Quasi static ensemble variational data assimilation

Anthony Fillion[1,2], Marc Bocquet[1], and Serge Gratton[3]

[1]CEREA, Joint Laboratory École des Ponts ParisTech and EDF R&D, Université Paris-Est, Champs-sur-Marne, France
[2]CERFACS, Toulouse, France
[3]INPT-IRIT, Toulouse, France

Correspondence to: Anthony Fillion (anthony.fillion@enpc.fr)

Dear Editor,

Please find in this document:

- A point-by-point response to Referee 1,
- A point-by-point response to Referee 2,
- A manuscript that indicates all the changes and corrections.

The main additions to the manuscript are:

- References to the Sequential QSVA (Pires et al., 1996),
- Discussions about the useful DAW length (Pires et al., 1996),
- Discussions about the imperfect model case,
- Extended comparisons with Goodliff et al. (2015),
- Details about the IEnKS-MDA (Bocquet and Sakov, 2014),
- Extended Figure discussions and colormap change.

Yours sincerely, The authors.

References

Bocquet, M. and Sakov, P.: An iterative ensemble Kalman smoother, Q. J. R. Meteorol. Soc., 140, 1521–1535, doi:10.1002/qj.2236, 2014.

Goodliff, M., Amezcua, J., and Van Leeuwen, P. J.: Comparing hybrid data assimilation methods on the Lorenz 1963 model with increasing non-linearity, Tellus A, 67, 26 928, doi:10.3402/tellusa.v67.26928, 2015.

Pires, C., Vautard, R., and Talagrand, O.: On extending the limits of variational assimilation in nonlinear chaotic systems, Tellus A, 48, 96–121, doi:10.3402/tellusa.v48i1.11634, 1996.

Response to reviews of "Quasi static ensemble variational data assimilation" (ID npg-2017-65) by A. Fillion, M. Bocquet and S. Gratton

A. Fillion, M. Bocquet and S. Gratton

February 20, 2018

Response to Carlos Pires

We are grateful to the Reviewer for his suggestions.

Comments

1. **In section 3 (Quasi Static algorithms) it is worth to mention and to put into context the Sequential Quasi Static Variational Assimilation (section 4.2 of (Pires et al. 1996)) as a variation of the QS scheme**

   Indeed. Thank you for pointing this out. We now refer to this *sequential* QSVA scheme in the revised manuscript. Note that this can be seen as an ancestor of the MDA IEnKS. However, as shown in Bocquet and Sakov (2014), a sequential QSVA cannot be transposed directly into an EnVar scheme without further modification because of the multiple assimilation of the same observations, hence the MDA IEnKS.

2. **In the discussion of upper triangles of Figs. 8 (L95) and 9 (L63), showing the average smoothing and filtering errors, the authors should discuss how far it is useful to increase the DAW length. In Pires et al (1996), it is presented the concept of useful assimilation window -ln(0.01/(2 Lambda-max), beyond which the DA is not useful anymore where Lambda-max is the Largest Lyapunov value. Giving the steps delta-t and lambda-max, the authors may provide the largest useful DAW length Lmax.**

   The idea of useful data assimilation length is a very nice concept introduced by Pires et al. (1996). For both low-order models, one obtains:

$$\begin{aligned} L_{\max}^{\mathrm{L95}} &= \frac{-\ln(0.01)}{2\lambda^{\mathrm{L95}}\Delta t^{\mathrm{L95}}}, \\ &= \frac{-\ln(0.01)}{2\times 1.7\times 0.05}, \\ &\simeq 27. \end{aligned} \qquad \begin{aligned} L_{\max}^{\mathrm{L63}} &= \frac{-\ln(0.01)}{2\lambda^{\mathrm{L63}}\Delta t^{\mathrm{L63}}}, \\ &= \frac{-\ln(0.01)}{2\times 0.91\times 0.02}, \\ &\simeq 127. \end{aligned}$$

   This concept is now recalled (twice) in the revised manuscript with a reference to Pires et al. (1996).

   Note, however, that it has some limitations. First, it applies to the filtering error (at present time), not to the smoothing error – at least not directly. Second, this result does not easily translates to an advanced cycled scheme such as the $\mathrm{IEnKS}_{\mathrm{QS}}$, where a lot of observations have already been assimilated and their information condensed in the background. Thus, the performance gain with the DAW length comes from the precision of this Gaussian background approximation – a precision that the linearized theory is not able to evaluate. We have shown in Bocquet and Sakov (2013, 2014) and in the present manuscript, that one can go very far in the past – well beyond $27\Delta t$ in the L95 case – and still improve the smoothing RMSE. Third, this useful length does not account for nonlinearities, the appearance of local minima, and correlatively potential saturation. As a result there is a somehow arbitrary constant in its definition (0.01 here). In Pires et al. (1996), it is related to the targeted error.

The length that we estimate in Sec. 2.4 can be seen to some degree as an improvement on Pires et al. (1996)' endeavor of a useful length by estimating the constant resorting to saturation and the occurence of local minima in the cost function.

3. **In the discussion and conclusions, the authors should add a small paragraph on the limitations of extending the DAW length in DA with nonperfect models (refer to Swanson et al 1998).**

   Thank you for the suggestion. Extending the DAW length is less relevant for significantly noisy models. Swanson et al. (1998) showed that the perfect model results are expected to extend to the imperfect model case provided that the growth rate of the model error is similar to that of the leading Lyapunov vectors of the system. This is discussed in the revised manuscript at the end of the conclusion and a reference to Swanson et al. (1998) has been added.

Response to referee 2

We thank the Reviewer for the questions and comments. There are a few points on which we partly or totally disagree and we justify why.

General comments

1. **The paper is restricted to the case of low order models with perfect model assumption. This limitation seems restrictive and this clearly diminishes the scope of the results. Indeed, the original papers on quasi-static variational assimilation (QSVA) were at least partly addressing the case of higher dimensions and model error (e.g., Swanson et al. 1998).**

   We disagree in two ways:

   (a) First, two thirds of the paper are on the theory of QSVA in an EnVar context, whose scope is broad, and significantly larger than numerically testing QSVA with EnVar/hybrid methods as in Bocquet and Sakov (2013, 2014); Goodliff et al. (2015). This is independent from the dimension of the problem, though it depends on the perfect-model assumption.

   (b) Secondly, the reviewer seems to assume that data assimilation methods based on perfect model assumptions cannot be applied to imperfect models. This is clearly not the case since strong-constraint 4D-Var has been applied in operational meteorological forecast for 20 years to imperfect models. Hence, of course, the algorithm proposed in this manuscript can be applied to imperfect models as well, with limitations that have been discussed in Swanson et al. (1998). Although it is important to mention this point, we consider it a rather distinct subject from our endeavor in the theory part of this manuscript.

   Note, that a mathematically consistent variant of the IEnKF/IEnKS with additive model error has been recently designed and tested (Sakov and Bocquet, 2018; Sakov et al., 2018), so that we could contemplate in a near future an extension of the present study to an IEnKS where model error is properly accounted for.

2. **The application of QSVA together with ensemble formulation has received some interest in the community (e.g., Goodlif et al. 2015) but the paper is in my opinion missing a discussion on how the results compare with the one of Goodlif et al. 2015.**

   To our knowledge, the first quasi-static algorithm in an EnVar/hybrid context has been proposed and tested in Bocquet and Sakov (2013, 2014), specifically the MDA IEnKS scheme. Another attempt came from M. Jardak and O. Talagrand at about the same time but reported in conferences, and it was only concerned with 4D-Var as it was applied to a *non-cycled* EDA scheme. The interested reader can have a look at their very recent 2018 submission in Nonlinear Processes in Geophysics.

   Goodlif et al. 2015 provides a numerical exploration of the impact of flow dependent background covariances and QS minimizations on the performance of hybrid schemes with the L63 model. QSVA is merely used as a tool following Pires et al. (1996). This impact of QSVA is just established on numerical experiments, which confirm the findings of Pires et al. (1996), or those of Bocquet and Sakov (2013, 2014) with the MDA IEnKS. It does not seem that there is much more to mention, as far as QSVA is concerned.

Here, by contrast, our goal is to justify theoretically and give insights about QSVA in the context of cycled EnVar data assimilation. This is later illustrated by algorithms and numerics.

A more thourough (but not really necessary in our opinion) would be for instance to compare our L63 numerical results with those of Goodliff et al. (2015):

(a) First, the critical and interesting Sec. 3.6 of Goodliff et al. (2015) is not sufficiently documented. For instance, we do not know if the algorithms are cycled or just averages over several instances. The definition of their RMSE Eq. (29) does tell the reader how the average is actually done (something must be missing in the definition), and it mixes filtering and smoothing RMSEs which makes any interpretation more difficult.

(b) Second, Goodliff et al. (2015) showed that the ETKS outperforms all the schemes in their study. Since the IEnKS systematically outperforms the ETKS in all conditions (and in particular L63) as long as the DAW length is not overwhelmingly long (for a chaotic model), then one concludes that our RMSEs would be systematically equal or smaller that those reported for any hybrid scheme in Goodliff et al. (2015).

We have increased the discussion/comparison on Bocquet and Sakov (2013, 2014); Goodliff et al. (2015), and made a more detailed reference to those in the introduction of the revised manuscript.

3. **The paper is sometimes hard to follow : in a first section, theoretical developments are used to compare the performance of 4D-Var and of the IEnKS in a linear and highly simplified context. This is interesting but then 4D-Var is dropped out of the DA schemes that are considered and it is not obvious why. The limitations of the standard IEnKS with increasing DA windows are well illustrated and lead to section 3 with quasi-static versions of the IEnKS compared to standard ones. Here, a novel algorithm is discussed, the MDA. I would recommend to focus on IEnKS only; dropping the 4D-Var and the MDA versions to make the paper more focused.**

   The discussion about 4D-Var is here to illustrate the impact of an improper modeling of the prior pdf in a simplified linear context. The analogy with the improper prior modeling of the IEnKS in a non-linear context becomes then clearer. The 4D-Var is dropped in the numerical experiments because the proof that ensemble variational methods are numerically more efficient than variational methods has already been established (Bocquet and Sakov, 2013).

   That is why 4D-Var is replaced by another quasi-static ensemble variational method: the MDA IEnKS. This is not a novel method (Bocquet and Sakov, 2013, 2014), and it is the first documented quasi-static EnVar method (with $S = 1$ at least). Note also that the question of how long the data assimilation should or could be in an EnVar context that we addressed in this paper was first formulated in Bocquet and Sakov (2014) and discussed in their conclusion as an open question.

   Hence, we are not convinced that the manuscript would benefit from your present suggestions.

4. **Figures are generally clear, with the exception of Fig. 10 and 11 where the third panel (about the number of ensemble propagations) is put on the same level as he two other ones (RMSE) which is confusing at first glance. There is yet a general problem with the colours that do not render well in gray scale and thus likely confusing for colour blinded people : the authors may consider using better colour maps for this purpose.**

   Thank you very much for the suggestion. In the revised manuscript, we choose to use a colormap that renders properly in grayscale. To keep color variability with small RMSEs, values beyond a certain limit have the same color. Also, each axis of Fig. 10, 11 has its own title to avoid confusion.

Specific comments

1. **I would recommend that the paper is more clear about the limitations of the study. It is definitively in the text but not in the title and in the abstract. I would mention the perfect model assumption in the abstract. Also, the title is too general. I would make it more specific, for instance Performances of the quasi-static formulation of the iterative Kalman Smoother on low-order models, or A quasi-static version of the strong constraint iterative Kalman Smoother for instance**

As we explained, we do not believe that the findings of this paper are as limited as you claim they are. That said, we can certainly mention the perfect model assumption in the abstract. We did so in the revised manuscript.

We believe our title was not too general. But it can surely help the reader to make it more focused. The titles that you propose do not reflect the generality of our findings. Indeed, the IEnKS is the archetype of a deterministic EnVar method and we use it as such in this manuscript (as derived in Bocquet and Sakov, 2014). We expect any good (or close to optimal) EnVar method to reach the same conclusions.

We believe "Quasi-static ensemble variational data assimilation: a theoretical and numerical study with the iterative ensemble Kalman smoother" now perfectly reflects the content of the manuscript.

2. **Page 2, line 18. Is it a known fact that the number of local minima increases exponentially with the data assimilation window ? If yes, please provide a proof or quote, if not please be more vague**

   This statement comes from Swanson et al. (1998) p.377 and is justified by Pires et al. (1996) p.106. A "may" mitigates the statement, since this may have only been proven for emblematic chaotic model (such as the baker map). These references have been added at this point in the revised manuscript. Thank you for this clarification enquiry.

3. **Page 2, lines 20-25. There are other methods that address the convergence of minimization despite the non-linearity of the operators by using globalization methods, even published by the authors (e.g. Preconditioning and globalizing conjugate gradients in dual space for quadratically penalized nonlinear-least squares problems by Gratton et. al.). Please add and comment references with alternative minimization algorithms to address non-linearity.**

   Please check your definition of *globalization methods*. They do not aim at finding the global minimum but are meant to obtain convergence of the iterates for every initial guess. About finding the global minimum, we gave references to Ye et al. (2015); Judd et al. (2004) which are the only one we can think of in the geophysical data assimilation context (besides QSVA).

   The reason why methods looking for a global minimum are seldomly used in data assimilation was given p.2 line 12.

4. **Page 2, last paragraph. You mention that your paper is designed to be a more complete analytical and numerical investigation, but you do not comment on the main results of the paper you are citing. Please provide a better discussion of your paper with the existing literature.**

   The existing literature as far as quasi-static hybrid/EnVar methods are concerned is Bocquet and Sakov (2013, 2014); Goodliff et al. (2015). A discussion is given in our response to question 2 of the general comments, and to some extent included in the revised manuscript.

5. **Page 3, line 19. Your paper is about low dimensional and perfect model, such that I would change the sentence to not meant to improve high-dimensional nor imperfect models.**

   The sentence has been corrected. Thank you for spotting this mistake.

6. **Page 6, line 16 : please detail in which sense the inverse square root of the matrix is taken , as it is ambiguous.**

   We mention in the revised manuscript that it is the unique symmetric definite positive square root matrix of a symmetric definite positive matrix (which is by far the most common definition).

7. **Page 11, line 5 : I do not understand the qualitative explanation that is given, please reformulate**

   The explanation means that to assimilate the same number of observations, an algorithm using a greater value for the DAW parameter S need less cycles. Because a cost function approximation is made on the background term each cycle, this algorithm relies less often on this approximation making the analysis more accurate. The sentence has been reformulated in the revised manuscript. Thank you.

8. **Page 20 : the description of Figs. 8, 9 and 10 is very short, with only a few lines to comment 10 panels. Please consider discussing more the results or simplifying the figures by showing only what you tell.**

   We fully agree that the discussion was too short. Thank you for pointing out this weakness.

   A description of the performance variation with the DAW parameters has been added about Figs. 8, 9. Then the IEnKS$_{\text{QS}}$ filtering RMSE invariance with L is discussed and compared to the 4D-Var filtering performance in a linear context. A missing discussion of the performance with L63 has been added as well, about Fig. 10.

9. **Page 23 and 24 : I do not understand what the number of ensemble propagation is, and the paper is missing an explanation of why in Fig 11 we observe different behaviors between L63 and L95, and also why it has non-monotonic evolution with parameter NQ.**

   The number of ensemble propagations is the total number of times an ensemble is propagated with a time step of Δt in the future, divided by the total number of observations assimilated. The similarity with an Heaviside function comes from the 2 main regimes for the RMSE. When N_Q is too small the methods does not locate the global minimum and the RMSE is close to the climatological variance. When N_Q is sufficiently high, the method locates the global minimum and the RMSE is low. The difference of number of ensemble propagation between the L63 model and the L95 model comes from the minimization. When it misses the global minimum, it does not converge with L95 leading to a large number of iteration and ensemble propagations. It converges with L63 but to a local extremum leading to few iterations and few ensemble propagations.

   This discussion has been added to the revised manuscript. Thank you for pointing out to this weakness in the original manuscript.

References

Bocquet, M., Sakov, P., 2013. Joint state and parameter estimation with an iterative ensemble Kalman smoother. Nonlin. Processes Geophys. 20, 803–818. doi:10.5194/npg-20-803-2013.

Bocquet, M., Sakov, P., 2014. An iterative ensemble kalman smoother. Q. J. R. Meteorol. Soc. 140, 1521–1535. doi:10.1002/qj.2236.

Goodliff, M., Amezcua, J., Van Leeuwen, P.J., 2015. Comparing hybrid data assimilation methods on the lorenz 1963 model with increasing non-linearity. Tellus A 67, 26928. doi:10.3402/tellusa.v67.26928.

Judd, K., Smith, L., Weisheimer, A., 2004. Gradient free descent: shadowing, and state estimation using limited derivative information. Physica D 190, 153–166. doi:http://doi.org/10.1016/j.physd.2003.10.011.

Pires, C., Vautard, R., Talagrand, O., 1996. On extending the limits of variational assimilation in nonlinear chaotic systems. Tellus A 48, 96–121. doi:10.3402/tellusa.v48i1.11634.

Sakov, P., Bocquet, M., 2018. Asynchronous data assimilation with the EnKF in presence of additive model error. Tellus A 70, 1414545. doi:10.1080/16000870.2017.1414545.

Sakov, P., Haussaire, J.M., Bocquet, M., 2018. An iterative ensemble Kalman filter in presence of additive model error. Q. J. R. Meteorol. Soc. 0, 0–0. doi:10.1002/qj.3213. accepted for publication.

Swanson, K., Vautard, R., Pires, C., 1998. Four-dimensional variational assimilation and predictability in a quasi-geostrophic model. Tellus A 50, 369–390. doi:10.1034/j.1600-0870.1998.t01-4-00001.x.

