A multi-model ensemble Kalman filter for data assimilation and forecasting

JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2022MS003123

Special Section:
Data assimilation for Earth system models

A Multi-Model Ensemble Kalman Filter for Data Assimilation and Forecasting

Eviatar Bach[1,2] and **Michael Ghil**[1,3]

[1]Geosciences Department and Laboratoire de Météorologie Dynamique (CNRS and IPSL), École Normale Supérieure and PSL University, Paris, France, [2]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA, [3]Department of Atmospheric and Oceanic Sciences, University of California at Los Angeles, Los Angeles, CA, USA

Key Points:

- Multiple models and observations can be optimally combined for data assimilation (DA) and forecasting using multi-model DA
- We formulate a multi-model ensemble Kalman filter (MM-EnKF), which incorporates model error and is appropriate for high-dimensional models
- Using numerical experiments, we show that the MM-EnKF can significantly outperform the best model and an unweighted multi-model ensemble

Correspondence to:
E. Bach,
eviatarbach@protonmail.com

Citation:
Bach, E., & Ghil, M. (2023). A multi-model ensemble Kalman filter for data assimilation and forecasting. *Journal of Advances in Modeling Earth Systems*, *15*, e2022MS003123. https://doi.org/10.1029/2022MS003123

Received 5 APR 2022
Accepted 28 DEC 2022
Corrected 13 FEB 2023

This article was corrected on 13 FEB 2023. See the end of the full text for details.

Abstract Data assimilation (DA) aims to optimally combine model forecasts and observations that are both partial and noisy. Multi-model DA generalizes the variational or Bayesian formulation of the Kalman filter, and we prove that it is also the minimum variance linear unbiased estimator. Here, we formulate and implement a multi-model ensemble Kalman filter (MM-EnKF) based on this framework. The MM-EnKF can combine multiple model ensembles for both DA and forecasting in a flow-dependent manner; it uses adaptive model error estimation to provide matrix-valued weights for the separate models and the observations. We apply this methodology to various situations using the Lorenz96 model for illustration purposes. Our numerical experiments include multiple models with parametric error, different resolved scales, and different fidelities. The MM-EnKF results in significant error reductions compared to the best model, as well as to an unweighted multi-model ensemble, with respect to both probabilistic and deterministic error metrics.

Plain Language Summary Forecasts that combine multiple imperfect models of a system are used in many fields, including the physical, natural and socio-economic sciences. In particular, data assimilation (DA), the process by which observations are integrated with model forecasts, is critical in the prediction of chaotic systems. Multi-model DA (MM-DA) unifies multi-model forecast combination and DA into a single process. Here, we significantly improve on previous formulations of MM-DA by accounting for model error, and formulate a multi-model ensemble Kalman filter appropriate for high-dimensional systems.

1. Introduction

Combining multiple forecasts from imperfect models of reality can often lead to forecasts that are better than any single model. Such multi-model forecasts have been enormously successful in weather and climate prediction (Hagedorn et al., 2005; Krishnamurti et al., 2016); economics (Clemen, 1989); epidemiological forecasting, including that of COVID-19 (Cramer et al., 2022); hydrology (Okuno et al., 2019; Xue & Zhang, 2014); tracking and navigation (Bar-Shalom et al., 2001); space weather (Schunk et al., 2016); air quality forecasting (Mallet et al., 2009; Sengupta et al., 2020); and numerous other application areas (Clemen, 1989; Fragoso et al., 2018; Trenkler & Gotu, 1998).

Data assimilation (DA) is the process of combining model forecasts with observations to obtain a state estimate of a system. DA is an essential part of forecasting in a wide variety of scientific and engineering fields, with most methods based nowadays on the Kalman filter (Asch et al., 2016). In the meteorological literature, the need for methods to take noisy, possibly sparse observations and produce an initial condition suitable for a numerical model has been recognized since the first numerical weather forecast (Panofsky, 1949), and the Kalman filter was proposed for this purpose in Ghil et al. (1981). Ensemble Kalman filters (EnKFs), which approximate the evolution of the probability distribution using a Monte Carlo approach (Evensen, 2003), have become popular for geophysical and other problems (Hamilton et al., 2016).

