Controlling model error of underdamped forecast models in sparse observational networks using a variance limiting Kalman filter

Lewis Mitchell [a] * and Georg A. Gottwald [a]*
[a]*School of Mathematics and Statistics, University of Sydney, NSW 2006, Australia*
*Correspondence to: E-mail: georg.gottwald@sydney.edu.au; lewism@maths.usyd.edu.au

The problem of controlling covariance overestimation due to underdamped forecast models and sparsity of the observational network in an ensemble Kalman filter setting is considered. It is shown in a variational setting that limiting the analysis error covariance to stay below the climatological value and driving the mean towards the climatological mean for the unobserved variables can significantly improve analysis skill over standard ensemble Kalman filters. These issues are explored for a Lorenz-96 system. It is shown that for large observation intervals the climatological information assures that the statistical properties of the unobserved variables are recovered, providing superior analysis skill. This skill improvement is increased for larger observational noise.

Key Words: Data assimilation; Ensemble Kalman filter; model error

Received . . .

Citation: . . .

1. Introduction

To account for uncertainties in both the initial conditions and forecast model and to combat the associated forecast error and flow-dependent predictability, ensemble methods have become popular for producing numerical weather forecasts (Molteni and Palmer 1993; Toth and Kalnay 1993) and for performing data assimilation (Evensen 1994). The idea behind ensemble based methods is that the nonlinear chaotic dynamics of the underlying forecast model and the associated sensitivity to initial conditions cause an ensemble of trajectories to explore sufficiently large parts of the phase space in order to deduce meaningful statistical properties of the dynamics. However, all currently operational ensemble systems are underdispersive, meaning that the true atmospheric state is on average outside the statistically expected range of the forecast or analysis (e.g. Buizza *et al.* (2005); Hamill and Whitaker (2011)). This detrimental underdispersiveness can be linked to two separate sorts of model error: a *dynamical* model error due to a misrepresentation of the physics of the forecast model and a *numerical* model error due to the choice of the numerical method used to simulate those forecast models.
It has been argued (Palmer 2001) that dynamical model error is largely due to a misrepresentation of unresolved sub-grid scale processes. The very active field of stochastic parametrization is aimed at reintroducing statistical fluctuations of the unresolved degrees of freedom and a direct coupling between the resolved and unresolved degrees of freedom into the equations of motion (see Palmer and Williams (2010) for a recent review on current trends). Indeed, using a low-dimensional toy model Mitchell and Gottwald (2012) showed that stochastic parametrizations can lead to superior skill performance in ensemble data assimilation schemes in situations where the dynamics involves rapid large amplitude excursions such as in regimes switches or in intermittent dynamics.
Numerical model error is often introduced deliberately to control numerical instabilities arising in the simulation of geophysical fluids, such as in numerical weather prediction or climate dynamics. In such simulations numerical codes tend to produce large errors linked to numerical instabilities at the grid resolution level, which limits the reliability of the forecast over the forecast window. In a data assimilation context this numerical "noise" can lead to an unrealistic overestimation of forecast error covariances at small scales. The standard approach to dealing with these numerical instabilities is to add various forms of

artificial dissipation to the model (Durran 1999). This however has several well documented drawbacks. In the context of ensemble data assimilation artificial viscosity diminishes the spread of the forecast ensemble, causing detrimental underestimation of the error covariances (Houtekamer and Mitchell 1998; Burgers *et al.* 1998; Anderson and Anderson 1999; Houtekamer *et al.* 2005; Charron *et al.* 2010). Most notably, artificial dissipation implies unrealistic and excessive drainage of energy out of the system (Shutts 2005), effecting processes ranging from frontogenesis to large scale energy budgets. For example, Palmer (2001) reports on the unphysical dissipation of kinetic turbulent energy caused by mountain gravity wave drag parametrizations. Skamarock and Klemp (1992) introduce divergence damping to control unwanted gravity wave activity in hydrostatic and non-hydrostatic elastic equations to stabilize the numerical scheme, leading to numerical energy dissipation (see also Durran (1999)). Blumen (1990) controls the collapse of near-surface frontal width by artificially introducing momentum diffusion in a semi-geostrophic model of frontogenesis, with the effect of producing unrealistic energy dissipation rates. In semi-Lagrangian advection schemes Côté and Staniforth (1988) and Ritchie (1988) find unrealistic energy dissipation rates that are caused by interpolation. Besides effecting the small-scale dynamics, numerical dissipation also has a major effect on the large scales and their energy spectra (Palmer 2001; Shutts 2005). In ideal fluids, which often serve as a good approximation of the large scale dynamics (Salmon 1998), numerical dissipation destroys dynamically important conservation laws which may be undesirable in long time simulations such as in climate modelling (Thuburn 2008). Using statistical mechanics Dubinkina and Frank (2007, 2010) show how the choice of the conservation laws respected by a numerical scheme effects the statistics of the large scale fields.
There exists a plethora of methods to remedy the unphysical energy loss due to artificial diffusion and reinject energy back into the numerical model, for example via energy backscattering (Frederiksen and Davies 1997; Shutts 2005) or systematic stochastic model reductions (Majda *et al.* 1999).

Here we will look at these issues in the context of ensemble data assimilation. Rather than controlling numerical instabilities of the dynamic core via appropriate discretization schemes of the numerical forecast model, we will modifiy the actual ensemble data assimilation procedure, avoiding excessive numerical dissipation. Underdamping in dynamic cores causes the ensemble to acquire unrealistically large spread. This overdispersion is exacerbated in sparse observational networks where for the unobserved variables unrealistically large forecast covariances are poorly controlled by observational data (Liu *et al.* 2008; Whitaker *et al.* 2009). The questions we address in this article are: *can one use numerical forecasting models which are not artificially damped, yet still control the resulting overestimation of forecast covariances within the data assimilation procedure?*, and, *is it possible to use the increased spread induced by numerical instabilities in a controlled fashion to improve the skill of ensemble data assimilation schemes?*
In a recent article Gottwald *et al.* (2011) proposed a variance limiting ensemble Kalman filter (VLKF) which controls overestimation in sparse observational networks by including climatological information of the unobserved variables. In a perfect model data assimilation scheme unrealistic overestimation of the error covariances occurs in sparse observational networks as a finite size effect; it was shown that for short observational intervals of up to 10 hours the VLKF produces superior skill when compared to the standard ensemble transform Kalman filter (ETKF) (Bishop *et al.* 2001; Tippett *et al.* 2003; Wang *et al.* 2004). Here we will apply the variance limiting Kalman filter in the case of an imperfect, underdamped forecast model, and for large observation intervals. Unlike in the perfect model case where for large observation intervals the forecast error covariances approach the climatic variance, when forecasting with underdamped forecast models the error covariances will be larger than the climatic variance. This will be controlled by the VLKF, invoking a weak constraint onto the mean and the variance of the unobserved variables, leading to an improvement in the analysis skill. We will study this effect using the standard Lorenz-96 model (Lorenz 1996), where we assume that the truth is evolving to the standard set of parameters. To investigate whether underdamped models can be used as forecast models we then use forecast models with smaller values of the linear damping term of the Lorenz-96 system to generate ensembles with forecast spreads exceeding the true climatology.

Our main result is, that VLKF produces significantly better skill than ETKF, increasing with the strength of the model error, the sparsity of the observational network and the observational noise covariance. We will establish that a variance limiting weak constraint may be used as an adaptive alternative to artificial viscosity damping, counteracting model error on the fly when underdamping of the numerical forecast model causes an unrealistic overestimation of the forecast ensemble. However, we will see that there is a trade-off as for large model error the skill of VLKF becomes comparable to the skill of an analysis involving no forecast but using an analysis comprised of the observations and the climatological mean for the observed and unobserved variables, respectively. It will be shown that this is important for large observation times when the analysis is not tracking anymore due to either model error or the chaotic nature of the underlying dynamics, and one can only improve the analysis, assuring a reliable reproduction of statistical properties of the dynamics.

In the next section we briefly describe the variance limiting Kalman filter. In Section 3 we introduce the Lorenz-96 model and describe the underdamped forecast model used. In Section 4 we present results showing under what conditions the VLKF produces better skill than the usual ensemble transform Kalman filter. We conclude with a discussion and outlook in Section 5.

2. The variance limiting Kalman filter

The variance limiting Kalman filter (VLKF) is a variation of a standard ensemble Kalman filter designed to control overestimation of error covariances caused by finite ensemble sizes in sparse observational grids (Gottwald *et al.* 2011). Whereas usually only direct observations and model forecasts are combined to produce an analysis in data assimilation, it was proposed to incorporate climatological information to fill the data voids in sparse observational

grids. In the atmospheric sciences data voids are a huge obstruction to climate reanalyses (Bengtsson *et al.* 2007; Whitaker *et al.* 2004; Compo *et al.* 2011) when assimilating data from pre-radiosende era, or when attempting to include the mesospehere into numerical weather forecasts where the energetics is dominated by unobserved gravity waves (Polavarapu *et al.* 2005; Sankey *et al.* 2007; Eckermann *et al.* 2009). Climatic information is used to limit the posterior error covariance of the unobserved variables to not exceed their climatic covariance. It was shown that this limitation of the analysis error covariance controlled undesirable overestimation of error covariances which are caused by the finite size of the forecast ensembles. Furthermore, the VLKF drives the mean of the unobserved variables towards the climatological mean significantly increasing the analysis skill. We remark that one may use other values than the climatic covariance to control the analysis error covariance if one interprets the variance constraint as a numerical tool only to control overestimation of the error covariances. We present a brief summary of the algorithm here. For a detailed derivation the reader is referred to Gottwald *et al.* (2011).

Given an N-dimensional dynamical system

$$\dot{\mathbf{z}}_t = \mathbf{F}(\mathbf{z}_t) \,, \tag{1}$$

which is observed at discrete times $t_i = i\Delta t_{\rm obs}$, data assimilation aims at producing the best estimate of the current state given a typically chaotic, possibly inaccurate model $\dot{\mathbf{z}} = f(\mathbf{z})$ and noisy observations of the true state $\mathbf{z}_t$ (Kalnay 2002). We assume that the state space is decomposable according to $\mathbf{z} = (\mathbf{x}, \mathbf{y})$ with *observables* $\mathbf{x} \in \mathbb{R}^n$, for which direct observations are available, and *pseudo-observables* $\mathbf{y} \in \mathbb{R}^m$, for which only some integrated or statistical climatological information is available. Here we will assume $n + m = N$, however this is not essential.

Observations $\mathbf{y}_{\rm o} \in \mathbb{R}^n$ are expressed as a perturbed truth according to

$$\mathbf{y}_{\rm o}(t_i) = \mathbf{H}\mathbf{z}_t(t_i) + \mathbf{r}_{\rm o} \,,$$

where the observation operator $\mathbf{H} : \mathbb{R}^N \to \mathbb{R}^n$ maps from the whole space into observation space, and $\mathbf{r}_{\rm o} \in \mathbb{R}^n$ is assumed to be independent identically distributed observational Gaussian noise with associated error covariance matrix $\mathbf{R}_{\rm o}$.

We further introduce an operator $\mathbf{h} : \mathbb{R}^N \to \mathbb{R}^m$ which maps from the whole space into the space of the pseudo-observables spanned by $\mathbf{y}$. We assume that the pseudo-observables have variance $\mathbf{A}_{\rm clim} \in \mathbb{R}^{m\times m}$ and constant mean $\mathbf{a}_{\rm clim} \in \mathbb{R}^m$. This is the only information available for the pseudo-observables, and may be estimated, for example, from historical data or long time simulations. The error covariance of those pseudo-observations is denoted by $\mathbf{R}_{\rm w} \in \mathbb{R}^{m\times m}$.

