This manuscript was compiled using the Copernicus template for preprint submission to Nonlinear Processes in Geophysics.

Long-window tandem variational data assimilation methods for chaotic climate models tested with the Lorenz 63 system

Philip David Kennedy[1], Abhirup Banerjee[1], Armin Köhl[1], and Detlef Stammer[1]

[1]Universität Hamburg, Fakultät für Mathematik, Informatik und Naturwissenschaften, Fernerkundung & Assimilation, Bundesstr. 53, 20146 Hamburg, Deutschland

Correspondence: Philip David Kennedy (philipdkennedy@physics.org)

Abstract. ~~A new~~ 4D-variational data assimilation ~~method for chaotic climate models is introduced using~~ is applied to the Lorenz '63 model ~~. This~~ to introduce a new method for parameter estimation in chaotic climate models. The approach aims to optimise an Earth system model (ESM), for which no adjoint exists, by utilising ~~an adjoint model~~ the adjoint of a different, potentially simpler ESM. ~~The technique relies on~~ This relies on the synchronisation of the model to observed ~~time series dataemploying the dynamical~~ data. Dynamical state and parameter estimation (DSPE) ~~method~~ is used to stabilise the tangent linear system by reducing all positive Lyapunov exponents to negative values~~. Therefore, long windows can be used to improve parameter estimation . In this new extension~~, thereby improving parameter estimation by enabling long assimilation windows. The method introduces a second layer of synchronisation ~~is added~~ between the two models, with and without an adjoint, to facilitate linearisation around the trajectory of the model ~~without an adjoint . The method is conceptually demonstrated~~ for which no adjoint exists. This is achieved by synchronising two Lorenz '63 systems, ~~representing two ESMs,~~ one with and the other without an adjoint model. Results are presented for an idealised case of identical, perfect models and for a more realistic case in which they differ from one another. If employed ~~with a coarser ESM with an adjoint~~ on a high-resolution ESM for which a coarse resolution adjoint exists, the method will save computational ~~power~~ resources as only one forward run with the full high-resolution ESM per iteration ~~needs to be carried out~~ is needed. It is demonstrated that there is negligible error and uncertainty change compared to the ~~'traditional' optimisation of~~ traditional optimisation of a full ESM with an adjoint. ~~In a variation of the method outlined,~~ Stemming from this approach, it is shown that the synchronisation between two identical models can be used to filter noisy data ~~. This reduces optimised parametric model uncertainty~~ in a dynamical way which reduces the parametric uncertainty of the optimised model by approximately one third. Such a precision gain could prove valuable for seasonal, annual, and decadal predictions.

1 Introduction

The time evolution of the Earth system can be ~~realistically described using numerical models that capture it's processes and time evolution.~~ simulated using numerical Earth system models (ESMs). Provided these models skilfully describe the system's time

evolution and observed processes, they can be used to forecast future states of the system ~~provided that the past and present are known~~as long as accurate initial conditions exist. Data assimilation is a powerful tool ~~that combines observations~~ to bring ESMs into agreement with the observed climatic state by combining data with the numerical model ~~to represent the~~ while preserving dynamic principles governing the system. ~~This generates an estimation of its state (Wunsch and Heimbach, 2006; Nichols, 2010) in an attempt~~ (Wunsch and Heimbach, 2006; Nichols, 2010) while also attempting to further improve the ESM's predictive skills.

There are two common ~~approaches~~ assimilation approaches typically used to incorporate observations into a model~~, a~~: *sequential*~~data assimilation scheme (Bertino et al., 2003) and~~ and *variational*~~data assimilation (Le Dimet and Talagrand, 1986).~~ data assimilation schemes (Wunsch, 1996). *Sequential*~~data assimilation~~ data assimilation (Bertino et al., 2003) involves the application of a filter, most commonly Kalman filters (Kalman, 1960; Evensen, 1994, 2003; Tippett and Chang, 2003; Houtekamer and Mitchell, 2001). This technique ~~performs assimilation sequentially at each~~ merges a predicted state with observations at each analysis time step by ~~inputting the physical model's output state variable and updating it using the observations. This is done by~~ estimating a joint probability distribution between the two ~~which also takes~~ by taking into account their respective modelling and observational uncertainties. ~~Other filtering techniques, such as~~ Variants of the Kalman filter technique include: extended Kalman filters, ensemble Kalman filters, and square-root filters~~,~~ (Bar-Shalom et al., 2004; Simon, 2006; Evens. They all share a similar basic procedure ~~with variations in methodology often being case specific ; i.e. for a scenario with sparse observational data~~while differing in case specific variations of the methodology. The strength of all filtering techniques is ~~the sequential procedure which~~ that the *sequential* procedure allows for real-time assimilation of observations. ~~In contrast~~, for example in initialised numerical weather forecasting.

In contrast, *variational*~~data assimilation is performed~~ data assimilation (Le Dimet and Talagrand, 1986) estimates a joint probability distribution over an extended period of time by minimising a scalar *~~cost function which is~~* cost function, defined as the quadratic misfit between the model trajectory and all available observations ~~and model trajectory within a given~~ within a defined time window. ~~4DEnVar (Desroziers et al., 2014).In this work, we use the~~ The most common approaches include four-dimensional *variational* assimilation (4D-var.) (Rabier and Liu, 2003), three-dimensional *variational*~~data assimilation approach. This method is commonly known as~~ *four-dimensional variational assimilation* (4D-Var~~data assimilation (3D-var.) (Rabier and Li, which~~(Gustafsson et al., 2001), weak and strong constraint 4D-var (Tremolet, 2006; Fisher et al., 2011), and ensemble *variational* filters including 4DEnVar (Desroziers et al., 2014). *Variational* data assimilation is a useful technique for solving both initial value and parameter estimation problems (Evensen et al., 2022; Goodliff et al., 2015; Ruiz et al., 2013; Zou et al., 1992). It will be exclusively used in this study.

The 4D-var. approach utilises an adjoint of the model to iteratively minimise the model-data misfit by adjusting control variables (Tett et al., 2017; Lyu et al., 2018; Köhl and Willebrand, 2002; Allaire, 2015; Navon, 2009). ~~Aside from data assimilation purposes, adjoint models have~~ The adjoint equations of a fully non-linear model are derived from the forward equations using integration by parts. In the data assimilation context this can be used to numerically calculate the gradient of the cost function which is subsequently used to find the cost functions minimum in an iterative procedure. Adjoint models have also been widely used for sensitivity analysis in meteorology and oceanography (Hall et al., 1982; Hall and Cacuci, 1983; Hall,

60   1986; Marotzke et al., 1999; Stammer et al., 2016); this includes ~~,~~ calculating sensitivity with respect to lateral boundary ~~condition~~conditions (Gustafsson et al., 1998), estimating the sensitivity of the 2m surface temperature with respect to the sea surface temperature, sea ice, and sea surface salinity (Stammer et al., 2018). In practice, the primary limitation in finding the minimum of the cost function is the large amount of computational resources required due to non-linear or chaotic elements of the system.

65   ~~Due to the non-linearities within ESMs, the relevance of adjoint methods can be~~ In the context of a full non-linear ESM, the use of adjoint models faces several challenges. Applying an adjoint model to a state-of-the-art Earth system problem is primarily limited by the very large number of state variables $\mathcal{O}(10^7 - 10^8)$ which requires significant computational resources and observational constraints. However, more fundamental is the fact that non-linear dynamics of the system limit the applicability of adjoint methods to the Earth system predictability time scale. This can lead to ~~spikes in the~~ exponentially growing adjoint

70 sensitivities as a result of multiple local minima in the cost-function. Under such circumstances spikes occur in the estimated gradients and the cost function becomes very rough by showing an increasing number of local minima (Köhl and Willebrand, 2002; Lea et al., 2000). Fortunately, the problem can be mitigated ~~by~~ through synchronisation which removes the non-linear or chaotic dynamics ~~from the adjoint model~~ leading to a smooth cost function~~(Abarbanel et al., 2010; Sugiura et al., 2014)~~ (Abarbanel et al., 2010). This method allows for the extension of the assimilation window beyond the predictability time-

75 scale, provided that sufficient observations are available. ~~Another challenge of applying~~ However, this solution comes at the expense of a violation of the original model equations.

  The creation of an adjoint model ~~to any state-of-the-art system is that these ESMs have a very large number of state variables $\mathcal{O}(10^7 - 10^8)$, and the computation of its adjoint would be very time consuming and require significant volumes of memory (Stammer et al., 2018)~~code from the forward code usually requires considerable effort. Automatic differentiation

80 tools, such as Giering and Kaminski (1998); Hascoet and Pascual (2013) were developed to aid in this step. But substantial changes to the forward model code are required unless it was already developed with the adjoint modelling in mind. Stammer et al. (2018) created the first adjoint of an intermediate complexity fully coupled earth system model that is automatically created from the forward model by automatic differentiation using the TAF compiler, called the Centrum für Erdsystemforschung und Nachhaltigkeit (CEN) Earth System Assimilation Model (CESAM). The adjoint of this intermediate-complexity model is

85 intended to be utilised for tuning more complex CMIP-type models through parameter estimation since the basic underlying physics is very similar. Otherwise this is a manual process with considerable ambiguity in the choice of parameters (Mauritsen et al., 2012).

  ~~To mitigate both problems~~Therefore, we propose a novel framework ~~where~~ in which we use two climate models both coupled through synchronisation, one with a high ~~resolution~~ complexity and the other ~~with coarse resolution~~ of intermediate

90 complexity for which an adjoint exists to address the second problem. The technique also has a much wider range of additional applications, since resolutions using the adjoint method lag behind those applications featuring simpler assimilation methods as variational methods are typically a factor of 100 more costly than running the associated forward model. ~~In this context we only use the adjoint of~~ For example, the ~~low-resolution model to estimate the parameters that optimise the high-resolution model; i.e. , that brings the model closer to the assimilated observations. A pre-requisite of this method is that the parameters~~

95 ~~being optimised must be the same in both models. The~~ global GECCO3 ocean reanalysis based on the adjoint method (Köhl, 2020) features only a nominal resolution of 0.4°, while for instance the GOFS 3.1 (Laboratory, 2016) based on 3D-Var (Cummings and Smedstad, 2013) features 1/12° resolution. Employing coarser versions of the adjoint while still running the forward model with full resolution could significantly reduce the cost of the assimilation effort. Therefore, the objective of this paper is to ~~quantify~~ investigate the accuracy and precision ~~benefit~~ of such a synchronised data assimilation approach. We

100 perform this test using a Lorenz '63 model~~system~~.

The Lorenz '63 system (Lorenz, 1963) is a ~~well established proxy model to study~~ well-established proxy model of chaotic fluid systems, such as the atmosphere (Gauthier, 1992; Miller et al., 1994; Pires et al., 1996; Stensrud and Bao, 1992; Kravtsov and Tsonis, 2021; Huai et al., 2017; Yang et al., 2006; Daron and Stainforth, 2015; Errico, 1997). The advantage is that it can be used to ~~quantitatively evaluate the parameter dependence of the system prior to~~ rapidly evaluate parameter estimation

105 techniques in data assimilation schemes prior to their application in a full ~~model~~ESM with low computational resource requirements. New modelling techniques can thus be trialled in ~~relatively~~ fast experiments (Pasini and Pelino, 2005; Tandeo et al., 2015; Goodliff et al., 2020; Marzban, 2013; Yin et al., 2014). It can also be used in a wide range of other applications including, but not limited to, data assimilation, stochastic modelling terms, and predictions (Du and Shiue, 2021; Cameron and Yang, 2019; Pelino and Maimone, 2007). The system generates a ~~three dimensional, time varying~~ three-dimensional,

110 time-varying trajectory which with variation of both model parameters and/or initial conditions will produce very different trajectories. Thus, it is an ideal test bed for non-linear modelling in a number of fields (Hirsch et al., 2013). The Lyapunov exponent of the Lorenz '63 model is directly dependent upon its parameters making it ideal for climatological parameter estimation experiments. For our specific case these properties make it ideal to evaluate our techniques merits. In a previous study~~(Lyu et al., 2018) using~~, Lyu et al. (2018) used the Lorenz '63 model ~~, the assimilation primarily focused on the fit of~~

115 and its adjoint to fit a single parameter ρ and the initial conditions (x, y, z) ~~in a single model with an adjoint. This study will expand on the previous to~~ to observations. This present study builds on Lyu et al. (2018) to simultaneously fit all three model parameters ~~simultaneously~~ and use a model with an adjoint to optimise the parameters of ~~a~~ another model without one.