Ye, J., Rey, D., Kadakia, N., Eldridge, M., Morone, U., Rozdeba, P., Abarbanel, H., Quinn, J., 2015. Systematic variational method for statistical nonlinear state and parameter estimation. Phys. Rev. E 92, 052901. doi:10.1103/PhysRevE.92.052901.

Quasi static ensemble variational data assimilation: a theoretical and numerical study with the iterative ensemble Kalman smoother

Anthony Fillion[1,2], Marc Bocquet[1], and Serge Gratton[3]

[1]CEREA, Joint Laboratory École des Ponts ParisTech and EDF R&D, Université Paris-Est, Champs-sur-Marne, France
[2]CERFACS, Toulouse, France
[3]INPT-IRIT, Toulouse, France

Correspondence to: Anthony Fillion (anthony.fillion@enpc.fr)

Abstract. The analysis in nonlinear variational data assimilation is the solution of a non-quadratic minimization. Thus, the analysis efficiency relies on its ability to locate a global minimum of the cost function. If this minimization uses a Gauss-Newton (GN) method, it is critical for the starting point to be in the attraction basin of a global minimum. Otherwise the method may converge to a *local* extremum, which degrades the analysis. With chaotic models, the number of local extrema often increases with the temporal extent of the data assimilation window, making the former condition harder to satisfy. This is unfortunate because the assimilation performance also increases with this temporal extent. However, a quasi-static (QS) minimization may overcome these local extrema. It consists in gradually injecting the observations in the cost function. This method was introduced by Pires et al. (1996) in a 4D-Var context.

We generalize this approach to four-dimensional ~~nonlinear EnVar~~ strong-constraint nonlinear ensemble variational (EnVar) methods, which are based on both a nonlinear variational analysis and the propagation of dynamical error statistics via an ensemble. This forces to consider the cost function minimizations in the broader context of cycled data assimilation algorithms. We adapt this QS approach to the iterative ensemble Kalman smoother (IEnKS), an exemplar of nonlinear deterministic ~~4D~~ four-dimensional EnVar methods. Using low-order models, we quantify the positive impact of the QS approach on the IEnKS, especially for long data assimilation windows. We also examine the computational cost of QS implementations and suggest cheaper algorithms.

1 Introduction

1.1 Context

Data assimilation (DA) aims at gathering knowledge about the state of a system from acquired observations. In the Bayesian framework, this knowledge ~~can be~~ is represented by the posterior probability density function (pdf) of the system state given the observations. A specificity of sequential DA is that observations are not directly available; they become available as time goes by. Thus, the posterior pdf should be regularly updated.

In order to do so, one usually proceeds in two steps: the analysis and the propagation (or forecast). During the analysis step, a background pdf is used as a prior together with the observation likelihood to construct the (often approximate) posterior pdf,

following Bayes' theorem. During the propagation step, this posterior pdf is propagated in time with the model to yield the prior pdf of the next assimilation cycle.

In general these posterior and prior pdfs are not easily computable. In the Kalman filter, assumptions are notably made on the linearity of operators, to keep these pdfs Gaussian. This way, they are characterized by their mean and covariance matrix. Linear algebra is then sufficient to enforce both Bayes' theorem and the propagation step into operations on means and covariances.

However, with nonlinear models, the Kalman filter assumptions do not hold. The posterior and prior pdfs are not Gaussian anymore. A possibility in this case is to enforce Gaussianity with approximations. This requires the selection of mean and covariances intended for the Gaussian surrogate pdfs. With the Kullback-Leibler divergence, the best Gaussian approximation of a pdf is achieved by equating the mean and covariances (see, e.g., Bishop, 2006). However, the integrations necessary to evaluate these moments are also prohibitive.

In the 4D-Var algorithm (see, e.g., Lorenc, 2014, and references therein), Laplace's approximation gives us a way to work around the problem by replacing the posterior mean with the presumed unique global minimizer of the cost function. A model propagation is then sufficient to estimate the prior pdf mean. This approach calls for efficient global optimization routines. However, in practice, solving a global optimization problem is challenging when the number of unknowns is large, and local methods like Gauss-Newton are often preferred in practice (see, e.g., Björck, 1996).

Unfortunately, Gauss-Newton methods' ability to locate the global minimum depends on the minimization starting point and on the cost function properties. Furthermore, missing this global minimum is likely to cause a quick divergence (from the truth) of the sequential DA method. Thus, it is critical for the assimilation algorithm to keep the minimization starting point in a global minimum basin of attraction.

1.2 Quasi-static variational data assimilation

This requirement is constraining because, with a chaotic model, the number of local minima may increase exponentially with the data assimilation window (DAW) time extent L ~~(Miller et al., 1994; Pires et al., 1996)~~(Swanson et al., 1998; Pires et al., 1996). Unfortunately, assuming a perfect, chaotic – and hence unstable – model, this is also for the longest time extents that the assimilation performs best. Several strategies have been investigated to go beyond this restriction. Pires et al. (1996) propose the quasi-static (QS) minimization in a 4D-Var context: as the observations are progressively added to the cost function, the starting point (or first guess) of the 4D-Var minimization is also gradually updated. This lead to the method known as QSVA for quasi-static variational assimilation. Ye et al. (2015) propose to gradually increase the model error covariances in the weak-constraint 4D-Var cost function in a minimization over an entire trajectory; this way the model nonlinearity is gradually introduced into the cost function (see also Judd et al., 2004). They also propose to parallelize this minimization over multiple starting points to increase the chance to locate the global minimum.

On the one hand, ~~the~~ 4D-Var benefits from the QS approach to approximate the posterior and prior means. On the other hand, with traditional 4D-Var, the prior covariance matrix is taken as static. This is appropriate when ~~, as in Pires et al. (1996), Swanson et al. (1998) or Ye et al. (2015),~~ only one cycle of assimilation is considered. But this limits the dynamical transfer of

error statistics from one cycle to the next, for instance when Pires et al. (1996) propose to gradually move their fixed-lag data assimilation window in order to build a sequential QSVA.

~~In contrast,~~

1.3 Ensemble variational methods

By contrast, four-dimensional (4D) ensemble variational (EnVar) schemes allow to both perform a nonlinear variational analysis and a propagation of dynamical errors via the ensemble (see Chapter 7 of Asch et al., 2016). The improvement brought by QS minimizations on these schemes has been suggested and numerically evaluated in ~~Bocquet and Sakov (2013, 2014); Goodliff et al. (2015)~~Bocquet and Sakov (2013, 2014); Goodliff et al. (2015). This motivates a more complete analytical and numerical investigation.

The iterative ensemble Kalman smoother (IEnKS) (Bocquet and Sakov, 2014; Bocquet, 2016) is the archetype of such 4D nonlinear EnVar scheme, where the ensemble parts of the algorithm are deterministic. Using low-order models (usually toy-models), it was shown to significantly outperform 4D-Var, the EnKF or the ensemble Kalman smoother in terms of accuracy.

The IEnKS improves the DA cycling by keeping track of the pdfs mean and covariance matrix. To do this, Laplace's approximation is used to replace the posterior mean and covariance matrix with the minimizer of the cost function and an approximation of the inverse Hessian at the minimizer, respectively. These moments are then used to update the ensemble statistics. The updated ensemble is then propagated to estimate the prior mean and covariance matrix. Hence, it is also critical for the IEnKS to locate the global minimum of the cost function.

Here, we are interested in the application of the QS minimization to the IEnKS. ~~It will be shown that the bigger the DAW, the better the performance as long as the global minimum is found. Because the QS minimization improves this detection, the IEnKS will benefit from it~~One of the variant of the IEnKS called the multiple data assimilation (MDA) IEnKS was shown (Bocquet and Sakov, 2014) to be quasi-static by design and can be seen as the EnVar generalization of the sequential QSVA by Pires et al. (1996). It was first tested in Bocquet and Sakov (2013). However, the MDA IEnKS is a specific variant of the IEnKS whereas we see here the IEnKS as an exemplar of deterministic nonlinear 4D EnVar methods.

Goodliff et al. (2015) have applied QSVA numerically to a collection of hybrid and EnVar techniques on the Lorenz 1963 model (Lorenz, 1963), where they vary the magnitude of nonlinearity. Nonetheless the focus of their study was not cycling and the transfer of information from one cycle to the next, which is critical to EnVar methods. They also showed that the ensemble transform Kalman smoother outperforms all of the benchmarked methods. It turns out that we have shown that this smoother was systematically outperformed by the IEnKS by design, which was also demonstrated on numerics (Bocquet and Sakov, 2013). This strengthens our claim that the IEnKS can be used here as an exemplar of deterministic nonlinear 4D EnVar methods.

1.4 Outline

The rest of the paper is organized as follows. In section 2, the performance dependency of 4D-Var and IEnKS algorithms on the DAW parameters is investigated. The emphasis is on the transfer of information from one cycle to next, which distinguishes

the IEnKS from 4D-Var. In order to do so, a brief presentation of 4D-Var and the IEnKS algorithms is given. Then we define a measure of performance for assimilation algorithms. This definition is used to give analytic expressions for the accuracy of both algorithms with a linear, diagonal, autonomous model. This quantifies the impact of cycling on the algorithms. After these preliminaries, the nonlinear, chaotic case is studied. In section 3, we provide and describe the algorithms of the quasi-static IEnKS (~~IEnKS-QS~~IEnKS$_{QS}$). Section 4 is dedicated to numerical experiments with two Lorenz low-order models and to the improvement of the ~~IEnKS-QS~~ numerical efficiency. Conclusions are given in section 5.

We emphasize that the algorithmic developments of this study are not meant to improve neither high-dimensional ~~,~~ nor imperfect model, data assimilation techniques. Even if Miller et al. (1994) show some similarities between the perfect and imperfect settings of a model of intermediate complexity, model error would generally forbid the use of very long DAWs as sometimes considered in this study. Instead, the objective of this paper is to better understand the interplay between chaotic dynamics, ensemble variational data assimilation schemes and their cycling, irrespective of whether they could be useful in high-dimensional systems.

2 The data assimilation window and the assimilation performance

After reviewing 4D-Var (in a constant background matrix version) and the IEnKS algorithms, we will investigate the dependency of assimilation performance on the DAW key parameters. This will illustrate the cycling improvement brought in by the IEnKS compared to 4D-Var. We shall see that, with chaotic models, the longer the DAW is, the better the accuracy of these algorithms. Which highlights the QSVA relevance for cycled data assimilation.

The evolution and observation equations of the system are assumed of the form:

$$\mathbf{y}_l = \mathcal{H}(\mathbf{x}_l) + \boldsymbol{\varepsilon}_l, \tag{1a}$$

$$\mathbf{x}_{l+1} = \mathcal{M}(\mathbf{x}_l), \tag{1b}$$

where the unknown state $\mathbf{x}_l$ at time t_l is propagated to t_{l+1} with the model resolvent $\mathcal{M} : \mathbb{R}^m \to \mathbb{R}^m$. The model is assumed to be perfect so that there are no errors in Eq. (1b) and autonomous ($\mathcal{M}$ does not depend on time). The observation operator $\mathcal{H} : \mathbb{R}^m \to \mathbb{R}^d$ relates the state $\mathbf{x}_l$ to the observation vector $\mathbf{y}_l$. The observation errors $(\boldsymbol{\varepsilon}_l)_{l\geq 0}$ are assumed Gaussian with mean $\mathbf{0} \in \mathbb{R}^d$ and covariance matrix $\mathbf{R} \in \mathbb{R}^{d\times d}$; they are uncorrelated in time.

2.1 4D-Var and IEnKS algorithms

Both 4D-Var and the IEnKS use a variational minimization in their analysis step. The objective of this minimization is to locate the global maximum of the posterior pdf $p(\mathbf{x}_0|\mathbf{y}_{L:K})$ of the system past state $\mathbf{x}_0$ given the observations $\mathbf{y}_{L:K} = [\mathbf{y}_K, \ldots, \mathbf{y}_L]$ at times $t_{L:K} = [t_K, \ldots, t_L] \in \mathbb{R}^S$. The system state and observations are seen as random vectors with values in $\mathbb{R}^m$ and $\mathbb{R}^d$, respectively. The posterior pdf quantifies how our knowledge on the state $\mathbf{x}_0$ changes with realizations of $\mathbf{y}_{L:K}$. Thus, its maximum is the most probable state after assimilating the observations. The DAW is displayed in Fig. 1. The parameters K and L are the time index of the DAW first and last assimilated observation batch, respectively. The number of ~~observations is~~

~~S and satisfies $S = L - K + 1$ in our case (i.e., no overlap between the DAWs)~~observation vectors used within the DAW is $L - K + 1$. To specify this posterior pdf, we have to make further assumptions on $\mathbf{x}_0$.

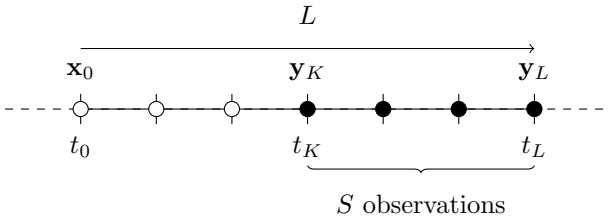

Figure 1. Schematic of a DAW. The state variable at t_0 is $\mathbf{x}_0$, the observation vector $\mathbf{y}_K$ at time t_K is the first of the DAW to be assimilated and the observation $\mathbf{y}_L$ at present time t_L is the last one~~, $S = L - K + 1$ is the number of observations to assimilate in the current cycle~~. These observations, possibly observation vectors, are represented by black dots.

The initial state $\mathbf{x}_0$ is assumed to be Gaussian with mean $\mathbf{x}_0^{\mathrm{b}} \in \mathbb{R}^m$ and covariance matrix $\mathbf{B} \in \mathbb{R}^{m \times m}$:

$$p(\mathbf{x}_0) = \mathcal{N}\left(\mathbf{x}_0 | \mathbf{x}_0^{\mathrm{b}}, \mathbf{B}\right). \tag{2}$$

With these assumptions, an analytic expression can be obtained for the posterior smoothing pdf $p(\mathbf{x}_0|\mathbf{y}_{L:K})$ at the first cycle, or for the cost function associated with this pdf. The latter is defined as

$$G(\mathbf{x}_0|\mathbf{y}_{L:K}) = -\ln p(\mathbf{x}_0|\mathbf{y}_{L:K}). \tag{3}$$

The notation G is used rather than the traditional J to refer to an exact cost function, i.e., a cost function defined from an exact posterior pdf. Bayes' theorem yields:

$$\begin{aligned} G(\mathbf{x}_0|\mathbf{y}_{L:K}) = &-\ln p(\mathbf{x}_0) - \ln p(\mathbf{y}_{L:K}|\mathbf{x}_0) \\ &+ \ln p(\mathbf{y}_{L:K}), \end{aligned} \tag{4}$$

and with the Gaussian assumption on the background and observation errors we have

$$\begin{aligned} G(\mathbf{x}_0|\mathbf{y}_{L:K}) = &\frac{1}{2}\left\|\mathbf{x}_0^b - \mathbf{x}_0\right\|^2_{\mathbf{B}^{-1}} + c_0 \\ &+ \frac{1}{2}\sum_{l=K}^{L}\left\|\mathbf{y}_l - \mathcal{H}\circ\mathcal{M}^l(\mathbf{x}_0)\right\|^2_{\mathbf{R}^{-1}}, \end{aligned} \tag{5}$$

where $\|\mathbf{x}\|^2_{\mathbf{A}} = \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$ is ~~,~~the norm of $\mathbf{x}$ associated with a symmetric positive definite matrix $\mathbf{A}$; c_0 is a normalization constant ensuring $\int e^{-G(\mathbf{x}_0|\mathbf{y}_{L:K})}\mathrm{d}\mathbf{x}_0 = 1$; $\mathcal{M}^l$ stands for l compositions of $\mathcal{M}$.

The propagation corresponds to a time shift of S time steps. Thus, at the k-th assimilation cycle, the posterior pdf is $p(\mathbf{x}_{kS}|\mathbf{y}_{kS+L:K})$. Using Bayes' theorem the k-th cycle cost function is:

$$G(\mathbf{x}_{kS}|\mathbf{y}_{kS+L:K}) = -\ln p\left(\mathbf{x}_{kS}|\mathbf{y}_{(k-1)S+L:K}\right) + \frac{1}{2}\sum_{l=K}^{L}\left\|\mathbf{y}_{kS+l} - \mathcal{H}\circ\mathcal{M}^{l}(\mathbf{x}_{kS})\right\|^{2}_{\mathbf{R}^{-1}} + c_{kS}, \tag{6}$$

where $-\ln p\left(\mathbf{x}_{kS}|\mathbf{y}_{(k-1)S+L:K}\right)$ is the background term. If the model and observation operators are nonlinear, an analytical expression for this prior is not accessible and one needs to approximate it. The 4D-Var and the IEnKS algorithms are solutions based on distinct approximation strategies.