In an ensemble Kalman filter (EnKF) (Evensen 2006) an ensemble with k members $\mathbf{z}_k$

$$\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_k] \in \mathbb{R}^{N\times k}$$

is propagated by the model dynamics according to

$$\dot{\mathbf{Z}} = \mathbf{f}(\mathbf{Z}) \,, \quad \mathbf{f}(\mathbf{Z}) = [\mathbf{f}(\mathbf{z}_1), \mathbf{f}(\mathbf{z}_2), \ldots, \mathbf{f}(\mathbf{z}_k)] \in \mathbb{R}^{N\times k} \,.$$

This forecast ensemble is split into its mean $\bar{\mathbf{z}}_f$ and ensemble deviation matrix $\mathbf{Z}'_f$. The ensemble deviation matrix $\mathbf{Z}'_f$ can be used to approximate the ensemble forecast covariance matrix via

$$\mathbf{P}_f(t) = \frac{1}{k-1}\mathbf{Z}'(t)\left[\mathbf{Z}'(t)\right]^T \in \mathbb{R}^{N\times N} \,.$$

Note that $\mathbf{P}_f(t)$ is rank-deficient for $k < N$ which is the typical situation in numerical weather prediction where N is of the order of 10^9 and k of the order of 100.

Given the forecast mean $\bar{\mathbf{z}}_f$, the forecast error covariance $\mathbf{P}_f$ and an observation $\mathbf{y}_{\rm o}$, an analysis mean is given by

$$\bar{\mathbf{z}}_a = \bar{\mathbf{z}}_f - \mathbf{K}_{\rm o}\left[\mathbf{H}\bar{\mathbf{z}}_f - \mathbf{y}_{\rm o}\right] - \mathbf{K}_{\rm w}\left[\mathbf{h}\bar{\mathbf{z}}_f - \mathbf{a}_{\rm clim}\right] \,, \tag{2}$$

where

$$\begin{aligned} \mathbf{K}_{\rm o} &= \mathbf{P}_a\mathbf{H}^T\mathbf{R}_{\rm o}^{-1} \\ \mathbf{K}_{\rm w} &= \mathbf{P}_a\mathbf{h}^T\mathbf{R}_{\rm w}^{-1} \,, \end{aligned}$$

with analysis error covariance matrix

$$\mathbf{P}_a = \left(\mathbf{P}_f^{-1} + \mathbf{H}^T\mathbf{R}_{\rm o}^{-1}\mathbf{H} + \mathbf{h}^T\mathbf{R}_{\rm w}^{-1}\mathbf{h}\right)^{-1} \,.$$

The variance of the unresolved pseudo-observables $\mathbf{hz}$ is controlled by setting

$$\mathbf{h}\mathbf{P}_a\mathbf{h}^T = \mathbf{A}_{\rm clim} \,. \tag{3}$$

This yields upon using the Sherman-Morrison-Woodbury formula (see for example Golub and Loan (1996)) the equation for the as yet unspecified error covariance matrix for the pseudo-observables $\mathbf{R}_{\rm w}$

$$\mathbf{R}_{\rm w}^{-1} = \mathbf{A}_{\rm clim}^{-1} - \left(\mathbf{h}\boldsymbol{\mathcal{P}}\mathbf{h}^T\right)^{-1} \,,$$

where

$$\boldsymbol{\mathcal{P}} = \left(\mathbf{P}_f^{-1} + \mathbf{H}^T\mathbf{R}_{\rm o}^{-1}\mathbf{H}\right)^{-1}$$

is the analysis error covariance of an equivalent EnKF without variance constraint. We remark that the matrix $\mathbf{R}_{\rm w}$ itself is never required in the VLKF algorithm, so an expensive inversion of $\mathbf{R}_{\rm w}^{-1}$ is not explicitly needed. Furthermore, the matrix $\boldsymbol{\mathcal{P}}$ can also be calculated without computing inverses of the forecast covariance matrix by means of the Sherman-Morrison-Woodbury formula. It is readily seen that for sufficiently small background error covariance $\mathbf{P}_f$, the error covariance $\mathbf{R}_{\rm w}$ is not positive semi-definite. To ensure that the error covariance is non-negative we diagonalize $\mathbf{R}_{\rm w}^{-1} = \mathbf{V}\mathbf{D}\mathbf{V}^T$; introducing $\bar{\mathbf{D}}$ with $\bar{\mathbf{D}}_{ii} = \mathbf{D}_{ii}$ for $\mathbf{D}_{ii} \geq 0$ and $\bar{\mathbf{D}}_{ii} = 0$ for $\mathbf{D}_{ii} < 0$, we set $\mathbf{R}_{\rm w}^{-1} = \mathbf{V}\bar{\mathbf{D}}\mathbf{V}^T$. Hence the variance limiter acts only on those eigendirections of $\mathbf{h}\boldsymbol{\mathcal{P}}\mathbf{h}^T$ whose corresponding singular eigenvalues are larger than those of the prescribed (climatic) covariance $\mathbf{A}_{\rm clim}$. The diagonalisation of $\mathbf{R}_{\rm w}^{-1}$ poses an additional computational complexity when compared to ETKF.

To determine an ensemble $\mathbf{Z}_a$ which is consistent with the analysis error covariance $\mathbf{P}_a$ and satisfies

$$\mathbf{P}_a = \frac{1}{k-1}\mathbf{Z}'_a\left[\mathbf{Z}'_a\right]^T \,,$$

we use the method of ensemble square root filters (Simon 2006). In particular we use the ensemble transform Kalman filter (ETKF) (Bishop *et al.* 2001; Tippett *et al.* 2003; Wang *et al.* 2004), which seeks a transformation $\mathbf{S} \in \mathbb{R}^{k\times k}$ such that the analysis deviation ensemble $\mathbf{Z}'_a$ is given as a deterministic perturbation of the forecast ensemble $\mathbf{Z}_f$ via $\mathbf{Z}'_a = \mathbf{Z}'_f\mathbf{S}$. Details of the implementation can be found in (Bishop *et al.* 2001; Tippett *et al.* 2003; Wang *et al.* 2004). Alternatively one could choose the ensemble adjustment filter (Anderson 2001) in which the ensemble deviation matrix $\mathbf{Z}'_f$ is pre-multiplied with an appropriately determined matrix $\mathbf{A} \in \mathbb{R}^{N\times N}$. Yet another method to construct analysis deviations was proposed in Bergemann *et al.* (2009) where the continuous Kalman-Bucy filter is used to calculate $\mathbf{Z}'_a$ without using any computations of matrix inverses, which may be advantageous in high-dimensional systems.
A new forecast is obtained by propagating $\mathbf{Z}_a$ with the nonlinear forecast model to the next observation time, where a new analysis cycle will be started.

In the perfect model case, when $\mathbf{A}_{\rm clim}$ and $\mathbf{a}_{\rm clim}$ are given by the climatology, the forecast model will acquire comparable forecast error covariance $\mathbf{P}_f$ for large observational intervals $\Delta t_{\rm obs}$ using either the ETKF or VLKF. Hence both filters will have approximately the same input parameters prior to the analysis step, and as was shown in Gottwald *et al.* (2011) the filters perform equally well. This is not true for imperfect models with an underdamping forecast model; when the variance limiting constraint is switched on, the background covariance of the pseudo-observables is set to $\mathbf{A}_{\rm clim} = \sigma^2_{\rm clim}\mathbf{I}$ for the VLKF, much smaller than that for the ETKF when using an underdamped model. Note that this implies that the forecast error covariance is also typically much smaller for the VLKF than ETKF over short observation intervals.

3. The Lorenz-96 model

Consider the following modification of the Lorenz-96 model (Lorenz 1996)

$$\frac{dz_j}{dt} = z_{j-1}(z_{j+1} - z_{j-2}) - \gamma z_j + F \quad j = 1, ..., D \quad (4)$$

with $\mathbf{z} = (z_1, ..., z_D)^T$ and periodic $z_j = z_{j+D}$, where we have introduced a variable linear damping parameter γ. The Lorenz-96 model is a standard test bed for data assimilation (see for example Ott *et al.* (2004); Fisher *et al.* (2005)) as it is computationally manageable but still incorporates crucial ingredients of real midlatitude atmospheric flows such as nonlinear energy conservation, advection, forcing and linear damping. We set the number of degrees of freedom as $D = 40$ and $F = 8$ for the forcing. For $\gamma = 1$, the classical value of the damping in the Lorenz-96 system, this corresponds to a time scaling of 5 days per model time unit, an e-folding time of 2.1 days and a distance between adjacent grid points of 750 km, roughly equalling the Rossby radius of deformation (Lorenz 1996).

Figure 1 illustrates how the climatic mean $\mu_{\rm clim}$ and variance $\sigma^2_{\rm clim}$ depend on the value of the damping parameter γ. The climatic values are obtained from simulations of one long time trajectory (1000 days). Note that in the Lorenz-96 system all variables z_j share the same statistical behaviour. Furthermore, we show the variances σ^2 when calculated from simulations running only over a finite length of time, such as analysis cycle length. As expected, the variance increases as the dynamics gets less and less damped for fixed length of time, and increase with the simulation length $\Delta t_{\rm obs}$ for fixed damping parameter γ (asymptoting to the climatic variance).

Two characteristic time scales of the Lorenz-96 system (4) will be important in the following, the e-folding time, measured here by the inverse of maximal Lyapunov exponent $\lambda_{\max}$ and the advection time $T_{\rm adv}$, which is a measure for the time taken for a front to propagate from one site j to the next $j + 1$. In Figure 2 we show how these characteristic time scales vary with the damping parameter γ. Firstly, we show the inverse of the maximal Lyapunov exponent $\lambda_{\max}$ which is a measure for the predictability time of the Lorenz-96 dynamics and correlates with the decorrelation time. For the original Lorenz-96 model with $\gamma = 1$ we reproduce the e-folding time of 50 hours, i.e. 2.1 days. In accordance with the observed increase in variance, the predictability time decreases with decreasing damping. As a second characteristic time scale we show the advective time $T_{\rm adv}$. Since the nonlinear term in (4) can be interpreted as a finite difference discretization of an advection term zz_x, information from one grid point j is mainly transported to its right neighbours in an asymmetric fashion. We estimate $T_{\rm adv}$ by initializing an integration of the Lorenz-96 model with only one non-zero variable and then measuring the time taken for the resulting front to propagate across sites. The characteristic time scale of this information transport is important when discussing the influence of the sparsity of an observational network on the filtering skill. Figure 2 shows that this time scale decreases with increasing model error.

4. Numerical results

We now investigate how and when the variance limiting Kalman filter can be used to control the increased covariances of an underdamped forecast model. In particular, we will study the situation of a sparse observational grid.

We will address two issues. First, we will investigate under what conditions the VLKF as described in Section 2 produces better analysis skill than the ETKF. Second, we will investigate when the VLKF produces better skill than a worst-case analysis which does not involve any forecast but simply constructs the analysis by taking the observational data and the climatic mean for the observed and the unobserved analysis variables, respectively (a "poor-man's analysis"). This latter unfortunate situation can occur, for example, for moderate observation intervals $\Delta t_{\rm obs}$, as we will see. Besides the degree of underdamping γ, the skill will be seen to depend on the sparsity of the observational grid used as well as on the observation interval and the observational noise level.

