Nonlin. Processes Geophys., 21, 417–426, 2014
www.nonlin-processes-geophys.net/21/417/2014/
doi:10.5194/npg-21-417-2014

Nonlinear Processes in Geophysics
Open Access

Controlling balance in an ensemble Kalman filter

G. A. Gottwald

School of Mathematics and Statistics, University of Sydney, Sydney, NSW 2006, Australia

Correspondence to: G. A. Gottwald (georg.gottwald@sydney.edu.au)

Received: 7 May 2013 – Revised: 13 December 2013 – Accepted: 7 February 2014 – Published: 14 March 2014

Abstract. We present a method to control unbalanced fast dynamics in an ensemble Kalman filter by introducing a weak constraint on the imbalance in a spatially sparse observational network. We show that the balance constraint produces significantly more balanced analyses than ensemble Kalman filters without balance constraints and than filters implementing incremental analysis updates (IAU). Furthermore, our filter with the weak constraint on imbalance produces good rms error statistics which outperform those of ensemble Kalman filters without balance constraints for the fast fields.

1 Introduction

In data assimilation one seeks to find the best estimation of the state of a dynamical system given a forecast model with possible model error and noisy observations at discrete observation intervals (Kalnay, 2002). This estimate is coined the *analysis*. This procedure, however, does not necessarily produce dynamically consistent analyses. In particular, the analysis may contain unbalanced gravity waves, which are absent in the true atmospheric state and which may spoil the subsequent forecast initialised with these dynamically inconsistent states. Ever since the early days of numerical weather prediction the creation of imbalance has been central to the problem of producing reliable forecasts (see for example Daley, 1993, Chapter 6, and Lynch, 2006 for a historical account). The heuristic reasoning behind the occurrence of unbalanced analyses is that there may be several states of the fast variables which are compatible with the observations of the slow state variables, most of them corresponding to unbalanced states. Furthermore, unbalanced states can be generated by the discontinuous nature of the data assimilation procedure, leading to unphysical readjustment processes of analyses by the subsequent nonlinear forecast model (Bloom et al., 1996; Ourmières et al., 2006). Examples of the creation of imbalance in variational data assimilation schemes are, for example, Bloom et al. (1996) and Lorenc (2003b). In the context of ensemble filters, unbalanced analyses are further created by the procedure of localisation which was introduced by Houtekamer and Mitchell (1998, 2001), Hamill et al. (2001), Ott et al. (2004) and Szunyogh et al. (2005) to mitigate spurious cross-correlations in the covariance matrices due to finite ensemble sizes. Localisation of any type can potentially cause imbalance in the initial conditions (Cohn et al., 1998; Lorenc, 2003a; Mitchell et al., 2002; Houtekamer and Mitchell, 2005; Oke et al., 2007; Kepert, 2009; Greybush et al., 2011).

There exist several strategies to combat undesired unbalanced analyses. These strategies can be divided into those which employ a re-balancing procedure after the data assimilation, and those which try to create balanced analyses within the data assimilation process itself. Post-processing methods include digital filtering (Lynch and Huang, 1992) and normal mode initialisation (Machenhauer, 1977; Baer and Tribbia, 1977). Within variational data assimilation algorithms balance constraints can be implemented to ensure sufficient balance (Thépaut and Courtier, 1991; Polavarapu et al., 2000; Gauthier and Thépaut, 2001; Neef et al., 2006; Watkinson et al., 2007; Cotter, 2013). To militate against the effects of intermittent discontinuous assimilations, several filtering approaches have been introduced to render the assimilation procedure more continuous. Bloom et al. (1996) introduced the method of incremental analysis updates (IAU) for 3D-Var in which the analyses increments are distributed over a fixed time window. It has since been applied to ensemble filters, see for example Polavarapu et al. (2004), and has found numerous applications in atmospheric and oceanic contact (Zhu et al., 2003; Weaver et al., 2003; Ourmières et al., 2006). Bergemann and Reich (2010a) create balanced analyses by using a continuous formulation of the Kalman

analysis step (Bergemann et al., 2009; Bergemann and Reich, 2010b). Kepert (2009) modified the covariance localisation procedure so that it respects balance.

Here we will present a novel approach to generating balanced analyses within an intermittent discontinuous data assimilation procedure. We will incorporate prior information on the amount of imbalance to augment given observational information for the slow variables. This implementation of a balance constraint within the data assimilation step eliminates unwanted spurious imbalance, leading to physical analyses states and to an improved analysis skill as measured by the rms error of the fast variables.

In the next section we briefly describe the framework of variance limiting Kalman filters developed in Gottwald et al. (2011) which will form the basis of our imbalance limiting filter. In Sect. 3 we present a modified slow-fast Lorenz-96 model which incorporates balanced dynamics, introduced in Bergemann and Reich (2010a). In Sect. 4 we present results showing how controlling unbalance can produce better skill than current ensemble Kalman filters. We conclude with a summary in Sect. 5.

2 The variance limiting Kalman filter

Gottwald et al. (2011) introduced a variation of the ensemble Kalman filter, coined variance limiting Kalman filter (VLKF). This filter was designed to control overestimation of error covariances caused by finite ensemble sizes in sparse observational grids. The filter imposes weak constraints on unobserved variables and data voids using climatological information. The effect of the weak constraint was shown to drive the analysis of the unobserved variables towards their climatic mean and furthermore to limit the posterior error covariance of the unobserved variables to not exceed their climatic covariance. This yielded a remarkable increase in the skill, even in the observed variables. The filter has since been used in Mitchell and Gottwald (2012) to control noise at the grid resolution scale caused by model error.

It is our aim here to employ the VLKF to control undesirable imbalance. In general the instantaneous amount of imbalance is not available through direct observations. We assume prior knowledge of the climatological mean and of the climatological covariance of imbalance. This statistical information may be available through historical observational data or through free running simulations. We will use the weak constraint in VLKF on imbalance to drive the analysis towards balance, inhibiting excessive unphysical unbalanced fast energy.

The filter described in Gottwald et al. (2011) and Mitchell and Gottwald (2012) was formulated for large ensemble sizes, ensuring invertibility of the forecast error covariance (a situation not satisfied for data assimilation in operational numerical weather forecast centres). We recast the VLKF here in a form which allows for small ensemble sizes, and redo the derivation in a slightly different manner.

Given a D-dimensional dynamical system

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

which is observed at discrete times $t_i = i\,\Delta t_{\text{obs}}$, data assimilation aims at producing the best estimate of the current state given a typically chaotic, possibly inaccurate model $\dot{z} = \boldsymbol{f}(z)$ and noisy observations of the true state z_t (Kalnay, 2002). We assume that we are given observations

$$\boldsymbol{y}_{\text{o}}(t_i) = \mathbf{H} z_t(t_i) + \boldsymbol{r}_{\text{o}},$$

where the observation operator $\mathbf{H} : \mathbb{R}^D \to \mathbb{R}^{D_{\text{o}}}$ maps from the whole space into observation space, and $\boldsymbol{r}_{\text{o}} \in \mathbb{R}^{D_{\text{o}}}$ is assumed to be i.i.d. observational Gaussian noise with associated error covariance matrix $\mathbf{R}_{\text{o}}$. Additionally we incorporate climatological information of D_{w} pseudo-observables, in particular their mean $\boldsymbol{a}_{\text{clim}} \in \mathbb{R}^{D_{\text{w}}}$ and their covariances $\mathbf{A}_{\text{clim}} \in \mathbb{R}^{D_{\text{w}} \times D_{\text{w}}}$. In general, it is not advisable to incorporate simultaneously direct observations and climatological information for a variable, as this may spoil the generally more accurate information of the direct observations. In our application here the climatological information will be the mean and covariances of some measure of imbalance, but pseudo-observables may be any subset of unobserved variables or their integrated quantities such as their energy. We assume that we can determine those quantities prior to the data assimilation procedure either through historical data or through long-time numerical simulations. We remark that one may use values other than the climatic covariance to control the analysis error covariance if one interprets the variance constraint merely as a numerical tool to stabilise and regularise the filter. Furthermore, in non-equilibrium situations, when climatological information is irrelevant, such as during strong fronto-genesis in a weather forecasting context for example, we may estimate the mean and the covariance of the unobserved pseudo-observables via a running average of the analysis (this requires the analysis to be tracking).