The 4D-Var cost function at the k-th cycle, based on the static error covariance matrix $\mathbf{B}$, is defined by

$$J\left(\mathbf{x}_{kS};\mathbf{y}_{kS+L:kS+K},\mathbf{x}^{\mathrm{b}}_{kS}\right) = \frac{1}{2}\left\|\mathbf{x}^{b}_{kS} - \mathbf{x}_{kS}\right\|^{2}_{\mathbf{B}^{-1}} + \frac{1}{2}\sum_{l=K}^{L}\left\|\mathbf{y}_{kS+l} - \mathcal{H}\circ\mathcal{M}^{l}(\mathbf{x}_{kS})\right\|^{2}_{\mathbf{R}^{-1}}. \tag{7}$$

The analysis of 4D-Var consists in minimizing Eq. (7), yielding $\mathbf{x}^{\mathrm{a}}_{kS}$ at t_{kS}. Because this cost function depends on realizations of the random observations, $\mathbf{x}^{\mathrm{a}}_{kS}$ is also a random variable. This analysis is then propagated at time $t_{(k+1)S}$ with the resolvent of the model to produce the next cycle background state:

$$\mathbf{x}^{\mathrm{b}}_{(k+1)S} = \mathcal{M}^{S}\left(\mathbf{x}^{\mathrm{a}}_{kS}\right). \tag{8}$$

In general, $G(\mathbf{x}_0|\mathbf{y}_{L:K})$ and $J\left(\mathbf{x}_0;\mathbf{y}_{L:K},\mathbf{x}^{\mathrm{b}}_{0}\right)$ only coincide at the first cycle of an assimilation because of the assumption Eq. (2). Subsequently, the background term of the 4D-Var cost function $\frac{1}{2}\left\|\mathbf{x}^{\mathrm{b}}_{kS} - \mathbf{x}_{kS}\right\|^{2}_{\mathbf{B}^{-1}}$ is a Gaussian approximation of the exact background term $-\ln p\left(\mathbf{x}_{kS}|\mathbf{y}_{(k-1)S+L:K}\right)$. By definition, the background error covariance matrix $\mathbf{B}$ of the traditional 4D-Var cost function is the same for each cycle. This is not the case for the IEnKS.

The IEnKS (Bocquet and Sakov, 2014) is an ensemble method with a variational analysis. Two versions of the algorithm exist: the singular data assimilation (SDA) version where observations are assimilated only once; the multiple data assimilation (MDA) version where they are assimilated several times. We focus on the SDA version in the theoretical development. The MDA version ~~is mentioned~~, which can be seen as the first published quasi-static EnVar scheme is used in the numerical experiments ~~.~~ for comparison. Note that, for the SDA IEnKS, the number observations is $L-K+1=S$.

At the k-th cycle of the IEnKS, the background ensemble at t_{kS} is obtained by a propagation from the previous cycle. The ensemble members are the columns of the matrix $\mathbf{E}^{\mathrm{b}}_{kS}$, which is seen as a random matrix with values in $\mathbb{R}^{m\times n}$. It is used to estimate the prior mean and covariance matrix:

$$\mathbb{E}\left[\mathbf{x}_{kS}|\mathbf{y}_{(k-1)S+L:K}\right] \simeq \bar{\mathbf{x}}^{\mathrm{b}}_{kS}, \tag{9}$$

$$\mathbb{C}\left[\mathbf{x}_{kS}|\mathbf{y}_{(k-1)S+L:K}\right] \simeq \mathbf{X}^{\mathrm{b}}_{kS}\mathbf{X}^{\mathrm{bT}}_{kS}, \tag{10}$$

where $\mathbb{E}$ and $\mathbb{C}$ are the expectation and covariance operators, respectively; $\bar{\mathbf{x}}^{\mathrm{b}}_{kS}$ and $\mathbf{X}^{\mathrm{b}}_{kS}$ are the empirical mean and normalized anomaly of $\mathbf{E}^{\mathrm{b}}_{kS}$, respectively:

$$\bar{\mathbf{x}}^{\mathrm{b}}_{kS} = \mathbf{E}_{kskS}{}^{\mathrm{b}} \frac{\mathbf{1}_n}{n}, \tag{11}$$

$$\mathbf{X}^{\mathrm{b}}_{kS} = \mathbf{E}_{kskS}{}^{\mathrm{b}} \frac{\mathbf{I}_n - \frac{\mathbf{1}_n \mathbf{1}_n^{\mathrm{T}}}{n}}{\sqrt{n-1}}, \tag{12}$$

with $\mathbf{1}_n = [1, \ldots, 1]^{\mathrm{T}} \in \mathbb{R}^n$ a vector of ones and $\mathbf{I}_n$ is the identity of $\mathbb{R}^n$. Note that Eqs (9, 10) are ~~not exact~~ approximations because of sampling errors. If the state vector is of the form:

$$\mathbf{x}_{kS} = \bar{\mathbf{x}}^{\mathrm{b}}_{kS} + \mathbf{X}^{\mathrm{b}}_{kS} \mathbf{w}_{kS}, \tag{13}$$

with $\mathbf{w}_{kS} \in \mathbb{R}^n$ the control variable in the ensemble space, the IEnKS cost function is defined in the ensemble space by

$$J\left(\mathbf{w}_{kS}; \mathbf{y}_{kS+L:kS+K}, \mathbf{E}^{\mathrm{b}}_{kS}\right) = \frac{1}{2} \|\mathbf{w}_{kS}\|^2 + \frac{1}{2} \sum_{l=K}^{L} \left\| \mathbf{y}_{kS+l} - \mathcal{H} \circ \mathcal{M}^l \left(\bar{\mathbf{x}}^{\mathrm{b}}_{kS} + \mathbf{X}^{\mathrm{b}}_{kS} \mathbf{w}_{kS}\right) \right\|^2_{\mathbf{R}^{-1}}, \tag{14}$$

where $\|\cdot\| = \|\cdot\|_{\mathbf{I}}$ with $\mathbf{I}$ the identity matrix. The analysis of the IEnKS ~~also~~ consists in minimizing this cost function. It yields an ~~analyzed~~ updated ensemble $\mathbf{E}^{\mathrm{a}}_{kS}$ at time t_{kS} verifying:

$$\bar{\mathbf{x}}^{\mathrm{a}}_{kS} = \bar{\mathbf{x}}^{\mathrm{b}}_{kS} + \mathbf{X}^{\mathrm{b}}_{kS} \mathbf{w}^{\mathrm{a}}_{kS}, \tag{15}$$

$$\mathbf{X}^{\mathrm{a}}_{kS} = \mathbf{X}^{\mathrm{b}}_{kS} \left[\nabla^2 J\left(\mathbf{w}^{\mathrm{a}}_{kS}; \mathbf{y}_{kS+L:kS+K}, \mathbf{E}^{\mathrm{b}}_{kS}\right)\right]^{-1/2} \mathbf{U}, \tag{16}$$

with $\bar{\mathbf{x}}^{\mathrm{a}}_{kS}$ and $\mathbf{X}^{\mathrm{a}}_{kS}$ ~~respectively~~ the empirical mean and normalized anomalies of the analyzed ensemble, respectively, $\mathbf{w}^{\mathrm{a}}_{kS}$ the cost function minimizer, $\nabla^2 J\left(\mathbf{w}^{\mathrm{a}}_{kS}; \mathbf{y}_{kS+L:kS+K}, \mathbf{E}^{\mathrm{b}}_{kS}\right)$ its Hessian at this minimum (usually approximated to avoid computations of the model second derivatives), the exponent $-1/2$ refers to the unique symmetric definite positive inverse square root of a symmetric definite positive matrix and $\mathbf{U} \in \mathbb{R}^{n \times n}$ an orthogonal matrix such that $\mathbf{U}\mathbf{1}_n = \mathbf{1}_n$. This analyzed ensemble is also propagated to time $t_{(k+1)S}$ to produce the next cycle background ensemble

$$\mathbf{E}^{\mathrm{b}}_{(k+1)S} = \mathcal{M}^S\left(\mathbf{E}^{\mathrm{a}}_{kS}\right). \tag{17}$$

The cycle is completed by using in the next analysis the cost function Eq. (14) with time indexes incremented by S. 4D-Var and the IEnKS are sketched in Fig. 2.

2.2 Performance of assimilation

In order to evaluate the efficiency of 4D-Var and the IEnKS with DAW parameters S and L, two measures of accuracy are investigated here: the usual empirical RMSE, and a theoretical counterpart.

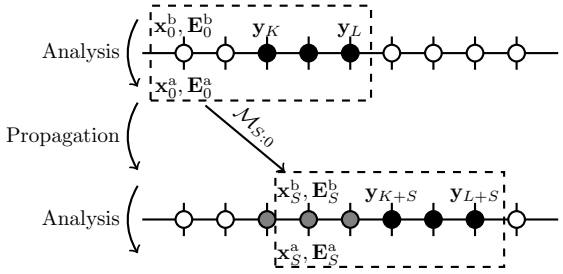

Figure 2. Chaining of the 4D-Var and IEnKS ~~two~~ first two cycles with $S = 3, L = 4$. The first 4D-Var analysis ~~use~~ uses the background $\mathbf{x}_0^{\mathrm{b}}$ at t_0 and the observations $\mathbf{y}_{L:K}$ to ~~produce~~ give the analysis $\mathbf{x}_0^{\mathrm{a}}$ at t_0. It is propagated S steps forward in time to produce the new background at time t_S where another analysis can be performed. The IEnKS does the same but with an ensemble. The dashed rectangle symbolizes the current DAW, black dots represent the observations assimilated in the current cycle, gray dots represent already assimilated observations and white dots represent observations not assimilated.

At the k-th cycle, the algorithm generates at time t_{kS} an analysis $\mathbf{x}_{kS}^{\mathrm{a}}$ from the observations. This analysis is propagated with the model l steps forward in time to yield the analysis $\mathbf{x}_{kS+l}^{\mathrm{a}}$ meant to approximate the system true state $\mathbf{x}_{kS+l}$. A traditional measure of an assimilation performance is the root mean square error (RMSE). It is defined by

$$\mathrm{RMSE} = \frac{1}{\sqrt{m}} \left\| \mathbf{x}_{kS+l} - \mathbf{x}_{kS+l}^{\mathrm{a}} \right\|. \tag{18}$$

The RMSE takes different names depending on the time it is computed at. If $l = L$, it is called filtering RMSE; if $0 \leq l \leq L-1$, it is the smoothing RMSE with lag $L - l$. In the following, the smoothing RMSE will correspond to the one with (maximum) lag L.

The RMSE rigorously depends on the random variable realizations, and thus it is also a random variable. In our numerical experiments, as is usually done, the RMSE is averaged over the cycles to mitigate this variability:

$$\mathrm{aRMSE}_N = \frac{1}{N} \sum_{k=0}^{N-1} \frac{1}{\sqrt{m}} \left\| \mathbf{x}_{kS+l} - \mathbf{x}_{kS+l}^{\mathrm{a}} \right\|. \tag{19}$$

Let us assume that there is a random couple $\left(\mathbf{x}_{\infty S+l}, \mathbf{x}_{\infty S+l}^{\mathrm{a}}\right)$ whose distribution is invariant and ergodic with respect to the shift transformation:

$$T : \left(\mathbf{x}_{kS+l}, \mathbf{x}_{kS+l}^{\mathrm{a}}\right) \mapsto \left(\mathbf{x}_{(k+1)S+l}, \mathbf{x}_{(k+1)S+l}^{\mathrm{a}}\right). \tag{20}$$

Then by Birkhoff's ergodic theorem (see Walters, 1982) the sequence $(\mathrm{aRMSE}_N)_N$ converges when $N \to \infty$ and its limit aRMSE verifies:

$$\mathrm{aRMSE} = \frac{1}{\sqrt{m}} \mathbb{E}\left[\left\| \mathbf{x}_{\infty S+l} - \mathbf{x}_{\infty S+l}^{\mathrm{a}} \right\| \right], \tag{21}$$

where the expectation $\mathbb{E}$ is taken over $p\left(\mathbf{x}_{\infty S+l}, \mathbf{x}^{\mathrm{a}}_{\infty S+l}\right)$. In this case, the aRMSE measures the long time impact of the cycling on the assimilation accuracy. This limit is difficult to exploit algebraically. That is why, in the theoretical developments, we will prefer the expected MSE (eMSE), denoted by P:

$$P_{kS+l} = \mathbb{E}\left[\left\|\mathbf{x}_{kS+l} - \mathbf{x}^{\mathrm{a}}_{kS+l}\right\|^2\right],$$

where the expectation is taken over $p\left(\mathbf{x}_{kS+l}, \mathbf{x}^{\mathrm{a}}_{kS+l}\right)$. In the following subsection, we will focus on the long term impact of the cycling on P. Simplifying assumptions will be made to express $P_{\infty S+l} = \lim_{k\to\infty} P_{kS+l}$ as a function of S and L.

2.3 Performance in the linear, diagonal, autonomous case

In order to obtain analytical expressions of the eMSE for 4D-Var and the IEnKS, we make drastic simplifying assumptions.

First, the model is assumed to be the resolvent of a linear, diagonal, autonomous ordinary differential equation. Thus, it can be expressed as

$$\mathcal{M}^l(\mathbf{x} + \delta\mathbf{x}) = \mathcal{M}^l(\mathbf{x}) + \mathbf{M}^l \delta\mathbf{x}, \tag{22}$$

where $\mathbf{M} = \mathrm{diag}\left(\alpha_i\right)_{i=1..m}$ is diagonal and does not depends on $\mathbf{x}$. We further assume $\mathcal{H} = h\mathbf{I}_m$, $\mathbf{B} = b\mathbf{I}_m$, $\mathbf{R} = r\mathbf{I}_m$, where $\mathbf{I}_m$ is the identity matrix of $\mathbb{R}^m$ and $h, r, b > 0$. With these assumptions, appendix A provides an expression for the 4D-Var asymptotic eMSE in the univariate case. The generalization to the diagonal multivariate case is obtained by summing up the eMSEs of each direction:

$$P^{\text{4D-Var}}_{\infty S+l} = \sum_{i=1}^{m} \begin{cases} \infty & \text{if } \Delta_i \geq 1 \\ \frac{b^2 \Sigma^L_{K,i}}{\alpha_i^{2(S-l)}} \frac{\Delta_i}{1-\Delta_i} & \text{otherwise} \end{cases}, \tag{23a}$$

$$\Sigma^L_{K,i} = \frac{h^2}{r} \frac{\alpha_i^{2(L+1)} - {\alpha_i}^{2K}}{\alpha_i^2 - 1}, \tag{23b}$$

$$\Delta_i = \frac{\alpha_i^{2S}}{\left(1 + b\Sigma^L_{K,i}\right)^2}. \tag{23c}$$

The case $\Delta_i \geq 1$ means that too much credit is given to the background variance, which is approximated for all cycles by the constant b in our 4D-Var scheme. Therefore, the information carried by the observations is not sufficient to mitigate the exponential growth of errors in the propagation.

Concerning the IEnKS, the anomalies are assumed to be full rank to avoid any complication due to singular covariance matrices. Moreover, the linearity of the model is employed to express the background statistics:

$$\bar{\mathbf{x}}^{\mathrm{b}}_{(k+1)S} = \mathcal{M}^S\left(\bar{\mathbf{x}}^{\mathrm{a}}_{kS}\right), \tag{24a}$$

$$\mathbf{X}^{\mathrm{b}}_{(k+1)S} = \mathbf{M}^S \mathbf{X}^{\mathrm{a}}_{kS}, \tag{24b}$$

$$\bar{\mathbf{x}}^{\mathrm{b}}_0 = \mathbf{x}^{\mathrm{b}}_0, \tag{24c}$$

$$\mathbf{X}^{\mathrm{b}}_0 = \mathbf{B}^{1/2}. \tag{24d}$$

This way, sampling errors are avoided as they are not the focus of this study. The background ensemble $\mathbf{E}^{\mathrm{b}}_{kS}$ becomes a notational shortcut for the pair $\left(\bar{\mathbf{x}}^{\mathrm{b}}_{kS}, \mathbf{X}^{\mathrm{b}}_{kS}\right)$. ~~This~~ Actually, this simplified IEnKS is a Kalman smoother (Cosme et al., 2012; Bocquet and Carrassi, 2017). With these assumptions, appendix B gives an expression for the IEnKS asymptotic eMSE in the univariate case. The optimality of the IEnKS eMSE is also proven. The generalization to the diagonal multivariate case is also obtained by summing up the eMSEs of each direction:

$$P^{\mathrm{IEnKS}}_{\infty S+l} = \sum_{i=1}^{m} \begin{cases} 0 & \text{if } |\alpha_i| \leq 1 \\ \frac{r}{h^2 \alpha_i^{2(L-l)}} \frac{\alpha_i^2 - 1}{\alpha_i^2} & \text{otherwise} \end{cases}. \tag{25}$$

This expression shows that the eMSE components on the stable directions are null. Indeed one expects the IEnKS to be at least more efficient than a free-run, whose errors in the stable directions tend to zero[1]~~.~~.[1] This is not the case for 4D-Var since the static background covariance matrix introduces spurious variance in the stable directions as seen in Eq. (23). In Trevisan et al. (2010), 4D-Var error variances in the stable directions are forced to zero to improve the accuracy of the assimilation.

In the following, we study the eMSE dependency on the DAW parameters, S and L, for both algorithms. We focus on a bivariate case with $\alpha_1 = 1.2$, $\alpha_2 = 0.8$ in order to have one stable and one unstable direction; $h = b = r = 1$. Using Eqs (23,25), the asymptotic smoothing and filtering eMSEs are displayed as a function of L, S in Fig. 3. The eMSE components on the stable and unstable directions are also shown. Those graphs are interpreted in the following.

Specifically, the eMSE expression for the IEnKS is of the form

$$P^{\mathrm{IEnKS}}_{\infty S+l} = c_1 \alpha_1^{2(l-L)}, \tag{26}$$

where c_1 does not depend on S, L, l. The contribution to $P^{\mathrm{IEnKS}}_{\infty S+l}$ on the stable direction is zero. It depends only on the lag $L-l$; it does not depend on S. Moreover, the filtering eMSE ($l = L$) is constant: the propagation compensates for the analysis. The smoothing eMSE ($l = 0$) decreases exponentially with L.

Concerning 4D-Var, we assume K fixed and $S \to \infty$ to give an asymptotic eMSE expression:

$$P^{\mathrm{4D-Var}}_{\infty S+l} = P^{\mathrm{IEnKS}}_{\infty S+l} \left(1 + o\left(1\right)\right) + c_2 \alpha_2^{2l} \left(1 + o\left(1\right)\right), \tag{27}$$

where c_2 is constant with S, l and $o\left(1\right) \to 0$ when $S \to \infty$. The unstable component is close to the IEnKS overall eMSE. The biggest difference with it concerns the eMSE on the stable component. The inexact background variance modeling adds to the eMSE a detrimental term.