The main result will be that the VLKF framework can indeed provide skill improvement over the standard ETKF. However, as the level of model damping is decreased we find that there is a trade-off between increasing improvement over ETKF on the one hand, and decreasing

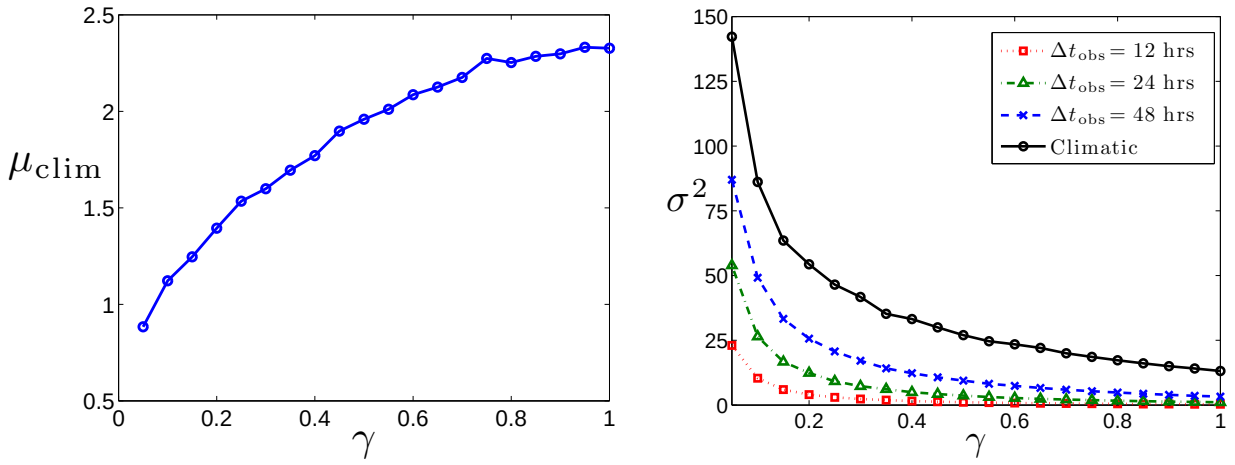

Figure 1. Dependence of climatic mean (left) and variance (right) on damping parameter γ. As well as the climatic variance (black, circles, solid) we show the average variance calculated over forecast intervals $\Delta t_{\rm obs} = 12$ hours (red, squares, dotted), $\Delta t_{\rm obs} = 24$ hours (green, triangles, dash-dotted) and $\Delta t_{\rm obs} = 48$ hours (blue, crosses, dashed).

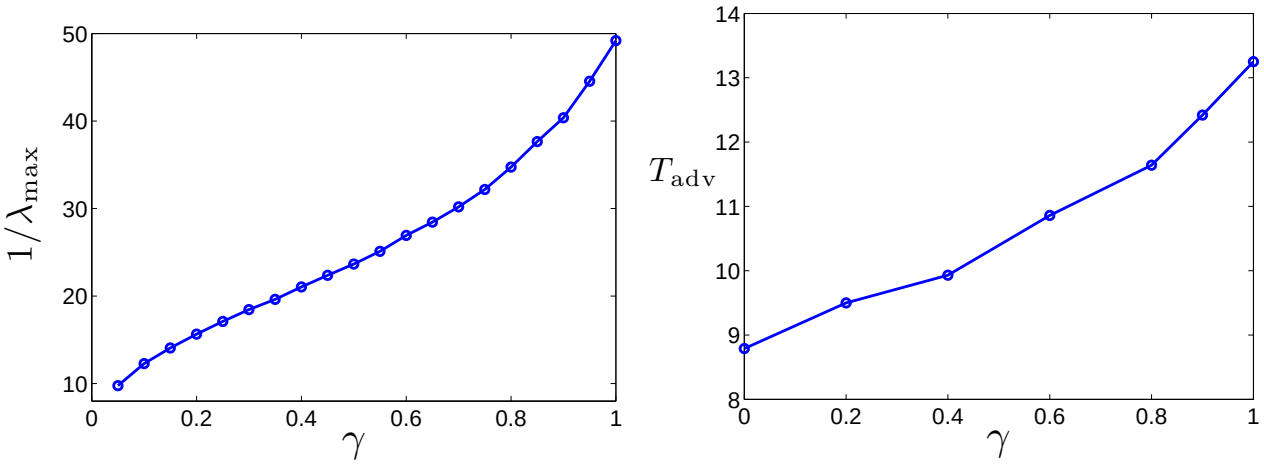

Figure 2. Characteristic time scales for the Lorenz-96 system (4) as a function of the damping parameter γ. Left: Inverse of the maximal Lyapunov exponent $\lambda_{\rm max}$ in hours. Right: Advection time $T_{\rm adv}$ in hours.

improvement over the poor-man's analysis on the other. The skill improvement can be explained by the improved ability of the VLKF to ensure the correct statistical properties of the analysis, as opposed to pathwise tracking of the truth, for small forecast model damping. We find analogous results as the observational sparsity $N_{\rm obs}$ or the magnitude of the observational noise is increased.

We generate a truth $\mathbf{z}_t$ from a realisation of the Lorenz-96 system (4) with $\gamma = 1$. Observations $\mathbf{y}_{\rm o}$ are then generated for the observables $\mathbf{Hz}$ with Gaussian error covariance $\mathbf{R}_{\rm o} = (\eta\sigma_{\rm clim})^2\mathbf{I}$, and are available after each observation interval $\Delta t_{\rm obs}$ at every $N_{\rm obs}^{\rm th}$ site. The observations are then assimilated in an ensemble filter framework using the Lorenz-96 model (4) with damping parameter $\gamma < 1$ as forecast model, mimicking the situation of underdamped forecast models. We perform twin experiments where the same truth and observations are used for data assimilation with both the VLKF and ETKF. Both filters are run with $k = 41$ ensemble members and the same forecast model (4) with $\gamma < 1$, making the forecasts underdamped with respect to the truth $\mathbf{z}_t$. Both filters are spun up for 100 cycles to ensure statistical stationarity of the ensembles. We employ 5% constant global multiplicative covariance inflation in both filters.

We use an implicit mid-point rule (see for example Leimkuhler and Reich (2005)) to simulate the system (4) to a time $T = 60$ time units with a time step $dt = 1/240$. The total integration time corresponds to 300 days, and the integration timestep dt corresponds to half an hour. The climatic mean and standard deviation over all variables with $\gamma = 1$ are estimated as temporal averages from a long trajectory as $\mu_{\rm clim} = 2.34$ and $\sigma_{\rm clim} = 3.63$ respectively, which determines our true climatology.

In Figures 3 and 4 we show sample analyses using both filters, ETKF and VLKF, with $N_{\rm obs} = 4$ and damping factor $\gamma = 0.5$ for both $\Delta t_{\rm obs} = 24$ and $\Delta t_{\rm obs} = 48$ hours, exemplifying moderate and large observation intervals respectively. We show analyses for an arbitrary unobserved component and an arbitrary observed component of the Lorenz-96 model. Both filters track the truth for the observed components reasonably well for both moderate

and large observation intervals, with the analysis tending more towards the observations for large $\Delta t_{\rm obs}$. However, for the pseudo-observables with moderate observation interval $\Delta t_{\rm obs} = 24$ hours, the ETKF tracks the truth better than the VLKF despite the inclusion of additional climatological information. For large observation intervals $\Delta t_{\rm obs} = 48$ hours, neither filter tracks the truth for the pseudo-observables well. However, the additional input of the climatological information into the VLKF leads to an analysis with statistics which better resemble the true climatology, thereby improving the overall RMS errors and skill. It is also clearly seen that the main contribution to the error stems from the pseudo-observables.

Figures 3 and 4 suggest that the notion of what constitutes a good performance of a filter depends on the length $\Delta t_{\rm obs}$ of the observation interval. When aiming for improved performance in the case of moderate observation intervals, one is interested in tracking the truth for both the observables and pseudo-observables. In the case of large observation intervals, one can only expect to track the truth for the observed variables and produce analyses which exhibit the same statistical behaviour as the truth for the pseudo-observables. We remark that for very large observational sparsity $N_{\rm obs}$ tracking cannot be demanded even for moderate observation intervals; for example with $\Delta t_{\rm obs} = 6$ hours the analysis of an observational grid with $N_{\rm obs} = 20$ does not produce tracking analyses for any variable $\mathbf{z}_j$.

In Figure 5 we show the average maximal singular values s_H and s_h of the forecast and analysis error covariances projected onto the observation and pseudo-observation subspaces respectively. As expected the covariances increase with increasing model error, as well as with increasing sparsity. The variance constraint of the VLKF markedly controls the analysis covariance of the unobserved variables to always assume values below the climatic variance. This causes the forecast covariances to be limited as well, albeit at values greater than the climatic variance $\sigma^2_{\rm clim}$. For large observational sparsity $N_{\rm obs}$, the analysis covariance of the unobserved variables is set to the climatic variance for the VLKF, whereas for the ETKF the analysis covariances grow larger than $\sigma^2_{\rm clim}$ with increasing sparsity. Interestingly, the VLKF also limits the covariances for the observed variables. As we will see this may have negative effects and may lead to an underdispersive ensemble which can spoil the VLKF analysis for the observed variables.

In Figure 6 we show how for long observation intervals $\Delta t_{\rm obs} = 48$ hours the true climatic mean and variance are better preserved by the VLKF analysis than by the ETKF analysis. The true climatic mean $\mu_{\rm clim}$ is reasonably well preserved by the VLKF analysis whereas the ETKF greatly underestimates it. The VLKF analysis underestimates the true climatic variance $\sigma^2_{\rm clim}$, but not as severely as the ETKF does. We will see that this leads to a marked increase in skill, especially for large model error $\gamma < 1$.
The fact that the variance of the analysis $\langle(\bar{\mathbf{z}}_a - \langle\bar{\mathbf{z}}_a\rangle)^2\rangle$ decreases with decreasing damping may seem counterintuitive at first sight. This can be justified by considering a simple two-dimensional system where x is observed and y is unobserved. The Kalman analysis update (cf. Equation (2)) for the unobserved variables is $\bar{y}_a = \bar{y}_f - P_{f,xy}/(P_{f,xx} + R_{\rm o})(\bar{x}_f - x_{\rm o})$ where $P_{f,xx} > 0$ and $P_{f,xy}$ measures the forecast error covariance between the unobserved and observed variable. Note that due to finite size ensemble sizes one can have $P_{f,xy} \neq 0$ even if dynamically there are no correlations between observed and unobserved variables. We argue that the decrease in variance of the analysis $\langle(\bar{\mathbf{z}}_a - \langle\bar{\mathbf{z}}_a\rangle)^2\rangle$ with decreasing damping is due to the analysis being damped stronger for smaller values of γ in the case of the Lorenz-96 system (4). For the simple system this is the case when $P_{f,xy}/(P_{f,xx} + R_{\rm o})(\bar{x}_f - x_{\rm o}) > 0$ on average, provided that $\bar{y}_f > 0$ also on average. Since the climatic mean is positive and decreases with decreasing damping γ (cf. Figure 1), $\bar{\mathbf{z}}_f > 0$ and $\mathbf{H}\mathbf{z}_f - \mathbf{y}_{\rm o} < 0$ on average. Moreover, as all variables z_j are statistically equal due to the periodicity of the Lorenz-96 system, the forecast error covariances between observed and unobserved variables $\mathbf{H}\mathbf{P}_f\mathbf{h}^T$ (equivalent to $P_{f,xy}$ in the simple model) must also be negative on average. Hence the variance of the analysis time series $\langle(\bar{\mathbf{z}}_a - \langle\bar{\mathbf{z}}_a\rangle)^2\rangle$ has the same decreasing tendency with decreasing damping γ as the climatic mean.