We introduce a pseudo-observation operator $\mathbf{h} : \mathbb{R}^D \to \mathbb{R}^{D_{\text{w}}}$ which maps from the whole space into the space of the pseudo-observables. The (as yet unknown) error covariance of those pseudo-observations is denoted by $\mathbf{R}_{\text{w}}$. Gottwald et al. (2011) considered $D_{\text{o}} + D_{\text{w}} = D$, which will be relaxed here.

The Kalman filter can be formulated as a minimisation problem of the following cost function (e.g. Kalnay, 2002; Simon, 2006) with a given background z_{f} and associated error forecast covariance $\mathbf{P}_{\text{f}}$ as

$$\begin{aligned} J(z) = &\frac{1}{2}(z - z_{\text{f}})^T \mathbf{P}_{\text{f}}^{-1}(z - z_{\text{f}}) \\ &+ \frac{1}{2}(\mathbf{H}z - \boldsymbol{y}_{\text{o}})^T \mathbf{R}_{\text{o}}^{-1}(\mathbf{H}z - \boldsymbol{y}_{\text{o}}) \\ &+ \frac{1}{2}(\mathbf{h}z - \boldsymbol{a}_{\text{clim}})^T \mathbf{R}_{\text{w}}^{-1}(\mathbf{h}z - \boldsymbol{a}_{\text{clim}}). \end{aligned} \tag{2}$$

The error covariance matrix $\mathbf{R}_\mathrm{w}$ is so far undetermined. We will invoke below a constraint on the analysis error covariance, namely that the analysis error covariance projected onto the subspace spanned by the pseudo-observations equals the climatological covariance $\mathbf{A}_\mathrm{clim}$. In anticipation of the analytical results below which reveal that such a constraint cannot be imposed on the whole D_w-dimensional unobserved subspace whilst simultaneously ensuring positive definiteness of $\mathbf{R}_\mathrm{w}$, but only on a $\hat{D}_\mathrm{w} \le D_\mathrm{w}$-dimensional subspace of the unobserved subspace, we introduce here a (so far undetermined) transformation matrix $\mathbf{S}_\mathrm{w} \in \mathbb{R}^{D_\mathrm{w} \times \hat{D}_\mathrm{w}}$. The transformation matrix satisfies $\mathbf{S}_\mathrm{w}^T \mathbf{S}_\mathrm{w} = \mathbf{I}_{\hat{D}_\mathrm{w}}$ (but not necessarily $\mathbf{S}_\mathrm{w} \mathbf{S}_\mathrm{w}^T = \mathbf{I}_{D_\mathrm{w}}$). We will formulate the filter restricted to this subspace and introduce the transformed pseudo-observation operator

$$\hat{\mathbf{h}} = \mathbf{S}_\mathrm{w}^T \mathbf{h},$$

and the transformed error covariances

$$\hat{\mathbf{R}}_\mathrm{w}^{-1} = \mathbf{S}_\mathrm{w}^T \mathbf{R}_\mathrm{w}^{-1} \mathbf{S}_\mathrm{w}$$
$$\hat{\mathbf{A}}_\mathrm{clim} = \mathbf{S}_\mathrm{w}^T \mathbf{A}_\mathrm{clim} \mathbf{S}_\mathrm{w},$$

as well as the transformed climatological mean of the pseudo-observations $\hat{\boldsymbol{a}}_\mathrm{clim} = \mathbf{S}_\mathrm{w}^T \boldsymbol{a}_\mathrm{clim}$. We now combine direct observations and pseudo-observations, and write the cost function in the more compact form

$$J(z) = \frac{1}{2} (z - z_\mathrm{f})^T \mathbf{P}_\mathrm{f}^{-1} (z - z_\mathrm{f}) + \frac{1}{2} \left(\hat{\mathbf{H}} z - \hat{y}\right)^T \hat{\mathbf{R}}^{-1} \left(\hat{\mathbf{H}} z - \hat{y}\right), \quad (3)$$

where we introduced combined observations $\hat{y}$, the observation operator $\hat{\mathbf{H}}$ and the error covariance matrix $\hat{\mathbf{R}}$ with

$$\hat{y} = \begin{pmatrix} y_\mathrm{o} \\ \hat{\boldsymbol{a}}_\mathrm{clim} \end{pmatrix} \in \mathbb{R}^{D_\mathrm{o} + \hat{D}_\mathrm{w}},$$
$$\hat{\mathbf{H}} = \begin{pmatrix} \mathbf{H} \\ \hat{\mathbf{h}} \end{pmatrix} \in \mathbb{R}^{(D_\mathrm{o} + \hat{D}\mathrm{w}) \times D},$$
$$\hat{\mathbf{R}}^{-1} = \begin{pmatrix} \mathbf{R}_\mathrm{o}^{-1} & 0 \\ 0 & \hat{\mathbf{R}}_\mathrm{w}^{-1} \end{pmatrix} \in \mathbb{R}^{(D_\mathrm{o} + \hat{D}_\mathrm{w}) \times (D_\mathrm{o} + \hat{D}_\mathrm{w})}.$$

The analysis is given as the minimum of the cost function $J(z)$ and is readily calculated as

$$z_\mathrm{a} = z_\mathrm{f} - \hat{\mathbf{K}} \left[\hat{\mathbf{H}} z_\mathrm{f} - \hat{y}\right], \quad (4)$$

where the Kalman gain matrix is given by

$$\hat{\mathbf{K}} = \mathbf{P}_\mathrm{a} \hat{\mathbf{H}}^T \hat{\mathbf{R}}^{-1}, \quad (5)$$

with the error covariance matrix of the analysis given by

$$\mathbf{P}_\mathrm{a} = \left(\mathbf{P}_\mathrm{f}^{-1} + \hat{\mathbf{H}}^T \hat{\mathbf{R}}^{-1} \hat{\mathbf{H}}\right)^{-1}. \quad (6)$$

Using the matrix identity $(\mathbf{P}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} = \mathbf{P} - \mathbf{P}\mathbf{H}^T (\mathbf{R} + \mathbf{H}\mathbf{P}\mathbf{H}^T)^{-1} \mathbf{H}\mathbf{P}$ (see for example Simon, 2006) the analysis error covariance is recast in a form which does not involve the inverse of the forecast error covariance $\mathbf{P}_\mathrm{f}$ as

$$\mathbf{P}_\mathrm{a} = \left[\mathbf{I} - \hat{\mathbf{K}} \hat{\mathbf{H}}\right] \mathbf{P}_\mathrm{f}, \quad (7)$$

and the Kalman gain matrix can be rewritten in the computationally more convenient form

$$\hat{\mathbf{K}} = \mathbf{P}_\mathrm{f} \hat{\mathbf{H}}^T \left(\hat{\mathbf{H}} \mathbf{P}_\mathrm{f} \hat{\mathbf{H}}^T + \hat{\mathbf{R}}\right)^{-1}, \quad (8)$$

which involves only taking the inverse of $((D_\mathrm{o} + \hat{D}_\mathrm{w}) \times (D_\mathrm{o} + \hat{D}_\mathrm{w}))$ matrices rather than of $(D \times D)$ matrices.

We remark that one can explicitly separate the updates according to the deviations from the observations and the pseudo-observations in the analysis and have

$$z_\mathrm{a} = z_\mathrm{f} - \mathbf{K}_\mathrm{o} \left[\mathbf{H} z_\mathrm{f} - y_\mathrm{o}\right] - \hat{\mathbf{K}}_\mathrm{w} \left[\hat{\mathbf{h}} z_\mathrm{f} - \hat{\boldsymbol{a}}_\mathrm{clim}\right], \quad (9)$$

with

$$\mathbf{K}_\mathrm{o} = \mathbf{P}_\mathrm{a} \mathbf{H}^T \mathbf{R}_\mathrm{o}^{-1} \quad \text{and} \quad \hat{\mathbf{K}}_\mathrm{w} = \mathbf{P}_\mathrm{a} \hat{\mathbf{h}}^T \hat{\mathbf{R}}_\mathrm{w}^{-1}.$$

This shows that weighted by the error covariance of the weak constraint $\hat{\mathbf{R}}_\mathrm{w}$ the analysis of the pseudo-observables is driven towards their climatic mean. However, due to the generically global nature of the Kalman gain matrices the inclusion of climatological information of the pseudo-observables also affects the observed degrees of freedom.