To qualify the long term impact of the cycling on the errors, the filtering eMSE is more instructive than the smoothing eMSE. Indeed, the smoothing eMSE is improved with L as it adds future observations (with respect to the analysis time) in the DAW. This improvement dominates the potentially detrimental impact of the Gaussian background approximations. In the filtering eMSE, this improvement is balanced by the propagation of the analysis at the end of the DAW. In Fig. 3, the filtering eMSE stable component is ~~damped~~ mitigated such that it has little effect. However, on the unstable component, the

[1] ~~The fact that the errors lie in the unstable subspace is more general (Bocquet and Carrassi, 2017)~~

[1] The fact that the errors lie in the unstable subspace is more general (Bocquet and Carrassi, 2017)

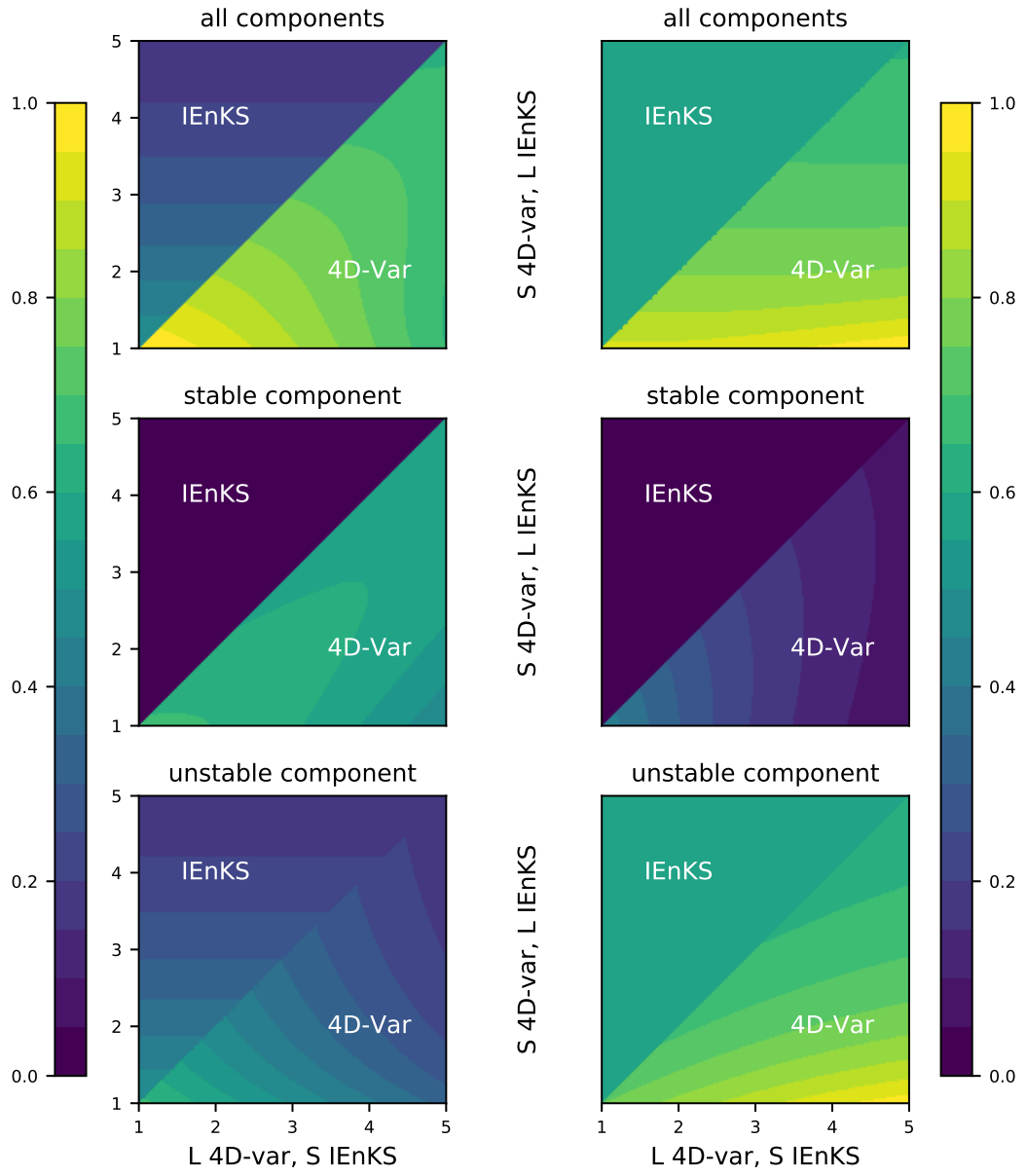

Figure 3. The asymptotic smoothing (on the left column) and filtering (on the right column) eMSEs of 4D-Var and the IEnKS. They are displayed as functions of L, S with their components on the stable, unstable subspaces. The L parameter is on the abscissa axis for ~~the~~ 4D-Var and on the ordinate axis for the IEnKS. The S parameter is on the abscissa axis for the IEnKS and on the ordinate axis for ~~the~~ 4D-Var.

parameter S improves the filtering eMSE. The bigger S , the closer $P_{\infty S+l}^{\text{4D}-\text{Var}}$ is to $P_{\infty S+l}^{\text{IEnKS}}$, which is optimal (cf. appendix B). A qualitative explanation is that to assimilate the same numbers of observations, a 4D-Var using high values of S need

fewer cycles. Therefore, it relies less often on the Gaussian background approximation. ~~Making~~ , making the analysis more trustworthy.

Figure 4 displays the asymptotic eMSEs of both algorithms as a function of the lag for $S = L = 5$. These curves are similar

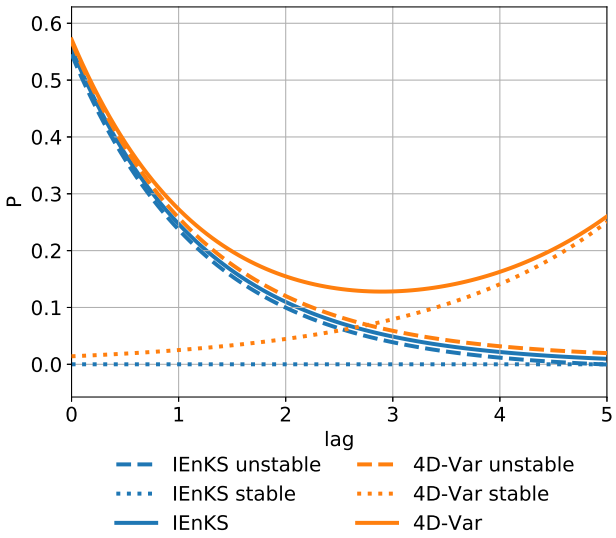

Figure 4. The asymptotic eMSEs as a function of the lag ($\mathrm{lag} = 0$ is the filtering performance, $\mathrm{lag} = L$ is the smoothing performance). The superposing curves have been slightly translated for better readability, $S = L = 5$, $h = b = r = 1$.

to those of Trevisan et al. (2010). Concerning 4D-Var, the eMSE can be written as a function of the lag $L - l$:

$$P^{\mathrm{4D-Var}}_{\infty S+l} \simeq c_1 \alpha_1^{-2\mathrm{lag}} + c_2 \alpha_2^{2(L-\mathrm{lag})}. \tag{28}$$

Thus, the unstable component of the eMSE is an exponentially decreasing function of the lag and the stable component is an exponentially increasing function of the lag. The sum is therefore decreasing when the unstable component is dominant; it is increasing when the stable component is dominant.

In this section, we have studied the accuracy of both the *cycled* IEnKS and 4D-Var eMSEs as a function of the DAW parameters under linear, autonomous, diagonal assumptions. We found that the DAW parameter L improves the smoothing eMSE and S improves the filtering eMSE. These properties will be discussed and numerically investigated in a nonlinear context in the next section.

2.4 Performance in the nonlinear, chaotic case

The results of section 2.3 foster the use of the largest possible L to improve the 4D-Var and IEnKS smoothing eMSEs. For filtering, the error propagation at the end of the DAW balances the gain in eMSE due to the assimilation of future observations. Thus, the filtering eMSE is not improved by the assimilation of observations distant in time; it is rather affected by

the background pdf approximation. To assimilate the same number of observations, algorithms using high values of S need fewer cycles. Therefore, they rely less often on the background approximation. That is why the 4D-Var filtering performance is improved with S. For the IEnKS as in section 2.3, the prior pdf approximation is exact so that there is no dependence on S. This is no longer true in the nonlinear case. Hence, with a similar reasoning, one actually expects an improvement on the IEnKS filtering performance with S. As a matter of fact, it has been shown that with a nonlinear chaotic model, the filtering accuracy increases with L in most cases (see Bocquet and Sakov, 2014, and section 4 of the present paper).

2.4.1 Multiple local minima

However, with a chaotic model, Pires et al. (1996) showed that the 4D-Var cost function number of local extrema increases with L, making minimization problematic. We show in this section that the IEnKS cost function suffers from the same problem.

This behavior will be illustrated with the Lorenz 95 (L95) model (Lorenz and Emanuel, 1998). It represents a mid-latitude zonal circle of the atmosphere and is described by a set of m nonlinear differential equations:

$$\frac{\mathrm{d}x_j}{\mathrm{d}t} = (x_{j+1} - x_{j-2})\,x_{j-1} - x_j + F, \tag{29}$$

where x_j is the j-th modulo m component of $\mathbf{x}$, $m = 40$ and $F = 8$. This equation is integrated using a fourth-order Runge-Kutta scheme with a time-step of $\delta t = 0.05$. The dynamics of L95 are chaotic; the L95 largest Lyapunov exponent is $\lambda \simeq 1.7$.

Figure 5 shows a typical IEnKS cost function profile in one direction of the analyzed ensemble space for multiple values of the DAW parameters. The system is observed at every time step and $\mathcal{H} = \mathbf{B} = \mathbf{R} = \mathbf{I}_m$. The curves have more and more local extrema when L increases. The curves with the highest amplitudes of the ripples are found for small values of S. Indeed, an averaging effect may settle in as the number of observations increases.

This hilly shape causes minimization problems. A possible minimization procedure for the IEnKS is the Gauss-Newton (GN) algorithm (e.g., Björck, 1996). GN is not a global procedure meaning that, depending on the starting point and the cost function properties, the algorithm can converge towards any local extremum, take many iterations or even diverge. However, if the cost function is quadratic, the global minimum is reached in one iteration. The cost function non-quadraticity is induced by the model nonlinearity. In the following, we will give a heuristic argument yielding a bound on the S parameter beyond which the GN method probably misses the global minimum in the configuration $\mathcal{H} = \mathbf{B} = \mathbf{R} = \mathbf{I}_m$. To some extent, our argument can be seen as an improvement on the notion of *useful DAW length* introduced by Pires et al. (1996) beyond which the performance gain is negligible. In contrast with the useful DAW length, we account in the following for both the cycling and the nonlinearity.

2.4.2 Effective data assimilation window length

First, the GN convergence properties are drastically simplified. We assume the method converges to the global minimum if and only if the minimization starting point is in a neighborhood of the global minimizer where the IEnKS cost function is almost quadratic.

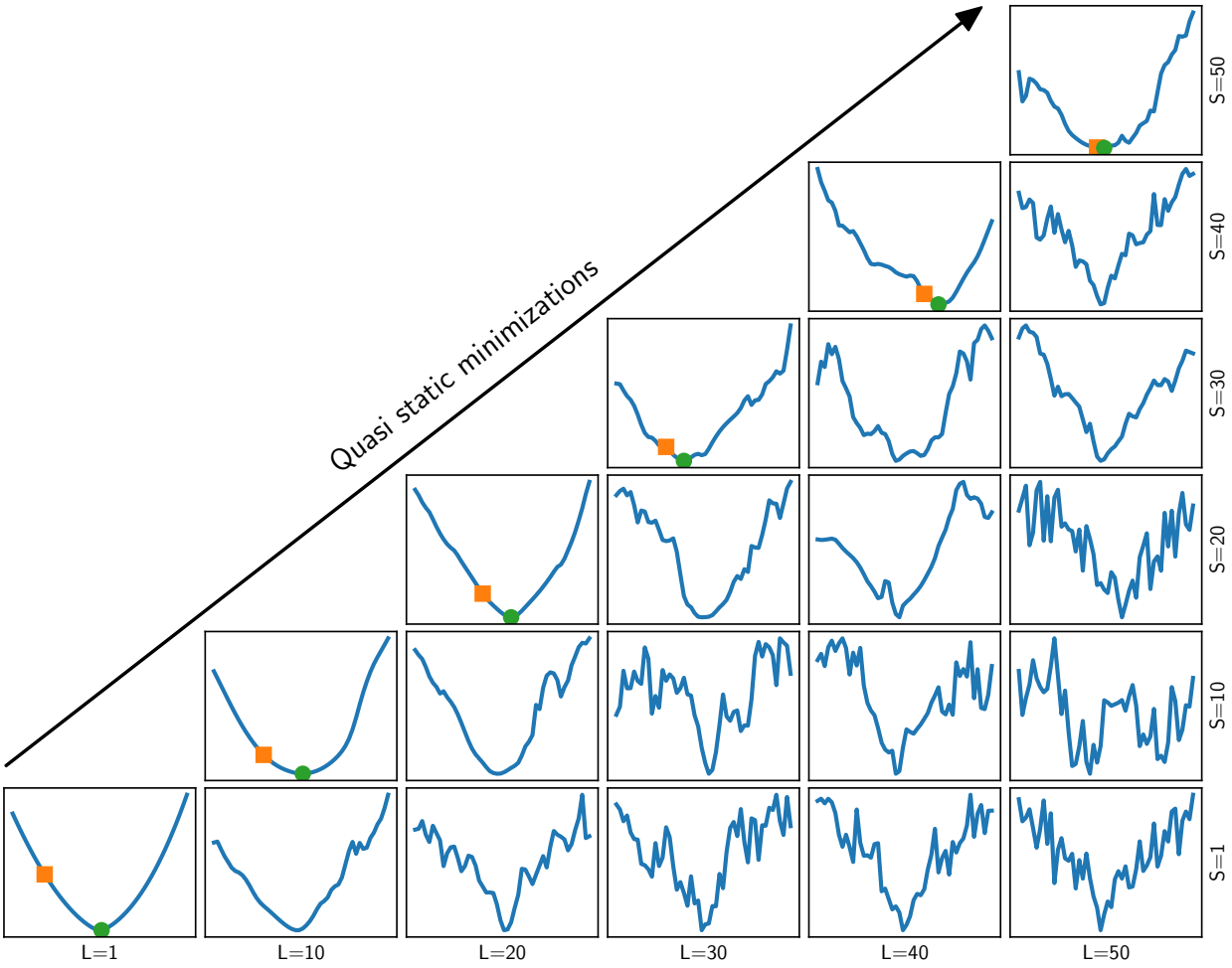

Figure 5. Cost functions of the IEnKS projected in one direction of the analyzed ensemble (hence centered and normalized) with various DAW parameters S, L. A quasi-static minimization can be visualized from these panels. It begins with the bottom-left cost function; the orange ~~dot~~ square is the starting point and the green ~~one~~ dot is at the minimum. From the bottom-left cost function up to the top-right cost function, batches of 9 then 10 observation vectors are progressively added to the DAW, and the minimizer (green dot) is updated accordingly.

Unfortunately, this minimizer is unknown because the cost function depends on realizations of many random variables. In order to eliminate this variability, Pires et al. (1996) introduced a so-called error-free cost function. We will rather use an averaged cost function $J_{\infty S}$ defined by

$$J_{\infty S}(\mathbf{w}) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} J\left(\mathbf{w}; \mathbf{y}_{kS+L:kS+K}, \mathbf{E}^{\mathrm{b}}_{kS}\right). \tag{30}$$

Relying on an ergodicity assumption, appendix C proves that this averaged cost function verifies

$$J_{\infty S}(\mathbf{w}) = \frac{1}{2}\|\mathbf{w}\|^2 + \frac{dS}{2} + \frac{1}{2}\sum_{l=K}^{L}\mathbb{E}\left[\left\|\delta\mathbf{x}^{\mathrm{b}}_{\infty S+l}\right\|^2\right], \tag{31a}$$

$$\delta\mathbf{x}^{\mathrm{b}}_{\infty S+l} = \mathcal{M}^l\left(\bar{\mathbf{x}}^{\mathrm{b}}_{\infty S} + \mathbf{X}^{\mathrm{b}}_{\infty S}\mathbf{w}\right) - \mathcal{M}^l\left(\mathbf{x}_{\infty S}\right), \tag{31b}$$

where the ergodic random variables $\bar{\mathbf{x}}_{\infty S}$, $\mathbf{X}^{\mathrm{b}}_{\infty S}$ and $\mathbf{x}_{\infty S}$ have been defined in section 2.2 and Appendix C. As seen in Eq. (31), a sufficient condition for the starting point $\mathbf{w} = \mathbf{0}$ to be in a neighborhood of the global minimizer where the cost function is assumed almost quadratic is to require that $\bar{\mathbf{x}}^{\mathrm{b}}_{\infty S}$ be in a neighborhood of $\mathbf{x}_{\infty S}$ where all the $\left(\mathcal{M}^l\right)_{K\leq l\leq L}$ are almost linear.