We now quantify the performance of the VLKF and the standard ETKF by the site-averaged RMS error

$$\mathcal{E} = \sqrt{\langle \frac{1}{LD_G} \sum_{l=1}^{L} \|\bar{\mathbf{z}}_a(l\Delta t_{\rm obs}) - \mathbf{z}_t(l\Delta t_{\rm obs})\|^2_{\mathbf{G}} \rangle} \qquad (5)$$

between the truth $\mathbf{z}_t$ and the ensemble mean $\bar{\mathbf{z}}_a$, where $L = \lfloor T/\Delta t_{\rm obs} \rfloor$ is the number of analysis cycles and D_G denotes the number of variables involved. We introduced the norm $\|\mathbf{a}\|^2_{\mathbf{G}} = \mathbf{a}^T\mathbf{G}\mathbf{a}$ to investigate the error over all variables $\mathcal{E}$ using $\mathbf{G} = \mathbf{I}$, the error $\mathcal{E}_o$ distributed over the observed variables using $\mathbf{G} = \mathbf{H}^T\mathbf{H}$ and the error $\mathcal{E}_{uo}$ distributed over the unobserved variables using $\mathbf{G} = \mathbf{h}^T\mathbf{h}$. We also introduce a measure $\hat{\mathcal{E}}$ to quantify how well a filter performs against the poor-man's analysis of using the observations and the climatic mean for the observed and unobserved components of the analysis respectively, implying $\mathbf{H}(\bar{\mathbf{z}}_a - \mathbf{z}_t) \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_{\rm o})$ and $\mathbf{h}(\bar{\mathbf{z}}_a - \mathbf{z}_t) \sim \mathcal{N}(\mathbf{0}, \mathbf{A}_{\rm clim})$. Assuming that $\mathbf{H}(\bar{\mathbf{z}}_a - \mathbf{z}_t)$ and $\mathbf{h}(\bar{\mathbf{z}}_a - \mathbf{z}_t)$ are uncorrelated and setting $\mathbf{A}_{\rm clim} = \sigma^2_{\rm clim}\mathbf{I}_m$, we can calculate the RMS error $\hat{\mathcal{E}}$ in the poor-man's analysis as

$$\hat{\mathcal{E}}^2 = \frac{1}{LD} \sum_{i=1}^{L} \Big[||\mathbf{H}(\bar{\mathbf{z}}_a(t_i) - \mathbf{z}_t(t_i))||^2 + ||\mathbf{h}(\bar{\mathbf{z}}_a(t_i) - \mathbf{z}_t(t_i))||^2 \Big]$$

where $L = \lfloor T/\Delta t_{\rm obs} \rfloor$ is the number of analysis cycles as in (5), giving the RMS error

$$\hat{\mathcal{E}} = \sqrt{\frac{1}{N_{\rm obs}}\eta^2\sigma^2_{\rm clim} + \left(1 - \frac{1}{N_{\rm obs}}\right)\sigma^2_{\rm clim}} = \sigma_{\rm clim}\sqrt{1 + \frac{(\eta^2 - 1)}{N_{\rm obs}}}\,, \qquad (6)$$

with $\mathbf{R}_{\rm o} = (\eta\sigma_{\rm clim})^2\mathbf{I}$.

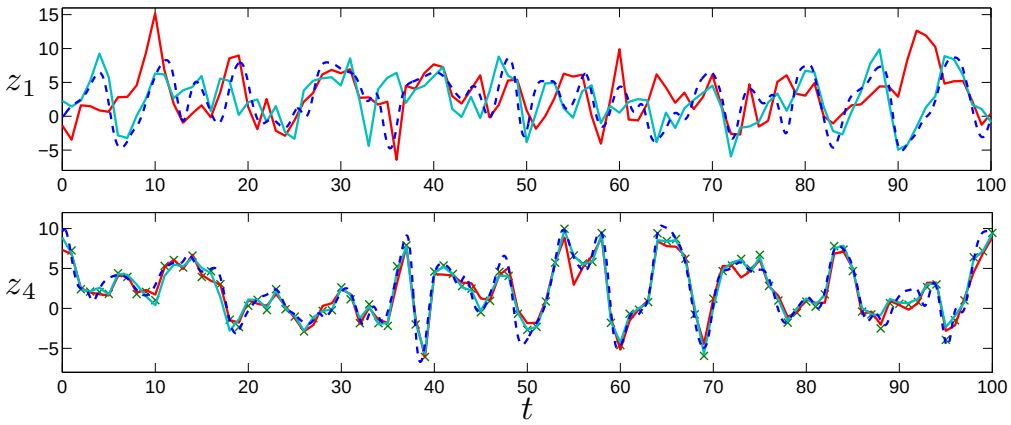

Figure 3. Sample ETKF (blue-green, solid line) and VLKF (red, solid line) analyses over 100 days for observed z_4 (bottom panel) and unobserved z_1 (top panel) component for $\Delta t_{\rm obs} = 24$ hours and model error $\gamma = 0.5$. The dashed line is the truth, the crosses are observations. Parameters used were $N_{\rm obs} = 4$ and $\mathbf{R}_{\rm o} = (0.25\sigma_{\rm clim})^2\,\mathbf{I}$.

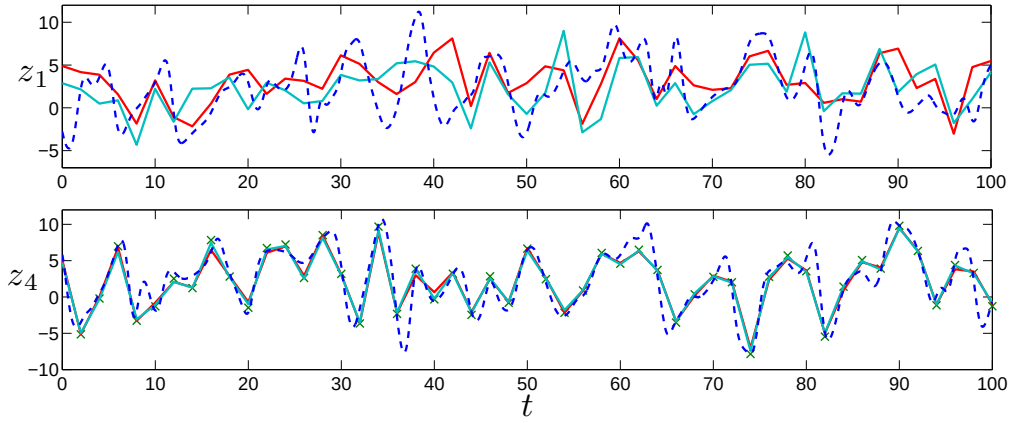

Figure 4. As in Figure 3, but with $\Delta t_{\rm obs} = 48$ hours.

To quantify the proportional improvement in RMS errors for the VLKF over ETKF we introduce the analysis skill

$$\mathcal{S} = \frac{\mathcal{E}^{\rm E}}{\mathcal{E}^{\rm V}}\ ,$$

where the RMS errors $\mathcal{E}$ are defined by (5). Values of $\mathcal{S} > 1$ indicate an improvement of the VLKF over ETKF. We also define the corresponding skills for proportional improvement of the ETKF and VLKF over the poor man's analysis

$$\hat{\mathcal{S}}^E = \frac{\hat{\mathcal{E}}}{\mathcal{E}^{\rm E}} \qquad \text{and} \qquad \hat{\mathcal{S}}^V = \frac{\hat{\mathcal{E}}}{\mathcal{E}^{\rm V}}\ .$$

We first investigate the dependence of skill upon the observation interval $\Delta t_{\rm obs}$. In Figure 7 the skills $\mathcal{S}$, $\hat{\mathcal{S}}^E$ and $\hat{\mathcal{S}}^V$ are shown as a function of the model error γ for different observation intervals $\Delta t_{\rm obs}$ with fixed sparsity $N_{\rm obs} = 4$. The figures reveal an intricate dependence of the skill on both the observation interval and the model error. We have checked that for $\Delta t_{\rm obs} > 2.1$ days, the e-folding time for the Lorenz-96 model with $\gamma = 1$, the skill curves $\mathcal{S}(\gamma)$ collapse onto one curve, independent of the observation interval. This indicates that the forecast error covariance, which saturates for large observational times to the climatic covariance, is the main factor driving filter performance.

We will treat the near-perfect model case γ close to one and the strong model error case with small γ separately. The perfect model case with $\gamma = 1$ was already considered in Gottwald *et al.* (2011). We confirm here that the VLKF is superior to the standard ETKF, its skill increasing with decreasing observation intervals $\Delta t_{\rm obs}$. For smaller observation intervals skill improvements of more than 20% can be achieved, and the analysis of the pseudo-observables track the truth surprisingly well (cf. Gottwald *et al.* (2011)). Note that both filters perform better than the poor man's analysis with $\hat{\mathcal{S}}^{\rm E} > 1$ and $\hat{\mathcal{S}}^{\rm V} > 1$ for $\gamma = 1$. The variance limiting filter here successfully controls the overestimation of the analysis error covariance caused by the combination of finite ensemble sizes and sparse observations.

The case of model error with γ significantly smaller than 1 – our main objective in this work – shows an interesting dependence on both the observation interval and the model error. For the discussion below it is important to note that for sufficiently large model error the analysis error covariances projected onto the pseudo-observables exceed the climatic variance $\sigma^2_{\rm clim}$; hence the background

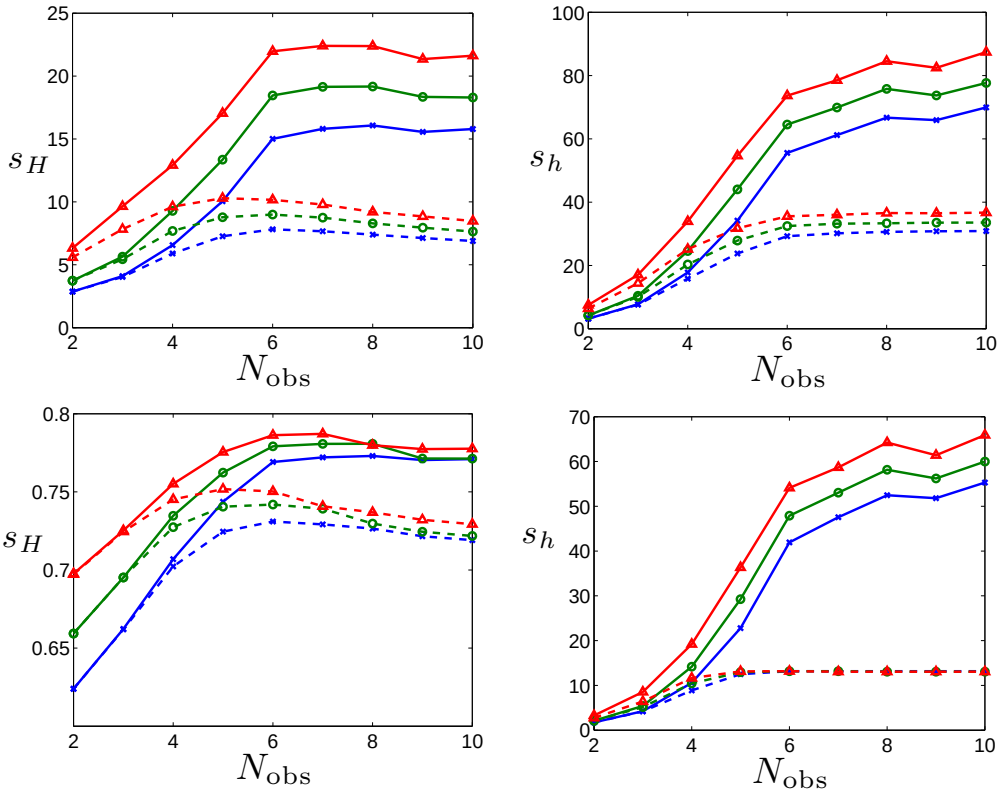

Figure 5. Average maximal singular values s_H (left panels) and s_h (right panels) for $\mathbf{HP}_f\mathbf{H}^T$ and $\mathbf{hP}_f\mathbf{h}^T$ (top) and $\mathbf{HP}_a\mathbf{H}^T$ and $\mathbf{hP}_a\mathbf{h}^T$ (bottom), respectively, as a function of the sparsity $N_{\rm obs}$. The solid lines depict values obtained from the ETKF and the dashed lines from the VLKF. We use $\gamma = 1$ (blue, crosses), $\gamma = 0.9$ (green, circles) and $\gamma = 0.8$ (red, triangles). We used $\mathbf{R}_{\rm o} = (0.25\sigma_{\rm clim})^2\mathbf{I}$ and $\Delta t_{\rm obs} = 48$ hours.