So far the error covariance $\hat{\mathbf{R}}_\mathrm{w}$ associated with the weak constraint is undetermined. We will now determine $\hat{\mathbf{R}}_\mathrm{w}$ and thereby control the variance of the unresolved pseudo-observables $\mathbf{h}z$ by requiring that the analysis error covariance, projected onto the pseudo-observables, equals the climatological covariance, i.e.

$$\mathbf{h} \mathbf{P}_\mathrm{a} \mathbf{h}^T = \mathbf{A}_\mathrm{clim}. \quad (10)$$

We rewrite the analysis error covariance (Eq. 6) as

$$\mathbf{P}_\mathrm{a} = \left(\mathbf{P}^{-1} + \hat{\mathbf{h}}^T \hat{\mathbf{R}}_\mathrm{w}^{-1} \hat{\mathbf{h}}\right)^{-1}, \quad (11)$$

where we introduced the analysis error covariance $\mathbf{P}$ for a standard Kalman filter without any weak constraint which only combines the forecast with direct observations

$$\mathbf{P} = \left(\mathbf{P}_\mathrm{f}^{-1} + \mathbf{H}^T \mathbf{R}_\mathrm{o}^{-1} \mathbf{H}\right)^{-1} = [\mathbf{I} - \mathbf{K}_\mathrm{o} \mathbf{H}] \mathbf{P}_\mathrm{f}. \quad (12)$$

Upon using the Sherman–Morrison–Woodbury formula $(\mathbf{P}^{-1} + \hat{\mathbf{h}}^T \hat{\mathbf{R}}_\mathrm{w}^{-1} \hat{\mathbf{h}})^{-1} \hat{\mathbf{h}}^T \hat{\mathbf{R}}_\mathrm{w}^{-1} = \mathbf{P}\hat{\mathbf{h}}^T (\hat{\mathbf{R}}_\mathrm{w} + \hat{\mathbf{h}}\mathbf{P}\hat{\mathbf{h}}^T)^{-1}$ (see for example Simon, 2006), the error covariance matrix for the pseudo-observables $\hat{\mathbf{R}}_\mathrm{w}$ is found from the constraint (Eq. 10) to be

$$\hat{\mathbf{R}}_\mathrm{w}^{-1} = \hat{\mathbf{A}}_\mathrm{clim}^{-1} - \left(\hat{\mathbf{h}} \mathbf{P} \hat{\mathbf{h}}^T\right)^{-1}. \quad (13)$$

Expanding Eq. (13) we obtain

$$\left(\mathbf{S}_\mathrm{w}^T\mathbf{R}_\mathrm{w}\mathbf{S}_\mathrm{w}\right)^{-1} = \left(\mathbf{S}_\mathrm{w}^T\mathbf{A}_\mathrm{clim}\mathbf{S}_\mathrm{w}\right)^{-1} - \left(\mathbf{S}_\mathrm{w}^T\mathbf{h}\mathbf{P}\mathbf{h}^T\mathbf{S}_\mathrm{w}\right)^{-1}. \quad (14)$$

This makes apparent the role of the transformation matrix $\mathbf{S}_\mathrm{w}$. $\mathbf{S}_\mathrm{w} \in \mathbb{R}^{D_\mathrm{w}\times\hat{D}_\mathrm{w}}$ can be chosen such that $\hat{\mathbf{R}}_\mathrm{w} = \mathbf{S}_\mathrm{w}^T\mathbf{R}_\mathrm{w}\mathbf{S}_\mathrm{w}$, being an error covariance, is positive definite: the transformation matrix $\mathbf{S}_\mathrm{w}$ projects onto the subspace of the space of pseudo-observables which in a standard Kalman filter would experience an analysis error covariance $\mathbf{h}\mathbf{P}\mathbf{h}^T$ exceeding the climatological covariance $\mathbf{A}_\mathrm{clim}$. All other $D_\mathrm{w} - \hat{D}_\mathrm{w}$ pseudo-observations are discarded in order to ensure a positive definite and invertible error covariance matrix $\hat{\mathbf{R}}_\mathrm{w} \in \mathbb{R}^{\hat{D}_\mathrm{w}\times\hat{D}_\mathrm{w}}$. In Appendix A we provide an algorithm to compute $\mathbf{S}_\mathrm{w}$.

We formulate the filter in the framework of ensemble Kalman filters (EnKF) (Evensen, 2006; Hamill, 2006) where an ensemble with k members z_k

$$\mathbf{Z} = [z_1,\ z_2,\ \ldots,\ z_k] \in \mathbb{R}^{D\times k}$$

is propagated by the model dynamics according to the model

$$\dot{\mathbf{Z}} = \boldsymbol{F}(\mathbf{Z}),\ \mathbf{F}(\mathbf{Z}) = \left[\boldsymbol{f}(z_1),\ \boldsymbol{f}(z_2),\ \ldots,\ \boldsymbol{f}(z_k)\right] \in \mathbb{R}^{D\times k}.$$

The forecast ensemble is split into its mean $\overline{z}_\mathrm{f}$ and ensemble deviation matrix $\mathbf{Z}'_\mathrm{f}$. The ensemble deviation matrix $\mathbf{Z}'_\mathrm{f}$ can be used to provide a Monte Carlo estimate for the ensemble forecast covariance matrix via

$$\mathbf{P}_\mathrm{f}(t) = \frac{1}{k-1}\mathbf{Z}'(t)\left[\mathbf{Z}'(t)\right]^T \in \mathbb{R}^{D\times D}.$$

Note that $\mathbf{P}_\mathrm{f}(t)$ is rank-deficient for $k < D$, 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.

At the end of each analysis cycle an ensemble $\mathbf{Z}_\mathrm{a}$ is generated which must be consistent with the analysis error covariance $\mathbf{P}_\mathrm{a}$, and satisfies

$$\mathbf{P}_\mathrm{a} = \frac{1}{k-1}\mathbf{Z}'_\mathrm{a}\left[\mathbf{Z}'_\mathrm{a}\right]^T.$$

In previous work Gottwald et al. (2011) and Mitchell and Gottwald (2012) used the ensemble transform Kalman filter (ETKF) (Bishop et al., 2001; Tippett et al., 2003; Wang et al., 2004), which seeks a transformation $\mathbf{T} \in \mathbb{R}^{k\times k}$ such that the analysis deviation ensemble $\mathbf{Z}'_\mathrm{a}$ is given as a deterministic perturbation of the forecast ensemble $\mathbf{Z}_\mathrm{f}$ via $\mathbf{Z}'_\mathrm{a} = \mathbf{Z}'_\mathrm{f}\mathbf{T}$. In order to incorporate localisation needed for small ensemble sizes easily, we will implement for our VLKF here an approximate square root filter (DEnKF) proposed by Sakov and Oke (2008) where the analysis deviations are determined according to

$$\mathbf{Z}'_\mathrm{a} = \left(\mathbf{I} - \frac{1}{2}\mathbf{K}\mathbf{H}\right)\mathbf{Z}'_\mathrm{f}. \quad (15)$$

A new forecast is obtained by propagating $\mathbf{Z}_\mathrm{a}$ with the nonlinear forecast model to the next observation time, where a new analysis cycle will be started.

We will use here diagonal target matrices $\mathbf{A}_\mathrm{clim}$ where the diagonal entries are set to the mean value of the diagonal entries of the full climatic covariance. We found that otherwise the variance constraint is not "switched on" sufficiently often to drive the dynamics to the mean $\boldsymbol{a}_\mathrm{clim}$ (due to a lack of simultaneous diagonalisability of $\mathbf{A}_\mathrm{clim}$ and $\mathbf{h}\mathbf{P}\mathbf{h}^T$; cf. Appendix A). This suggests that the variance constraint is a numerical tool to regularise the filter, with the advantage however that the regularisation is performed in a dynamically consistent way, performed *within* the data assimilation procedure using only dynamical quantities such as measured imbalance.