The structure of the remaining paper is as follows: In Section 2 we ~~introduce the model,~~ outline the methodology of ~~how these models are synchronised, show how~~ synchronisation, the cost function and the adjoint method~~is applied, describe our proposed~~

120 ~~multi-model setup, and detail the fitting procedure, including the minimisation algorithm~~. Section 3 introduces the Lorenz '63 model, describes our reference setup, before introducing our two novel multi-model methods, describing our minimisation algorithm, and detailing our statistical metrics for evaluating results. Section 4 shows and discusses the results of our multi-model setups, using a single model setup as a baseline for comparison. The results of introducing a mismodelling term to the adjoint model are also included. ~~Our results are summarised and conclusions discussed in~~ A summary and concluding remarks

125 are given Section 5.

2 Methodology

2.1 Synchronisation

In chaotic systems, integrating over periods longer than the predictability time scale creates problems for accurate parameter estimation. This is due to exponentially growing gradients, and a maximum likelihood estimate with an increasing number of local maxima (Köhl and Willebrand, 2002; Lea et al., 2000). The non-linear or chaotic dynamics, which detrimentally effect the maximum likelihood estimate, can be removed by synchronisation (Abarbanel et al., 2010; Sugiura et al., 2014) which transforms the chaotic model into one with linear dynamics without positive Lyapunov exponents leading to maximum likelihood functions with one unique maxima. This can be implemented into a generic model of ordinary differential equations,

$$\dot{\boldsymbol{x}}(t) = f(\boldsymbol{x}(t), \boldsymbol{\theta}, t) \tag{1}$$

where $\boldsymbol{x}(t)$ is the state vector, $\boldsymbol{\theta}$ is the parameter vector, and t is the time, synchronisation can be incorporated by adding a term which penalises the difference between the model and observations. This term is simply added to the equations

$$\dot{\boldsymbol{x}}(t) = f(\boldsymbol{x}(t), \boldsymbol{\theta}, t) + \alpha(\boldsymbol{x}_o(t) - \boldsymbol{x}(t)) \tag{2}$$

where α is the synchronisation coefficient and $\boldsymbol{x}_o(t)$ is the observation state vector.

According to the law of large numbers both with perfect models and in the presence of noise, the precision of the recovered parameters will improve with increasing window length since more data is integrated into the estimation. Similar benefits could be achieved by averaging estimates obtained over short windows, for which no synchronisation is necessary. However, underlying restrictions differ. For synchronisation, noise affects the state over the entire window, whereas for short windows noise effects are transported. Short window assimilation can be of benefit in perfect model settings from the error growth as suggested by the quasi-static variational assimilation (QSVA) framework (Pires et al., 1996) due to fact that sensitivities increase exponentially with time in chaotic models. The analogue of this QSVA effect in the Dynamical State and Parameter Estimation (DSPE) method (Abarbanel et al., 2009) is the attempt to reduce the synchronisation parameter as the optimisation progresses and parameters move closer to their true values. Since errors and sensitivities grow exponentially, feasible window lengths in QSVA have a maximum value due to limited numerical precision. Similarly, synchronisation parameters cannot approach zero for assimilation windows much larger than the predictability limit, because synchronisation will eventually fail if positive Lyapunov exponents exist (Quinn et al., 2009). We note that the reasoning for the need of long assimilation windows is somewhat different in the context of full ESM, for which it is essential to resolve long time scale physical mechanisms impacted by the specific choice of parameters, such as air-sea interactions of advection time scales in the ocean.

2.2 The cost function

As previously mentioned, in the context of variational data assimilation a cost function, J, must be introduced which measures the quadratic misfit between the model trajectory and observations. For the case of perfectly known initial conditions, but

uncertain parameters $\boldsymbol{\theta}$, J takes the generic form

$$J = \frac{1}{2N}\left(\boldsymbol{\theta} - \boldsymbol{\theta}_b\right)^T \frac{1}{\sigma_{\boldsymbol{\theta}_b}^2}\left(\boldsymbol{\theta} - \boldsymbol{\theta}_b\right) + \frac{1}{2N}\int_0^N dt\left(\boldsymbol{x}_o(t) - \boldsymbol{h}(\boldsymbol{x}(t))\right)^T \frac{1}{\sigma_{\boldsymbol{x}_o}^2}\left(\boldsymbol{x}_o(t) - \boldsymbol{h}(\boldsymbol{x}(t))\right) \tag{3}$$

where N is the total integration time, $\sigma_{\boldsymbol{x}_o}$ is the known uncertainty associated with the observational noise, $\boldsymbol{h}(\boldsymbol{x}(t))$ is the measurement function on model's predicted state vector $\boldsymbol{x}$, and $\boldsymbol{x}_o$ is the observation state vector. The prior parameter information with the associated uncertainty is denoted by $\boldsymbol{\theta}_o$ and $\sigma_{\boldsymbol{\theta}_b}$, respectively. The global minimum of this function is the maximum likelihood estimate of the model parameter values relative to the observations and prior information.

2.3 The adjoint method and the cost function gradient

To aid in the minimisation of the cost function, it is standard practice to calculate its gradient and use this to iteratively adjust control parameters. The adjoint model is introduced to provide these cost function gradients and requires the generation of the adjoint of the forward model equations. The resulting adjoint model can then be integrated in the reverse direction to give the gradient of the cost function. The background term in Eq. 3 can be omitted assuming a well-posed problem, without prior information on the parameter. Therefore, the gradient of the cost function with respect to the parameters is

$$\boldsymbol{\nabla}_{\boldsymbol{\theta}} J = -\frac{1}{N}\int_0^N dt \boldsymbol{\lambda}(t)\partial_{\boldsymbol{\theta}} f(\boldsymbol{x}(t),\boldsymbol{\theta},t), \tag{4}$$

where N is again the total integration time period, $\boldsymbol{\lambda}(t)$ is the adjoint vector at time t, and $\partial_{\boldsymbol{\theta}} f(\boldsymbol{x}(t),\boldsymbol{\theta},t)$ is the partial differential of the model with respect to the model parameters at time t.

3 Experimental setup

3.1 Lorenz '63 model

In this study, we use the ~~coupled~~ Lorenz '63 system ~~as described by Yang et al. (2006)~~ for all our experiments (Lorenz, 1963). The model is defined by the equations:

$$\frac{dx}{dt} = \sigma(y - x), \tag{5a}$$

$$\frac{dy}{dt} = \rho x - y - xz, \tag{5b}$$

$$\frac{dz}{dt} = xy - \beta z \tag{5c}$$

where ~~(x, y, z)~~ $\boldsymbol{x} = (x, y, z)$ are the state variables at each given time step and ~~σ, ρ, and β~~ $\boldsymbol{\theta} = (\sigma, \rho, \beta)$ are the model parameters. Throughout this article, we integrate all our models using the fourth-order Runge-Kutta method with a step size of $\Delta t = 0.01$ and total time period of 100 time units [TUs]. This system of equations will be subsequently referred to as

the *true* model for which we will apply the standard values for $\sigma_t = 10$, $\rho_t = 28$, and $\beta_t = 8/3$. This *true* true model with the parameters $\boldsymbol{\theta}_t = (10, 28, 8/3)$. This true model is used to generate pseudo-observation which will be used ~~to synchronise our physical models in Sub-section 2.1~~for synchronisation, data assimilation, and parameter estimation. Noise is included in these pseudo-observations by adding random values from a Gaussian distribution centred at zero ~~to the *true* values~~relative

185 to the true trajectory. The random noise ~~value~~ magnitudes are bounded ~~by a given percentage relative to the systems's standard distribution~~to 25% of the Lorenz '63 system's standard deviation. These pseudo-observations will be labelled as ~~(x_o, y_o, z_o)~~$\boldsymbol{x}_o = (x_o, y_o, z_o)$

3.2 ~~Adjoint method~~Reference setup (single model)

~~For the~~ To quantify the efficacy of our novel method we outline a reference setup for a synchronised Lorenz '63 ~~system outlined in Eq. (5) the adjoint matrix can be derived by transposing its tangent liner matrix (TLM). The TLM follows from~~

190 ~~Eq. (5) as:~~

$$M^T \equiv \frac{\partial \dot{\boldsymbol{x}}_i}{\partial \boldsymbol{x}_j} = \begin{pmatrix} -\sigma & \sigma & 0 \\ \rho - z & -1 & -x \\ y & x & -\beta \end{pmatrix}.$$

~~The adjoint M^* is then given by $M^* \equiv M^T$. As part of the adjoint model assimilation process, the adjoint equations are integrated backwards in time to calculate the gradient of the cost function with respect to control parameters which are used subsequently in an iterative process to adjust the control parameters such that the system is brought into consistency with~~

195 ~~observations.~~

3.3 ~~Synchronisation~~

~~A fundamental limitation of the adjoint method arises when integrating over periods that are longer than the predictability time scale of a system, leading to exponentially growing gradients and a cost function with an increasing number of local minima (Köhl and Willebrand, 2002; Lea et al., 2000). The problem can be mitigated by synchronisation which removes the~~

200 ~~non-linear or chaotic dynamics which effect the cost function (Abarbanel et al., 2010; Sugiura et al., 2014). To incorporate a synchronisation technique, we~~ framework similar to those described in Yang et al. (2006); Lyu et al. (2018). We expand the Lorenz '63 model ~~by adding so-called nudging~~ (Eq. 5) by adding synchronisation terms which then reads:

$$\frac{dx_a}{dt}\frac{dx}{dt} = \sigma(y_a - x_a \sim y - x) + \alpha(x_o - x_{\underset{\sim}{a}}), \tag{6a}$$

$$\frac{dy_a}{dt}\frac{dy}{dt} = \rho x_a - y_a - x_a z_a \sim x - y - xz + \alpha(y_o - y_{\underset{\sim}{a}}), \tag{6b}$$

205 $$\frac{dz_a}{dt}\frac{dz}{dt} = x_a y_a - xy - \beta z_{\underset{\sim}{a}} + \alpha(z_o - z). \tag{6c}$$

Here α is the synchronisation ~~constant, sub-script a denotes the model with an adjoint, (x_o, y_o, z_o)~~ coefficient and $\boldsymbol{x}_o = (x_o, y_o, z_o)$ are the pseudo-observations generated from the *true* model. For this synchronised model we will apply the standard values used

in the ~~*true* system~~ true model. We set the initial parameter values of this model at the start of the optimisation as the true system values plus a 10%-error~~, which~~. This gives $\sigma_a = 11$, $\rho_a = 30.8$, and $\beta_a = 44/15$ ~~. These~~ which will act as our initial ~~conditions~~ values for the parametric fit. ~~The adjoint matrix for the synchronised model in Eq. (6) is given by:~~

$$M_a^* = \begin{pmatrix} -(\sigma + \alpha) & \rho - z_a & y_a \\ \sigma & -(1 + \alpha) & x_a \\ 0 & -x_a & -\beta \end{pmatrix}.$$

~~The synchronisation constant is the only difference from the adjoint of Eq. (??)~~The initial conditions will remain unchanged compared to the true model as our interests are exclusively in climatic parameter estimation. Synchronisation will occur at every time step in all our setups and its coefficient will also be present in the adjoint equations. The significance of this ~~change will be shown~~ will be discussed in Section 4, as α has a critical role in the precision and accuracy to which parameters can be estimated, due to its influence on both the cost function and its gradient.

