Systems, 16, 323–340, 1998.

Raanes, P. N., Bocquet, M., and Carrassi, A.: Adaptive covariance inflation in the ensemble Kalman filter by Gaussian scale mixtures, Quarterly Journal of the Royal Meteorological Society, 145, 53–75, https://doi.org/https://doi.org/10.1002/qj.3386, 2019.

Rainwater, S. and Hunt, B. R.: Ensemble data assimilation with an adjusted forecast spread, Tellus A: Dynamic Meteorology and Oceanography, https://doi.org/10.3402/tellusa.v65i0.19929, 2013.

Roth, M., Hendeby, G., Fritsche, C., and Gustafsson, F.: The Ensemble Kalman Filter: A Signal Processing Perspective, EURASIP Journal on Advances in Signal Processing, 2017, 56, https://doi.org/10.1186/s13634-017-0492-x, 2017.

Salon, S., Cossarini, G., Bolzon, G., Feudale, L., Lazzari, P., Teruzzi, A., Solidoro, C., and Crise, A.: Novel metrics based on Biogeochemical Argo data to improve the model uncertainty evaluation of the CMEMS Mediterranean marine ecosystem forecasts, Ocean Science, 15, 997–1022, https://doi.org/https://doi.org/10.5194/os-15-997-2019, publisher: Copernicus GmbH, 2019.

Scheffler, G., Carrassi, A., Ruiz, J., and Pulido, M.: Dynamical effects of inflation in ensemble-based data assimilation under the presence of model error, Quarterly Journal of the Royal Meteorological Society, 148, 2368–2383, https://doi.org/https://doi.org/10.1002/qj.4307, 2022.

Spada, S., Teruzzi, A., Maset, S., Salon, S., Solidoro, C., and Cossarini, G.: GHOSH v1.0.0: a novel Gauss-Hermite High-Order Sampling Hybrid ensemble filter for computationally efficient data assimilation in geosciences - Part 2, https://doi.org/10.5194/gmd-2023-170, preprint version, 2024.

Stordal, A., Karlsen, H., Nævdal, G., Skaug, H., and Vallès, B.: Bridging the ensemble Kalman filter and particle filters: The adaptive Gaussian mixture filter, Computational Geosciences, 15, https://doi.org/10.1007/s10596-010-9207-1, 2011.

Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M., and Whitaker, J. S.: Ensemble Square Root Filters, Monthly Weather Review, 131, 1485 – 1490, https://doi.org/10.1175/1520-0493(2003)131<1485:ESRF>2.0.CO;2, 2003.

Triantafyllou, G., Hoteit, I., and Petihakisa, G.: A singular evolutive interpolated Kalman filter for efficient data assimilation in a 3-D complex physical-biogeochemical model of the Cretan Sea, Journal of Marine Systems, 40–41, 213–231, 2003.

Tödter, J. and Ahrens, B.: A Second-Order Exact Ensemble Square Root Filter for Nonlinear Data Assimilation, Tödter and Ahrens, 140, 1347–1369, 2015.

van Leeuwen, P. J., Künsch, H. R., Nerger, L., Potthast, R., and Reich, S.: Particle filters for high-dimensional geoscience applications: A review, Quarterly Journal of the Royal Meteorological Society, 145, 2335–2365, https://doi.org/https://doi.org/10.1002/qj.3551, 2019.

Vetra-Carvalho, S., Van Leeuwen, P. J., Nerger, L., Barth, A., Altaf, M. U., Brasseur, P., Kirchgessner, P., and Beckers, J.-M.: State-of-the-art stochastic data assimilation methods for high-dimensional non-Gaussian problems, Tellus A: Dynamic Meteorology and Oceanography, https://doi.org/10.1080/16000870.2018.1445364, 2018.

Wan, E. and Van Der Merwe, R.: The unscented Kalman filter for nonlinear estimation, in: Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373), pp. 153–158, https://doi.org/10.1109/ASSPCC.2000.882463, 2000.