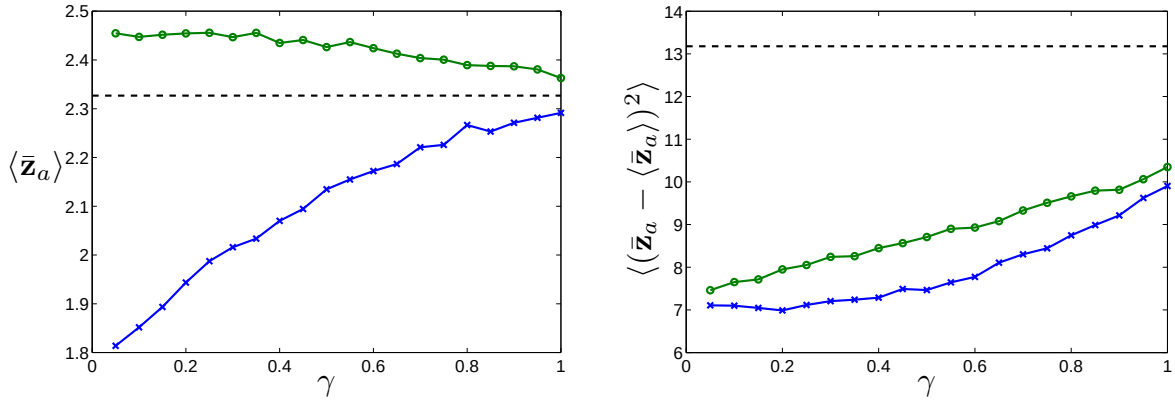

Figure 6. Mean $\langle \bar{\mathbf{z}}_a \rangle$ (left) and variance $\langle (\bar{\mathbf{z}}_a - \langle \bar{\mathbf{z}}_a \rangle)^2 \rangle$ (right) of the analysis for ETKF (blue, crosses) and VLKF (green, circles), averaged over time and sites. The dashed black lines shows the climatic mean and variance $\mu_{\rm clim}$ and $\sigma^2_{\rm clim}$ of the truth with $\gamma = 1$. We used $\mathbf{R}_{\rm o} = (0.25\sigma_{\rm clim})^2\mathbf{I}$, $N_{\rm obs} = 4$, $\Delta t_{\rm obs} = 48$ hours.

error covariance matrices which are used to initialize an analysis cycle are larger than $\sigma^2_{\rm clim}$ for the ETKF. On the other hand, the VLKF limits the analysis error covariance (projected onto the pseudo-observables) from above to $\sigma^2_{\rm clim}$, independent of the observation interval.

It is seen in Figure 7 that for moderate observation intervals $\Delta t_{\rm obs}$ (24 hours), the VLKF performs worse than the ETKF with skills $\mathcal{S} < 1$ decreasing with increasing model error (i.e. decreasing values of γ). On the contrary, for large observation intervals with $\Delta t_{\rm obs} > 36$ hours, the skill $\mathcal{S} > 1$ increases with increasing model error. Furthermore, for moderate observation intervals the VLKF performs increasingly worse than the poor man's analysis with increasing model error, but outperforms it for all values of γ for large observation intervals. The ETKF, however, always performs worse than the poor man's analysis for all observation intervals in the presence of sufficiently large model error. We first explain the behavior of the skill observed for moderate observation intervals and then for large observation intervals.

Over moderate observation intervals the forecast ensemble acquires only little spread, as already encountered

in Figure 1. This small spread in conjunction with model error can lead to instances where the analysis mean performs large excursions away from the truth, thereby spoiling the analysis. This effect is stronger for the VLKF as it is initialized with a smaller ensemble spread associated with the constrained analysis error covariance as discussed above. Increasing the model error by decreasing γ increases the probability of the ensemble mean diverging from the truth, thereby decreasing the skill. This explains the poor performance of the VLKF over ETKF with $\mathcal{S} < 1$ for moderate $\Delta t_{\rm obs}$, and the poor performance of both the ETKF and VLKF over the poor man's analysis with $\hat{\mathcal{S}}^E < 1$ and $\hat{\mathcal{S}}^V < 1$ as illustrated in Figure 7.

For large observation intervals the situation is different. Large observation intervals are characterized by the forecast dynamics having sufficiently decorrelated so that the forecast ensemble will have equilibrated with (γ-dependent) covariance, having lost memory of the background initialization. This has the effect that both filters will acquire comparable forecast error covariance $\mathbf{P}_f$ over the forecast interval $\Delta t_{\rm obs}$. The large spread which is acquired by the forecast ensemble during large observation intervals prohibits the analysed pseudo-observables tracking the truth well (cf. Figure 4). Skill improvements can only be achieved by 1.) tracking the truth for the observed variables as guided by the observations, and 2.) assuring the correct climatological behaviour for the pseudo-observables. This latter point is where the variance limiting method is beneficial, as illustrated in Figure 6. The more accurate statistical properties of the VLKF analysis results in improved skill with $\mathcal{S} > 1$ for all values of γ as shown in Figure 7.

However, Figure 7 also shows that the increased skill $\mathcal{S}$ for large observation intervals with increasing model error comes at the price of the VLKF becoming less able to outperform the poor man's analysis with error $\hat{\mathcal{E}}$. As expected, $\hat{\mathcal{S}}^V$ asymptotes to 1 from above for decreasing values of γ, whereas $\hat{\mathcal{S}}^E$ becomes less than 1 for sufficiently large model error. We found that the VLKF performs better than the poor man's analysis for all values of γ for $\Delta t_{\rm obs} > 36$ hours.

We conclude that the VLKF may be beneficial for ensemble based data assimilation with large underdamping model error and large observation intervals $\Delta t_{\rm obs}$. For smaller observation intervals $\Delta t_{\rm obs} < 18$ hours with sufficient model error (not shown) the standard ETKF performs better than the VLKF. However, for small $\Delta t_{\rm obs}$ with a near-perfect model ($\gamma \approx 1$) it is beneficial to use the VLKF, as shown in Gottwald *et al.* (2011). For large observation intervals the overall skill increases with increasing model error γ due to the imposed control of the increased error covariances. However, there is a trade-off between increased skill improvement of the VLKF over ETKF, and of the skill improvement of the VLKF over the poor man's analysis.

We now look at the dependence of the skill on the sparsity of the observational grid $N_{\rm obs}$ in the presence of model error. Numerical simulations show that the critical observation interval above which we observe skill improvement with $\mathcal{S} > 1$ for all values of γ decreases with increasing sparsity $N_{\rm obs}$. Note that this impacts on the notion of what is considered to be moderate or large observation interval in the discussion above. This dependence of skill on observational sparsity can be explained by noticing that sparser observations imply an increased forecast error covariance and ensemble spread for a given observation interval, as illustrated in Figure 5.
We have already seen that the main contribution to the overall RMS error stems from the pseudo-observables. Therefore one may be tempted to conclude that skill improvement consistently gets better for increasing $N_{\rm obs}$. This is correct when only comparing the VLKF and ETKF. However, as in the case of increasing model error for fixed sparsity discussed above, the increased sparsity for fixed model error causes the VLKF to be less effective in outperforming the poor-man's analysis. Figure 8 illustrates how increased sparsity affects the overall RMS error $\mathcal{E}$. It is seen that the ETKF increasingly performs worse than the poor-man's analysis with increasing model error and increasing $N_{\rm obs}$. In contrast, the RMS errors for the VLKF are below the RMS error of the poor man's analysis $\hat{\mathcal{E}}$ for all values of γ, and they asymptote to $\hat{\mathcal{E}}$ for large sparsity and large model error, in accordance with our observations in Figure 5 on the singular values of the error covariances of the VLKF. The trade-off between gain in skill of the VLKF over ETKF and less good skill with respect to the poor-man's analysis for increasing sparsity $N_{\rm obs}$ can be explained as follows. For sparse observations information from the observables cannot propagate to distant pseudo-observables within the observation interval $\Delta t_{\rm obs}$ (cf. Figure 2). This causes the forecast and analysis for those pseudo-observables to be essentially random variables with temporal mean $\mu_{\rm clim}$ and variance $\sigma^2_{\rm clim}$ for the VLKF, and $\mu(\gamma)$ and $\sigma^2(\gamma)$ for the ETKF. Hence increased sparsity leads to $\mathcal{E}^V$ asymptoting from below to $\hat{\mathcal{E}}$ for all values of γ, with increasingly worse performance for decreasing damping γ. In contrast, for the ETKF $\mathcal{E}^E > \hat{\mathcal{E}}$ for sufficiently large sparsity, for all values of γ.

It is illuminating to see how the RMS errors are distributed over the observed variables and the unobserved variables. In Figure 9 we depict the RMS errors $\mathcal{E}_o$ of the observed variables and $\mathcal{E}_{uo}$ of the unobserved variables, as defined in (5). We restrict our discussion to large observation intervals $\Delta t_{\rm obs} = 48$ hours. Consistent with our observation that the overall skill improvement is dominated by the superior analysis of the pseudo-observables, $\mathcal{E}_{uo}$ exhibits the same behaviour and trends as the overall RMS error $\mathcal{E}$ (cf. Figures 7 and 8). The error for the observed variables shows different behaviour. It is seen that the standard ETKF produces better or equal RMS errors than the VLKF for all values of the model error γ. The inferior performance of the VLKF over ETKF for the observed variables is an instance of underestimation of the error covariances – the familiar situation in ensemble data assimilation. As seen in Figure 5, where we show the average maximal singular value s_H of $\mathbf{H}\mathbf{P}_a\mathbf{H}^T$ for the ETKF and VLKF, the variance constraint of the VLKF for the pseudo-observations also significantly decreases the covariance in the observed subspace. The observed poor performance of the VLKF for the observed variables suggests that the limited covariance in the observed subspace is insufficient to provide enough ensemble spread for a reliable analysis.