For each state variable in Eq. 6 a synchronisation term is included. There are seven possible combinations of ~~the three~~ these state variables which can be synchronised. The effect of each of the possible choices on the root mean squared error (RMSE) between the ~~*true*~~ true and adjoint systems ~~is shown in Fig. 1~~ by varying the synchronisation constant α from 0 to ~~30.~~ 30 is shown in Fig. 1. Noise was added (with zero mean and $\sqrt{2}$ standard deviation) to the true when constructing the pseudo-observations. The figure demonstrates that synchronising the z-component is ineffective at reducing the RMSE (Yang et al., 2006). In contrast, synchronising both x and y prove effective, with y leading the lowest RMSE values of the single variable for all values of α. Synchronising xyz and xy achieve the most effective reduction in RMSE for the lowest value of α. It can be seen in the figure that synchronising z can lead to model instability. Thus, we choose to only synchronise x and y in the following ~~, as seen in Eq. 6,~~ research to achieve more stable and accurate results with negligible precision loss.

The Lorenz '63 attractors for the trajectories of the ~~*true*~~ true model and that with an adjoint are shown in Fig. 2a without synchronisation. A large divergence is visible between the trajectories. However, if synchronisation is introduced the trajectories become very similar, as shown in Fig. 2b~~, the trajectories become very similar~~. There is now significant overlap between their kernel density estimations (KDEs). ~~We choose to use KDEs throughout this paper as they~~ KDEs represent a smoothed estimate of the PDF for the ~~trajectory~~model trajectory over a given time period. This allows for convenient visual comparison of trajectories. A more numerically rigorous method to check for effective synchronisation will be discussed in ~~section~~Section 4.

For all subsequent experiments with our setups, a parallel experiment will be performed with this reference setup. The differences in the results can then be compared to evaluate the advantages and disadvantages of the novel techniques.

3.3 Multi-model data assimilation

~~This technique~~ A multi-model tandem technique is now considered, which consecutively synchronises two forward models before running the adjoint of the second model backward in time. For this purpose, Eq. 6 must be modified to incorporate a consecutive synchronisation. A schematic of this setup is provided in Fig.3 and the implications of the two possible ways to calculate the cost function are discussed in the subsequent subsections.

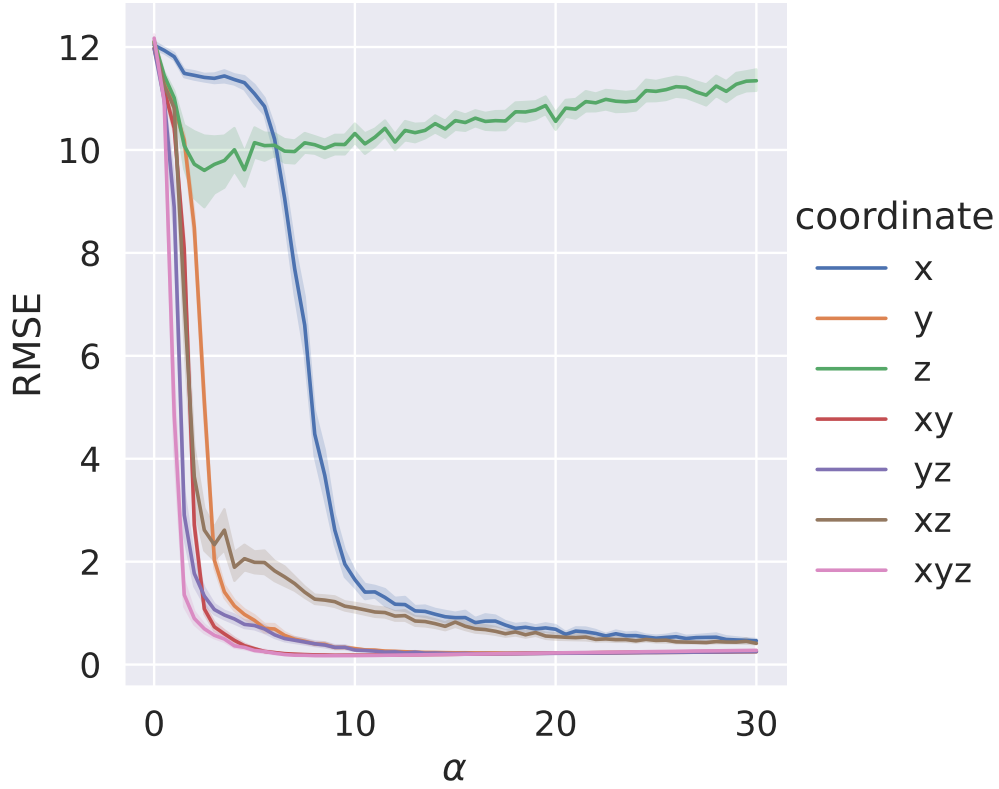



Figure 1. ~~Different coupling schemes are trialled. The~~ RMSE ~~is calculated~~ between the ~~*truth*~~ true and ~~*model*~~ reference model trajectories ~~after~~ in seven different synchronisation ~~to pseudo-observations. Noise was added (with zero mean and $\sqrt{2}$ standard deviation) to the *truth* when constructing the pseudo-observations~~scenarios. The synchronisation constant α is varied from 0 to 30 in steps of 0.5. The solid line is the median and the shaded area is the 68% percentile interval for the ensemble of 100 experiments carried out.

The first model has no adjoint equations and is the target model for which we wish to optimise the parameters. The equations of model 1, which is run only in forward mode, are

$$\frac{dx_f}{dt} = \sigma(y_f - x_f) + \alpha(x_o - x_f), \tag{7a}$$

$$\frac{dy_f}{dt} = \rho x_f - y_f - x_f z_f + \alpha(y_o - y_f), \tag{7b}$$

$$\frac{dz_f}{dt} = x_f y_f - \beta z_f \tag{7c}$$

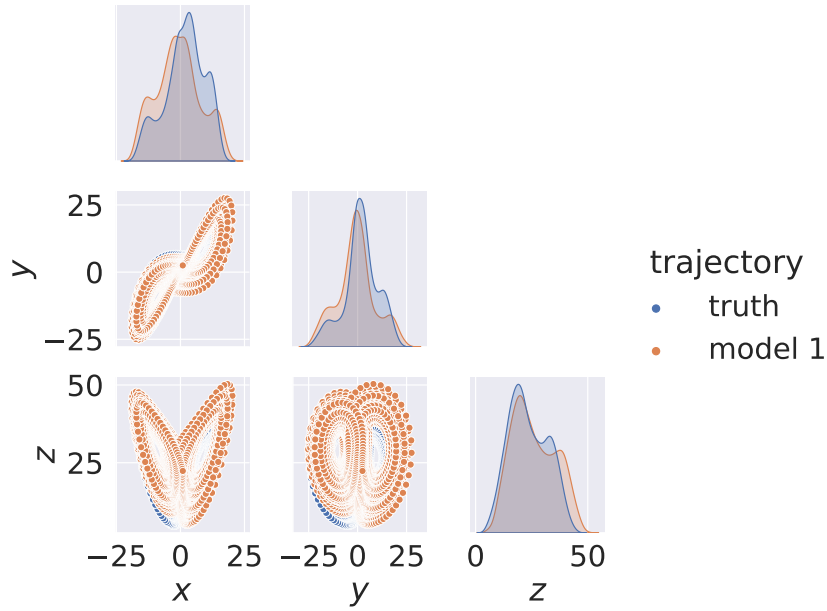
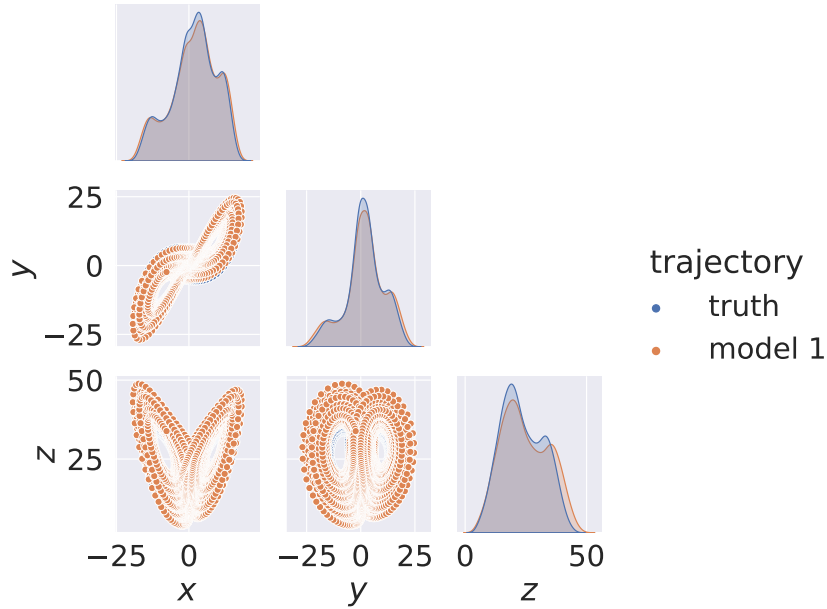
where the sub-script f denotes the forward run of model 1 and sub-script o denotes observations generated from ~~*truth*~~ true model. The system of equations for the model 2 which has an adjoint will now be modified to synchronise with the forward-

(a) $\alpha = 0$.

(b) $\alpha = 10$.

Figure 2. The bottom left quadrants show the Lorenz '63 ~~truth~~ true and ~~model~~ model attractors from the main three variable orientations. The diagonal plots show kernel density estimations (KDEs). Fig. 2a shows the trajectories without synchronisation. Fig. 2b shows the trajectories with synchronisation.

only model and not the observations:

$$\frac{dx_a}{dt} = \sigma(y_a - x_a) + \alpha(x_f - x_a), \tag{8a}$$

$$\frac{dy_a}{dt} = \rho x_a - y_a - x_a z_a + \alpha(y_f - y_a), \tag{8b}$$

$$\frac{dz_a}{dt} = x_a y_a - \beta z_a \tag{8c}$$

250 where the sub-script a denotes model 2 which has an adjoint. This model synchronises with the model 1 but never directly with the observations.