In the univariate case, if the model behavior is almost linear and unstable, we can use Eq. (25) to estimate the terms in the sum in Eq. (31) at the starting point $\mathbf{w} = \mathbf{0}$:

$$\mathbb{E}\left[\left\|\delta\mathbf{x}^{\mathrm{b}}_{\infty S+l}\right\|^2\right] \simeq \alpha^{2(S+l-L)}\frac{\alpha^2-1}{\alpha^2}, \tag{32}$$

where α is the model linear part and the extra S accounts for the propagation. But in a necessarily bounded physical system, the right-hand side of Eq. (32) cannot grow indefinitely with $l+S$. Such model saturation imposes

$$\mathbb{E}\left[\left\|\delta\mathbf{x}^{\mathrm{b}}_{\infty S+l}\right\|^2\right] \leq B, \tag{33}$$

with B a bound. Hence, Eqs (32,33) yield the following inequality on S:

$$S \leq S_{\max}, \tag{34a}$$

$$S_{\max} = \frac{\ln(B) - \ln\left(1-\alpha^{-2}\right)}{2\ln(\alpha)}. \tag{34b}$$

We choose $l = L$ because it corresponds to the most constraining case. When Eq. (34a) is violated, $\bar{\mathbf{x}}^{\mathrm{b}}_{\infty S}$ departs from $\mathbf{x}_{\infty S}$ such that the nonlinearities of $\mathcal{M}^L$ are significant.

To apply this inequality to the L95 model we choose:

$$\alpha = \lim_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N-1}\sigma\left(\frac{\mathrm{d}\mathcal{M}}{\mathrm{d}\mathbf{x}}(\mathbf{x}_k)\right), \tag{35}$$

σ being the mean of the singular values greater than 1. This corresponds to an average of the error amplification by $\frac{\mathrm{d}\mathcal{M}}{\mathrm{d}\mathbf{x}}(\mathbf{x}_{\infty S})$ in the local unstable subspace. We also choose for the bound B the average squared norm between two long trajectories:

$$B = \lim_{N\to\infty}\frac{1}{mN}\sum_{k=0}^{N-1}\left\|\mathbf{x}_{kS} - \mathcal{M}^{kS}\left(\bar{\mathbf{x}}^{\mathrm{b}}_0\right)\right\|^2. \tag{36}$$

This quantity is greater than $\mathbb{E}\left[\left\|\delta\mathbf{x}^{\mathrm{b}}_{\infty S+l}\right\|^2\right]$ because the IEnKS asymptotic performance is at least better than a free run. From the values of α and B we find $S_{\max} = 14$.

Figure 6 shows the filtering and smoothing aRMSEs of an IEnKS $L = S$ with L95 as a function of S, the performance strongly deteriorates for $S > 16$, which is remarkably consistent with our estimation. Figure 6 also shows another difference

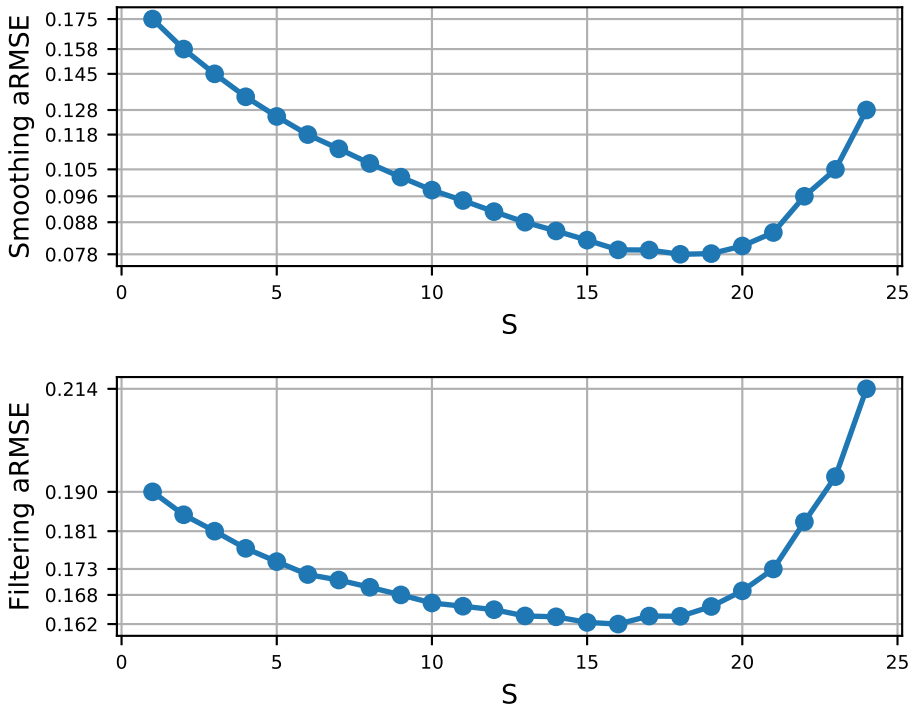

Figure 6. Smoothing aRMSE (top) and filtering aRMSE (bottom) of a Gauss-Newton IEnKS $L = S$ as a function of S with L95 model over 5×10^5 cycles (logarithmic scale). We used the finite-size version (Bocquet, 2011) to account for sampling errors and avoid the need for inflation, $\mathcal{H} = \mathbf{R} = \mathbf{B} = \mathbf{I}_m$, $n = 20$.

with the linear case: the IEnKS filtering aRMSE depends on S. The former discussion on the local extrema explains this dependency for large values of S.

However, for small values of S, the decreasing aRMSE has not been explained. This is a consequence of the Gaussian background approximation. At each cycle, the IEnKS uses the background ensemble $\mathbf{E}^{\mathrm{b}}_{kS}$ to estimate the first two moments of the background pdf and makes the approximation:

$$p\left(\mathbf{x}_{kS}|\mathbf{y}_{(k-1)S+L:K}\right) \simeq \mathcal{N}\left(\mathbf{x}_{kS}|\bar{\mathbf{x}}^{\mathrm{b}}_{kS}, \mathbf{X}^{\mathrm{b}}_{kS}\left(\mathbf{X}^{\mathrm{b}}_{kS}\right)^{\mathrm{T}}\right). \tag{37}$$

Because the model is nonlinear, this pdf is unlikely to be Gaussian. Therefore, this approximation results in a loss of information and the more it is used, the farther from G the IEnKS cost function is. This is exactly what happens when S is small: to assimilate the same number of observations, the IEnKS uses more cycles so that it relies more on the Gaussian background approximation.

3 Quasi-static algorithms

We have seen in the previous section that the effective DAW length ~~should be limited~~ is constrained by the cost function non-quadraticity. In this section we review and propose algorithms able to overcome these minimization issues and reach longer DAWs.

Quasi static algorithms have been introduced by Pires et al. (1996) in a 4D-Var context. The idea behind ~~quasi-staticism~~ QSVA is to control the way observations are added to the cost function in order to keep the minimization starting point in the basin of attraction containing the global minimizer. The method consists in repeating the minimization with ~~increasing parameter S~~ an increasing number of observations: the first minimization is performed using the cost function with ~~$S=1$ then S~~ a single observation vector then the number of observation vectors is increased and another minimization can be performed with the former minimizer as the new starting point. The process is then repeated until ~~S reaches its final value~~ all observations are accounted for.

This procedure is directly applicable to the IEnKS cost function minimization. The left panel in Fig. 7 is a schematic of a QS minimization and Algorithm 1 gives the pseudo-code of an IEnKS with a QS minimization. The new parameters $(L_q)_{q<N_Q}$ control the number of observations added at each minimization, where N_Q is the total number of batches of observation vectors. The three first lines initialize the minimization starting point, the ensemble mean and anomaly matrix. The **for** loop in lines 4-23 repeats the QS minimization. The **while** loop in lines 6-22 is the Gauss-Newton minimization. Lines 7 and 8 center the ensemble on the current minimizer. Lines 9 and 10 initialize the cost function gradient and the approximate Hessian. The **for** loop in lines 12-18 ~~construct~~ compute the observation terms of the cost function~~gradient and~~, the gradient and the approximate Hessian. Lines 13 and 14 use a finite difference formula to compute the tangent linear and adjoint of $\mathbf{w} \mapsto \mathcal{H} \circ \mathcal{M}^l \left(\bar{\mathbf{x}}_0^{\mathrm{b}} + \mathbf{X}_0^{\mathrm{b}} \mathbf{w}\right)$. Lines 15 and 16 use this adjoint to update the gradient and approximate Hessian. Lines 19 and 20 solve the linear system of the ~~Guauss-Newton~~ Gauss-Newton algorithm to update the current minimizer. When GN convergence is reached, this minimizer will be used as a starting point for the next QS minimization. Line 24 updates the ensemble. Line 25 propagates the updated ensemble to the next assimilation cycle.

~~In Fig. 5 ,~~ Figure 5 illustrates the QS scheme ~~corresponds to minimizing a cost function in the bottom row then using its minimum~~ on a single analysis, as described in the caption.

To cycle the scheme, the DAW is then shifted with a small S. This ensures minimal cost function deformation, since few vectors of observation enter and few leave the DAW. This new cost function is then minimized using the forecast of the preceding minimizer as a starting point~~for the minimization of the top right cost function and so on~~.

~~Note that Bocquet and Sakov (2014) already qualified the SDA IEnKS with~~

To keep this cost function deformation statistically consistent, Bocquet and Sakov (2014) advocated to use $S = 1$ ~~as QS. The reason is that, at each propagation, only one vector of observations enters and only one leaves~~ and assimilate one observation vector at the end of the DAW, ~~slightly deforming the cost function~~ which avoids multiple assimilation of the same observations. It is also the easiest way to ensure inequality (34a). However, this method has been shown to be suboptimal because of the

Algorithm 1 One cycle of the ~~IEnKS-QS~~IEnKS$_{\text{QS}}$

Require: $\mathbf{E}_0^{\text{b}}$ the background ensemble at t_0; λ the inflation; $(L_q)_{q<N_Q}$ a list of DAW time indexes; ε the finite differences step; $\delta, j_{\max}$ GN end of loop parameters; $\mathbf{1} = (1,\ldots,1)^{\mathrm{T}} \in \mathbb{R}^n$ and $\mathbf{I}$ is the identity of $\mathbb{R}^n$.

$\mathbf{w}_0^{\text{a}} := \mathbf{0}$
$\bar{\mathbf{x}}_0^{\text{b}} = \mathbf{E}_0^{\text{b}}\mathbf{1}/n$
$\mathbf{X}_0^{\text{b}} = \lambda\left(\mathbf{E}_0^{\text{b}} - \bar{\mathbf{x}}_0^{\text{b}}\mathbf{1}^{\mathrm{T}}\right)$
for $q = 0...N_Q - 1$ **do**
  $j := 0$
  repeat
    $\bar{\mathbf{x}}_0^{\text{a}} = \bar{\mathbf{x}}_0^{\text{b}} + \mathbf{X}_0^{\text{b}}\mathbf{w}_0^{\text{a}}$
    $\mathbf{E}_0 := \bar{\mathbf{x}}_0^{\text{a}}\mathbf{1}^{\mathrm{T}} + \epsilon\mathbf{X}_0^{\text{b}}$
    $\nabla J := (n-1)\mathbf{w}_0^{\text{a}}$
    $\tilde{\nabla}^2 J := (n-1)\mathbf{I}$
    $\mathbf{E}_K := \mathcal{M}^K(\mathbf{E}_0)$
    for $l = K,...,L_q$ **do**
      $\bar{\mathbf{y}}_l := \mathcal{H}(\mathbf{E}_l)\,\mathbf{1}/n$
      $\mathbf{Y}_l := \left(\mathcal{H}(\mathbf{E}_l) - \bar{\mathbf{y}}_l\mathbf{1}^{\mathrm{T}}\right)/\epsilon$
      $\nabla J := \nabla J - \mathbf{Y}_l^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y}_l - \bar{\mathbf{y}}_l)$
      $\tilde{\nabla}^2 J := \tilde{\nabla}^2 J - \mathbf{Y}_l^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{Y}_l$
      $\mathbf{E}_{l+1} := \mathcal{M}(\mathbf{E}_l)$
    end for
    solve $\tilde{\nabla}^2 J\delta\mathbf{w} := \nabla J$
    $\mathbf{w}_0^{\text{a}} := \mathbf{w}_0^{\text{a}} - \delta\mathbf{w}$
    $j := j + 1$
  until $||\delta\mathbf{w}|| \leq \delta$ or $j \geq j_{\max}$
end for
$\mathbf{E}_0^{\text{a}} := \mathbf{x}_0^{\text{a}}\mathbf{1}^{\mathrm{T}} + \sqrt{n-1}\mathbf{X}_0^{\text{b}}\tilde{\nabla}^2 J^{-1/2}$
$\mathbf{E}_S^{\text{b}} := \mathcal{M}^S(\mathbf{E}_0^{\text{a}})$

frequent Gaussian background approximations. Moreover, it ultimately fails to be QS since there is only one distant observation vector per analysis.

An alternative is to keep $S = 1$ ~~with many~~ but using all observations in the DAW ~~is~~ and, consequently, to relax the condition $K = L - S + 1$. This way, observations are assimilated several times. This is done in the MDA IEnKS (Bocquet and Sakov, 2014). ~~But to~~ To keep the statistics consistent, at least in the linear/Gaussian case, the observations error covariances ~~are~~ should be adequately altered. ~~Because $S = 1$, the cost function is slightly modified between each assimilation. That is why the~~ Hence, the MDA IEnKS is ~~qualified as QS~~ truly a QS scheme and makes a good reference scheme for our numerical

experiments. However, the multiple assimilation of the observations introduces spurious correlations in the nonlinear/non-Gaussian case, which entail sub-optimality. A scheme similar to the ~~IEnKS-QS~~ IEnKS$_{\text{QS}}$ has also been successfully used in ~~Carrassi et al. (2017)~~ Carrassi et al. (2017) to compute model evidence. Indeed, the efficient computation of model evidence as an integral over the state space depended on the proper identification of a global maximum of the integrand. However, its implementation was based on the update of the ensemble whenever an observation batch is added to the cost function, which is not as numerically efficient as the scheme presented here.

The success of the QS minimization lies in the fact that, when an observation is successfully assimilated, the eMSE is reduced. Thus, the analysis probability mass concentrates around the true state. The analysis is then more likely to be in a neighborhood of the true state where the model is linear. The cost function non-quadraticity can then be increased by adding a new term in it. This is confirmed by the following argument. Let $P^{(q)}$ be the IEnKS asymptotic eMSE at the q-th step of a QS minimization. With the notations and assumptions of section 2.3, i.e. in a linear context, we have the recurrence relation:

$$\left(P^{(q+1)}\right)^{-1} = \left(P^{(q)}\right)^{-1} + \Sigma_{L_q+1}^{L_{q+1}}, \tag{38a}$$

$$P^{(-1)} = \alpha^{2S}\frac{\alpha^2-1}{\alpha^{2\left(L_{N_Q-1}+1\right)}}. \tag{38b}$$

Thus, $P^{(q+1)} < P^{(q)}$ and we can increase the cost function non-quadraticity by adding new terms in it as long as the propagation of errors does not exceed the bound

$$\alpha^{2L_{q+1}}P^{(q)} \leq B. \tag{39}$$

This implies

$$L_{q+1} \leq L_q + S_{\max}, \tag{40a}$$

$$L_0 \leq L_{N_Q-1} + S_{\max} - S, \tag{40b}$$

which yields $S \leq N_Q S_{\max}$. Therefore, the QS ~~minimization allows~~ minimizations allow for a N_Q times longer DAW.

Unfortunately, these QS minimizations are very expensive. Indeed, they add a third outer loop repeating N_Q GN minimizations. The GN iterations used to compute the intermediate starting points give unnecessary precision; all that is required for these starting points is to be in a neighborhood of $\mathbf{x}_{\infty S}$ where the model is almost linear. Thus, one can restrain the number of intermediate GN loops and save the full convergence to the last minimization. This is done in the *quasi-convergent* IEnKS (~~IEnKS-QC~~IEnKS$_{\text{QC}}$) with the parameters $(j_q)_{q<N_Q}$. They correspond to the numbers of GN loops in the intermediate QS minimizations. They are typically equal to 1 except for the last one. Algorithm 2 gives the pseudo code of the ~~IEnKS-QC~~ IEnKS$_{\text{QC}}$ and the right panel in Fig. 7 is a schematic for it.