3 The modified Lorenz-96 model

The Lorenz-96 model (Lorenz, 1996; Lorenz and Emanuel, 1998)

$$\dot{x}_j = x_{j-1}\left(x_{j+1} - x_{j-2}\right) - x_j + F \quad j = 1,\ \ldots,\ d \quad (16)$$

with periodic $x_j = x_{j+d}$ is a standard test bed for data assimilation as it is computationally manageable but still incorporates crucial ingredients of real mid-latitude atmospheric flows such as nonlinear energy conservation, advection, forcing and linear damping. Recently, Bergemann and Reich (2010a) introduced a modification of the standard Lorenz-96 model by coupling it to a purely dispersive fast wave equation mimicking the influence of fast gravity waves on slow Rossby waves in a quasi-geostrophic regime. The modified Lorenz system reads as

$$\begin{aligned}\dot{x}_j &= (1-\eta)\,x_{j-1}\left(x_{j+1} - x_{j-2}\right) - x_j + F \\ &\quad + \eta\left(x_{j-1}h_{j+1} - x_{j-2}h_{j-1}\right) \quad (17)\\ \varepsilon^2\ddot{h}_j &= -h_j + \alpha^2\left(h_{j-1} - 2h_j + h_{j+1}\right) + x_j. \quad (18)\end{aligned}$$

The fast wave part (Eq. 18) is purely dispersive; if the dissipation and the forcing in the slow $\boldsymbol{x}$ equation (Eq. 17) is ignored the system conserves the total energy

$$\begin{aligned}H = \frac{\eta}{2}\sum_{j=1}^{d}\Big(&\frac{\eta-1}{\eta}x_j^2 + \epsilon^2\dot{h}_j^2 + h_j^2 \\ &+\alpha^2\left(h_{j+1} - h_{j-1}\right)^2 - 2x_j h_j\Big) \quad (19)\end{aligned}$$

with $0 \le \eta \le 1$. The modified Lorenz-96 system (Eqs. 17–18) contains an approximate slow manifold given by

$$x_j = h_j - \alpha^2\left(h_{j-1} - 2h_j + h_{j+1}\right), \quad (20)$$

which is obtained by formally setting $\varepsilon = 0$ in Eq. (18). Higher order balance relations could be derived by employing asymptotic theory.

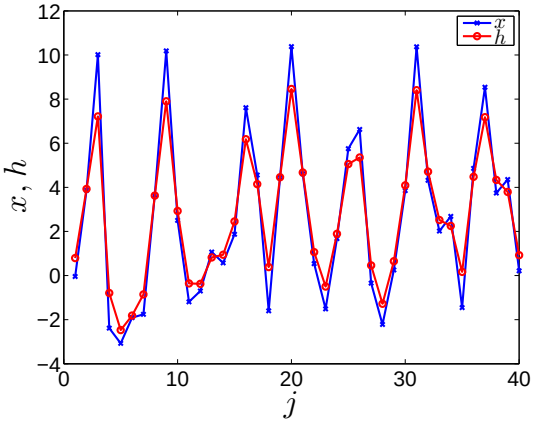

Fig. 1. Typical balanced fields $\{x_j\}$ (blue) and $\{h_j\}$ (red) for the modified slow-fast Lorenz-96 model (Eqs. 17–18).

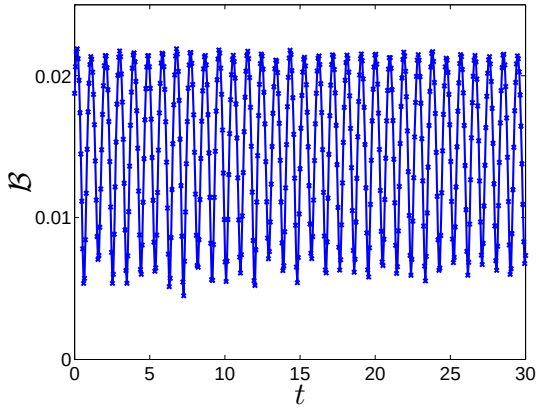

Fig. 2. Temporal evolution of the imbalance $\mathcal{B}$ for initially balanced fields of the modified slow-fast Lorenz-96 model (Eqs. 17–18).

We set the number of degrees of freedom to $d = 40$ and $F = 8$ for the forcing. We consider here weak coupling with $\eta = 0.1$, implying sufficiently nonlinear behaviour of the slow x variables. The "Rossby number" is set to $\varepsilon = 0.0025$ and the "Burgers number" is set to $\alpha^2 = 0.25$. In Fig. 1 we show typical initially balanced fields. Note that the balance relation (Eq. 20) implies that the balanced field $\{h_j\} = (h_1, h_2, \ldots, h_d)$ is smoother than $\{x_j\} = (x_1, x_1, \ldots, x_d)$ (h is obtained from x via the application of an inverse Helmholtz operator). Note that this is different to the situation in realistic atmospheric models where the fast variables are small scale and rapidly oscillate around the slow manifold.

We introduce the imbalance operator $\mathbf{B}$ which acts on z with $z_j = (x_j, h_j, \dot{h}_j)$ as

$$(\mathbf{B}z)_j = x_j - h_j + \alpha^2 \left(h_{j-1} - 2h_j + h_{j+1}\right), \qquad (21)$$

which according to Eq. (20) is zero to leading order if initially so. Figure 2 shows the temporal evolution of the site-averaged imbalance

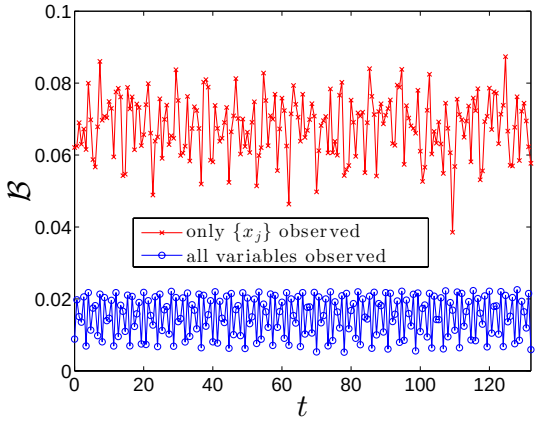

Fig. 3. Imbalance $\mathcal{B}$ of the analysis as a function of analysis cycles for ETKF with 1000 ensemble members and an observation interval of $\Delta t_{\mathrm{obs}} = 2$ h and observational noise error variance $R_0 = 0.84$, without covariance inflation and localisation. Results are shown for the case when all variables $\{x_j\}$, $\{h_j\}$ and $\{\dot{h}_j\}$ of the modified slow-fast Lorenz-96 model (Eqs. 17–18) are observed (blue) and for the case of a spatially sparse observations when only $\{x_j\}$ are observed (red).

$$\mathcal{B}(t) = \sqrt{\frac{1}{d}\sum_{j=1}^{d} (\mathbf{B}z)_j^2} \qquad (22)$$

for initially balanced fields with a small value of $\epsilon = 0.0025$. The figure clearly illustrates that balance is approximately preserved by the dynamics, provided the timescale separation is sufficiently large, i.e. ϵ sufficiently small. This justifies the terminology of Eq. (20) defining a *slow manifold*, as the initially generated imbalance does not interact with the slow variables on long timescales. The situation is very different when the dynamics is interrupted by data assimilation cycles where the data assimilation procedure introduces imbalance. In Cohn et al. (1998), Lorenc (2003a), Mitchell et al. (2002), Houtekamer and Mitchell (2005), Oke et al. (2007), Kepert (2009) and Greybush et al. (2011) the imbalance was associated with the procedure of covariance localisation. In Fig. 3 we show that ensemble filters can generate unbalanced analyses in sparse observational grids due to the intermittent discontinuous analyses updates, even without localisation. We present results for an ETKF with a large ensemble of 1000 members where only the slow $\{x_j\}$ variables are observed, and compare it to the case when all variables $\{x_j\}$, $\{h_j\}$ and $\{\dot{h}_j\}$ are observed. Whereas in the fully observed case the imbalance $\mathcal{B}$ exhibits the actual physical imbalance (cf. Fig. 2), increased imbalance is clearly seen in the sparser observational grid. We remark that this is not a finite size effect and cannot be mitigated by larger ensembles (we tested ensemble sizes of 3000), consistent with results for 3D-VAR by Bloom et al. (1996). We remark that for smaller observational noise with $R_0 = 0.21$ the imbalance exhibits the same mean values as in Fig. 3 with $R_0 = 0.84$. In the next section

we explore how this spurious imbalance can be controlled by using the VLKF framework established in Sect. 2.