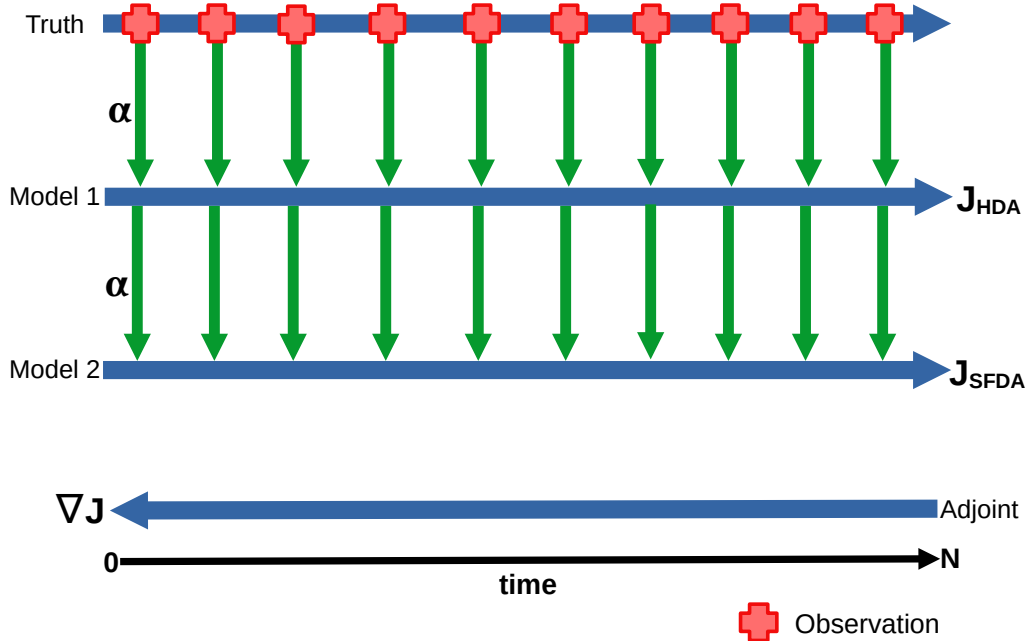



Figure 3. Illustration of the multi-model setup where each pseudo-observation generated from the ~~truth~~ true model includes random additive Gaussian noise. The cost function can measure the difference between the observations and either model 1 or 2 depending on the assumptions made. Both options are discussed in the text.

~~To assimilate the data, we fit one of the synchronised models to the observations by optimising the model parameters. A cost function is constructed to calculate the misfit between observations and the model of interest. The gradient of the cost function, with respect to the model parameters, is always calculated using the adjoint method associated with model 2.~~
255 ~~However, the form of the adjoint will vary between the two methods we subsequently present. The adjoint model is numerically evaluated by automatic differentiation (AD) of the forward model 2 with respect to either model 1 or 2 depending on our~~

chosen setup. This is done in the python package `JAX` ~~which numerically evaluates the vector Jacobian product of the model with respect to its state variable vector (Bradbury et al., 2018). This is then integrated using an inverse Runge-Kutta scheme. Our code stores the state variables and adjoint vectors at each time step. It is also possible to carry out the entire integration using~~ `JAX`~~. The process of synchronising all models, calculating the cost function and its gradient, and then adjusting model parameters is carried out iteratively by our chosen minimisation algorithm. Throughout all steps the parameter value of forward-only and adjoint models are identical and optimised simultaneously.For the present study we have chosen to use the minimisation package~~ `iminuit` ~~(Dembinski and et al., 2020). The version currently used is the~~ `Minuit2` ~~algorithm~~ `MIGRAD` ~~which is a variable-metric method including an inexact line search, a stable metric updating scheme, and it checks for positive-definiteness. This is chosen as it is a robust and stable minimiser that both utilises the gradient we derive using adjoint equations, and calculates accurate uncertainty estimates for all fitted parameters. The procedure stops when the estimated-distance-to-minim (EDM) stopping criteria is met. This measures the difference between the true cost function value at the estimated minimum and the estimated cost function value at the true minimum. The EDM criteria is given by maximum EDM < tolerance · 0.002, where we use the default tolerance = 0.1. This is appropriate for accurate uncertainty recovery at no adverse cost to computational speed. Further information on the specific minimisation algorithm details are given in James and Roos (1975).~~

3.3.1 Setup 1 - state filtered data assimilation (SFDA)

Assuming that both ~~Lorenz models~~ model 1 and 2 can be thought of as representing two identical climate models, the cost function can be placed on the model 2. This allows model 1 to filter out some of the background noise on the observations before they are given to the cost function attached to model 2. ~~This favourably increases the signal to background ratio synchronised into model 2.~~ Such a filtering setup would theoretically reduce parametric uncertainty below that of traditional single model data assimilation because model 1 should act to reduce the amount of noise synchronised into model 2. We will subsequently refer to this setup as state filtered data assimilation (SFDA.)

In SFDA the cost function acts to constrain model 2. The cost function is

$$J_{\text{SFDA}} = \frac{1}{2N} \sum \int_{t=0}^{N} dt \, (\boldsymbol{x}_o(t) - \boldsymbol{x}_a(t))^T \frac{1}{\sigma_{\boldsymbol{x}_o}^2} (\boldsymbol{x}_o(t) - \boldsymbol{x}_a(t)) \tag{9}$$

N is again the total number of time steps of the assimilation window, and $\sigma_{\boldsymbol{x}_o}$ is the uncertainty associated with the observation noise. The adjoint matrix includes ~~additional terms , compared with Eq. ??,~~ terms arising from the second ~~synchronisation step~~ tandem layer of synchronisation for model 2. This is given by:

$$M M_{\text{SFDA}}^* = \begin{pmatrix} -(\sigma + \alpha) & \rho - z_a & y_a & 0 & 0 & 0 \\ \sigma & -(1+\alpha) & x_a & 0 & 0 & 0 \\ 0 & -x_a & -\beta & 0 & 0 & 0 \\ \alpha & 0 & 0 & -(\sigma + \alpha) & \rho - z_f & y_f \\ 0 & \alpha & 0 & \sigma & -(1+\alpha) & x_f \\ 0 & 0 & 0 & 0 & -x_f & -\beta \end{pmatrix} \tag{10}$$

which in practice is numerically evaluated using ~~AD~~the automatic differentiation (AD) package `jax` to calculate the vector-Jacobian product.

The adjoint equation for SFDA is given by:

$$\dot{\boldsymbol{\lambda}}_{\text{SFDA}}(t) = \frac{1}{\sigma_{\boldsymbol{x}_o}^2}\left((\boldsymbol{x}_o(t),0,0,0) - (\boldsymbol{x}_a(t),0,0,0)\right) \tag{11a}$$

$$-\boldsymbol{M}^*_{\text{SFDA}}(t)\boldsymbol{\lambda}_{\text{SFDA}}(t) \text{ for } t = N,...,0$$

with $\boldsymbol{\lambda}_{\text{SFDA}}(N) = \boldsymbol{0}.$ \hfill (11b)

These equations were derived using the method detailed in Talagrand (2010). The gradient can then be calculated with respect to the parameters (σ,ρ,β) notated by the subscript $\boldsymbol{\theta}$. This yields

$$\boldsymbol{\nabla}_{\boldsymbol{\theta}} J_{\text{SFDA}} = \frac{1}{N}\int\limits_{t=N}^{0} dt \boldsymbol{\lambda}_{\text{SFDA}}(t) \begin{pmatrix} y_a(t) - x_a(t) \\ x_a(t) \\ -z_a(t) \\ y_f(t) - x_f(t) \\ x_f(t) \\ -z_f(t) \end{pmatrix} \tag{12}$$

which is a component-wise multiplication at each time step.

3.3.2 Setup 2 - ~~hybrid~~ Tandem data assimilation (~~HDA~~TDA)

~~Here~~ In this section, we want to explore if using an existing adjoint from one model could be utilised to optimise a different target model without adjoint. This will be referred to as ~~hybrid~~ tandem data assimilation (~~HDA.)In HDA~~ TDA). In TDA we assume that both models may differ in resolution or numerical formulation but are governed by the same ~~equations~~continuum dynamics. Instead of interpolating or transforming the original model variables onto the adjoint model grid, formulation of the adjoint model through synchronisation would provide a simpler means to do this ~~,~~ as only essential parameters need to be interpolated. Auxiliary variables and parameters will be generated by the synchronised model, including those that may not exist in the target model~~, will be generated by the synchronised model~~.

The cost function of ~~HDA is~~TDA is

$$J_{\text{HDA}\text{TDA}} = \frac{1}{2N}\sum\int_{t=0}^{N}dt\,(\boldsymbol{x}_o(t) - \boldsymbol{x}_f(t))^T \frac{1}{\sigma_{\boldsymbol{x}_o}^2}(\boldsymbol{x}_o(t) - \boldsymbol{x}_f(t)). \tag{13}$$

N is the total number of time steps of the assimilation window and $\sigma_{\boldsymbol{x}_o}$ is the uncertainty associated with the observation noise. This measures the quadratic misfit between the forward-only model 1 and the observations. Model 1, $\boldsymbol{x}_f$, will be constrained by this cost function and its gradient will be calculated using the adjoint of model 2, $\boldsymbol{x}_a$. In this formulation, the two systems are no longer considered as a single synchronised one but as two separate models, one for the calculation of the trajectory and

the other for calculating the gradient from its adjoint. The algorithm is also no longer exact as we only assume that model 2 ajoint will provide a good approximation as long as its trajectory follows model 1 closely and is driven by the model-data differences of model 1. The adjoint matrix is

$$M_{\text{HDA}}M_{\text{TDA}}^{*} = \begin{pmatrix} -(\sigma + \alpha) & \rho - z_a & y_a \\ \sigma & -(1+\alpha) & x_a \\ 0 & -x_a & -\beta \end{pmatrix}. \tag{14}$$

which is numerically evaluated using AD. The synchronisation with model 1 ensures that the trajectory of model 2 x_a closely follows that of model 1.

3.3.3 Cost function gradient

The gradient of the cost function, with respect to the state variables at $t = 0$, is given by

$$\nabla_{x_a} J = \frac{1}{N}\lambda(0),$$

where λ is the adjoint vector. The adjoint vector is calculated by integration of the differential adjoint equation in the reverse time direction. This will differ between our two setups.

3.3.3 SFDA (setup 1)

This adjoint equation for SFDA is given by:

$$\begin{aligned} \dot{\lambda}_{\text{SFDA}}(t) &\equiv \frac{1}{\sigma_{x_o}^2}(x_o(t) - x_a(t)) \\ &\quad -M_{\text{SFDA}}^*(t)\lambda_{\text{SFDA}}(t) \text{ for } t = N, ..., 0 \end{aligned}$$

$$\text{with } \lambda_{\text{SFDA}}(N) \equiv 0.$$

These equations were derived using the method detailed in Talagrand (2010). The gradient can then be calculated with respect to the parameters (σ, ρ, β) notated by the subscript θ. This yields

$$\nabla_{\theta} J_{\text{SFDA}} = \frac{1}{N}\sum_{t=N}^{0} \lambda_{\text{SFDA}}(t) \begin{pmatrix} y_a(t) - x_a(t) \\ x_a(t) \\ -z_a(t) \\ y_f(t) - x_f(t) \\ x_f(t) \\ -z_f(t) \end{pmatrix}$$

which is a component-wise multiplication at each time step.

3.3.4 ~~HDA (setup 2)~~

~~This~~ The adjoint equation for ~~HDA~~ TDA is given by:

$$\dot{\boldsymbol{\lambda}}_{\text{HDA TDA}}(t) = \frac{1}{\sigma_{\boldsymbol{x}_o}^2}\left(\boldsymbol{x}_o(t) - \boldsymbol{x}_f(t)\right) \tag{15a}$$

$$-M_{\text{HDA}} - \boldsymbol{M}_{\text{TDA}}{}^*(t)\boldsymbol{\lambda}_{\text{HDA TDA}}(t) \text{ for } t = N,...,0$$

with $\boldsymbol{\lambda}_{\text{HDA TDA}}(N) = \boldsymbol{0}.$ (15b)

The gradient with respect to the parameters $\boldsymbol{\theta} = (\sigma, \rho, \beta)$ is calculated to be:

$$\boldsymbol{\nabla}_{\boldsymbol{\theta}} J_{\text{HDA TDA}} = \frac{1}{N}\sum\int_{t=N}^{0} dt\, \boldsymbol{\lambda}_{\text{HDA TDA}}(t) \begin{pmatrix} y_a(t) - x_a(t) \\ x_a(t) \\ -z_a(t) \end{pmatrix} \tag{16}$$

which is again a component-wise multiplication at each time step. For ~~HDA~~ TDA and SFDA, the trajectories of both models and adjoint vectors are stored for evaluation of the gradient.