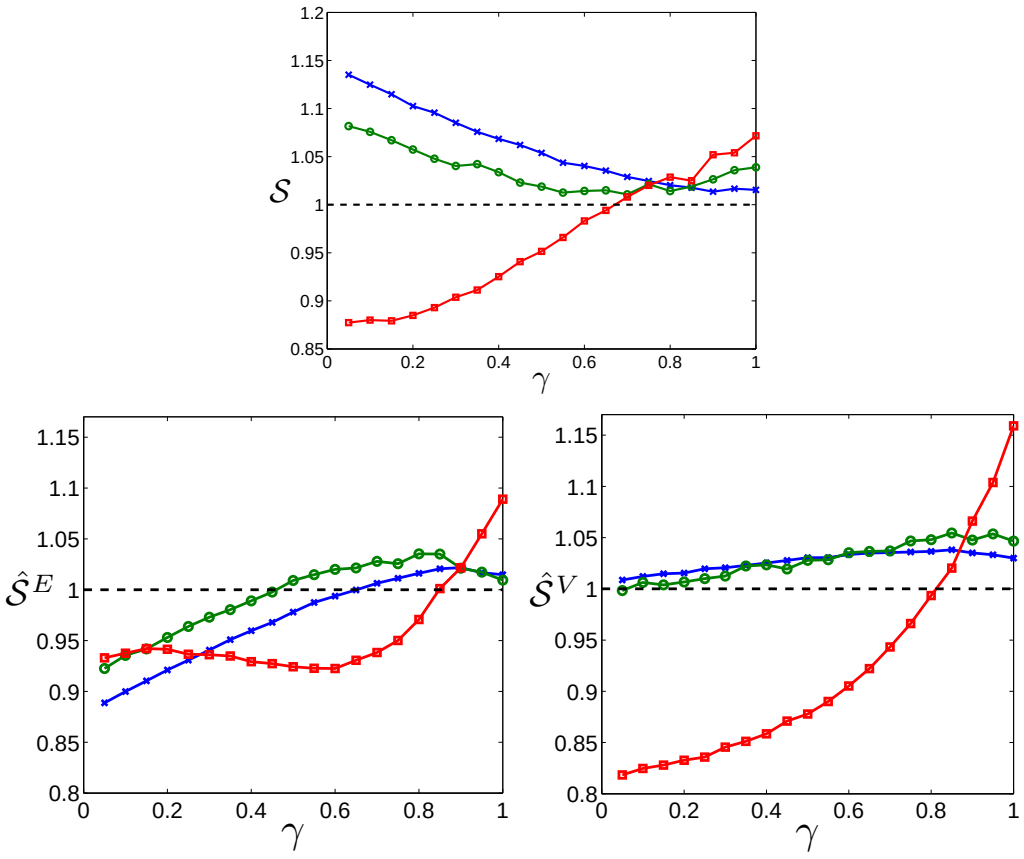

Figure 7. Skill $\mathcal{S}$ as a function of damping γ for $\Delta t_{\rm obs} = 24$ hours (red, squares), $\Delta t_{\rm obs} = 36$ hours (green, circles) and $\Delta t_{\rm obs} = 48$ hours (blue, crosses). We used $\mathbf{R}_{\rm o} = (0.25\sigma_{\rm clim})^2\mathbf{I}$, $N_{\rm obs} = 4$ and averaged all results over 200 realizations where one realization was performed over 300 days. With $\Delta t_{\rm obs} = 48$ hours, the RMS error in the ETKF analysis $\hat{\mathcal{E}}^E$ ranges from 3.13 (for $\gamma = 1$) to 3.57 (for $\gamma = 0.05$), while the RMS error in the VLKF analysis $\hat{\mathcal{E}}^V$ ranges from 3.08 (for $\gamma = 1$) to 3.15 (for $\gamma = 0.05$).

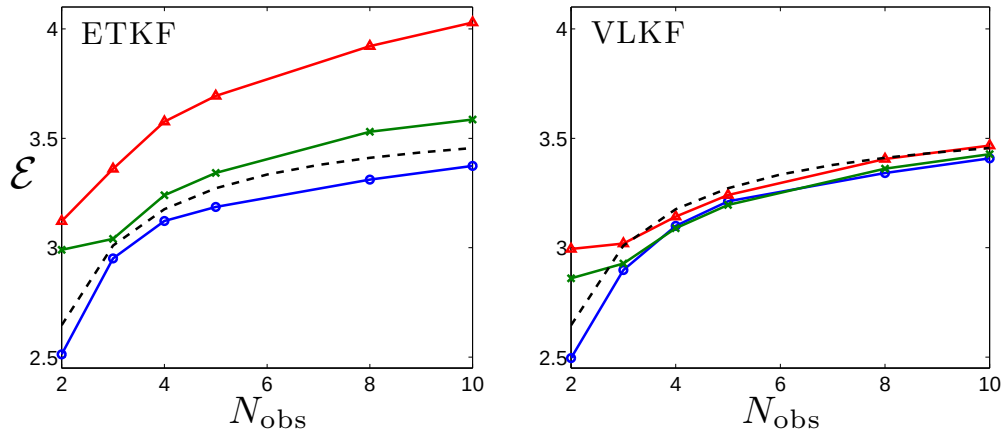

Figure 8. Overall RMS error $\mathcal{E}$ as a function of the observational grid sparsity $N_{\rm obs}$ for different values of γ. We show the perfect model case $\gamma = 1$ (circles, blue), $\gamma = 0.5$ (crosses, green) and $\gamma = 0.05$ (triangles, red). The dashed line denotes $\hat{\mathcal{E}}$. The left panel shows results for the ETKF, the right panel for the VLKF. We used $\mathbf{R}_{\rm o} = (0.25\sigma_{\rm clim})^2\mathbf{I}$ and $\Delta t_{\rm obs} = 48$ hours.

For small model error, both filters produce worse analyses than an analysis consisting of observations only. This is again linked to insufficient spread of the forecast ensembles. We have confirmed that for larger ensemble sizes $\mathcal{E}_o < \sqrt{\mathbf{R}_{\rm o}}$ for all γ with both the ETKF and VLKF. Underestimation of error covariances due to finite size effects in ensemble filters are usually controlled by some form of covariance inflation, such as multiplicative inflation whereby the prior forecast error covariance is multiplied by an inflation factor $\delta > 1$ (Anderson and Anderson 1999). In Figure 10 we show how the RMS error of the observed variables $\mathcal{E}_o$ decreases below the RMS error of the observed variables for $\gamma = 1$ and increasing inflation factor δ. We note that for very large values of δ in the perfect model situation the ETKF and VLKF produce analyses of equal RMS error. Increasing the forecast covariance inflation

has the same effect as increasing the model error (i.e. decreasing γ) in producing larger forecast ensemble spread.

Further, it is seen in Figure 9 that for sufficiently large model error, the RMS errors of the observables $\mathcal{E}_o$ for the VLKF and ETKF converge to the same γ-independent value $\tilde{\mathcal{E}}_o < \sqrt{\mathbf{R}_\mathrm{o}}$. This can be explained by the greater inflation for less damped forecast models; less model damping increases the forecast error covariance $\mathbf{P}_f$, and so the analysis tends to trust the observations more. As such, $\mathcal{E}_\mathrm{o} \to \sqrt{\mathbf{R}_\mathrm{o}}$ as $\gamma \to 0$ for both filters, albeit slightly reduced by the Kalman filtering. Due to the variance constraint the RMS error of the observed variables $\mathcal{E}_o$ asymptotes to $\tilde{\mathcal{E}}_o$ at larger model error for a given observational sparsity.
The value of the model damping γ below which the forecast variance is sufficiently large for the filter to trust the observations depends on the sparsity N_obs. The larger the sparsity N_obs, the larger the value of γ for which the RMS error $\mathcal{E}_\mathrm{o}$ of the two filters collapses onto the constant value $\tilde{\mathcal{E}}_o$. For $N_\mathrm{obs} = 8$, $\mathcal{E}_\mathrm{o} \approx \tilde{\mathcal{E}}_o$ for all values of γ for the ETKF and $\mathcal{E}_\mathrm{o}$ assumes this value at roughly $\gamma < 0.7$ for the VLKF. For $N_\mathrm{obs} = 3$, $\mathcal{E}_\mathrm{o} \approx \tilde{\mathcal{E}}_o$ for $\gamma < 0.7$ for the ETKF and for $\gamma < 0.25$ for the VLKF. This is linked again to the increased forecast error covariances caused by the sparsity, as illustrated in Figure 5. The top left panel in Figure 5 shows that for a given observation interval and model error γ, the forecast error covariance increases with sparsity N_obs, causing the observations to be weighted more by the analysis.

We remark that while for the observed variables there is generally little or no improvement in the RMS error of the VLKF over ETKF, the VLKF still creates more reliable ensembles, as characterised by the ranked probability histogram (also known as Talagrand diagram) (Anderson 1996; Hamill and Colucci 1997; Talagrand *et al.* 1997). In such a histogram the forecast ensemble is sorted $\mathbf{X}_f = [x_{f,1}, x_{f,2}, ..., x_{f,k}]$ and bins $(-\infty, x_{f,1}]$, $(x_{f,1}, x_{f,2}]$, ... , $(x_{f,k}, \infty)$ are generated at each forecast step. We then increment whichever bin the truth falls into at each forecast step to produce a histogram of probabilities P_i of the truth being in bin i. A reliable ensemble is considered to be one where the truth and the ensemble members can be viewed as drawn from the same distribution. A flat probability density histogram therefore is seen as indicating a reliable ensemble for which each ensemble member has equal probability of being nearest to the truth. A convex probability density histogram indicates a lack of spread of the ensemble, while a concave diagram indicates an ensemble which exhibits too large spread. In Figure 11 we show ranked probability histograms for an observed variable with $\gamma = 0.5$, and all other parameters as in Figure 9. The figure shows that for $N_\mathrm{obs} = 3, 4$ and 8 the VLKF makes the forecast ensemble more reliable by controlling the spread.

In Figure 12 we show how the overall skill, as well as the climatic skills $\hat{\mathcal{S}}^\mathrm{E}$ and $\hat{\mathcal{S}}^\mathrm{V}$, vary with the observational noise η, where $\mathbf{R}_\mathrm{o} = (\eta\sigma_\mathrm{clim})^2\mathbf{I}$. The VLKF outperforms both the ETKF and the observations/climatology for large enough noise $\eta \geq 0.25$. Large observational noise gives preference to the pseudo-observables, whose unrealistically large error covariances as seen in Figure 5 are limited by the VLKF. This is particularly true for large model error. It is seen that for sufficiently large observational noise $\eta \geq 0.25$, the VLKF performs better than a poor man's analysis for all values of the model error γ.

5. Discussion

Our numerical results suggest that including climatological information of the mean and variance of the unobserved variables in an ensemble filter can be beneficial in the case when the forecast error covariances are overestimated. In particular, this occurs when sparse observations are taken at large observation intervals and the forecast model is sufficiently underdamped. However, we have also encountered situations where the additional variance constraint provided by the VLKF causes error covariances to be underestimated, spoiling the analysis (albeit only for the observables and not, as has to be stressed, for the overall RMS error). This is the case for moderate observation intervals with large model error, and also for large observation intervals in the case of near-perfect models, but only if one is interested in the analysis of just the observables. This latter situation is not particularly limiting, as RMS errors for the observables are relatively small compared to those for pseudo-observables, and the finite size effects can be eliminated here by basic multiplicative covariance inflation. For sufficiently large observational noise levels, when the analysis puts more weight on the forecast, controlling the unrealistic overestimation caused by underdamping model error using the VLKF becomes the more effective strategy.

We argue that model error associated with an underdamped forecast model leads to an increased forecast error covariance and increased associated ensemble spread by the same mechanism by which covariance inflation is beneficial. The variance limiting filter controls this increased spread and keeps it within the bounds of the climatologically reasonable bounds, thereby leading to a lower RMS error and improving the overall skill.