In this paper, we consider multi-model DA (MM-DA), a generalization of DA which allows for multiple forecast models (Narayan et al., 2012). MM-DA combines multiple forecasts and observations, bridging the literature on forecast combination with that on DA. In this paper we make several contributions to the existing literature on MM-DA: (a) we incorporate and estimate model error in MM-DA, allowing models that are less accurate to have a lower weight, and allow the weights to differ for different variables; (b) we formulate several possible implementations of deterministic EnKFs for MM-DA (MM-EnKFs) and discuss computational issues; (c) we provide

an open-source software implementation of MM-EnKFs; (d) we test MM-EnKF with chaotic models for DA and forecasting in various scenarios; and, finally, (e) we prove linear minimum variance optimality of MM-DA.

The paper is laid out as follows. In Sections 1.1 and 1.2, we review the literature on combining forecasts and on MM-DA, respectively. In Section 2, we discuss the development and implementation of MM-EnKFs. In Section 4, we apply MM-EnKFs to chaotic systems. In Section 5, we draw conclusions and provide an outlook, including on applications to data-driven models. Finally, in Appendix A, we prove optimality of MM-DA in the linear minimum variance sense, and in Appendix B we detail the simple model error estimation method used in the numerical experiments.

1.1. Combining Multiple Model Forecasts

Bates and Granger (1969) were among the first to combine multiple distinct forecasts. They weighted multiple univariate forecasts according to past performance, and showed that the combined forecast resulted in lower error. Combining forecasts has become an important topic in statistics; see the historical overview Hoeting et al. (1999) and the bibliographies Clemen (1989), Trenkler and Gotu (1998).

The advantage of multi-model forecasts over single-model ones, at comparable total ensemble size, comes from distinct models having different model errors. The skill of the multi-model forecast will then be improved to the extent that the model errors compensate for each other (Hagedorn et al., 2005). Furthermore, when the multi-model forecast is probabilistic, these multiple model errors may lead to better spanning the true forecast uncertainty (Wilks, 2019). The need for weighting comes when some models have higher skill than others, implying that the former should have higher weight in the combined forecast than the latter. The general situation, however, is that one model may not be superior to all the others in all respects. More typically, some models may have superior skill in some variables, in the representation of particular processes, or at some forecast horizons. We discuss different weighting approaches in the following section.

1.1.1. Weighting Distinct Forecasts

Several approaches for weighting distinct model forecasts have been developed, with Bayesian model averaging (Hoeting et al., 1999) being one of the most common ones. This methodology estimates the posterior model probabilities based on past data, and assigns the models scalar weights based on these probabilities. Dynamic versions have also been developed, to allow the weights to evolve based on current conditions. Both these methods, as well as several others discussed in Narayan et al. (2012), are limited to scalar weights.

Other methods have been developed in the context of atmospheric prediction. Here, the models are high-dimensional, and each model may produce an ensemble of forecasts that attempts to capture the predictive uncertainty. Multi-model ensembles (MMEs), where multiple models are combined into a single ensemble without weighting, are used widely in climate prediction (Hagedorn et al., 2005). Multi-model superensembles, which weight the distinct model ensembles based on weights determined by multiple linear regression, have also been widely adopted (Krishnamurti et al., 2016).

The Dynamic Integrated Forecast System, developed by the National Center for Atmospheric Research, periodically nudges model weights in the direction of error decrease (Myers et al., 2011). Cross-pollination in time (CPT) uses the forecasts of each model as initial conditions for the other models, along with some pruning rule to avoid an exponential increase of trajectories with time (Du & Smith, 2017; Schevenhoven & Selten, 2017). In a connected supermodel, each model is nudged toward the others by introducing coupling terms in the evolution equations, and the supermodel is taken as an average of these coupled models (Duane et al., 2017; Selten et al., 2017). In a weighted supermodel, the individual models are not directly connected through coupling terms; rather, the supermodel tendency is taken to be a weighted average of the individual model tendencies, and the individual models compute their tendencies based on the supermodel state (Schevenhoven et al., 2019). CPT and weighted supermodels were compared by Schevenhoven et al. (2019) and Schevenhoven and Carrassi (2022).