4 ~~Results~~

~~Throughout the following section we will use the single model described in Lyu et al. (2018) as our benchmark to compare the new setups against. To understand the behaviour of the setups at different operating extremes, assimilations are carried out for variations of observational noise and~~ α~~. This will help establish the optimal synchronisation strength dependent on the noise amplitude. We will also be able to compare the errors and uncertainties of~~

3.1 Minimisation algorithm

To assimilate the data, we fit one of the synchronised models to the ~~single model with our multi-model setups~~ observations by optimising the model parameters. A cost function is constructed to calculate the misfit between observations and the model of interest. The gradient of the cost function, with respect to the model parameters, is always calculated using the adjoint method. However, the form of the adjoint will vary between the two methods we presented in Eqs. 11 and 15. The adjoint model is numerically evaluated by AD of model 2. This is done in the python package `JAX` which numerically evaluates the vector Jacobian product of the model with respect to its state variable vector (Bradbury et al., 2018). This is then integrated using an inverse Runge-Kutta scheme. Our code stores the state variables and adjoint vectors at each time step. It is also possible to carry out the entire integration using `JAX`. The process of synchronising all models, calculating the cost function and its gradient, and then adjusting model parameters is carried out iteratively by our chosen minimisation algorithm. Throughout all steps the parameter value of forward-only and adjoint models are identical and optimised simultaneously.

3.2 Statistical metrics

To get a more robust quantification of our setup's behaviour, it is necessary to repeat our study over a number data sets to
calculate medians and percentile intervals (PIs). This allows us to examine general traits of our model without an individual
noise event obscuring trends and features of significance. Here this is done by generating 100 pseudo-data sets and assimilating
each set independently. The plotting package is then ~~applied directly~~ directly applied to these 100 outputs to plot the median
and 68% ~~percentile intervals. The percentile intervals~~ PIs. The PIs are included to illustrate the statistical spread of the results
and reproducibility, not to explicitly indicate uncertainty. Hence, we choose 68% for our PIs to give a concise visualisation of
the central 1σ of results. The mean percentage error and uncertainty are plotted separately to allow for quantification of both
the accuracy and precision of our results. [1] These are calculated by:

$$\text{mean \%-error} = 100\% \cdot \sqrt{\frac{1}{3} \cdot \left[\left(\frac{\sigma - \sigma_t}{\sigma_t} \right)^2 + \left(\frac{\rho - \rho_t}{\rho_t} \right)^2 + \left(\frac{\beta - \beta_t}{\beta_t} \right)^2 \right]} \tag{17}$$

and

$$\text{mean \%-uncertainty} = 100\% \cdot \sqrt{\frac{1}{3} \cdot \left[\left(\frac{\Delta\sigma}{\sigma_t} \right)^2 + \left(\frac{\Delta\rho}{\rho_t} \right)^2 + \left(\frac{\Delta\beta}{\beta_t} \right)^2 \right]}. \tag{18}$$

The error is ~~simply~~ calculated by percentage difference between the fitted and ~~*true*~~ true parameter values. The parametric
uncertainty is calculated by the minimisation algorithm using a Hessian estimate.

After parameter estimation, the optimised parameters are used to initialise a free unsynchronised run of the model. The
attractors are plotted against the attractor of the true model. In all cases with synchronisation greater than or equal to the
optimum value, the attractors' KDE shows precise and consistent agreement with that of the true model. Results are not
displayed as there is no differences which are merit discussion. Thus, our focus in the subsequent results is to compare how the
examined setups differ in terms of accuracy and precision of optimised parameters recovered.

4 Results

Throughout the following section we will use the single model described in Lyu et al. (2018) as our benchmark to compare the
new setups against. To understand the behaviour of the setups at different operating extremes, assimilations are carried out for
variations of observational noise and α. This will help establish the optimal synchronisation strength dependent on the noise
amplitude. We will also be able to compare the errors and uncertainties of the single model with our multi-model setups.

---

[1] ~~$\text{mean \%-error} = 100\% \cdot \sqrt{\frac{1}{3} \cdot \left[\left(\frac{\sigma - \sigma_t}{\sigma_t} \right)^2 + \left(\frac{\rho - \rho_t}{\rho_t} \right)^2 + \left(\frac{\beta - \beta_t}{\beta_t} \right)^2 \right]}$~~

[1] ~~$\text{mean \%-uncertainty} = 100\% \cdot \sqrt{\frac{1}{3} \cdot \left[\left(\frac{\Delta\sigma}{\sigma_t} \right)^2 + \left(\frac{\Delta\rho}{\rho_t} \right)^2 + \left(\frac{\Delta\beta}{\beta_t} \right)^2 \right]}$~~

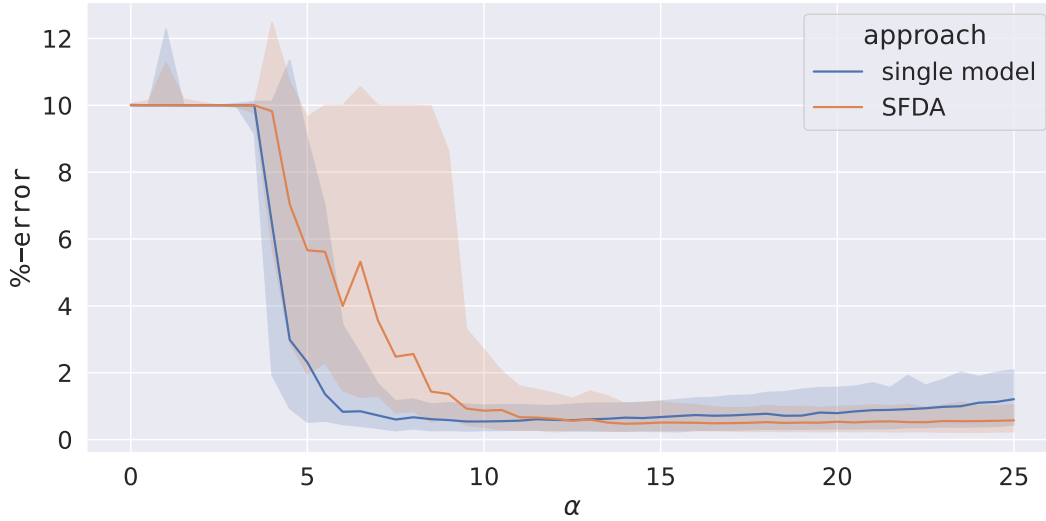



Figure 4. The percentage error between the ~~true~~ _true_ values of (σ, ρ, β) and the fitted value from SFDA. A single model assimilation is included for comparison. An ensemble of 100 assimilations is carried out over 100 different data sets. The median (lines) and 68% percentile intervals (shaded areas) are plotted. The noise level is 25%.

4.1 SFDA (setup 1)

The results from a scan of α are shown in Fig. 4. The single model scan has two main regions. The first, _for_ $\alpha \leq 7.5$, is where the system is poorly synchronised ~~giving~~ _leading to_ an inaccurate fit of the parameters _and an unstable median value_. The second, ~~above~~ _where_ $\alpha > 7.5$, is where the system is fully synchronised and recovers the ~~true~~ _true_ model parameters very effectively. SFDA has a higher onset of effective synchronisation than the single model setup, beginning at $\alpha = 11$. Above $\alpha = 12.5$, SFDA has consistently more accurate parameter recovery than the single model setup, _while the opposite holds below $\alpha = 12.5$. The minimum error at the respective optimal α values are nearly identical._

Fig. 5 shows the results of two fits carried out ~~with~~ _for data with an applied_ noise of 25% _relative to the systems' standard deviation_. The mean percentage uncertainty over the three parameters is plotted for both setups. Noticing in particular the spread of the percentile intervals, the single model setup is found to be synchronised and have a high precision from $\alpha = 7.5$ and SFDA from $\alpha = 11.5$. Once the SFDA setup is synchronised, it is found to have a reduced uncertainty compared with the single model. SFDA is found to be approximately one third more precise ~~better~~ than the single model setup for all value of α investigated. _However, since SFDA requires larger α, uncertainty values for the same α cannot be directly compared._

Fig. 4 suggest that the SFDA technique is more accurate than a standard single model setup at higher values of α. Fig. 5 additionally shows that SFDA is more precise than a single model for all values of α after the onset of effective synchronisation. In cases where accuracy and precision are desirable, and computational resources and time are available this would advocate for the use of SFDA. However, Fig. 4 suggests that the error estimates do not represent the actual achieved accuracy of the parameter estimation. Both metrics suggest that the accuracy is less sensitive to the choice of α for SFDA. It is also important

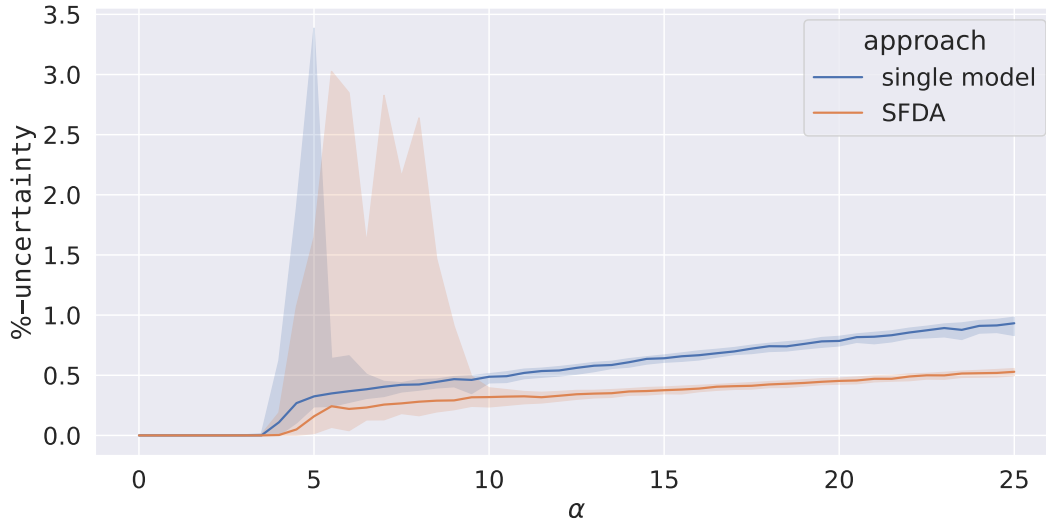



Figure 5. The average percentage uncertainty on the three parameters (σ, ρ, β) after SFDA from the minimisation algorithm for different values of α. A single model assimilation is included for comparison. An ensemble of 100 assimilations is carried out over 100 different data sets. The median (line) and 68% percentile intervals (shaded areas) are plotted.

395   to note that in the context where precision is the priority, the lowest value of α after effective synchronisation should be chosen. Increasing values of α increase the parametric uncertainty because of the associated declining sensitivity of the trajectory to parameter changes. This is also the reason why for the same α SFDA shows consistently better performance: the less efficient indirect constraint to the observations makes it more sensitive to the parameters.