4 Numerical experiments with low-order models

In the following, we perform numerical experiments with the Lorenz 1963 (L63) and Lorenz 1995 (L95) models. L95 has already been presented in section 2.4. L63 (Lorenz, 1963) is a simplified model for atmospheric convection. It is defined by

Algorithm 2 One cycle of the ~~IEnKS-QC~~IEnKS$_{QC}$

Require: $\mathbf{E}_0^{\mathrm{b}}$ the background ensemble at t_0; λ the inflation; $(L_i)_{q<N_Q}$ a list of DAW time indexes; $(j_q)_{q<N_Q}$ the number of intermediate GN loops; ε the finite differences step; δ GN end of loop parameter; $\mathbf{1} = (1,\dots,1)^{\mathrm{T}} \in \mathbb{R}^n$ and $\mathbf{I}$ is the identity of $\mathbb{R}^n$;

$\mathbf{w}_0^{\mathrm{a}} := \mathbf{0}$
$\bar{\mathbf{x}}_0^{\mathrm{b}} = \mathbf{E}_0^{\mathrm{b}}\mathbf{1}/n$
$\mathbf{X}_0^{\mathrm{b}} = \lambda\left(\mathbf{E}_0^{\mathrm{b}} - \bar{\mathbf{x}}_0^{\mathrm{b}}\mathbf{1}^{\mathrm{T}}\right)$
for $q = 0 \dots N_Q - 1$ **do**
  $j := 0$
  repeat
    $\bar{\mathbf{x}}_0^{\mathrm{a}} := \bar{\mathbf{x}}_0^{\mathrm{b}} + \mathbf{X}_0^{\mathrm{b}}\mathbf{w}_0^{\mathrm{a}}$
    $\mathbf{E}_0 := \bar{\mathbf{x}}_0^{\mathrm{a}}\mathbf{1}^{\mathrm{T}} + \epsilon\mathbf{X}_0^{\mathrm{b}}$
    $\nabla J := (n-1)\mathbf{w}_0^{\mathrm{a}}$
    $\tilde{\nabla}^2 J := (n-1)\mathbf{I}$
    $\mathbf{E}_K := \mathcal{M}^K(\mathbf{E}_0)$
    for $l = K, \dots, L_q$ **do**
      $\bar{\mathbf{y}}_l := \mathcal{H}(\mathbf{E}_l)\,\mathbf{1}/n$
      $\mathbf{Y}_l := \left(\mathcal{H}(\mathbf{E}_l) - \bar{\mathbf{y}}_l\mathbf{1}^{\mathrm{T}}\right)/\epsilon$
      $\nabla J := \nabla J - \mathbf{Y}_l^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y}_l - \bar{\mathbf{y}}_l)$
      $\tilde{\nabla}^2 J := \tilde{\nabla}^2 J - \mathbf{Y}_l^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{Y}_l$
      $\mathbf{E}_{l+1} := \mathcal{M}(\mathbf{E}_l)$
    end for
    solve $\tilde{\nabla}^2 J\delta\mathbf{w} := \nabla J$
    $\mathbf{w}_0^{\mathrm{a}} := \mathbf{w}_0^{\mathrm{a}} - \delta\mathbf{w}$
    $j := j + 1$
  until $||\delta\mathbf{w}|| \leq \delta$ or $j \geq j_q$
end for
$\mathbf{E}_0^{\mathrm{a}} := \mathbf{x}_0^{\mathrm{a}}\mathbf{1}^{\mathrm{T}} + \sqrt{n-1}\mathbf{X}_0^{\mathrm{b}}\tilde{\nabla}^2 J^{-1/2}$
$\mathbf{E}_S^{\mathrm{b}} := \mathcal{M}^S(\mathbf{E}_0^{\mathrm{a}})$

the ordinary differential equations:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \sigma(y - x), \tag{41a}$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \rho x - y - xz, \tag{41b}$$

$$\frac{\mathrm{d}z}{\mathrm{d}t} = xy - \beta z. \tag{41c}$$

These equations are integrated using a fourth-order Runge-Kutta scheme with a time-step of $\delta t = 0.01$ and $(\sigma, \rho, \beta) = (10, 28, 8/3)$. The dynamics of the L63 model are chaotic, with a largest Lyapunov exponent given by $\lambda \simeq 0.91$.

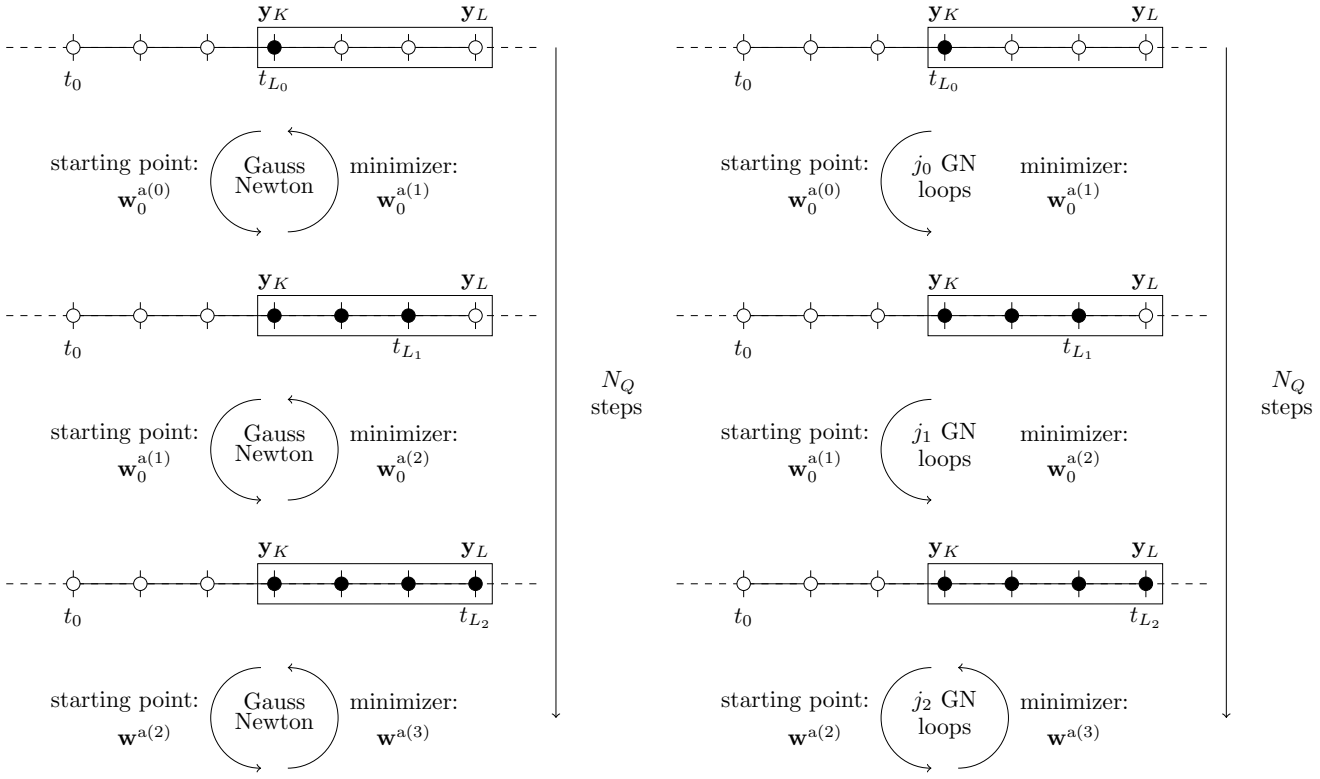

Figure 7. Schematics of an ~~IEnKS-QS~~ IEnKS$_{QS}$ minimization (on the left) and a ~~IEnKS-QC~~ IEnKS$_{QC}$ minimization (on the right). The rectangle contains the observations to be assimilated, the black dots represent the observations used in the current minimization. The "Gauss-Newton" surrounded by arrows represents the iterations of the Gauss-Newton procedure. The number of quasi-static steps is $N_Q = 3$. The flow of observations is controlled by the parameters $(L_0, L_1, L_2) = (3, 5, 6)$. For the QC IEnKS, the number of GN iterations is controlled by the (j_0, j_1, j_2) parameters.

Both models are assumed perfect. The truth run is generated from a random state space point. The initial ensemble is generated from the truth with $\mathbf{B} = \mathbf{I}_m$ where $m = 40, 3$ and $n = 20, 3$ for L95 and L63, respectively. Observation vectors are generated from the truth with $\mathcal{H} = \mathbf{R} = \mathbf{I}_m$ every $\Delta t = 0.05$ for L95 and every $\Delta t = 0.02$ for L63. A burn-in period of $5 \times 10^3 \times \Delta t$ is enforced in both cases.

The ~~algorithms~~ IEnKS parameters are $\varepsilon = 10^{-4}$, $\delta = 10^{-3}$, $j_{max} = 20$, $N_Q = 1$, $L_0 = L$~~for the IEnKS~~. For the ~~IEnKS-QS~~IEnKS$_{QS}$, the QS parameters are $\big(L_0, L_1, \ldots, L_{N_Q-1}\big) = K + (0, 1, \ldots, N_Q - 1) \times \frac{S-1}{N_Q-1}$. For the ~~IEnKS-QC~~IEnKS$_{QC}$, the QS parameters are the same and, in addition, $\big(j_0, \ldots, j_{N_Q-2}, j_{N_Q-1}\big) = (1, \ldots, 1, 20)$. Sampling errors are systematically accounted for using the IEnKS finite-size version (Bocquet, 2011; Bocquet and Sakov, 2012; Bocquet et al., 2015) which avoids the need for inflation and its costly tuning. Finally the aRMSE is averaged over a number of cycles which is determined by the num-

ber of observations assimilated. We use 5×10^5 observation vectors for L95 and 5×10^6 observation vectors for L63. ~~Unlike Goodliff et al. (2015)~~

Unlike Goodliff et al. (2015), our numerical experiments neither address increasing nonlinearity, nor do they ~~adress~~ address the use of climatological background error covariance matrices. Instead, we focus exclusively on the IEnKS performance dependence on the DAW key parameters L, S and the number N_Q of QS minimizations. Goodliff et al. (2015) merely numerically evaluates the QSVA approach with hybrid and EnVar techniques, similarly to Bocquet and Sakov (2013), and confirms the findings of Pires et al. (1996) albeit in a EnVar context. In terms of accuracy, Goodliff et al. (2015) show that the ensemble transform Kalman smoother (ETKS) outperforms all hybrid schemes in their numerical experiment. Since the IEnKS systematically outperforms the ETKS in all conditions (Bocquet and Sakov, 2013, 2014) as long as the DAW length is not excessively long (for a chaotic model), then one concludes that our RMSEs would be systematically equal or smaller that those reported for any hybrid scheme in Goodliff et al. (2015).

Figures 8, 9 show the aRMSE of the IEnKS and ~~IEnKS-QS~~ IEnKS$_{\text{QS}}$ for both L63 and L95 as a function of S and L. The smoothing and filtering performance of the IEnKS increases for small values of L, S then decreases for high values of L, S. This is due to the appearances of local minima. As noted by Goodliff et al. (2015) with L63, the QS variant allows to reach much longer DAWs, and improves the performance. However, ~~the~~ some limit of this method is visible with the L95 model when, for $L = 50$, best smoothing aRMSEs are reached for $S < 50$. However, the IEnKS$_{\text{QS}}$ filtering performance is invariant with L and improves with S as in the 4D-Var filtering performance of Fig. 3. As suggested by Pires et al. (1996), one can be tempted to estimate the useful DAW length due to past observations beyond which the performance gain is negligible. However, they estimated such length in a one-cycle 4D-Var context with a focus on the filtering RMSE. Hence, this useful DAW length is not directly relevant for a cycled IEnKS$_{\text{QS}}$. By contrast, here, a lot of observations have already been assimilated and condensed in the background approximation. Thus, the performance gain with the DAW length comes from the precision of this Gaussian background approximation; a precision that the linearized theory is not able to provide, as attested by the filtering ($l = L$) performance independence of Eq. (25) with the DAW parameters.

Figure 10 compares the smoothing aRMSE (first column), the filtering aRMSE (second column) and the number of ensemble propagations ~~of the IEnKS-QS~~ (third column) of the IEnKS$_{\text{QS}}$ ($N_Q = S, S = L$), the IEnKS ($S = L$) and the IEnKS-MDA ($S = 1$), for both L63 and L95 as a function of L. The number of ensemble propagations is the total number of ensemble propagations in units of Δt divided by the total number of assimilated observation vectors. For $L < 20$, all three algorithms show smoothing and filtering performance improvements with L. For $L > 20$, the IEnKS filtering and smoothing RMSE increase because of the multiple local extrema. For $L < 40$, the ~~IEnKS-QS~~ IEnKS$_{\text{QS}}$ has smaller aRMSEs than the IEnKS-MDA. ~~The IEnKS-QS is SDA so~~ Because the IEnKS$_{\text{QS}}$ is SDA by design, it does not suffer from suboptimality related to multiple assimilations and nonlinearity. Moreover, the ~~IEnKS-QS~~ IEnKS$_{\text{QS}}$ always requires less propagations of the ensemble, which improves the computational cost. However, for $L > 40$ with the L95 model, the quasi-static approach cannot sustain the ~~non-linearity~~ nonlinearity anymore and the ~~IEnKS-QS~~ IEnKS$_{\text{QS}}$ aRMSE degrades. Hence, the IEnKS-MDA $S = 1$ has still the best performance. With the L63 model, the IEnKS$_{\text{QS}}$ is always better than the IEnKS-MDA suggesting that L could still be increased.

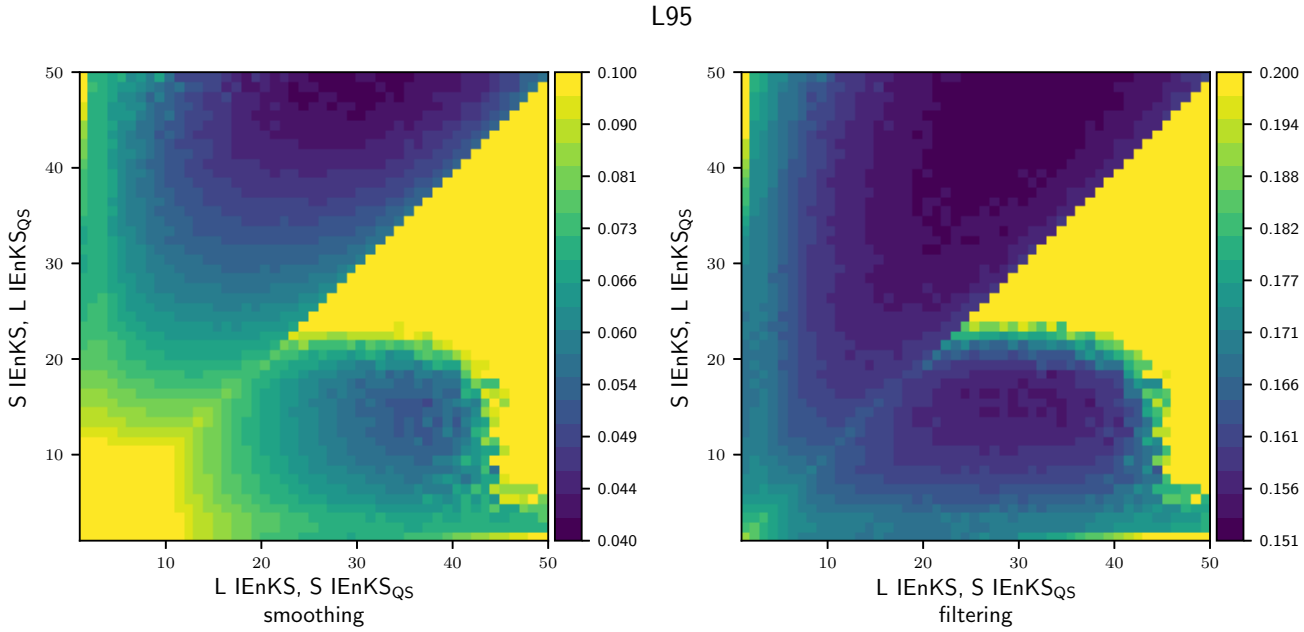

Figure 8. IEnKS (lower triangles) and ~~IEnKS-QS~~ IEnKS$_{QS}$ (upper triangles, $N_Q = S$) smoothing and filtering aRMSEs as a function of L and S with the L95 model. The L parameter is on the abscissa axis for the IEnKS and on the ordinate axis for the ~~IEnKS-QS~~IEnKS$_{QS}$. The S parameter is on the abscissa axis for the ~~IEnKS-QS~~ IEnKS$_{QS}$ and on the ordinate axis for the IEnKS. For readability, smoothing RMSE beyond 0.10 and filtering RMSE beyond 0.20 are in the same color and the scale is logarithmic.

Figure 11 compares the smoothing aRMSE (first column), the filtering aRMSE (second column) and the number of ensemble propagations ~~of the IEnKS-QS and IEnKS-QC~~ (third column) of the IEnKS$_{QS}$ and IEnKS$_{QC}$ as a function of the N_Q parameter. The ~~IEnKS-QS~~ IEnKS$_{QS}$ aRMSE decreases quickly after a point for both algorithms and for both models. ~~For the IEnKS-QS~~ Before this point, the algorithms fail to find the global minimum and the RMSE is close to the climatological variance. After this point, the algorithms succeed in finding the global minimum and the RMSE is low. For the IEnKS$_{QS}$ with the L95 model, this point can be estimated using results of section 3 by $S/S_{\max} \simeq 3.6$, in remarkable agreement with the experiments. This point comes later for the ~~IEnKS-QC~~ IEnKS$_{QC}$ but it demands less ensemble propagations making this algorithm numerically more efficient than the ~~IEnKS-QS~~IEnKS$_{QS}$. However those ensemble propagations have different behavior for the L95 and L63 models when the minimizations fail to find the global minimum. For the L95 model, the number of ensemble propagations is high meaning that the minimization takes a lot of iterations and fails to converge. For the L63 model, the number of ensemble propagations is low indicating that minimizations converge but to a non-global extrema.