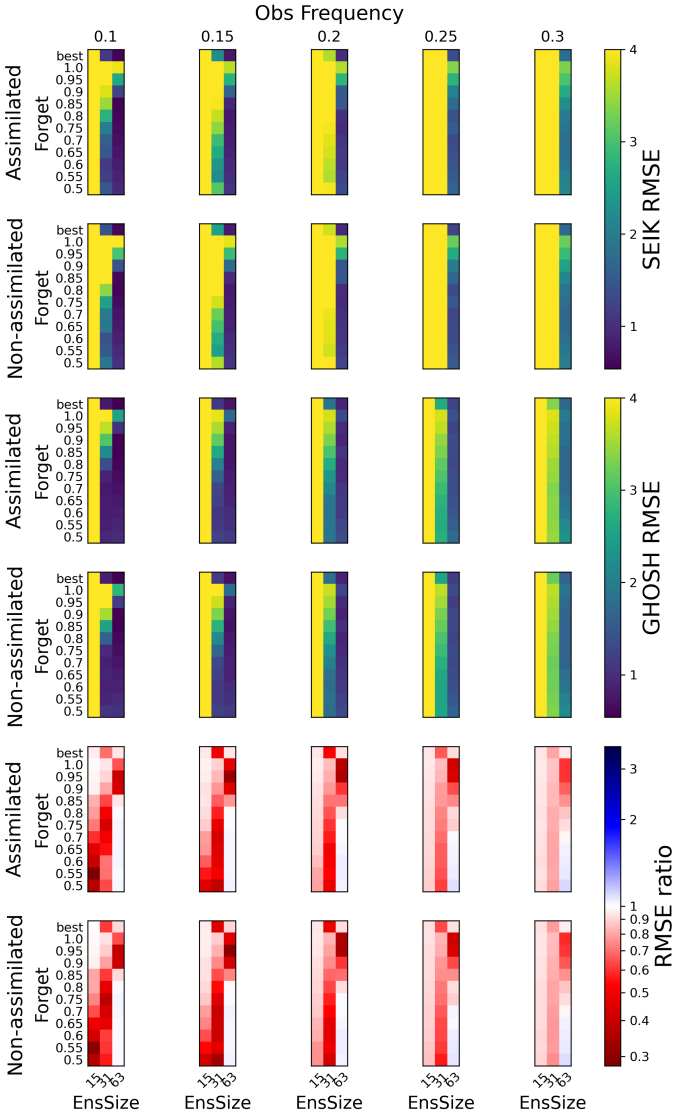

Figure 4. Result summary of 20-~~time-units-long~~ -time-unit-long twin experiments: each square in the colour-maps represents the aggregated results of 400 twin experiments, changing *truth*, observations and initial conditions. The results are summarized with colour-maps aggregated in six different rows, from top to bottom: SEIK RMSE of assimilated variables, SEIK RMSE of non-assimilated variables, GHOSH RMSE of assimilated variables, GHOSH RMSE of non-assimilated variables, the ratio of GHOSH RMSE over SEIK RMSE of assimilated variables, the ratio of GHOSH RMSE over SEIK RMSE of non-assimilated variables (red color implies that GHOSH is better than SEIK). Each column of colour-maps has different observation frequency, with the numbers on the top indicating the time elapsed between each observation/assimilation. Each colour-map shows different forgetting factors (Forget, along the y axis) and ensemble sizes (EnsSize, along the x axis).