Fig. 6 shows the results of varying the noise levels on the fitted parameter values. For all ~~noise levels studied~~ applied noise
400   levels the quality of the fit can be considered good as the median of the mean percentage uncertainty on the parameters remains below 0.5% even with noise levels of up to 50%. SFDA is found to have ~~similar~~ a mean error performance similar to the single model system across the range of noise levels tested. ~~The~~ However, the spread of the error is slightly improved in the double model setup at low noise ~~likely~~ due to the forward-only model ~~'smoothing '~~smoothing outlying observation better than a single model setup. The parametric uncertainty is found to be consistently reduced in the double model system for all noise
405   levels. This demonstrates the precision improvement achieved by running the forward model twice to ~~'smooth '~~smooth the observations before carrying out data assimilation. The consequences of this are that for smaller ~~or localised climate~~ models, where computational resources are available and improved precision ~~is~~ or accuracy are desirable, SFDA ~~could~~ can reduce error and particularly decrease uncertainty.

4.2 ~~HDA~~ TDA (setup 2)

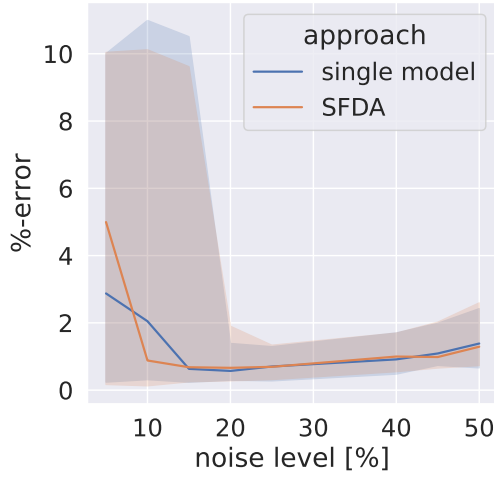
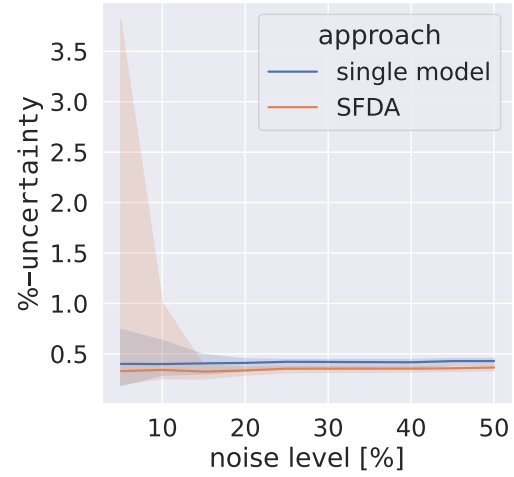
410   ~~The results from a scan of α~~ Similar to SFDA, TDA results are evaluated in terms of percentage error and uncertainty estimates against the single model with a data noise level of 25%. Error estimates are shown in Fig. 7 ~~. The single model scan has two~~

(a) Percentage error.

(b) Percentage uncertainty.

Figure 6. The percentage error between the ~~true~~ true values of (σ, ρ, β) and those from SFDA, as well as average percentage uncertainty on the SFDA parameters. A single model assimilation is included for comparison. An ensemble of 100 assimilations is carried out over 100 different data sets. The median (line) and 68% percentile intervals (shaded areas) are plotted. The noise level varies between 5% and 50% in steps of 5%.

415 ~~main regions. The first, $\alpha \leq 7.5$, is where the system is poorly synchronised giving an inaccurate fit of the parameters. The second,~~ as function of α. For $\alpha \geq 3$ synchronisation starts to set in and parameter estimation begins to improve. The system is only synchronising effectively for $\alpha \geq 7.5$ ~~, is where the system is synchronising very effectively and recovers~~ to consistently recover the true model parameters~~very accurately. The HDA~~, as is visible from the small spread of the error. The TDA scan follows the behaviour of the primary model very closely with no visible disadvantage.

~~Fig. 8 shows the results of two fits carried out with noise of 25%.~~ The mean percentage uncertainty over the three parameters is plotted in Fig. 8 for both setups. ~~HDA~~ TDA is found to have almost identical uncertainty to the single model. The plot consists of two regions. The first, for $\alpha < 7.5$, is the region where the model is not yet consistently synchronised producing

420 high variability depending on the specific noise. The second, for $\alpha \geq 7.5$, is where the system is consistently synchronised. The minimum median of the mean parametric uncertainty, after consistent synchronisation begins, is $\approx 0.35\%$ and achieved at $\alpha = 7.5$. The subsequent increase in uncertainty is due to the reduced the parametric sensitivity associated with the increased α~~both reducing and flattening~~, thereby reducing the curvature of the cost function ~~which reduces the parametric sensitivity of the fit~~at the minimum.

425 Fig. 9 shows the results of varying the noise levels on the fitted parameter values. For all noise levels studied~~the quality of~~, the fit can be considered ~~good~~ to be accurate as the mean percentage error on the parameters remains below 1% even with noise levels of 50%. The increased spread of the error ~~results~~ at low noise is thought to be ~~the adjoint experiencing reduced~~

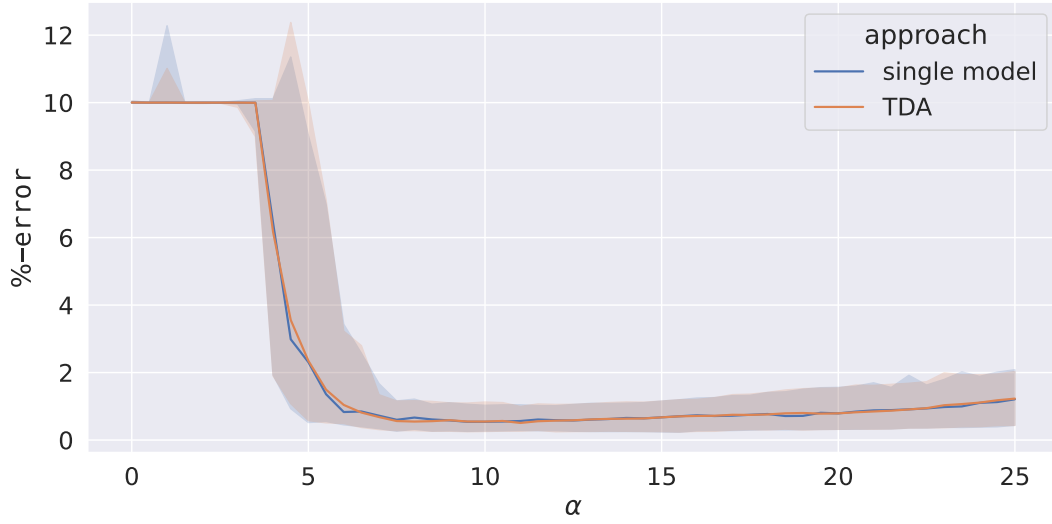



Figure 7. The percentage error between the ~~true~~ true values of (σ, ρ, β) and those from ~~HDA~~TDA. A single model assimilation is included for comparison. An ensemble of 100 assimilations is carried out over 100 different pseudo-data sets. The median (lines) and 68% percentile intervals (shaded areas) are plotted. The noise level is 25%.

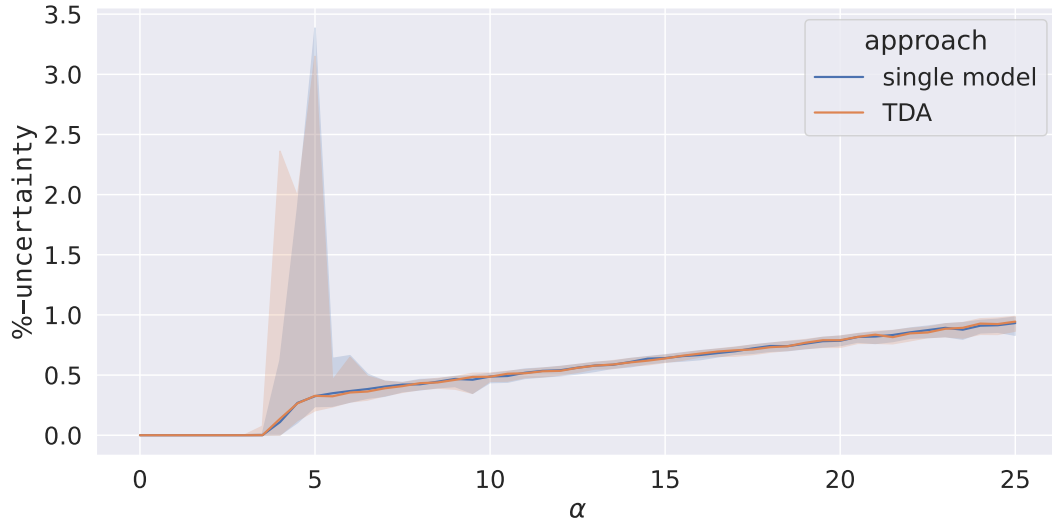



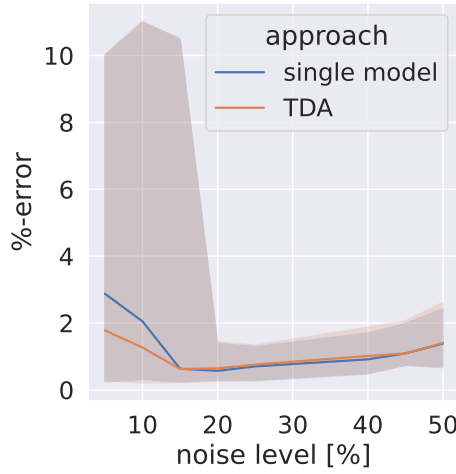
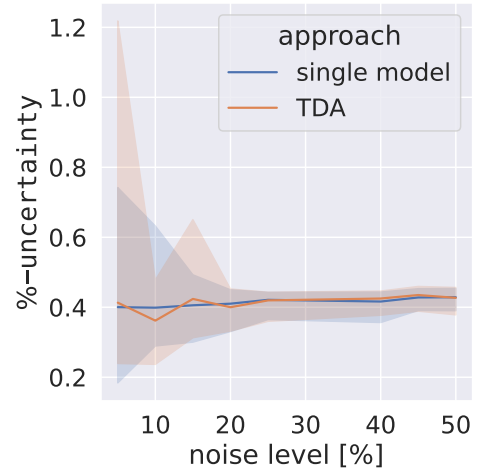
Figure 8. The percentage uncertainty, from the minimisation algorithm, averaged over all three parameters (σ, ρ, β) after ~~HDA~~TDA. A single model assimilation is included for comparison. An ensemble of 100 assimilations is carried out over 100 different data sets. The median (lines) and 68% percentile intervals (shaded areas) are plotted.

(a) Percentage error.

(b) Percentage uncertainty.

Figure 9. The percentage error between the ~~true~~ true values of (σ, ρ, β) and those from ~~HDA~~TDA. A single model assimilation is included for comparison. An ensemble of 100 assimilations is carried out over 100 different data sets. The median (lines) and 68% percentile intervals (shaded areas) are plotted. The noise level varies between 5% and 50% in steps of 5%.

~~effects of the noise compared to the single model. The length of the time window used is 100TUs, but if the window were~~ ~~increased the accuracy of the results would improve. The HDA~~ due to the fixed value of α used for all noise levels impacting the synchronisation of the system. The TDA setup is found to have extremely consistent uncertainty compared to the single model system.

The consistent results of TDA in Figs. 7, 8, and 9 relative to the standard single model setup show that transferring information via synchronisation does not compromise precision. Figs. 7, and 8 also concur with those of SFDA in suggesting that the optimal value of α is the smallest value after the onset of effective synchronisation. An increase in α beyond this point can lead to a significant reduction in the precision of the parameters. In cases where only a simpler, but similar, model with an adjoint is available results are likely to degrade. In the following section, we will study the potential impact of model inconsistencies on the the precision of the parameter estimation.