4 Numerical results

We now present results from numerical data assimilation cycles of Eqs. (17)–(18). We consider a sparse observational grid in which only every second slow $\{x_{2j}\}$ variable is observed; the variables $\{h_j\}$ and $\{\dot{h}_j\}$ are not observed. We use $D = 3d = 3 \times 40$, and therefore in the notation of Sect. 2 we have $D_o = 20$. We observe the system in equidistant observation intervals Δt_{obs} ranging from 1 to 6.5 h, adopting the timescales suggested by the standard Lorenz-96 system (Eq. 16), i.e. $t = 1/120$ roughly corresponds to 1 hour (see for example Lorenz and Emanuel, 1998). Observations are contaminated by Gaussian noise, with error variance $\mathbf{R}_o = (0.25\sigma_{x,\mathrm{clim}})^2 \mathbf{I}_{20} = 0.84\,\mathbf{I}_{20}$ where $\sigma^2_{x,\mathrm{clim}} = 13.50$ is the climatic variance of $\{x_j\}$. We perform 4000 analysis cycles after a spin-up period of 1000 analysis cycles. All simulations are initialised with balanced data using Eq. (20). To generate the observations and to propagate forward the forecast model (Eqs. 17–18) we employ an implicit midpoint rule with a time step of $dt = 0.0025$ (see, for example, Leimkuhler and Reich, 2005).

Besides the variance limiting Kalman filter VLKF-$\mathcal{B}$ where we impose a climatic constraint on the imbalance $\mathbf{B}z$, we also employ a variance limiting Kalman filter VLKF-$\dot{h}$ where we impose a climatic constraint on the unobserved fast variables $\{\dot{h}_j\}$. Both VLKF-$\mathcal{B}$ and VLKF-$\dot{h}$ have, in the notation of Sect. 2, $D_w = 40$. With $z = (\boldsymbol{x}, \boldsymbol{h}, \dot{\boldsymbol{h}}) \in \mathbb{R}^{120}$, for VLKF-$\mathcal{B}$ the pseudo-observation operator $\mathbf{h} : \mathbb{R}^{120} \to \mathbb{R}^{40}$ is

$$\mathbf{h} = (\mathbf{B}\,\mathbf{0}_{40}),$$

with $\mathbf{B} \in \mathbb{R}^{40\times 80}$ defined in Eq. (21), and for VLKF-$\dot{h}$ the pseudo-observation operator is

$$\mathbf{h} = (\mathbf{0}_{40}\,\mathbf{0}_{40}\,\mathbf{I}_{40}).$$

The climatic mean and variances of $\{\dot{h}_j\}$ and those of the imbalance $\{(\mathbf{B}z)_j\}$ were estimated through long-time simulations of the full modified Lorenz-96 system (Eqs. 17–18) with balanced initial data as $\mathbf{B}z = \mathbf{0}$ and $\sigma^2_{Bz,\mathrm{clim}} = 8.4 \times 10^{-4}$, and $\dot{h} = -0.01$ and $\sigma_{\dot{h},\mathrm{clim}} = 224.35$, respectively. We set $\boldsymbol{a}_{\mathrm{clim}} = \mathbf{0}$ and $\mathbf{A}_{\mathrm{clim}} = \sigma^2_{Bz,\mathrm{clim}}\mathbf{I}_{40}$ for VLKF-$\mathcal{B}$, and $\boldsymbol{a}_{\mathrm{clim}} = -\mathbf{0.01}$ and $\mathbf{A}_{\mathrm{clim}} = \sigma^2_{\dot{h},\mathrm{clim}}\mathbf{I}_{40}$ for VLKF-$\dot{h}$, respectively. We note that both the climatic covariance of $\mathbf{B}z$ and of $\dot{\boldsymbol{h}}$ are concentrated near the diagonal.

For comparison with our implementations of VLKF-$\mathcal{B}$ and VLKF-$\dot{h}$ we will employ a suite of ensemble filters. In particular, we will use the EnKF with perturbed observations as in Burgers et al. (1998) and the approximate square root filter DEnKF as in Sakov and Oke (2008). Furthermore, we implement an incremental analysis update (IAU) as in Bloom et al. (1996) and Polavarapu et al. (2004), where the analysis increments are calculated by a DEnKF.

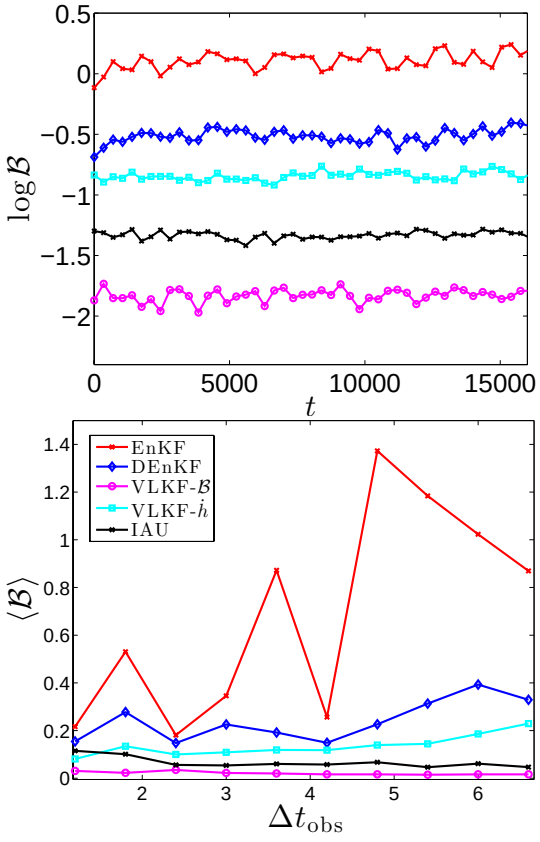

Fig. 4. Top panel: imbalance $\mathcal{B}$ of the analysis as a function of analysis cycles for $\Delta t_{obs} = 5.5$ h in a log plot. In order of increasing values of $\mathcal{B}$ are VLKF-$\mathcal{B}$ (magenta open circles), IAU (black crosses) VLKF-$\dot{h}$ (cyan squares), DEnKF (blue diamonds) and EnKF (red crosses). Bottom panel: temporally averaged imbalance $\langle\mathcal{B}\rangle$ of the analysis as a function of the observation interval Δt_{obs}.

All filter implementations use 10 ensemble members, which is smaller than the attractor dimension of the system (Eqs. 17–18). We employ covariance inflation whereby the prior forecast error covariance is increased by an inflation factor δ (Anderson and Anderson, 1999). Since the $\{h_j\}$ variables are not damped in the modified Lorenz system (Eq. 18), inflation of the unobserved $\{h_j\}$ variables would in general lead to an increasing growth in the associated forecast covariance. Inflation is therefore applied at each time step only to the $\{x_j\}$ components of the ensembles for DEnKF, EnKF and IAU, but to all components of the ensemble for VLKF-$\mathcal{B}$ and VLKF-$\dot{h}$ which explicitly constrain. This was found to be advantageous for all respective filters. Results are obtained for a wide range of inflation factors and only the optimal result for each particular formulation of the filter is reported here. We have optimised over 1000 equally spaced values of $\delta \in (1, 1.16)$. The small ensemble size chosen here requires localisation. We employ the localisation

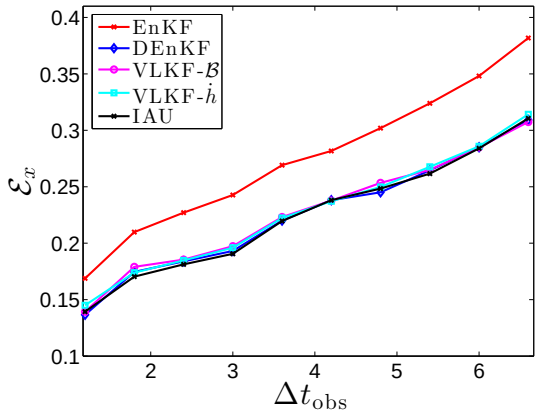

Fig. 5. Rms error of the analysis as a function of the observation interval $\Delta t_{\rm obs}$ (in hours) for the slow variables $\{x_j\}$.