Sengupta et al. (2020) used a Bayesian neural network to infer model weights. Sequential aggregation takes inspiration from online learning and game theory in weighting forecasts with rules that have theoretical performance guarantees (Gonzalez et al., 2021; Mallet et al., 2009; Thorey et al., 2017). Forecast weights can also be determined using Markov chain Monte Carlo (Dumont Le Brazidec et al., 2021). Several other methods were compared for meteorological applications in Young (2002), Gerding and Myers (2003).

1.1.2. DA With Multiple Models

Several methods have been proposed to weight models using DA methods, in particular relying on Kalman or particle filters. Anandalingam and Chen (1989) first recognized that a particular Bayesian forecast combination problem was equivalent to a Kalman filter. Du and Smith (2017) used DA in addition to CPT in combining forecasts. Chen and Stechmann (2019) and Counillon et al. (2022) used DA to synchronize distinct models by assimilating forecasts as pseudo-observations. Multiple parametric variations, or variations in physical parameterizations in an atmospheric model, have also been used in EnKFs without weighting, in order to capture the effect of model error (Houtekamer & Zhang, 2016; Wu et al., 2008).

Here, we are interested in the problem of generating an optimal state estimate using multiple model forecasts and observations. Section 10.2 in Simon (2006) proposes to run a Kalman filter for each model, and estimate its conditional probability given the observations from the innovations. These probabilities are then used as weights in combining the model forecasts. This approach is similar to the interacting multiple model (Bar-Shalom et al., 2001) filter, popular in tracking applications, and multiple model adaptive estimation (Akca & Efe, 2019).

Xue and Zhang (2014) combines a Bayesian model averaging approach with an EnKF, by recomputing the ensemble weights as new observations arrive. Coelho et al. (2015) estimates model weights using a separate filter. Otsuka and Miyoshi (2015) implements a multi-model EnKF by adjusting the number of ensemble members for each model at every cycle based on a Bayesian estimate of the model's probability. In the ensemble average, the model with more ensemble members is then weighted more heavily. Mallet (2010) combines the sequential aggregation approach with DA.

1.2. Multi-Model Data Assimilation (MM-DA)

In this paper, we consider a generalization of the Kalman filter formulation to multiple models. This generalization differs from the methods in the previous paragraphs in the models' and the observations' weights being determined as part of the filtering process itself, instead of being estimated separately. MM-DA, proposed by Logutov and Robinson (2005) and Narayan et al. (2012), is based on the variational or Bayesian formalisms from which the Kalman filter and related methods are derived, except that multiple models are included.

The MM-DA formulation was perhaps first studied by Logutov and Robinson (2005), who also proposed an expectation maximization algorithm for estimating the forecast error parameters along with the state estimate. The connection to the Kalman filter was not explicitly made in Logutov and Robinson (2005). The same formulation was independently developed by Narayan et al. (2012), who showed that it can be implemented by using an iterative method. We base our exposition on Narayan et al. (2012), but using the common DA notation of Ide et al. (1997); see Table 1 for a definition of symbols.

Suppose we have M models, with each model m having its own forecast state $\mathbf{x}_m^{\mathrm{f}} \in \mathbb{R}^{n_m}$ with forecast error covariance matrix $\mathbf{P}_m^{\mathrm{f}}$. Each model is assumed to be unbiased. One has to choose a space for the multi-model forecasts to reside in; for example, a given spatial grid in the case of an atmospheric model. We take this to be the space of one of the model states—although this is not necessary—and refer to this model as the reference model m_r; its choice will be discussed later.

We then define the operators $\mathbf{G}_m : \mathbb{R}^{n_{m_r}} \to \mathbb{R}^{n_m}$ which map from the reference model space to the model space of model m with dimension n_m. Clearly, $\mathbf{G}_{m_r} = \mathbf{I}$ and we assume for the moment that these operators are linear, although this assumption can be relaxed later.