4.3 Mismodelling in ~~HDA~~TDA

~~HDA~~ In this section, the tandem data assimilation (setup 2) ~~uses~~ is be used with different forward and adjoint models ~~with some~~ ~~shared physics . Therefore, it is important~~ that share common physics to examine the impact of ~~a difference in the physics of~~ ~~these two models~~introducing model discrepancies. We construct a test case where the equations of model 2, which has an adjoint, Eq. ~~(8)~~ 8 are modified to give an oscillatory difference to both the ~~truth~~ true model and model 1. This can be done in a number of ways. We ~~chose~~ choose to introduce a multiplicative sine function to the equations in such a way it is also included

in the adjoint matrix and thus modifies the gradient values returned to the fitting algorithm. Model 2 with an adjoint is now

445
$$\frac{dx_a}{dt} = \sigma(y_a - x_a) + \alpha(x_f - x_a), \tag{19a}$$

$$\frac{dy_a}{dt} = \rho x_a - y_a - x_a z_a + \alpha(y_f - y_a), \tag{19b}$$

$$\frac{dz_a}{dt} = x_a y_a - \beta z_a \cdot (1 - \epsilon \sin(2\pi t)) \tag{19c}$$

where ϵ is term which determines the strength of the oscillation term~~and t is the simulation time~~. The effect of this term on the attractor is shown in Fig. 10, without synchronisation~~and assimilation~~. When compared to Fig. 2a, it can be seen that this term
450 is ~~very effective at~~ successful in distorting the shape and probability density of the attractor.

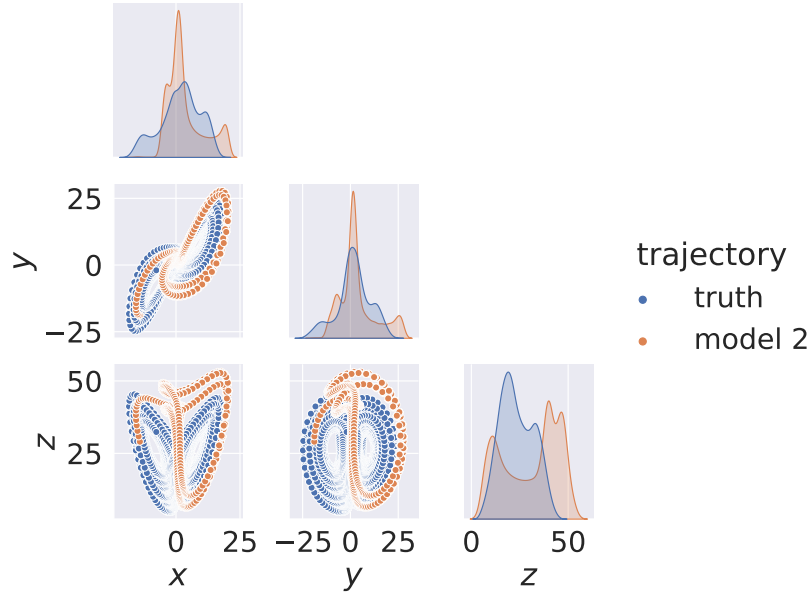



Figure 10. The Lorenz *truth* true model and model 2 attractors in the case of $\epsilon = 1.0$. The bottom left and top right quadrants shows the attractors from all possible co-ordinate orientations. The diagonal plots show kernel density estimations (KDEs). No noise is added.

The ~~consequence~~ consequences of varying this term on the accuracy of parameter optimisation after assimilation are shown in Fig. 11. With increasing ~~epsilon~~ ϵ the percentage error and uncertainty between fitted and ~~truth~~ true systems remains stable. In spite of the large impact this term has on the attractor shape, ~~this~~ the figure demonstrates a resilience of ~~HDA~~ TDA to modelling differences between the forward-only model 1 and model 2 with an adjoint.

455 ## 5 Conclusions

In this paper we have demonstrated the ability to constrain a Lorenz '63 model using a second model with similar physics and an adjoint by 4D-Var data assimilation. Such an approach removes the need to generate an adjoint for a forward model,

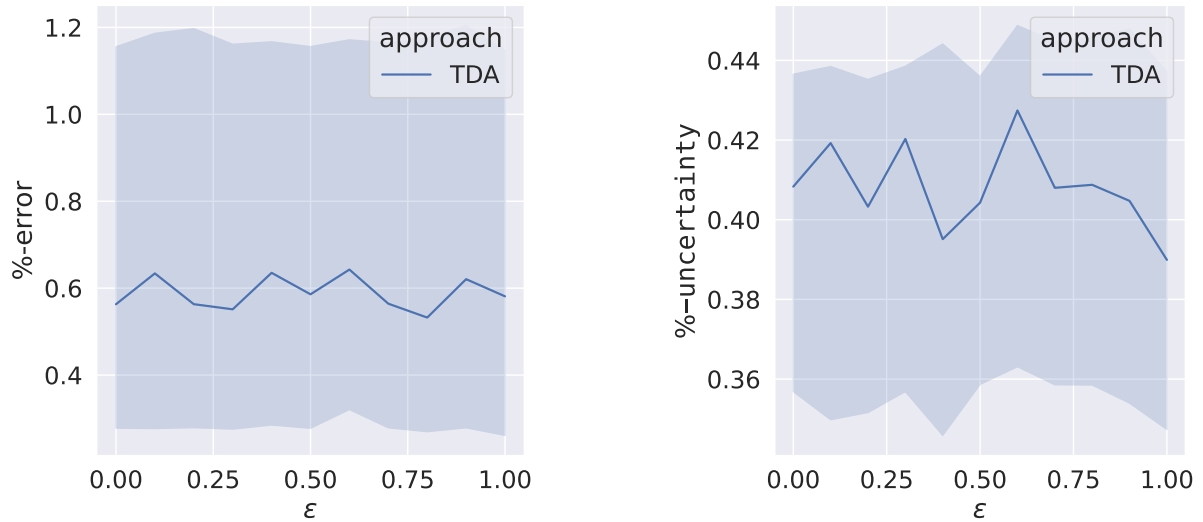



Figure 11. The percentage error (left) and uncertainty (right) between the ~~true~~ true values of (σ, ρ, β) and those from ~~HDA~~TDA. An ensemble of 100 assimilations is carried out over 100 different data sets. The median (line) and 68% percentile intervals (shaded areas) are plotted. The noise level 25% and, $\alpha = 7.5$.

if such an adjoint already exists for a separate, yet dynamically similar system. An important application of this technique in Earth system modelling would be a situation where a low-resolution ESM with an adjoint shares a parametrisation with a high-resolution ESM ~~without an adjoint . The~~ for which no adjoint exists. Moreover, using a lower-resolution version of the same model could computationally make data assimilation much faster. We have shown that in both cases the low-resolution ESM with an adjoint could, through synchronisation, ~~provide~~ follow the trajectory of the more complex and high-resolution model while at the same time providing all necessary variables to run its tangent linear adjoint model. This can then be utilised to estimate parameters in the ~~high-resolution ESM without an adjoint. Moreover, using a simpler lower resolution version could make data assimilation quicker and would use fewer computational resources than an adjoint of the~~ complex high-resolution ESM. We have also shown that running a forward model twice before beginning data assimilation can act to smooth the data and reduce the parametric uncertainty~~by roughly one third compared to a single forward run. Our overarching attention in this approach is~~. Our focus is in optimising the parameters of a full ESM, which will be tested as a next step. ~~Nevertheless, it~~ It would also be possible to optimise the ~~state variables which are~~ initial condition of the state variables, which is more applicable to weather forecasting techniques. Future work will examine the resilience of such setups to spacially and temporally sparse data.

Data availability. The pseudo-data samples used in this study are available on request from the authors.

Author contributions. PDK carried out the research, wrote the initial manuscript draft, and edited it. AB assisted with the research and edited the manuscript. AK and DS designed the research concept, supervised the work, and participated in the writing of the manuscript.

475 *Competing interests.* The authors declare no potential competing interest.

Acknowledgements. The authors would like to thank the two anonymous referees for providing very detailed and constructive reviews which significantly contributed to the final form of this manuscript. The figures in this research were plotted by the `seaborn` plotting package (Waskom, 2021) utilising our output which was managed using `pandas` (pandas development team, 2020). Automatic differentiation of our model was carried out using `JAX` (Bradbury et al., 2018). The minimisation of our cost function and uncertainty evaluation was
480 done by `iminuit` (Dembinski and et al., 2020). This research was supported, in part, through the Koselleck grant *Earth^{RA}* ~~from~~ funded by the Deutsche Forschungsgemeinschaft (grant number: 66492934.) ~~Contribution to the Centrum für Erdsystemforschung und Nachhaltigkeit (CEN) of Universität Hamburg.~~ This work used resources of the Deutsches Klimarechenzentrum (DKRZ) granted by its Scientific Steering Committee (WLA). Contribution to the Centrum für Erdsystemforschung und Nachhaltigkeit (CEN) of Universität Hamburg.

References

Abarbanel, H. D., Creveling, D. R., Farsian, R., and Kostuk, M.: Dynamical state and parameter estimation, SIAM Journal on Applied Dynamical Systems, 8, 1341–1381, 2009.

Abarbanel, H. D., Kostuk, M., and Whartenby, W.: Data assimilation with regularized nonlinear instabilities, Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, 136, 769–783, 2010.

Allaire, G.: A review of adjoint methods for sensitivity analysis, uncertainty quantification and optimization in numerical codes, Ingénieurs de l'Automobile, 836, 33–36, 2015.

Bar-Shalom, Y., Li, X. R., and Kirubarajan, T.: Estimation with applications to tracking and navigation: theory algorithms and software, John Wiley & Sons, 2004.

Bertino, L., Evensen, G., and Wackernagel, H.: Sequential data assimilation techniques in oceanography, International Statistical Review, 71, 223–241, 2003.

Bradbury, J., Frostig, R., Hawkins, P., Johnson, M. J., Leary, C., Maclaurin, D., Necula, G., Paszke, A., VanderPlas, J., Wanderman-Milne, S., and Zhang, Q.: JAX: composable transformations of Python+NumPy programs, http://github.com/google/jax, 2018.

Cameron, M. and Yang, S.: Computing the quasipotential for highly dissipative and chaotic SDEs an application to stochastic Lorenz'63, Communications in Applied Mathematics and Computational Science, 14, 207–246, 2019.

Cummings, J. A. and Smedstad, O. M.: Variational data assimilation for the global ocean, in: Data assimilation for atmospheric, oceanic and hydrologic applications (Vol. II), pp. 303–343, Springer, 2013.

Daron, J. and Stainforth, D. A.: On quantifying the climate of the nonautonomous Lorenz-63 model, Chaos: An Interdisciplinary Journal of Nonlinear Science, 25, 043 103, 2015.

Dembinski, H. and et al., P. O.: scikit-hep/iminuit, https://doi.org/10.5281/zenodo.3949207, 2020.

Desroziers, G., Camino, J.-T., and Berre, L.: 4DEnVar: link with 4D state formulation of variational assimilation and different possible implementations, Quarterly Journal of the Royal Meteorological Society, 140, 2097–2110, 2014.

Du, Y. J. and Shiue, M.-C.: Analysis and computation of continuous data assimilation algorithms for Lorenz 63 system based on nonlinear nudging techniques, Journal of Computational and Applied Mathematics, 386, 113 246, 2021.

Errico, R. M.: What is an adjoint model?, Bulletin of the American Meteorological Society, 78, 2577–2592, 1997.

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

Evensen, G.: The ensemble Kalman filter: Theoretical formulation and practical implementation, Ocean dynamics, 53, 343–367, 2003.

Evensen, G., Vossepoel, F. C., and Van Leeuwen, P. J.: Data assimilation fundamentals: A unified formulation of the state and parameter estimation problem, Springer Nature, 2022.