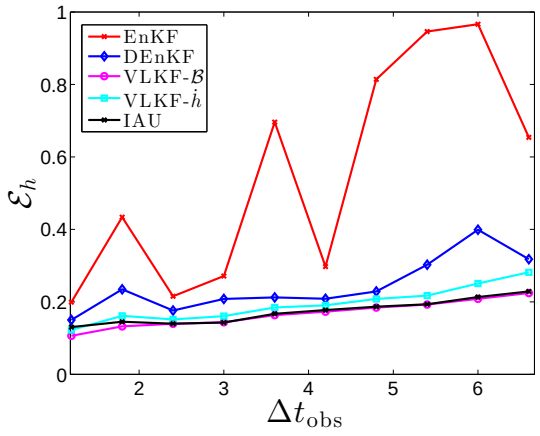

Fig. 6. Rms error of the analysis as a function of the observation interval $\Delta t_{\rm obs}$ (in hours) for the unobserved height variables $\{h_j\}$.

method along the line of Houtekamer and Mitchell (1998, 2001) and Hamill et al. (2001) whereby the forecast error covariance $\mathbf{P}_{\rm f}$ is Schur-multiplied with a localisation matrix $\mathbf{C}_{\rm loc}$. We use the compactly supported localisation function introduced by Gaspari and Cohn (1999) where correlations with distances larger than $2\rho_{\rm loc}$ are set to 0. We set the localisation radius to $\rho_{\rm loc} = 8$ for all filters.

In Fig. 4 we present results of the amount of imbalance as measured by the imbalance $\mathcal{B}(t)$ and by the temporally averaged imbalance

$$\langle \mathcal{B} \rangle = \frac{1}{N} \sum_{n=1}^{N} \mathcal{B}\,(n\,\Delta t_{\rm obs})\,,$$

accrued during the data assimilation procedure for our suite of filters. EnKF and DEnKF generate a significant amount of unphysical imbalance, with values much larger than those of the actual balanced toy model with $\langle \mathcal{B} \rangle \approx 0.018$ (cf. Fig. 2). The increased imbalance in EnKF may be due to sampling errors introduced through the perturbed observations. Given the particular nature of imbalance present in the toy model (Eqs. 17–18), this may not be an issue in realistic atmospheric models. The IAU implementation strongly reduces imbalance, albeit to levels significantly larger than those expected from the actual dynamics. Our VLKF-$\mathcal{B}$ filter is able to constrain imbalance very close to the the actual physical imbalance. Note that although the pseudo-observations used in VLKF-$\mathcal{B}$ were for the imbalance of each variable $\mathbf{B}\,z$ driving the dynamics towards $\mathbf{B}\,z = \mathbf{0}$, the analysis reproduces dynamically realistic values of the integrated measure of imbalance $\langle \mathcal{B} \rangle$. VLKF-$\dot{h}$ also achieves a pronounced reduction in imbalance, but to larger values than the IAU implementation, in particular for larger observation intervals $\Delta t_{\rm obs}$. Surprisingly, for filters which only constrain the climatic variance of the height variable $\{h_j\}$, and do not impose any explicit constraints on the imbalance, i.e. $\mathbf{h} = (\mathbf{0}_{40}\,\mathbf{I}_{40}\,\mathbf{0}_{40})$, one also observes a significant reduction in imbalance (not shown).

We now investigate rms error statistics. We consider the site-averaged rms error of variables z

$$\mathcal{E} = \sqrt{\langle \frac{1}{N\,D_G} \sum_{n=1}^{N} |\overline{z}_{\rm a}\,(n\,\Delta t_{\rm obs}) - z_t\,(n\,\Delta t_{\rm obs})|_G^2 \rangle} \qquad (23)$$

between the truth z_t and the ensemble mean $\overline{z}_{\rm a}$, where N is the number of analysis cycles and D_G denotes the number of variables involved. We introduce the norm $\|\mathbf{a}\|_G^2 = \mathbf{a}^T\,\mathbf{G}\,\mathbf{a}$ to investigate the error over all $\{x_j\}$ variables $\mathcal{E}_x$ using $\mathbf{G} = \delta_{ij}$ for $1 \le i \le 40$ and the error of the fast $\{h_j\}$ variables $\mathcal{E}_h$ using $\mathbf{G} = \delta_{ij}$ for $41 \le i \le 80$. Figure 5 shows $\mathcal{E}_x$ for our suite of filters. DEnKF, IAU and our VLKF-$\mathcal{B}$ and VLKF-$\dot{h}$ exhibit very similar rms errors, with values much smaller than the observational noise with $r_{\rm o} = \sqrt{0.84} = 0.91$. EnKF produces consistently worse rms errors, which again may be due to sampling errors stemming from the randomly perturbed observations.

Figure 6 shows that the correct balance statistics of VLKF-$\mathcal{B}$ manifests itself in superior rms errors for the unobserved fast height field $\{h_j\}$ when compared to filters which do not incorporate a balance constraint. IAU and VLKF-$\mathcal{B}$ exhibit comparable rms error statistics for the height field. Furthermore, it is seen that constraining the covariance of $\{\dot{h}_j\}$, as done in VLKF-$\dot{h}$, also generates comparably good rms errors for the height field. The variations of the error in the height field $\mathcal{E}_{\rm h}$ with the observation interval $\Delta t_{\rm obs}$ mirror exactly the imbalance shown in Fig. 4.

We found that EnKF, DEnKF and IAU exhibit instances of catastrophic filter divergence whereby the forecast model develops numerical instabilities (Gottwald and Majda, 2013). Instances of this type of filter divergence were not observed in the variance constraining filters VLKF-$\mathcal{B}$ and VLKF-$\dot{h}$.

5 Conclusions

We have presented here an implementation of an ensemble filter which explicitly limits the amount of imbalance *within* the data assimilation procedure. We were able to produce balanced analyses by incorporating statistical information such as mean and variance of imbalance, available through long-time integration or historical data, as pseudo-observations. This procedure not only successfully constraints imbalance to its climatic values, but also produces very good filter performance in terms of rms errors of the fast unobserved variables.

We presented a comparison between a filter which explicitly constraints the amount of imbalance, coined VLKF-$\mathcal{B}$, a filter which constraints the statistics of the fast variables but not the imbalance, coined VLKF-$\dot{h}$, and standard implementations of EnKF, DEnKF and IAU. It was found that our balance controlling filter VLKF-$\mathcal{B}$ is able to constrain the amount of imbalance to lie within the physically observed limits. Besides improved balance of the analyses this also implied very good error statistics for the unobserved height field. We tested our method against the widely used IAU implementation and found that it generates less unphysical imbalance and has very similar rms error statistics for the observed and the unobserved variables.

The variance constraint we employ requires the determination of the overestimating subspace – the eigenspace in which $\mathbf{P}_\mathrm{f}$ experiences covariances above the climatological – which was achieved in this work by singular vector decomposition. The extra computational cost implied has to be weighed against the cost of an additional application of the forecast model involved in IAU or of fast Fourier transforms when using digital filters, as well as whether the superior performance of VLKF-$\mathcal{B}$ in generating less unphysical imbalance is worth it.