We remark that one may choose to drive the system towards values other than the climatological mean and variance. The variance limitation can be viewed as a simple strategy to numerically stabilize the filtering process. As such it may be more beneficial to choose values differing from climatology which are given by numerical necessities rather than by dynamical and physical considerations. For example, in non-equilibrium situations where the dynamics is locally off the attractor, such as in the presence of strong fronts, the climatological values are not meaningful. Our approach still provides a means to control the dynamically natural overestimation to numerically stabilize the filtering procedure, however this would imply careful tuning of the limiters $\mathbf{A}_\mathrm{clim}$ and $\mathbf{a}_\mathrm{clim}$. Furthermore, it would be computationally impossible in a full-scale model to treat all of the unobserved variables as pseudoobservables, so one would need to choose carefully which of the unobserved variables to apply the VLKF to.

It was noted in Gottwald *et al.* (2011) that the RMS errors are not sensitive to uncertainties in the estimation of $\mathbf{a}_\mathrm{clim}$ and $\mathbf{A}_\mathrm{clim}$, which may be obtained from historical data or from long time simulations of numerical models. However, instead of damping all pseudoobservables towards the

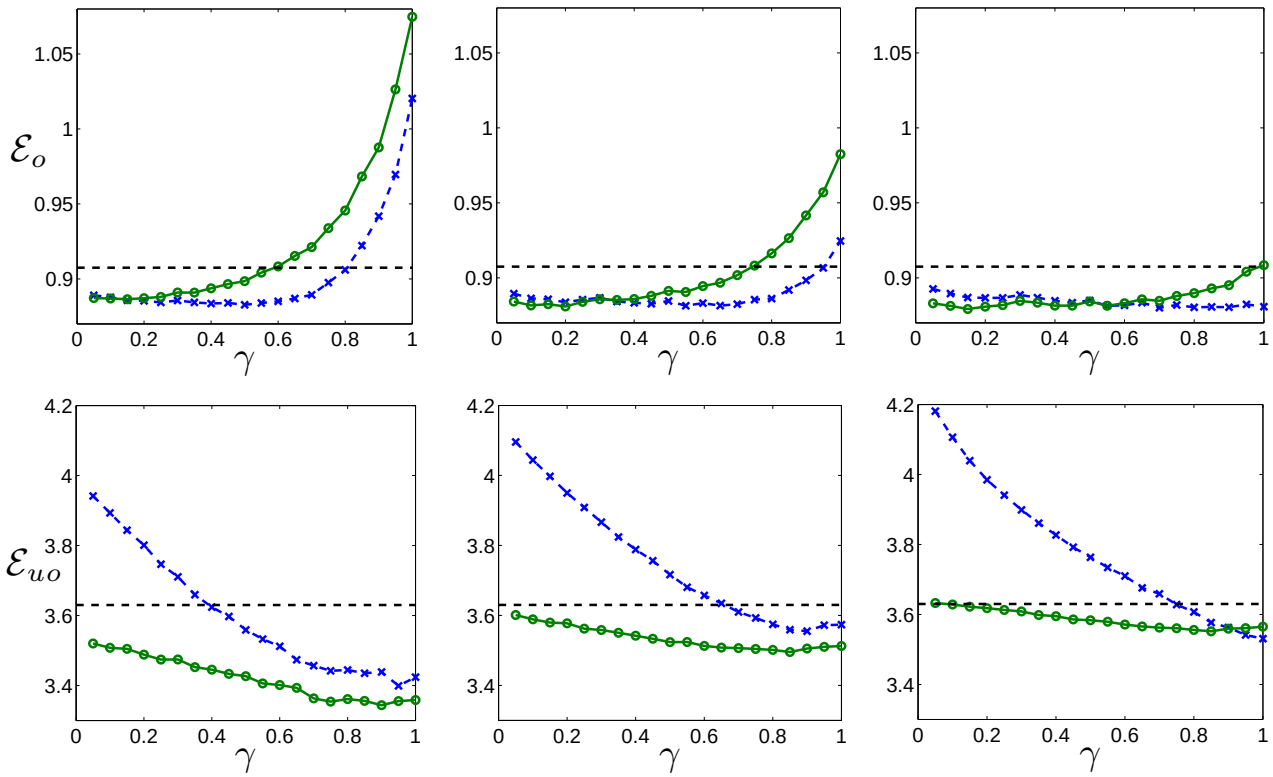

Figure 9. RMS errors of observed variables $\mathcal{E}_o$ (top) and of unobserved variables $\mathcal{E}_{uo}$ (bottom) for the ETKF (blue, dashed, crosses) and VLKF (green, dashed, circles) for $N_{\rm obs} = 3$ (left), $N_{\rm obs} = 4$ (centre) and $N_{\rm obs} = 8$ (right). Top and bottom horizontal dashed lines show $\sqrt{\mathbf{R}_{\rm o}}$ and $\sigma_{\rm clim}$, respectively. We used $\mathbf{R}_{\rm o} = (0.25\sigma_{\rm clim})^2\mathbf{I}$ and $\Delta t_{\rm obs} = 48$ hours.

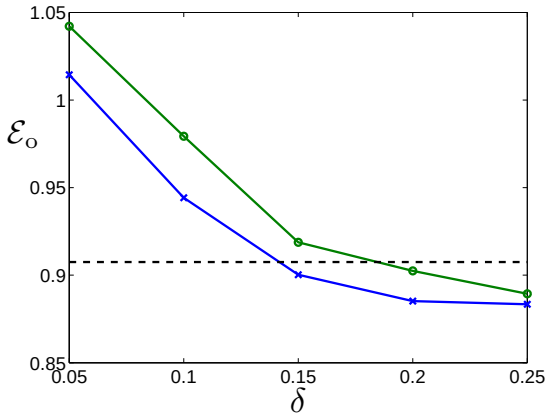

Figure 10. RMS errors $\mathcal{E}_o$ for ETKF (blue, dashed, crosses) and VLKF (green, dashed, circles) for increasing forecast covariance inflation δ. The dashed line shows $\sqrt{\mathbf{R}_{\rm o}}$. We used $\gamma = 1$, $\Delta t_{\rm obs} = 48$ hrs, $N_{\rm obs} = 3$.

same spatially homogeneous climatological value $\mathbf{a}_{\rm clim}$, one may wish to damp towards a spatially inhomogeneous finite-time statistics. One may, for example, use a running mean of the past history of analyses $\mathbf{z}_a$ instead of the constant $\mathbf{a}_{\rm clim}$ we have used here. However, for the long observation times used here, the analysis is not tracking (cf. Figure 4). Therefore driving the pseudoobservables towards non-climatological values of $\mathbf{a}_{\rm clim}$ will introduce on average larger RMS errors, degrading the skill. It is exactly the situation of large observational intervals where the quality of an analysis cannot be measured by how well it estimates the actual truth but by how well the statistics of the truth is reproduced. The situation of large observation intervals and sparse observational grids is particularly relevant for climate reanalysis in the pre-radiosonde era. In this case long time climatology was shown to be beneficial as additional information in data assimilation. Our results show that model error may provide the necessary ensemble spread needed to combat finite ensemble size effects, similarly to covariance inflation; VLKF controls this spread by limiting unrealistically large analysis error covariances.

The problem we addressed here is closely related to forecast bias correction schemes (Takacs 1996). Dee and Da Silva (1998) introduced a filtering strategy where a forecast bias is estimated using a running mean which

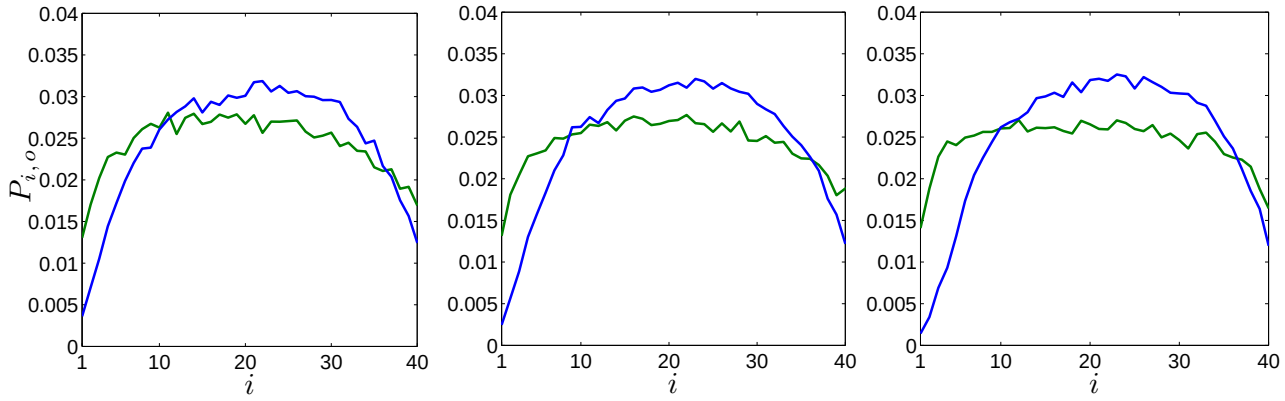

Figure 11. Ranked probability histograms for the observables for ETKF (blue) and VLKF (green), with $N_{\rm obs} = 3$ (left), $N_{\rm obs} = 4$ (centre) and $N_{\rm obs} = 8$ (right). We used $\gamma = 0.5$, all other parameters as in Figure 9.

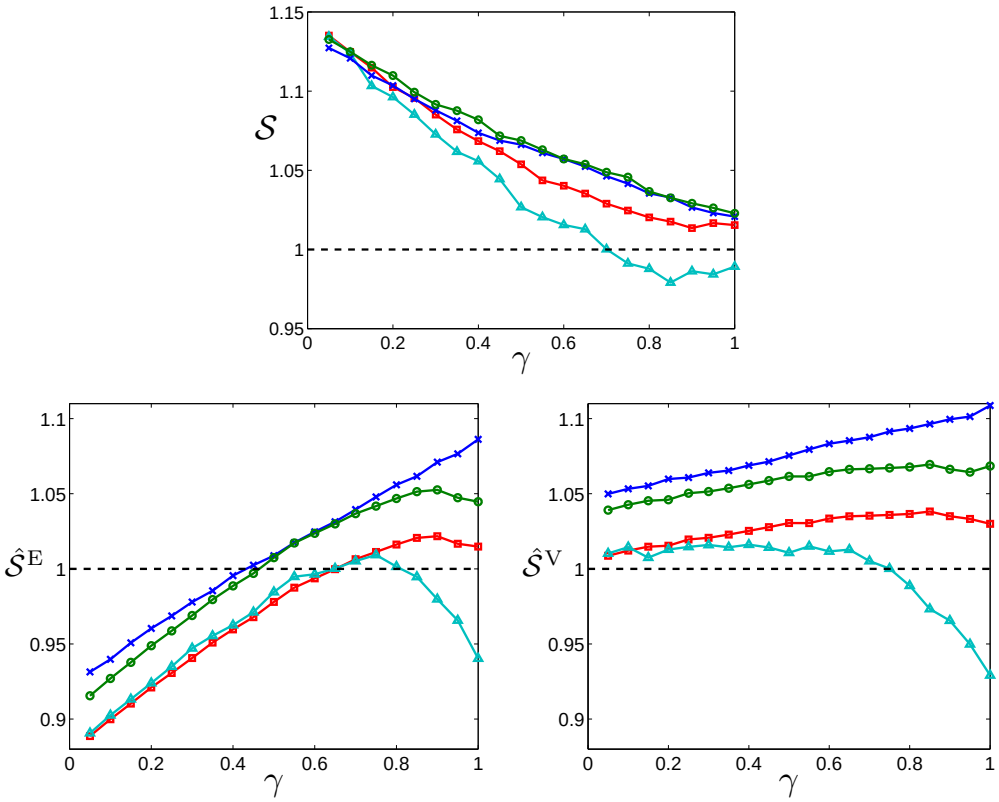

Figure 12. Overall skill $\mathcal{S}$ (top) and climatic skills $\mathcal{S}^{\rm E}$ (bottom left) and $\mathcal{S}^{\rm V}$ (bottom right) as a function of γ for different levels of observational noise η, where $\mathbf{R}_{\rm o} = (\eta\sigma_{\rm clim})^2\mathbf{I}$. Here $\eta = 1$ (blue, crosses), $\eta = 0.5$ (green, circles), $\eta = 0.25$ (red, squares) and $\eta = 0.15$ (blue-green, triangles). We used $\Delta t_{\rm obs} = 48$ hours and $N_{\rm obs} = 4$.

was then subtracted from the forecast prior to the analysis. This strategy was applied to the Goddard Earth Observing System moisture analysis (Dee and Todling 2000). We are dealing here with the additional problem of sparse observational grids. Both approaches, however, gain their skill improvement from driving the dynamics towards a mean which was estimated from the preceding dynamics using a running average rather than allowing the analysis to drift away due to forecast bias caused as in our case by model error.