Fisher, M., Tremolet, Y., Auvinen, H., Tan, D., and Poli, P.: Weak-constraint and long-window 4D-Var, ECMWF Technical Memoranda, 655, 47, 2011.

Gauthier, P.: Chaos and quadri-dimensional data assimilation: A study based on the Lorenz model, Tellus A: Dynamic Meteorology and Oceanography, 44, 2–17, 1992.

Giering, R. and Kaminski, T.: Recipes for adjoint code construction, ACM Transactions on Mathematical Software (TOMS), 24, 437–474, 1998.

520 Goodliff, M., Amezcua, J., and Van Leeuwen, P. J.: Comparing hybrid data assimilation methods on the Lorenz 1963 model with increasing non-linearity, Tellus A: Dynamic Meteorology and Oceanography, https://doi.org/10.3402/tellusa.v67.26928, 2015.

Goodliff, M., Fletcher, S., Kliewer, A., Forsythe, J., and Jones, A.: Detection of Non-Gaussian Behavior Using Machine Learning Techniques: A Case Study on the Lorenz 63 Model, Journal of Geophysical Research: Atmospheres, 125, e2019JD031 551, 2020.

Gustafsson, N., Källén, E., and Thorsteinsson, S.: Sensitivity of forecast errors to initial and lateral boundary conditions, Tellus A: Dynamic
525 Meteorology and Oceanography, 50, 167–185, 1998.

Gustafsson, N., Berre, L., Hörnquist, S., Huang, X.-Y., Lindskog, M., Navascues, B., Mogensen, K., and Thorsteinsson, S.: Three-dimensional variational data assimilation for a limited area model: Part I: General formulation and the background error constraint, Tellus A, 53, 425–446, 2001.

Hall, M. C.: Application of adjoint sensitivity theory to an atmospheric general circulation model, Journal of Atmospheric Sciences, 43,
530 2644–2652, 1986.

Hall, M. C. and Cacuci, D. G.: Physical interpretation of the adjoint functions for sensitivity analysis of atmospheric models, Journal of the atmospheric sciences, 40, 2537–2546, 1983.

Hall, M. C., Cacuci, D. G., and Schlesinger, M. E.: Sensitivity analysis of a radiative-convective model by the adjoint method, Journal of Atmospheric Sciences, 39, 2038–2050, 1982.

535 Hascoet, L. and Pascual, V.: The Tapenade automatic differentiation tool: Principles, model, and specification, ACM Transactions on Mathematical Software (TOMS), 39, 1–43, 2013.

Hirsch, M. W., Smale, S., and Devaney, R. L.: 14 - The Lorenz System, in: Differential Equations, Dynamical Systems, and an Introduction to Chaos (Third Edition), edited by Hirsch, M. W., Smale, S., and Devaney, R. L., pp. 305–328, Academic Press, Boston, third edition edn., ISBN 978-0-12-382010-5, https://doi.org/https://doi.org/10.1016/B978-0-12-382010-5.00014-2, 2013.

540 Houtekamer, P. L. and Mitchell, H. L.: A sequential ensemble Kalman filter for atmospheric data assimilation, Monthly weather review, 129, 123–137, 2001.

Huai, X.-W., Li, J.-P., Ding, R.-Q., Feng, J., and Liu, D.-Q.: Quantifying local predictability of the Lorenz system using the nonlinear local Lyapunov exponent, Atmospheric and Oceanic Science Letters, 10, 372–378, 2017.

James, F. and Roos, M.: Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations, Comput. Phys.
545 Commun., 10, 343–367, https://doi.org/10.1016/0010-4655(75)90039-9, 1975.

Kalman, R. E.: A New Approach to Linear Filtering and Prediction Problems, Journal of Basic Engineering, 82, 35–45, https://doi.org/10.1115/1.3662552, 1960.

Köhl, A.: Evaluating the GECCO3 1948–2018 ocean synthesis–a configuration for initializing the MPI-ESM climate model, Quarterly Journal of the Royal Meteorological Society, 146, 2250–2273, 2020.

550 Köhl, A. and Willebrand, J.: An adjoint method for the assimilation of statistical characteristics into eddy-resolving ocean models, Tellus A: Dynamic meteorology and oceanography, 54, 406–425, 2002.

Kravtsov, S. and Tsonis, A. A.: Lorenz-63 Model as a Metaphor for Transient Complexity in Climate, Entropy, 23, 951, 2021.

Laboratory, N. R.: Global HYCOM+CICE 1/12 degree page, https://www7320.nrlssc.navy.mil/GLBhycomcice1-12/, 2016.

Le Dimet, F.-X. and Talagrand, O.: Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects,
555 Tellus A: Dynamic Meteorology and Oceanography, 38, 97–110, 1986.

Lea, D. J., Allen, M. R., and Haine, T. W.: Sensitivity analysis of the climate of a chaotic system, Tellus A: Dynamic Meteorology and Oceanography, 52, 523–532, 2000.

Lorenz, E. N.: Deterministic nonperiodic flow, Journal of atmospheric sciences, 20, 130–141, 1963.

Lyu, G., Köhl, A., Matei, I., and Stammer, D.: Adjoint-based climate model tuning: Application to the planet simulator, Journal of Advances in Modeling Earth Systems, 10, 207–222, 2018.

Marotzke, J., Giering, R., Zhang, K. Q., Stammer, D., Hill, C., and Lee, T.: Construction of the adjoint MIT ocean general circulation model and application to Atlantic heat transport sensitivity, Journal of Geophysical Research: Oceans, 104, 29 529–29 547, 1999.

Marzban, C.: Variance-based sensitivity analysis: An illustration on the Lorenz'63 model, Monthly Weather Review, 141, 4069–4079, 2013.

Mauritsen, T., Stevens, B., Roeckner, E., Crueger, T., Esch, M., Giorgetta, M., Haak, H., Jungclaus, J., Klocke, D., Matei, D., et al.: Tuning the climate of a global model, Journal of advances in modeling Earth systems, 4, 2012.

Miller, R. N., Ghil, M., and Gauthiez, F.: Advanced data assimilation in strongly nonlinear dynamical systems, Journal of Atmospheric Sciences, 51, 1037–1056, 1994.

Navon, I. M.: Data assimilation for numerical weather prediction: a review, Data assimilation for atmospheric, oceanic and hydrologic applications, pp. 21–65, 2009.

Nichols, N. K.: Mathematical Concepts of Data Assimilation, pp. 13–39, Springer Berlin Heidelberg, Berlin, Heidelberg, ISBN 978-3-540-74703-1, https://doi.org/10.1007/978-3-540-74703-1_2, 2010.

pandas development team, T.: pandas-dev/pandas: Pandas, https://doi.org/10.5281/zenodo.3509134, 2020.

Pasini, A. and Pelino, V.: Can we estimate atmospheric predictability by performance of neural network forecasting? The toy case studies of unforced and forced Lorenz models, in: CIMSA. 2005 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications, 2005., pp. 69–74, IEEE, 2005.

Pelino, V. and Maimone, F.: Energetics, skeletal dynamics, and long-term predictions on Kolmogorov-Lorenz systems, Physical Review E, 76, 046 214, 2007.

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

Quinn, J. C., Bryant, P. H., Creveling, D. R., Klein, S. R., and Abarbanel, H. D.: Parameter and state estimation of experimental chaotic systems using synchronization, Physical Review E—Statistical, Nonlinear, and Soft Matter Physics, 80, 016 201, 2009.

Rabier, F. and Liu, Z.: Variational data assimilation: theory and overview, in: Proc. ECMWF Seminar on Recent Developments in Data Assimilation for Atmosphere and Ocean, Reading, UK, September 8–12, pp. 29–43, 2003.

Ruiz, J. J., Pulido, M., and Miyoshi, T.: Estimating model parameters with ensemble-based data assimilation: A review, Journal of the Meteorological Society of Japan. Ser. II, 91, 79–99, 2013.

Simon, D.: Optimal state estimation: Kalman, H infinity, and nonlinear approaches, John Wiley & Sons, 2006.

Stammer, D., Balmaseda, M., Heimbach, P., Köhl, A., and Weaver, A.: Ocean data assimilation in support of climate applications: status and perspectives, Annual review of marine science, 8, 491–518, 2016.

Stammer, D., Köhl, A., Vlasenko, A., Matei, I., Lunkeit, F., and Schubert, S.: A pilot climate sensitivity study using the CEN coupled adjoint model (CESAM), Journal of Climate, 31, 2031–2056, 2018.

Stensrud, D. J. and Bao, J.-W.: Behaviors of variational and nudging assimilation techniques with a chaotic low-order model, Monthly weather review, 120, 3016–3028, 1992.

Sugiura, N., Masuda, S., Fujii, Y., Kamachi, M., Ishikawa, Y., and Awaji, T.: A framework for interpreting regularized state estimation, Monthly Weather Review, 142, 386–400, 2014.

Talagrand, O.: Variational assimilation, Data assimilation: making sense of observations, pp. 41–67, 2010.

Tandeo, P., Ailliot, P., Ruiz, J., Hannart, A., Chapron, B., Cuzol, A., Monbet, V., Easton, R., and Fablet, R.: Combining analog method and ensemble data assimilation: application to the Lorenz-63 chaotic system, in: Machine Learning and Data Mining Approaches to Climate Science: proceedings of the 4th International Workshop on Climate Informatics, pp. 3–12, Springer, 2015.

Tett, S. F., Yamazaki, K., Mineter, M. J., Cartis, C., and Eizenberg, N.: Calibrating climate models using inverse methods: case studies with HadAM3, HadAM3P and HadCM3, Geoscientific Model Development, 10, 3567–3589, 2017.

Tippett, M. K. and Chang, P.: Some theoretical considerations on predictability of linear stochastic dynamics, Tellus A: Dynamic Meteorology and Oceanography, 55, 148–157, 2003.

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

Tremolet, Y.: Accounting for an imperfect model in 4D-Var, Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, 132, 2483–2504, 2006.

Van Der Merwe, R. and Wan, E. A.: The square-root unscented Kalman filter for state and parameter-estimation, in: 2001 IEEE international conference on acoustics, speech, and signal processing. Proceedings (Cat. No. 01CH37221), vol. 6, pp. 3461–3464, IEEE, 2001.

Waskom, M. L.: seaborn: statistical data visualization, Journal of Open Source Software, 6, 3021, https://doi.org/10.21105/joss.03021, 2021.

Wunsch, C.: The ocean circulation inverse problem, Cambridge University Press, 1996.

Wunsch, C. and Heimbach, P.: Estimated Decadal Changes in the North Atlantic Meridional Overturning Circulation and Heat Flux 1993–2004, Journal of Physical Oceanography, 36, 2012 – 2024, https://doi.org/https://doi.org/10.1175/JPO2957.1, 2006.

Yang, S.-C., Baker, D., Li, H., Cordes, K., Huff, M., Nagpal, G., Okereke, E., Villafane, J., Kalnay, E., and Duane, G. S.: Data assimilation as synchronization of truth and model: Experiments with the three-variable Lorenz system, Journal of the atmospheric sciences, 63, 2340–2354, 2006.

Yin, X., Wang, B., Liu, J., and Tan, X.: Evaluation of conditional non-linear optimal perturbation obtained by an ensemble-based approach using the Lorenz-63 model, Tellus A: Dynamic Meteorology and Oceanography, 66, 22 773, 2014.

Zou, X., Navon, I., and Ledimet, F.: An optimal nudging data assimilation scheme using parameter estimation, Quarterly Journal of the Royal Meteorological Society, 118, 1163–1186, 1992.