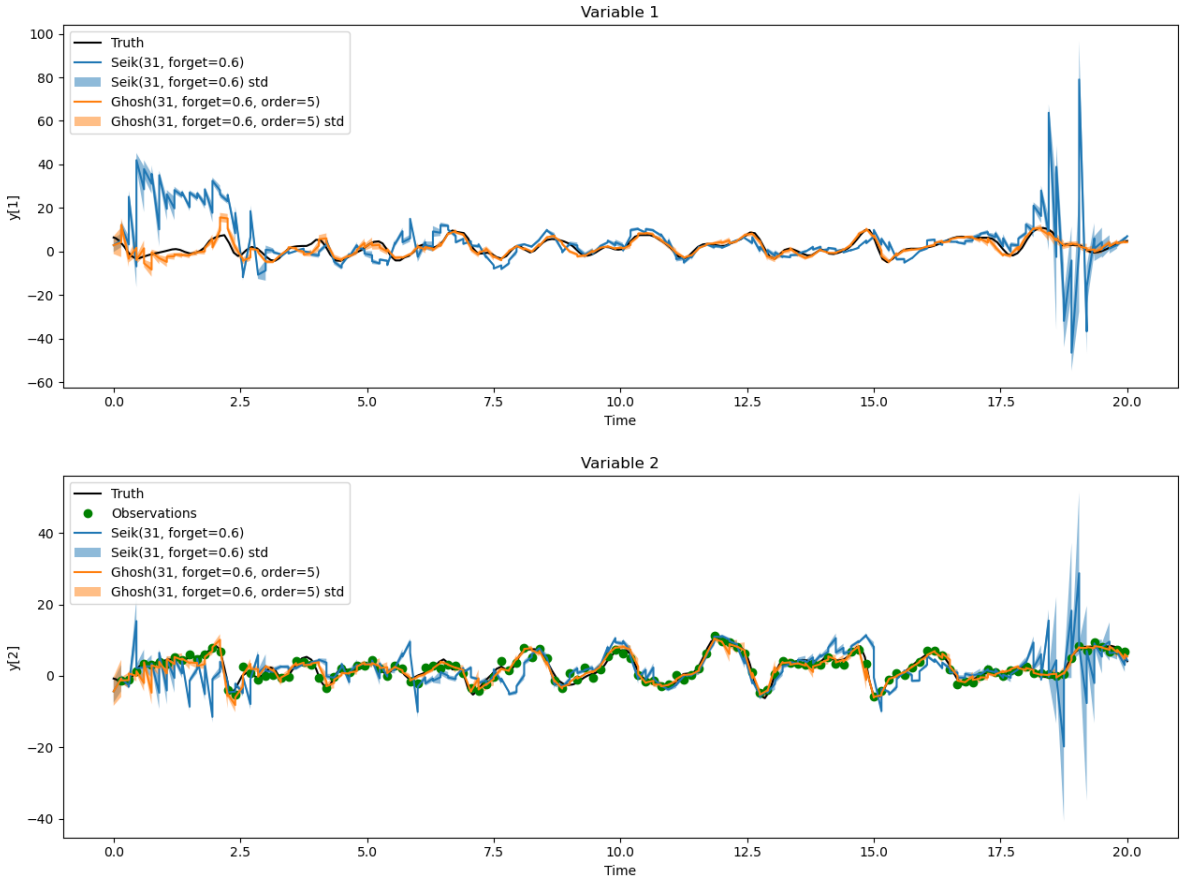

Figure 5. Example of a single twin experiment, with SEIK (blue line) showing diverging behaviour: time between observations (green dots) is 0.15, forgetting factor is 0.6, ensemble size is 31. Values over time of a non-assimilated (top) and an assimilated (bottom) variable.