We have used here underdamping as a model for numerically unstable dynamic cores; our forecast models still produce stable dynamics, albeit with significantly increased covariances. The effectiveness of our approach at controlling actual numerical instabilities remains an open question planned for further research. We mention the work by Grote and Majda (2006) who addressed a similar problem and designed a filter which assures an observability condition in the case of unstable dynamics in the dynamic core where the unstable dynamics is tamed by appropriately incorporating observations. Furthermore, we have not addressed the important issue raised in the introduction whether the proposed method is able to produce realistic energy spectra and control the energy

dissipation rates. This again may require tuning of the limiters and is planned for further research.

Acknowledgements

We thank Sebastian Reich for bringing the problem of underdamping model error to our attention. LM acknowledges the support of an Australian Postgraduate Award. GAG acknowledges support from the Australian Research Council.

References

Anderson JL. 1996. A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *Journal of Climate* **9**(7): 1518–1530.

Anderson JL. 2001. An ensemble adjustment Kalman filter for data assimilation. *Monthly Weather Review* **129**(12): 2884–2903.

Anderson JL, Anderson SL. 1999. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review* **127**(12): 2741–2758.

Bengtsson L, Arkin P, Berrisford P, Bougeault P, Folland CK, Gordon C, Haines K, Hodges KI, Jones P, Kallberg P, Rayner N, Simmons A, Stammer D, Thorne PW, Uppala S, Vose RS. 2007. The need for a dynamical climate reanalysis. *Bulletin of the American Meteorological Society* **88**(4): 495–501.

Bergemann K, Gottwald GA, Reich S. 2009. Ensemble propagation and continuous matrix factorization algorithms. *Quarterly Journal of the Royal Meteorological Society* **135**(643): 1560–1572.

Bishop CH, Etherton BJ, Majumdar SJ. 2001. Adaptive sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical aspects. *Monthly Weather Review* **129**(3): 420–436.

Blumen W. 1990. A semi-geostrophic Eady-wave frontal model incorporating momentum diffusion. Part II: Kinetic energy and enstrophy dissipation. *Journal of the Atmospheric Sciences* **47**: 2903–2908.

Buizza R, Houtekamer PL, Pellerin G, Toth Z, Zhu Y, Wei M. 2005. A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Monthly Weather Review* **133**(5): 1076–1097.

Burgers G, van Leeuwen PJ, Evensen G. 1998. Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review* **126**(6): 1719–1724.

Charron M, Pellerin G, Spacek L, Houtekamer PL, Gagnon N, Mitchell HL, Michelin L. 2010. Toward random sampling of model error in the Canadian ensemble prediction system. *Monthly Weather Review* **138**(5): 1877–1901.

Compo GP, Whitaker JS, Sardeshmukh PD, Matsui N, Allan RJ, Yin X, Gleason BE, Vose RS, Rutledge G, Bessemoulin P, Bronnimann S, Brunet M, Crouthamel RI, Grant AN, Groisman PY, Jones PD, Kruk MC, Kruger AC, Marshall GJ, Maugeri M, Mok HY, Nordli V, Ross TF, Trigo RM, Wang XL, Woodruff SD, Worley SJ. 2011. The Twentieth Century Reanalysis Project. *Quarterly Journal of the Royal Meteorological Society* **137**(654): 1–28.

Côté J, Staniforth A. 1988. A two-time-level semi-Lagrangian semi-implicit scheme for spectral models. *Monthly Weather Review* **116**(10): 2003–2016.

Dee DP, Da Silva AM. 1998. Data assimilation in the presence of forecast bias. *Quarterly Journal of the Royal Meteorological Society* **124**: 269–295.

Dee DP, Todling R. 2000. Data assimilation in the presence of forecast bias: The GEOS moisture analysis. *Monthly Weather Review* **128**: 3268–3282.

Dubinkina S, Frank J. 2007. Statistical mechanics of Arakawa's discretizations. *Journal of Computational Physics* **227**(2): 1286 – 1305.

Dubinkina S, Frank J. 2010. Statistical relevance of vorticity conservation in the Hamiltonian particle-mesh method. *Journal of Computational Physics* **229**(7): 2634 – 2648.

Durran DR. 1999. *Numerical Methods for Wave Equations in Geophysical Fluid Dynamics*. Springer: New York.

Eckermann SD, Hoppel KW, Coy L, McCormack JP, Siskind DE, Nielsen K, Kochenash A, Stevens MH, Englert CR, Singer W, Hervig M. 2009. High-altitude data assimilation system experiments for the northern summer mesosphere season 2007. *Journal of Atmospheric and Solar-Terrestrial Physics* **71**(3-4): 531–551.

Evensen G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research* **99**(C5): 10 143–10 162.

Evensen G. 2006. *Data Assimilation: The Ensemble Kalman Filter*. Springer: New York.

Fisher M, Leutbecher M, Kelly GA. 2005. On the equivalence between Kalman smoothing and weak-constraint four-dimensional variational data assimilation. *Quarterly Journal of the Royal Meteorological Society* **131**(613): 3235–3246.

Frederiksen JS, Davies AG. 1997. Eddy viscosity and stochastic backscatter parameterizations on the sphere for atmospheric circulation models. *Journal of the Atmospheric Sciences* **54**(20): 2475–2492.

Golub GH, Loan CFV. 1996. *Matrix Computations*. The Johns Hopkins University Press: Baltimore, 3rd edn.

Gottwald GA, Mitchell L, Reich S. 2011. Controlling overestimation of error covariance in ensemble Kalman filters with sparse observations: A variance limiting Kalman filter. *Monthly Weather Review* **139**(8): 2650–2667.

Grote MJ, Majda AJ. 2006. Stable time filtering of strongly unstable spatially extended systems. *Proceedings of the National Academy of Sciences* **103**(20): 7548–7553.

Hamill TM, Colucci SJ. 1997. Verification of Eta/RSM short-range ensemble forecasts. *Monthly Weather Review* **125**(6): 1312–1327.

Hamill TM, Whitaker JS. 2011. What constrains spread growth in forecasts initialized from ensemble Kalman filters. *Monthly Weather Review* **139**: 117–131.

Houtekamer PL, Mitchell HL. 1998. Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review* **126**(3): 796–811.

Houtekamer PL, Mitchell HL, Pellerin G, Buehner M, Charron M, Spacek L, Hansen B. 2005. Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Monthly Weather Review* **133**(3): 604–620.

Kalnay E. 2002. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press: Cambridge.

Leimkuhler B, Reich S. 2005. *Simulating Hamiltonian Dynamics*. Cambridge University Press: Cambridge.

Liu J, Fertig EJ, Li H, Kalnay E, Hunt BR, Kostelich EJ, Szunyogh I, Todling R. 2008. Comparison between local ensemble transform Kalman filter and PSAS in the NASA finite volume GCM - perfect model experiments. *Nonlinear Processes in Geophysics* **15**(4): 645–659.

Lorenz EN. 1996. Predictability - a problem partly solved. In: *Predictability*, Palmer T (ed), European Centre for Medium-Range Weather Forecast, Shinfield Park, Reading, UK.

Majda AJ, Timofeyev I, Vanden Eijnden E. 1999. Models for stochastic climate prediction. *Proceedings of the National Academy of Sciences* **96**(26): 14 687–14 691.

Mitchell L, Gottwald GA. 2012. Data assimilation in slow-fast systems using homogenized climate models. *Journal of the Atmospheric Sciences* **69**(4): 1359–1377.

Molteni F, Palmer TN. 1993. Predictability and finite-time instability of the northern winter circulation. *Quarterly Journal of the Royal Meteorological Society* **119**: 269–298.

Ott E, Hunt B, Szunyogh I, Zimin A, Kostelich E, Corrazza M, Kalnay E, Yorke J. 2004. A local ensemble Kalman filter for atmospheric data assimilation. *Tellus A* **56**(5): 415–428.

Palmer T, Williams P (eds). 2010. *Stochastic physics and climate modelling*. Cambridge University Press: Cambridge.

Palmer TN. 2001. A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models. *Quarterly Journal of the Royal Meteorological Society* **127**(572): 279–304.

Polavarapu S, Shepherd TG, Rochon Y, Ren S. 2005. Some challenges of middle atmosphere data assimilation. *Quarterly Journal of the Royal Meteorological Society* **131**(613): 3513–3527.

Ritchie H. 1988. Application of the semi-Lagrangian method to a spectral model of the shallow-water equations. *Monthly Weather Review* **116**: 1587–1598.

Salmon R. 1998. *Lectures on Geophysical Fluid Dynamics*. Oxford University Press: New York.

Sankey D, Ren S, Polavarapu S, Rochon Y, Nezlin Y, Beagley S. 2007. Impact of data assimilation filtering methods on the mesosphere. *Journal of Geophysical Research* **112**(D24): 104–118.

Shutts GJ. 2005. A stochastic kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quarterly Journal of the Royal Meteorological Society* **131**(612): 3079–3102.

Simon DJ. 2006. *Optimal State Estimation*. John Wiley & Sons, Inc.: New York.

Skamarock WC, Klemp JB. 1992. The stability of time-split numerical methods for the hydrostatic and the nonhydrostatic elastic equations. *Monthly Weather Review* **120**(9): 2109–2127.

Takacs LL. 1996. A simple bias correction algorithm for use in data assimilation. *DAO Office Note 96-25* .

Talagrand O, Vautard R, Strauss B. 1997. Evaluation of probabilistic prediction systems. In: *Proceedings of the ECMWF Workshop on Predictability*. ECMWF.

Thuburn J. 2008. Some conservation issues for the dynamical cores of NWP and climate models. *Journal of Computational Physics* **227**(7): 3715–3730.

Tippett MK, Anderson JL, Bishop CH, Hamill TM, Whitaker JS. 2003. Ensemble square root filters. *Monthly Weather Review* **131**(7): 1485–1490.

Toth Z, Kalnay E. 1993. Ensemble forecasting at NMC: The generation of perturbations. *Bulletin of the American Meteorological Society* **74**: 2317–2330.

Wang X, Bishop CH, Julier SJ. 2004. Which is better, an ensemble of positive-negative pairs or a centered spherical simplex ensemble? *Monthly Weather Review* **132**(7): 1590–1505.

Whitaker JS, Compo GP, Thépaut JN. 2009. A comparison of variational and ensemble-based data assimilation systems for reanalysis of sparse observations. *Monthly Weather Review* **137**(6): 1991–1999.

Whitaker JS, Compo GP, Wei X, Hamill TM. 2004. Reanalysis without radiosondes using ensemble data assimilation. *Monthly Weather Review* **132**(5): 1190–1200.