Appendix A

Construction of the transformation matrix $\mathbf{S}_\mathrm{w}$

We present here an algorithm of how to construct the transformation matrix $\mathbf{S}_\mathrm{w}$. This matrix projects into the subspace of those pseudo-observable subspaces which in a standard Kalman filter would produce an analysis whose analysis error covariance matrix exceeds the prescribed (climatological) error covariance $\mathbf{A}_\mathrm{clim}$. We need to find $\mathbf{S}_\mathrm{w}$ such that Eq. (14) produces a positive error covariance matrix $\mathbf{R}_\mathrm{w}$. For convenience we recall Eq. (14):

$$\left(\mathbf{S}_\mathrm{w}^T \mathbf{R}_\mathrm{w} \mathbf{S}_\mathrm{w}\right)^{-1} = \left(\mathbf{S}_\mathrm{w}^T \mathbf{A}_\mathrm{clim} \mathbf{S}_\mathrm{w}\right)^{-1} - \left(\mathbf{S}_\mathrm{w}^T \mathbf{h}\mathbf{P}\mathbf{h}^T \mathbf{S}_\mathrm{w}\right)^{-1}. \quad \text{(A1)}$$

Introducing $\hat{\mathbf{P}}_\mathrm{h} = \mathbf{S}_\mathrm{w}^T \mathbf{h}\mathbf{P}\mathbf{h}^T \mathbf{S}_\mathrm{w}$, $\hat{\mathbf{A}}_\mathrm{clim} = \mathbf{S}_\mathrm{w}^T \mathbf{A}_\mathrm{clim} \mathbf{S}_\mathrm{w}$ and $\hat{\mathbf{R}}_\mathrm{w} = \mathbf{S}_\mathrm{w}^T \mathbf{R}_\mathrm{w} \mathbf{S}_\mathrm{w}$, we may rewrite this as

$$\hat{\mathbf{R}}_\mathrm{w}^{-1} = \hat{\mathbf{A}}_\mathrm{clim}^{-1} - \hat{\mathbf{P}}_\mathrm{h}^{-1}.$$

Matrices $\mathbf{S}_\mathrm{w}$ satisfying Eq. (A1) can be determined provided $\mathbf{A}_\mathrm{clim}$ and $\mathbf{h}\mathbf{P}\mathbf{h}^T$ are simultaneously diagonalisable. We remark that Eq. (A1) can be readily converted into

$$\hat{\mathbf{R}}_\mathrm{w} = \hat{\mathbf{P}}_\mathrm{h} \left[\hat{\mathbf{P}}_\mathrm{h} - \hat{\mathbf{A}}_\mathrm{clim}\right]^{-1} \hat{\mathbf{A}}_\mathrm{clim}.$$

Simultaneous diagonalisation of $\mathbf{A}_\mathrm{clim}$ and $\mathbf{h}\mathbf{P}\mathbf{h}^T$ can be checked either by looking at the null space of the commutator $\mathbf{C} = \mathbf{A}_\mathrm{clim}\mathbf{h}\mathbf{P}\mathbf{h}^T - \mathbf{h}\mathbf{P}\mathbf{h}^T\mathbf{A}_\mathrm{clim}$, or by comparing the eigenspaces of the respective matrices directly. We note that if we prescribe a diagonal target covariance $\mathbf{A}_\mathrm{clim} = \lambda\mathbf{I}$, simultaneous diagonalisation is automatically asserted. Let the transformation matrix onto the subspace in which $\mathbf{A}_\mathrm{clim}$ and $\mathbf{h}\mathbf{P}\mathbf{h}^T$ are simultaneously diagonalisable be $\mathbf{M}_0^T \in \mathbb{R}^{D_\mathrm{w} \times D_{\mathrm{w}0}}$ with $D_{\mathrm{w}0} \le D_\mathrm{w}$.

Consider $\mathbf{A}_{\mathrm{clim}\,0} = \mathbf{M}_0^T \mathbf{A}_\mathrm{clim} \mathbf{M}_0$ and $\mathbf{P}_0 = \mathbf{M}_0^T \mathbf{h}\mathbf{P}\mathbf{h}^T \mathbf{M}_0$, and simultaneously diagonalise by writing (wlog),

$$\mathbf{A}_{\mathrm{clim}\,0} = \mathbf{S}_\mathrm{A} \mathbf{A}_{\mathrm{clim}\,0} \mathbf{S}_\mathrm{A}^T = \tilde{\mathbf{S}}_\mathrm{A} \tilde{\mathbf{S}}_\mathrm{A}^T,$$

with $\tilde{\mathbf{S}}_\mathrm{A} = \mathbf{S}_\mathrm{A} \mathbf{A}_{\mathrm{clim}\,0}^{\frac{1}{2}}$. Introducing

$$\tilde{\mathbf{P}}_0 = \tilde{\mathbf{S}}_\mathrm{A}^{-1} \mathbf{P}_0 \tilde{\mathbf{S}}_\mathrm{A}^{-T},$$

we transform with an orthogonal transformation $\mathbf{S}_\mathbf{P}$

$$\tilde{\mathbf{P}}_0 = \mathbf{S}_\mathbf{P} \hat{\tilde{\mathbf{P}}}_0 \mathbf{S}_\mathbf{P}^T.$$

Introducing $\mathbf{Q} = \tilde{\mathbf{S}}_\mathrm{A} \mathbf{S}_\mathbf{P}$ we may write

$$\mathbf{A}_{\mathrm{clim}\,0} = \mathbf{Q}\mathbf{Q}^T,$$

$$\mathbf{P}_0 = \mathbf{Q} \hat{\tilde{\mathbf{P}}}_0 \mathbf{Q}^T.$$

Using the diagonal matrices $\mathbf{I}$ and $\hat{\tilde{\mathbf{P}}}_0$, associated with $\mathbf{A}_\mathrm{clim}$ and $\mathbf{h}\mathbf{P}\mathbf{h}^T$, respectively, one can now readily check for overestimation of $\mathbf{R}_\mathrm{w}$. Transforming into the subspace in which $\mathbf{A}_\mathrm{clim}$ and $\mathbf{h}\mathbf{P}\mathbf{h}^T$ are both diagonal with the same eigenspaces, we obtain

$$\hat{\tilde{\mathbf{R}}}_\mathrm{w}^{-1} = \mathbf{Q}^T \mathbf{M}_0^T \mathbf{R}_\mathrm{w}^{-1} \mathbf{M}_0 \mathbf{Q} = \mathbf{I} - \hat{\tilde{\mathbf{P}}}_0^{-1},$$

and $\hat{\tilde{\mathbf{R}}}_\mathrm{w}$ can be calculated directly by inverting the diagonal $\hat{\tilde{\mathbf{R}}}_\mathrm{w}^{-1}$.

We have now determined a transformation $\mathbf{Q}^T \mathbf{M}_0^T$ which simultaneously diagonalises $\mathbf{A}_\mathrm{clim}$ and $\mathbf{h}\mathbf{P}\mathbf{h}^T$. Transforming renders $\hat{\tilde{\mathbf{R}}}_\mathrm{w} \in \mathbb{R}^{D_{\mathrm{w}0} \times D_{\mathrm{w}0}}$ diagonal, but not necessarily positive definite as required. However, $\hat{\tilde{\mathbf{R}}}_\mathrm{w}$ being diagonal allows us to readily determine a transformation matrix $\mathbf{S}_\mathrm{red} \in \mathbb{R}^{\hat{D}_\mathrm{w} \times D_{\mathrm{w}0}}$ which projects onto the $\hat{D}_\mathrm{w}$-dimensional overestimating subspace in which $\hat{\tilde{\mathbf{R}}}_\mathrm{w}$ is positive definite, with $\hat{D}_\mathrm{w} \le D_{\mathrm{w}0}$.

This concludes our algorithm for how to compute a positive definite invertible covariance matrix $\hat{\mathbf{R}}_\mathrm{w}$ and

also provides an expression for the transformation matrix $\mathbf{S}_{\mathrm{w}} \in \mathbb{R}^{\hat{D}_{\mathrm{w}} \times D_{\mathrm{w}}}$ as

$$\mathbf{S}_{\mathrm{w}} = \mathbf{S}_{\mathrm{red}} \mathbf{Q}^T \mathbf{M}_0^T .$$

We note that one may define formally an effective pseudo-observation operator $\hat{\mathbf{h}} \in \mathbb{R}^{\hat{D}_{\mathrm{w}} \times D_{\mathrm{w}}}$

$$\hat{\mathbf{h}} = \mathbf{S}_{\mathrm{red}} \mathbf{Q}^T \mathbf{M}_0^T \mathbf{h}.$$

Acknowledgements. I am grateful to Reiner Pope for speeding up my code and parallelising it. I am grateful to Sebastian Reich for fruitful discussions and to two anonymous referees for helpful comments. I acknowledge support from the Australian Research Council.

Edited by: G. Desroziers
Reviewed by: two anonymous referees

References

Anderson, J. L. and Anderson, S. L.: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts, Mon. Weather Rev., 127, 2741–2758, 1999.

Baer, F. and Tribbia, J.: On complete filtering of gravity modes through non-linear initialization, Mon. Weather Rev., 105, 1536–1539, 1977.

Bergemann, K. and Reich, S.: A mollified ensemble Kalman filter, Q. J. Roy. Meteorol. Soc., 136, 1636–1643, 2010a.