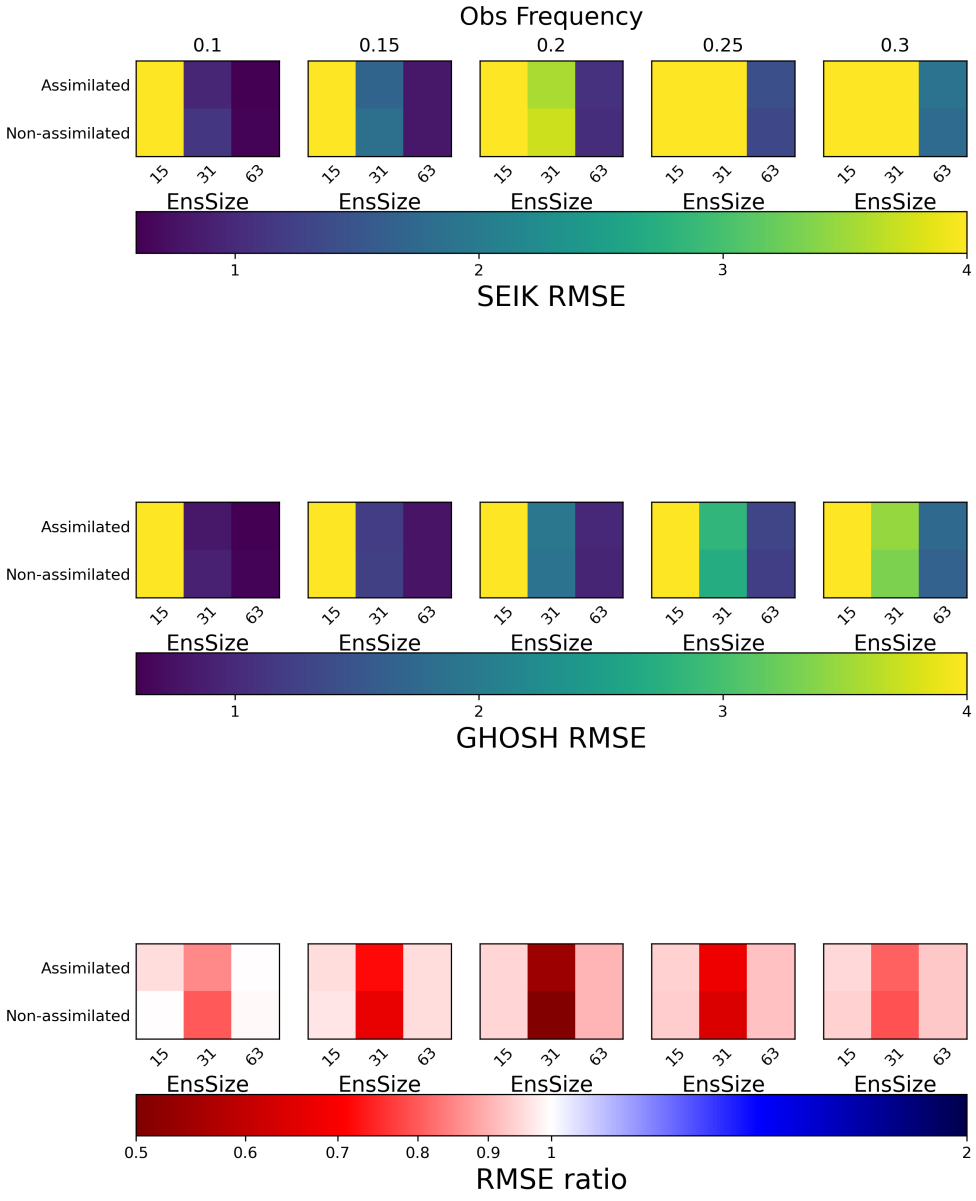

Figure 6. Result summary of 150 ~~time-units-long~~ time-unit-long twin experiments: each square in the colour-maps represents the aggregated results of 100 twin experiments, changing observations and initial conditions. The results are summarized with colour-maps aggregated in 3 different rows, from top to bottom: SEIK RMSE, GHOSH RMSE, ratio of GHOSH RMSE over SEIK RMSE (red color implies that GHOSH is better than SEIK). Each column of colour-maps has different observation frequency, with the numbers on the top indicating the time elapsed between each observation/assimilation. Each colour-map shows, for assimilated and non-assimilated variables, different ensemble sizes (EnsSize, along the x axis).