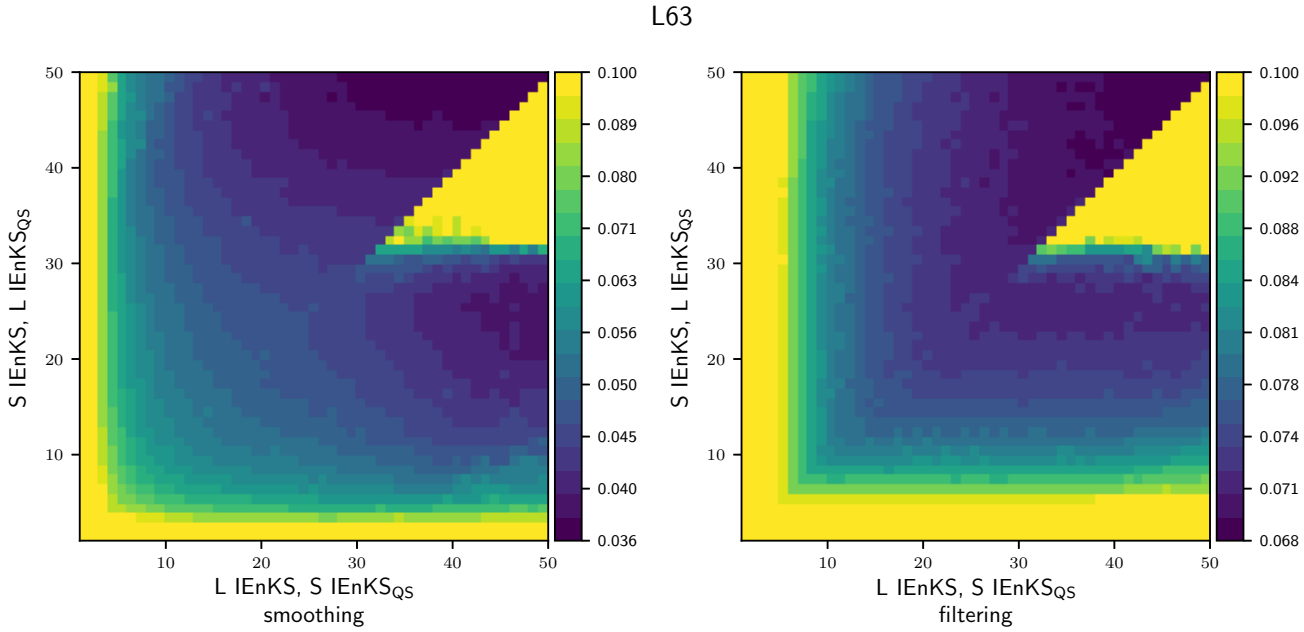

Figure 9. IEnKS (lower triangles) and ~~IEnKS-QS~~ IEnKS$_{QS}$ (upper triangles, $N_Q = S$) smoothing and filtering aRMSEs as a function of L and S with the L63 model. The L parameter is on the abscissa axis for the IEnKS and on the ordinate axis for the ~~IEnKS-QS~~IEnKS$_{QS}$. The S parameter is on the abscissa axis for the ~~IEnKS-QS~~ IEnKS$_{QS}$ and on the ordinate axis for the IEnKS. For readability, smoothing and filtering RMSE beyond 0.10 are in the same color and the scale is logarithmic.

5 Conclusions

In this paper, we have extended the study of Pires et al. (1996) on quasi-static variational data assimilation, focused on 4D-Var technique, to cycled data assimilation schemes and specifically four-dimensional nonlinear ensemble ~~variationnal~~ variational techniques, an exemplar of which being the iterative ensemble Kalman smoother (IEnKS).

The long term impact of cycling has been first investigated theoretically in a linear context for 4D-Var and the IEnKS, then numerically for the IEnKS in a nonlinear context. The way information is propagated between data assimilation cycles indeed makes up for the difference between 4D-Var and the IEnKS. Both reveal performance improvements with the DAW parameter S, which counts the number of observation ~~batches~~ vectors within the DAW, as well as the time shift between cycles. This is a consequence of the Gaussian background approximation: the larger S is, the less the assimilation relies on it.

However, it is observed that this improvement has a limit in the chaotic, perfect model case. The cost function global minimum basin of attraction appears to ~~shrinks with~~ shrink with increasing L. This causes the Gauss-Newton procedure to miss the cost function global minimum, which deteriorates the assimilation performance.

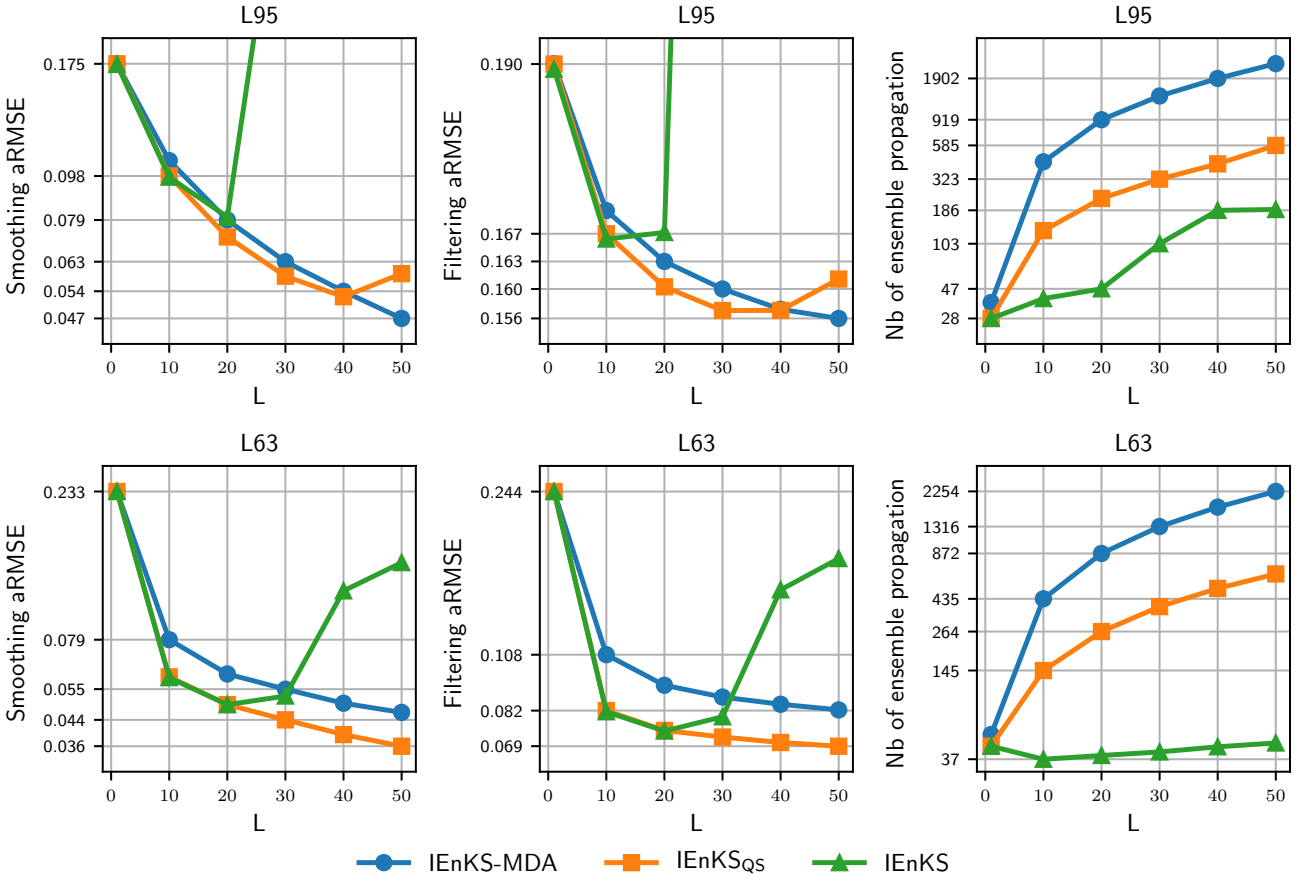

Figure 10. IEnKS-MDA ($S = 1$), ~~IEnKS-QS~~ IEnKS$_{QS}$ and IEnKS ($S = L$, $N_Q = S$) filtering and smoothing aRMSEs and number of ensemble propagations as a function of L for the L95 and L63 models (logarithmic scale). The ensemble propagations are in units of Δt. For instance, with $L = 50$, a single ensemble propagation through the DAW counts for 50.

Quasi-static minimizations lead slowly but surely to the global minimum by repeated cost function minimizations. As the DAW length L is gradually increased, the starting point of the minimization remains in the global minimum basin of attraction. For most S, L couples, the quasi-static IEnKS turns out ~~as~~ to be a more accurate substitute for the multiple data assimilation IEnKS (IEnKS-MDA).

Unfortunately, this method (IEnKS$_{QS}$) adds an outer loop which could significantly increase the numerical cost. Precision on intermediate minima being superfluous, one can limit the intermediate Gauss-Newton number of loops. The unavoidable space increments required to minimize the non-quadratic cost function are thus reported in time in the *quasi-convergent* IEnKS (IEnKS$_{QC}$).

We did not focus on the applicability of the methods to high-dimensional and imperfect models. In particular, we ~~consider~~ considered very long DAWs, which, even if of high mathematical interest or for low-order reliable models, is less relevant for

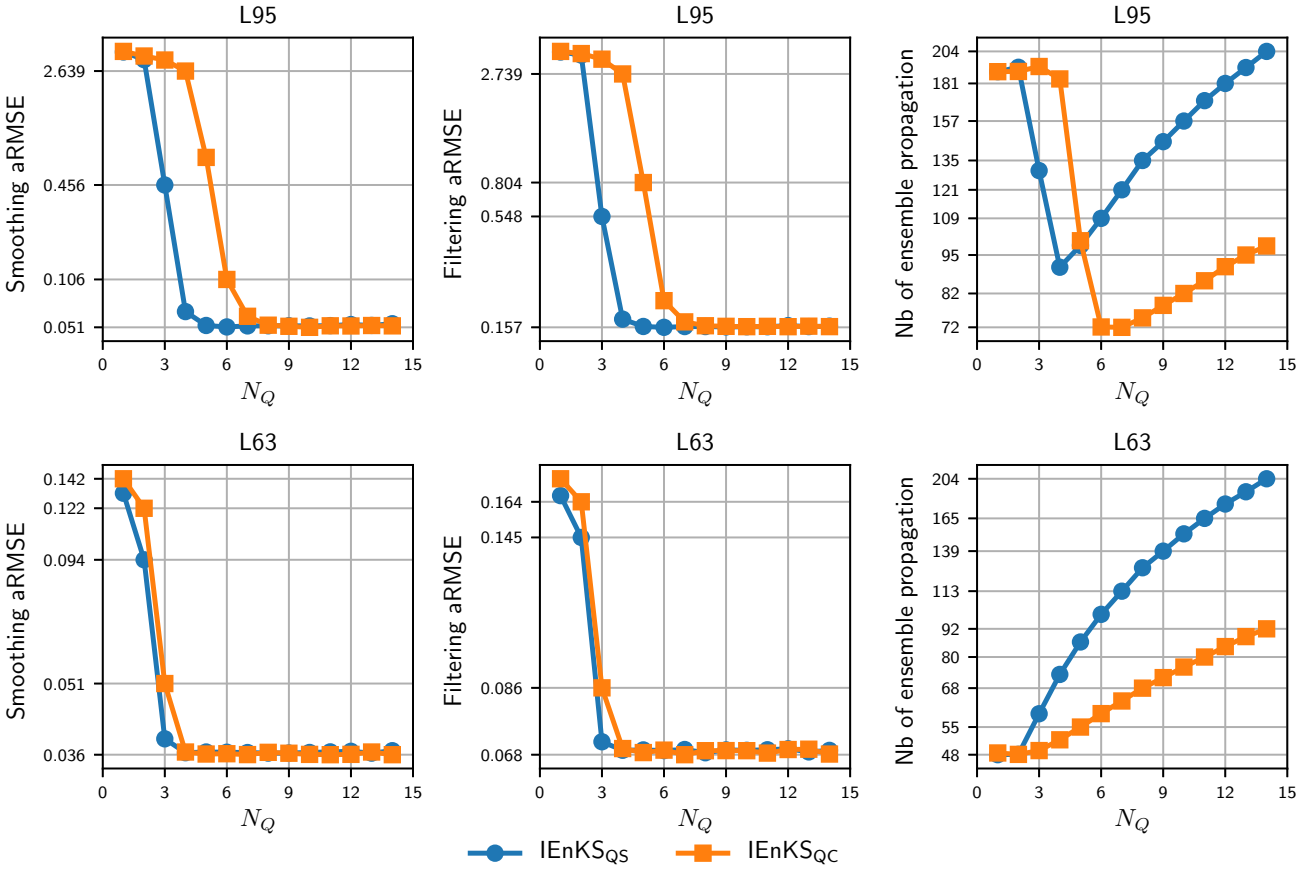

Figure 11. ~~IEnKS-QS~~ IEnKS$_{QS}$ and ~~IEnKS-QC~~ IEnKS$_{QC}$ ($S = L = 50$) filtering and smoothing aRMSEs and number of ensemble propagations as a function of ~~L~~ N_Q for the L95 and L63 models (logarithmic scale). The ensemble propagations are in units of Δt.

significantly noisy models. ~~An~~ However, we know from Swanson et al. (1998), that the perfect model results are expected to extend to the imperfect model case provided that the growth rate of the model error is similar to that of the leading Lyapunov vectors of the model. This is likely to apply as well to a (strong-constraint) IEnKS$_{QS}$. Beyond, an extension to this work would therefore ~~consists~~ consist in investigating the same ideas but using a weak constraint 4D-Var and IEnKS ~~(Trémolet, 2006; Sakov et al., 2018)~~(Trémolet, 2006; Sakov and Bocquet, 2018; Sakov et al., 2018).

Appendix A: Performance of 4D-Var in the linear, univariate case

The objective of this appendix is to establish a recurrence relation between the 4D-Var eMSE of each cycle. From this relation we will get an expression for the 4D-Var asymptotic eMSE.

We assume $m = 1$ and $\mathcal{M}^l(x) = \alpha^l x$. At the k-th cycle, the gradient and Hessian of the 4D-Var cost function Eq. (7) are ÷

$$\nabla J\left(x_{kS}; y_{:}, x^{\mathrm{b}}_{kS}\right) = -\frac{1}{b}\left(x^{\mathrm{b}}_{kS} - x_{kS}\right) - \frac{h}{r}\sum_{l=K}^{L}\alpha^l\left(y_{kS+l} - h\alpha^l x_{kS}\right), \tag{A1}$$

$$\nabla^2 J\left(x_{kS}; y_{:}, x^{\mathrm{b}}_{kS}\right) = \frac{1}{b} + \frac{h^2}{r}\sum_{l=K}^{L}\alpha^{2l}, \tag{A2}$$

where $y_{kS+L:kS+K}$ is temporarily denoted $y_{:}$. Because of the operators' linearity, J is quadratic with respect to x_{kS} and convex. Hence, its minimizer x^{a}_{kS} exists and is characterized by the null gradient equation:

$$\nabla J\left(x^{\mathrm{a}}_{kS}; y_{:}, x^{\mathrm{b}}_{kS}\right) = 0. \tag{A3}$$

With an exact Taylor expansion around the state x_{kS} we obtain:

$$0 = \nabla J\left(x_{kS}; y_{:}, x^{\mathrm{b}}_{kS}\right) + \nabla^2 J\left(x_{kS}; y_{:}, x^{\mathrm{b}}_{kS}\right) \times \left(x^{\mathrm{a}}_{kS} - x_{kS}\right). \tag{A4}$$

Note that $\nabla^2 J\left(x_{kS}; y_{:}, x^{\mathrm{b}}_{kS}\right)$, given by Eq. (A2), is not random and does not depend on $x_{kS}, y_{:}, x^{\mathrm{b}}_{kS}$. That is why it is simply noted $\nabla^2 J$. Using Eqs (1,A4), we have ÷

$$\nabla^2 J \cdot \left(x^{\mathrm{a}}_{kS} - x_{kS}\right) = \frac{1}{b}\left(x^{\mathrm{b}}_{kS} - x_{kS}\right) + \frac{h}{r}\sum_{l=K}^{L}\alpha^l \varepsilon_{kS+l}, . \tag{A5}$$

The random variable $x^{\mathrm{b}}_{kS} - x_{kS} = \alpha^S\left(x^{\mathrm{a}}_{(k-1)S} - x_{(k-1)S}\right)$ is independent from the errors $\varepsilon_{kS+K}, \ldots, \varepsilon_{kS+L}$. Thus, taking the expectation of the square of Eq. (A5) gives the following expression for the eMSE $P^{\text{4D-Var}}_{kS}$ of 4D-Var at time t_{kS}:

$$P^{\text{4D-Var}}_{kS} = \mathbb{E}\left[\left(x_{kS} - x^{\mathrm{a}}_{kS}\right)^2\right], = \left(\nabla^2 J\right)^{-2}\left(\frac{\alpha^{2S}}{b^2}P^{\text{4D-Var}}_{(k-1)S} + \frac{h^2}{r}\sum_{l=K}^{L}\alpha^{2l}\right), \tag{A6a}$$

$$P^{\text{4D-Var}}_{-S} = b. \tag{A6b}$$

Introducing the notation ÷

$$\Sigma^L_K = \frac{h^2}{r}\sum_{l=K}^{L}\alpha^{2l} = \frac{h^2}{r}\alpha^{2K}\frac{\alpha^{2S}-1}{\alpha^2-1}, \tag{A7}$$

$$\Delta = \frac{\alpha^{2S}}{\left(1 + b\Sigma^L_K\right)^2}, \tag{A8}$$

we obtain

$$P_{kS}^{\text{4D-Var}} = \Delta P_{(k-1)S}^{\text{4D-Var}} + \frac{\Sigma_K^L}{\left(\frac{1}{b} + \Sigma_K^L\right)^2}. \tag{A9}$$

Thus, $\left(P_{kS}^{\text{4D-Var}}\right)_k$ is an arithmetico-geometric sequence. Its limit $P_{\infty S}^{\text{4D-Var}}$ depends on the value of Δ in the following ~~manner~~way:

$$P_{\infty S}^{\text{4D-Var}} = \begin{cases} \infty & \text{if } \Delta \geq 1, \\ \frac{b^2 \Sigma_K^L}{\alpha^{2S}} \frac{\Delta}{1-\Delta} & \text{otherwise.} \end{cases} \tag{A10}$$

The generalization to any asymptotic eMSE with lag $L - l$ is straightforward:

$$P_{\infty S+l}^{\text{4D-Var}} = \begin{cases} \infty & \text{if } \Delta \geq 1, \\ \frac{b^2 \Sigma_K^L}{\alpha^{2(S-l)}} \frac{\Delta}{1-\Delta} & \text{otherwise.} \end{cases} \tag{A11}$$

In the multivariate, diagonal case the algebra can be conducted on each direction independently. The eMSE in this case is the sum of the univariate eMSEs of each direction.

Appendix B: Performance of the IEnKS in the linear, univariate case

The objective of this appendix is to establish a recurrence relation between the IEnKS eMSE of each cycle. From this relation we will get an expression for the IEnKS asymptotic eMSE.