Bergemann, K. and Reich, S.: A localisation technique for ensemble Kalman filters, Q. J. Roy. Meteorol. Soc., 136, 701–707, 2010b.

Bergemann, K., Gottwald, G. A., and Reich, S.: Ensemble propagation and continuous matrix factorization algorithms, Q. J. Roy. Meteorol. Soc., 135, 1560–1572, 2009.

Bishop, C. H., Etherton, B. J., and Majumdar, S. J.: Adaptive sampling with the Ensemble Transform Kalman Filter, Part I: Theoretical aspects, Mon. Weather Rev., 129, 420–436, 2001.

Bloom, S. C., Takacs, L. L., Da Silva, A. M., and Ledvina, D.: Data assimilation using incremental analysis updates, Mon. Weather Rev., 124, 1256–1271, 1996.

Burgers, G., van Leeuwen, P. J., and Evensen, G.: Analysis scheme in the ensemble Kalman filter, Mon. Weather Rev., 126, 1719–1724, 1998.

Cohn, S. E., Da Silva, A., Guo, J., Sienkiewicz, M., and Lamich, D.: Assessing the effects of data selection with the DAO Physical-Space Statistical Analysis System, Mon. Weather Rev., 126, 2913–2926, 1998.

Cotter, C.: Data assimilation on the exponentially accurate slow manifold, Philos. T. Roy. Soc. A, 371, 20120300, doi:10.1098/rsta.2012.0300, 2013.

Daley, R.: Atmospheric Data Analysis, Cambridge University Press, Cambridge, 1993.

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

Gaspari, G. and Cohn, S. E.: Construction of correlation functions in two and three dimensions, Q. J. Roy. Meteorol. Soc, 125, 723–757, 1999.

Gauthier, P. and Thépaut, J. N.: Impact of the digital filter as a weak constraint in the pre operational 4DVAR assimilation system of Météo-France, Mon. Weather Rev., 117, 1225–1254, 2001.

Gottwald, G. A. and Majda, A. J.: A mechanism for catastrophic filter divergence in data assimilation for sparse observation networks, Nonlin. Processes Geophys., 20, 705–712, doi:10.5194/npg-20-705-2013, 2013.

Gottwald, G. A., Mitchell, L., and Reich, S.: Controlling overestimation of error covariance in ensemble Kalman filters with sparse observations: A variance limiting Kalman filter, Mon. Weather Rev., 139, 2650–2667, 2011.

Greybush, S. J., Kalnay, E., Miyoshi, T., Ide, K., and Hunt, B. R.: Balance and ensemble Kalman filter localization techniques, Mon. Weather Rev, 139, 511–522, 2011.

Hamill, T. M.: Ensemble-based atmospheric data assimilation, in: Predictability of Weather and Climate, edited by: Palmer, T. and Hagedorn, R., Cambridge University Press, Cambridge, 124–156, 2006.

Hamill, T. M., Whitaker, J. S., and Snyder, C.: Distance-dependent filtering of background covariance estimates in an ensemble Kalman filter, Mon. Weather Rev., 129, 2776–2790, 2001.

Houtekamer, P. L. and Mitchell, H. L.: Data Assimilation Using an Ensemble Kalman Filter Technique, Mon. Weather Rev., 126, 796–811, 1998.

Houtekamer, P. L. and Mitchell, H. L.: A sequential ensemble Kalman filter for atmospheric data assimilation, Mon. Weather Rev., 129, 123–136, 2001.

Houtekamer, P. L. and Mitchell, H. L.: Ensemble Kalman filtering, Q. J. Roy. Meteorol. Soc., 131, 3269–3289, 2005.

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

Kepert, J. D.: Covariance localisation and balance in an ensemble Kalman filter, Q. J. Roy. Meteorol. Soc. 135, 1157–1176, 2009.

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

Lorenc, A. C.: The potential of the ensemble Kalman filter for NWP – a comparison with 4DVAR, Q. J. Roy. Meteorol. Soc., 129, 3183–3203, 2003a.

Lorenc, A. C.: Modelling of error covariances by 4D-Var data assimilation, Q. J. Roy. Meteorol. Soc., 129, 3167–3182, 2003b.

Lorenz, E. N.: Predictability – a problem partly solved, in: Predictability, edited by: Palmer, T., European Centre for Medium-Range Weather Forecast, Shinfield Park, Reading, UK, 1996.

Lorenz, E. N. and Emanuel, K. A.: Optimal sites for supplementary weather observations: simulation with a small model, J. Atmos. Sci., 55, 399–414, 1998.

Lynch, P.: The Emergence of Numerical Weather Prediction: Richardson's Dream, Cambridge University Press, Cambridge, 2006.

Lynch, P. and Huang, X.-Y.: Initialization of the HIRLAM model using a digital filter, Mon. Weather Rev., 120, 1019–1034, 1992.

Machenhauer, B.: On the dynamics of gravity oscillations in a shallow water model with applications to normal mode initialization, Contrib. Atmos. Phys., 50, 253–271, 1977.

Mitchell, H. L., Houtekamer, P. L., and Pellerin, G.: Ensemble Size, balance, and model-error representation in an Ensemble Kalman Filter, Mon. Weather Rev., 130, 2791–2808, 2002.

Mitchell, L. and Gottwald, G. A.: Controlling model error of underdamped forecast models in sparse observational networks using a variance-limiting Kalman filter, Q. J. Roy. Meteorol. Soc., 139, 212–225, 2012.

Neef, L. J., Polavarapu, S. M., and Shepherd, T. G.: Four-Dimensional Data Assimilation and Balanced Dynamics, J. Atmos. Sci., 63, 1840–1858, 2006.

Oke, P., Sakov, P., and Corney, S.: Impact of localization in the EnKF and EnOI: Experiments with a small model, Ocean Dynam., 57, 32–45, 2007.

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

Ourmières, Y., Brankart, J. M., Berline, L., Brasseur, P., and Verron, J.: Incremental analysis update implementation into a sequential ocean data assimilation system, J. Atmos. Ocean. Tech., 23, 1729–1744, 2006.

Polavarapu, S., Tanguay, M., and Fillion, L.: Four-dimensional variational data assimilation with digital filter initialisation, Mon. Weather Rev., 128, 2491–2510, 2000.

Polavarapu, S., Ren, S., Clayton, A. M., Sankey, D., and Rochon, Y.: On the relationship between incremental analysis updating and incremental digital filtering, Mon. Weather Rev., 132, 2495–2502, 2004.

Sakov, P. and Oke, P. R.: A deterministic formulation of the ensemble Kalman filter: an alternative to ensemble square root filters, Tellus A, 60, 361–371, 2008.

Simon, D. J.: Optimal State Estimation, John Wiley & Sons, Inc., New York, 2006.

Szunyogh, I., Kostelich, E., Gyarmati, G., Patil, D. J., Hunt, B., Kalnay, E., Ott, E., and Yorke, J.: Assessing a local ensemble Kalman filter: Perfect model experiments with the National Centers for Environmental Prediction global model, Tellus A, 57, 528–545, 2005.

Thépaut, J. N. and Courtier, P.: Four-dimensional variational data assimilation using the adjoint of a multilevel primitive equation model, Q. J. Roy. Meteorol. Soc., 129, 2089–2102, 1991.

Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M., and Whitaker, J. S.: Ensemble square root filters, Mon. Weather Rev., 131, 1485–1490, 2003.

Wang, X., Bishop, C. H., and Julier, S. J.: Which is better, an ensemble of positive-negative pairs or a centered spherical simplex ensemble?, Mon. Weather Rev., 132, 1590–1605, 2004.

Watkinson, L. R., Lawless, A. S., Nichols, N. K., and Roulstone, I.: Weak constraints in four-dimensional variational data assimilation, Meteorol. Z., 16, 767–776, 2007.

Weaver, A. T., Vialard, J., and Anderson, D. L. T.: Three- and four-dimensional variational assimilation with a general circulation model of the tropical Pacific Ocean, Part I: Formulation, internal diagnostics, and consistency checks, Mon. Weather Rev., 131, 1360–1378, 2003.

Zhu, Y., Todling, R., Guo, J., Cohn, S. E., Navon, I. M., and Yang, Y.: The GEOS-3 retrospective data assimilation system: The 6-hour lag case, Mon. Weather Rev., 131, 2129–2150, 2003.