First, it is ~~proved~~ proven by recurrence that for all $k \geq 0$, $G\left(x_{kS}|y_{kS+L:K}\right)$ is Gaussian with moments ÷

$$\mathbb{E}\left[x_{kS}|y_{kS+L:K}\right] = \bar{x}_{kS}^{\text{a}}, \tag{B1}$$

$$\mathbb{V}\left[x_{kS}|y_{kS+L:K}\right] = \left(X_{kS}^{\text{a}}\right)^2, \tag{B2}$$

where $\mathbb{V}$ is the variance of a random variable and $\bar{x}_{kS}^{\text{a}}, X_{kS}^{\text{a}}$ are defined by Eqs (15,16). Because of the assumptions Eqs (24c,24d) with Eq. (13) one gets ÷

$$G\left(x_0|y_{L:K}\right) = J\left(w_0; y_{L:K}, \mathbf{E}_0^{\text{b}}\right) + c_0, \tag{B3}$$

where c_0 is a constant independent from x_0 and w_0. Hence, $G\left(x_0|y_{L:K}\right)$ is Gaussian and its moments are given by Eq. (B1,B2) with $k = 0$. Now, assume $G\left(x_{kS}|y_{kS+L:K}\right)$ is Gaussian with moments given by Eqs (B1,B2) for a $k \geq 0$. Because $x_{(k+1)S} = \mathcal{M}^S\left(x_{kS}\right)$ with $\mathcal{M}$ affine, one gets

$$\mathbb{E}\left[x_{(k+1)S}|y_{kS+L:K}\right] = \mathcal{M}^S\left(\bar{x}_{kS}^{\text{a}}\right) = \bar{x}_{(k+1)S}^{\text{b}}, \tag{B4}$$

$$\mathbb{V}\left[x_{(k+1)S}|y_{kS+L:K}\right] = \alpha^{2S}\left(X_{kS}^{\text{a}}\right)^2 = X_{(k+1)S}^{\text{b}}, \tag{B5}$$

using the conditional expectation properties and Eqs (24a,24b). This result together with Eq. (6) and assumption Eq. (13) yields ÷

$$G\left(x_{(k+1)S}|y_{(k+1)S+L:K}\right) = J\left(w_{(k+1)S}; y_{(k+1)S+L:(k+1)S+K}, \mathbf{E}_{(k+1)S}^{\text{b}}\right) + c_{k+1}, \tag{B6}$$

where c_{k+1} is a constant independent from $x_{(k+1)S}$ and $w_{(k+1)S}$. Hence $G\left(x_{(k+1)S}|y_{(k+1)S+L:K}\right)$ is Gaussian with moments given by Eqs (B1,B2).

The conditional variance Eq. (B2) is therefore related to the IEnKS performance by the total law of expectation:

$$\begin{aligned} P_{kS}^{\text{IEnKS}} &= \mathbb{E}\left[\left(\bar{x}_{kS}^{\text{a}} - x_{kS}\right)^2\right], \\ &= \mathbb{E}\left[\mathbb{E}\left[\left(\bar{x}_{kS}^{\text{a}} - x_{kS}\right)^2 | y_{kS+L:K}\right]\right], \\ &= \mathbb{E}\left[\mathbb{V}\left[x_{kS}|y_{kS+L:K}\right]\right], \\ &= \mathbb{E}\left[\left(X_{kS}^{\text{a}}\right)^2\right]. \end{aligned} \tag{B7}$$

Then, from Eq. (16) we get the recurrence relation:

$$\left(X_{(k+1)S}^{\text{a}}\right)^{-2} = \alpha^{-2S}\left(X_{kS}^{\text{a}}\right)^{-2} + \frac{h^2}{r}\sum_{l=K}^{L}\alpha^{2l}, \tag{B8a}$$

$$\left(X_{0}^{\text{a}}\right)^{-2} = \frac{1}{b} + \frac{h^2}{r}\sum_{l=K}^{L}\alpha^{2l}. \tag{B8b}$$

Thus, X_{kS}^{a} is not random and $P_{kS}^{\text{IEnKS}} = \left(X_{kS}^{\text{a}}\right)^2$. ~~This last equation with Eq.~~Equation (B8) ~~tell~~ tells that the sequence of inverse IEnKS eMSEs is arithmetico-geometric:

$$\left(P_{kS}^{\text{IEnKS}}\right)^{-1} = \alpha^{-2S}\left(P_{(k-1)S}^{\text{IEnKS}}\right)^{-1} + \Sigma_K^L, \tag{B9a}$$

$$\left(P_{-S}^{\text{IEnKS}}\right)^{-1} = b^{-1}, \tag{B9b}$$

where the notation Eq. (A7) has been used. Properties of arithmetico-geometric sequences allow to obtain the IEnKS asymptotic eMSE:

$$P_{\infty S}^{\text{IEnKS}} = \begin{cases} 0 & \text{if } |\alpha| \leq 1, \\ \frac{r}{h^2\alpha^{2L}}\frac{\alpha^2-1}{\alpha^2} & \text{otherwise,} \end{cases} \tag{B10}$$

and the generalization to asymptotic eMSEs with lag $L-l$ is straightforward:

$$P_{\infty S+l}^{\text{IEnKS}} = \begin{cases} 0 & \text{if } |\alpha| \leq 1, \\ \frac{r}{h^2\alpha^{2(L-l)}}\frac{\alpha^2-1}{\alpha^2} & \text{otherwise.} \end{cases} \tag{B11}$$

Let us now show that the IEnKS eMSE is optimal. Let $x_{kS}^{\text{a}}\left(y_{kS+L:K}\right)$ be the 4D-Var analysis or any other function of $y_{kS+L:K}$. A bias-variance decomposition (e.g., Bishop, 2006) of this estimator yields ~~:~~

$$\begin{aligned} &\mathbb{E}_{x_{kS},y_{kS+L:K}}\left[\left(x_{kS} - x_{kS}^{\text{a}}\right)^2\right] \\ =&\mathbb{E}_{y_{kS+L:K}}\left[\mathbb{V}_{x_{kS}}\left[x_{kS}|y_{kS+L:K}\right]\right] \\ &+\mathbb{E}_{y_{kS+L:K}}\left[\left(\mathbb{E}_{x_{kS}}\left[x_{kS}|y_{kS+L:K}\right] - x_{kS}^{\text{a}}\right)^2\right]. \end{aligned} \tag{B12}$$

Replacing the moments with Eqs (B1,B2) yields ÷

$$P_{kS}^{\text{4D-Var}} = P_{kS}^{\text{IEnKS}} + \mathbb{E}\left[\left(\bar{x}_{kS}^{\text{a}} - x_{kS}^{\text{a}}\right)^2\right].$$
$$\geq P_{kS}^{\text{IEnKS}}. \tag{B13}$$

In the multivariate, diagonal case the algebra can be conducted on each direction independently. Thus, the eMSE in this case is the sum of the univariate eMSEs of each direction.

Appendix C: Expression of the averaged cost function

The IEnKS averaged cost function $J_{\infty S}$ is the N goes to ∞ limit of ÷

$$\frac{1}{N}\sum_{k=0}^{N-1} J\left(\mathbf{w}; \mathbf{y}_{kS+L:kS+K}, \mathbf{E}_{kS}^{\text{b}}\right) = \frac{1}{2}\|\mathbf{w}\|^2 +$$
$$\frac{1}{2N}\sum_{l=K}^{L}\sum_{k=0}^{N-1}\left\|\mathbf{y}_{kS+l} - \mathcal{H}\circ\mathcal{M}^l\left(\bar{\mathbf{x}}_{kS}^{\text{b}} + \mathbf{X}_{kS}^{\text{b}}\mathbf{w}\right)\right\|_{\mathbf{R}^{-1}}^2. \tag{C1}$$

Expanding the squared norm around $\mathcal{H}\circ\mathcal{M}^l\left(\mathbf{x}_{kS}\right)$ using Eq. (1) gives ÷

$$\frac{1}{2}\left\|\mathbf{y}_{kS+l} - \mathcal{H}\circ\mathcal{M}^l\left(\bar{\mathbf{x}}_{kS}^{\text{b}} + \mathbf{X}_{kS}^{\text{b}}\mathbf{w}\right)\right\|_{\mathbf{R}^{-1}}^2 =$$
$$\frac{1}{2}\|\varepsilon_{kS+l}\|_{\mathbf{R}^{-1}}^2 + \frac{1}{2}\|\delta\mathbf{y}_{kS+l}\|_{\mathbf{R}^{-1}}^2 + \varepsilon_{kS+l}^{\text{T}}\mathbf{R}^{-1}\delta\mathbf{y}_{kS+l}, \tag{C2}$$

where $\delta\mathbf{y}_{kS+l} = \mathcal{H}\circ\mathcal{M}^l\left(\mathbf{x}_{kS}\right) - \mathcal{H}\circ\mathcal{M}^l\left(\bar{\mathbf{x}}_{kS}^{\text{b}} + \mathbf{X}_{kS}^{\text{b}}\mathbf{w}\right)$. We assume that there exists random variables $(\varepsilon_{\infty S}, \mathbf{x}_{\infty S}, \mathbf{E}_{\infty S}^{\text{a}})$ whose distribution is invariant and ergodic with respect to the shift transformation:

$$T : (\varepsilon_{kS}, \mathbf{x}_{kS}, \mathbf{E}_{kS}^{\text{a}}) \mapsto$$
$$\left(\varepsilon_{(k+1)S}, \mathbf{x}_{(k+1)S}, \mathbf{E}_{(k+1)S}^{\text{a}}\right). \tag{C3}$$

Then because the $(\varepsilon_{kS})_k$ are mutually independent, independent from the $(\mathbf{x}_{kS}, \mathbf{E}_{kS}^{\text{a}})_k$ and identically distributed, $p\left(\varepsilon_{\infty S}, \mathbf{x}_{\infty S}, \mathbf{E}_{\infty S}^{\text{a}}\right) = p\left(\varepsilon_0\right)p\left(\mathbf{x}_{\infty S}, \mathbf{E}_{\infty S}^{\text{a}}\right)$. By Birkhoff's ergodic theorem (see Walters, 1982) we get ÷

$$\frac{1}{N}\sum_{k=0}^{N-1}\|\varepsilon_{kS+l}\|_{\mathbf{R}^{-1}}^2 \underset{N\to\infty}{\to} \mathbb{E}\left[\|\varepsilon_0\|_{\mathbf{R}^{-1}}^2\right] = d, \tag{C4}$$
$$\frac{1}{N}\sum_{k=0}^{N-1}\varepsilon_{kS+l}^{\text{T}}\delta\mathbf{y}_{kS+l} \underset{N\to\infty}{\to} \mathbb{E}\left[\varepsilon_0^{\text{T}}\right]^{\text{T}}\mathbb{E}\left[\delta\mathbf{y}_{\infty S+l}\right] = 0, \tag{C5}$$
$$\frac{1}{N}\sum_{k=0}^{N-1}\|\delta\mathbf{y}_{kS+l}\|_{\mathbf{R}^{-1}}^2 \underset{N\to\infty}{\to} \mathbb{E}\left[\|\delta\mathbf{y}_{\infty S+l}\|_{\mathbf{R}^{-1}}^2\right], \tag{C6}$$

where $\delta\mathbf{y}_{\infty S+l} = \mathcal{H}\circ\mathcal{M}^l\left(\mathbf{x}_{\infty S}\right) - \mathcal{H}\circ\mathcal{M}^l\left(\bar{\mathbf{x}}^{\mathrm{b}}_{\infty S} + \mathbf{X}^{\mathrm{b}}_{\infty S}\mathbf{w}\right)$ and $\bar{\mathbf{x}}^{\mathrm{b}}_{\infty S}, \mathbf{X}^{\mathrm{b}}_{\infty S}$ are respectively the mean and normalized anomaly of $\mathcal{M}^S\left(\mathbf{E}^{\mathrm{a}}_{\infty S}\right)$. Finally,

$$J_{\infty S}\left(\mathbf{w}\right) = \frac{1}{2}\left\|\mathbf{w}\right\|^2 + \frac{dS}{2} + \frac{1}{2}\sum_{l=K}^{L}\mathbb{E}\left[\left\|\delta\mathbf{y}_{\infty S+l}\right\|^2_{\mathbf{R}^{-1}}\right]. \tag{C7}$$

Acknowledgements. The authors are grateful to S. Gürol for her comments and suggestions on the manuscript. CEREA is a member of Institut Pierre-Simon Laplace (IPSL).

References

Asch, M., Bocquet, M., and Nodet, M.: Data assimilation: methods, algorithms, and applications, Fundamentals of Algorithms, Society for Industrial and Applied Mathematics, 2016.

Bishop, C.: Pattern Recognition and Machine Learning, Information Science and Statistics, Springer-Verlag New York, 2006.

Björck, Å.: Numerical methods for least squares problems, Society for Industrial and Applied Mathematics, doi:10.1137/1.9781611971484, 1996.

Bocquet, M.: Ensemble Kalman filtering without the intrinsic need for inflation, Nonlin. Processes Geophys., 18, 735–750, doi:10.5194/npg-18-735-2011, 2011.

Bocquet, M.: Localization and the iterative ensemble Kalman smoother, Q. J. R. Meteorol. Soc., 142, 1075–1089, doi:10.1002/qj.2711, 2016.

Bocquet, M. and Carrassi, A.: Four-dimensional ensemble variational data assimilation and the unstable subspace, Tellus A, 69, 1304 504, doi:10.1080/16000870.2017.1304504, 2017.

Bocquet, M. and Sakov, P.: Combining inflation-free and iterative ensemble Kalman filters for strongly nonlinear systems, Nonlin. Processes Geophys., 19, 383–399, doi:10.5194/npg-19-383-2012, 2012.

Bocquet, M. and Sakov, P.: Joint state and parameter estimation with an iterative ensemble Kalman smoother, Nonlin. Processes Geophys., 20, 803–818, doi:10.5194/npg-20-803-2013, 2013.

Bocquet, M. and Sakov, P.: An iterative ensemble Kalman smoother, Q. J. R. Meteorol. Soc., 140, 1521–1535, doi:10.1002/qj.2236, 2014.

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

Carrassi, A., Bocquet, M., Hannart, A., and Ghil, M.: Estimating model evidence using data assimilation, Q. J. R. Meteorol. Soc., 143, 866–880, doi:10.1002/qj.2972, 2017.

Cosme, E., Verron, J., Brasseur, P., Blum, J., and Auroux, D.: Smoothing Problems in a Bayesian Framework and Their Linear Gaussian Solutions, Mon. Wea. Rev., 140, 683–695, doi:10.1175/MWR-D-10-05025.1, 2012.

Goodliff, M., Amezcua, J., and Van Leeuwen, P. J.: Comparing hybrid data assimilation methods on the Lorenz 1963 model with increasing non-linearity, Tellus A, 67, 26 928, doi:10.3402/tellusa.v67.26928, 2015.

Judd, K., Smith, L., and Weisheimer, A.: Gradient free descent: shadowing, and state estimation using limited derivative information, Physica D, 190, 153–166, doi:http://doi.org/10.1016/j.physd.2003.10.011, 2004.

Lorenc, A.: Four-dimensional variational data assimilation, in: Advanced Data Assimilation for Geosciences, Lecture Notes of the Les Houches School of Physics: Special Issue, June 2012, Oxford University Press, Les Houches, France, 2014.

Lorenz, E.: Deterministic Nonperiodic Flow, J. Atmos. Sci., 20, 130–141, doi:10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2, 1963.

Lorenz, E. and Emanuel, K.: Optimal Sites for Supplementary Weather Observations: Simulation with a Small Model, J. Atmos. Sci., 55, 399–414, doi:10.1175/1520-0469(1998)055<0399:OSFSWO>2.0.CO;2, 1998.

Miller, R. N., Ghil, M., and Gauthiez, F.: Advanced Data Assimilation in Strongly Nonlinear Dynamical Systems, J. Atmos. Sci., 51, 1037–1056, doi:10.1175/1520-0469(1994)051<1037:ADAISN>2.0.CO;2, 1994.

Pires, C., Vautard, R., and Talagrand, O.: On extending the limits of variational assimilation in nonlinear chaotic systems, Tellus A, 48, 96–121, doi:10.3402/tellusa.v48i1.11634, 1996.

Sakov, P. and Bocquet, M.: Asynchronous data assimilation with the EnKF in presence of additive model error, Tellus A, 70, 1414 545, doi:10.1080/16000870.2017.1414545, 2018.

Sakov, P., Haussaire, J.-M., and Bocquet, M.: An iterative ensemble Kalman filter in presence of additive model error, Q. J. R. Meteorol. Soc., 0, 0–0, doi:10.1002/qj.3213, accepted for publication, 2018.

Swanson, K., Vautard, R., and Pires, C.: Four-dimensional variational assimilation and predictability in a quasi-geostrophic model, Tellus A, 50, 369–390, doi:10.1034/j.1600-0870.1998.t01-4-00001.x, 1998.

Trémolet, Y.: Accounting for an imperfect model in 4D-Var, Q. J. R. Meteorol. Soc., 132, 2483–2504, doi:10.1256/qj.05.224, 2006.

Trevisan, A., D'Isidoro, M., and Talagrand, O.: Four-dimensional variational assimilation in the unstable subspace and the optimal subspace dimension, Q. J. R. Meteorol. Soc., 136, 487–496, doi:10.1002/qj.571, 2010.

Walters, P.: An Introduction to Ergodic Theory, Graduate texts in mathematics, Springer-Verlag New York, 1982.

Ye, J., Rey, D., Kadakia, N., Eldridge, M., Morone, U., Rozdeba, P., Abarbanel, H., and Quinn, J.: Systematic variational method for statistical nonlinear state and parameter estimation, Phys. Rev. E, 92, 052 901, doi:10.1103/PhysRevE.92.052901, 2015.