We also have a p-dimensional observation vector $\mathbf{y}$ with observation error covariance matrix R. We define the observation operator $\mathbf{H} : \mathbb{R}^{n_{m_r}} \to \mathbb{R}^p$, which maps from the reference model space to the observation space.

Each model's state evolution operator is denoted by $\mathbf{M}_m$, and it is also assumed to be linear for the moment. Later, the nonlinear state evolution operator will be denoted by $\mathcal{M}_m$.

1.2.1. Variational Formulation and Direct Solution

1.2.1.1. The Formulation

For a single model forecast $\mathbf{x}^{\mathrm{f}}$ with covariance matrix $\mathbf{P}^{\mathrm{f}}$, the variational formulation of the optimal state estimation problem defines a cost function $\mathcal{I}[\mathbf{x}]$ for a control variable $\mathbf{x}$ as

Table 1
Definition of Symbols

$\mathbf{b}_m$	Bias of mth model
$\mathbf{B}$	Climatological forecast error covariance matrix
$\{\gamma, \delta\}$	Smoothing parameter for {inflation, model error} estimation
$\mathbf{E}^{\{\mathrm{f,a}\}}$	{forecast, analysis} ensemble
$\mathbf{E}^{\mathrm{f}'}_{1:m}$	Multi-model forecast ensemble after averaging over models 1 to m
$\{\mathbf{G}_m, \mathcal{G}_m\}$	{linear, nonlinear} mapping from reference model space m_r to the space of model m
$\left\{\mathbf{G}_{m_1\to m_2}, \mathcal{G}_{m_1\to m_2}\right\}$	{linear, nonlinear} mapping from space of model m_1 to space of model m_2
$\{\mathbf{H}, \mathcal{H}\}$	{linear, nonlinear} observation operator of reference model
$\{\mathbf{H}_m, \mathcal{H}_m\}$	{linear, nonlinear} observation operator of mth model
$\mathbf{K}$	Gain matrix
m_r	Reference model
M	Number of models
$\{\mathbf{M}, \mathcal{M}\}$	{linear, nonlinear} forecast model
n_m	Dimension of mth model
N_m	Ensemble size of mth model
$\boldsymbol{\rho}$	Localization matrix
$\mathbf{P}^{\{\mathrm{f,a}\}}$	{forecast, analysis} error covariance
$\mathbf{P}^{\mathrm{f}'}_{1:m}$	Multi-model forecast error covariance after averaging over models 1 to m
$\mathbf{Q}$	Model error covariance
$\mathbf{R}$	Observation error covariance
$\mathbf{x}^{\{\mathrm{t,f,a}\}}$	{true, forecast, analysis} state
$\mathbf{x}^{\mathrm{f}'}_{1:m}$	Multi-model forecast state after averaging over models 1 to m
$\left(\mathbf{x}^{\{\mathrm{f,a}\}}\right)_i$	ith member of {forecast, analysis} ensemble
$\overline{\mathbf{x}}^{\{\mathrm{f,a}\}}$	Mean of {forecast, analysis} ensemble
$\mathbf{X}^{\{\mathrm{f,a}\}}$	{forecast, analysis} ensemble anomalies
$\mathbf{y}$	Observation

Note. The superscript convention follows Ide et al. (1997).

$$\mathcal{I}[\mathbf{x}] = \|\mathbf{x} - \mathbf{x}^{\mathrm{f}}\|^2_{(\mathbf{P}^{\mathrm{f}})^{-1}} + \|\mathbf{H}\mathbf{x} - \mathbf{y}\|^2_{\mathbf{R}^{-1}}. \tag{1}$$

Here we use the short-hand weighted-norm notation $\|\mathbf{v}\|^2_{\mathbf{A}} \equiv \mathbf{v}^T\mathbf{A}\mathbf{v}$ for a quadratic form with symmetric positive semidefinite matrix $\mathbf{A}$. This cost function measures the sum of the squared Mahalanobis distances of $\mathbf{x}$ from the forecast $\mathbf{x}^{\mathrm{f}}$ and the observations $\mathbf{y}$. The minimizer of Equation 1 is the assimilation step of the Kalman filter.

Equation 1 can be generalized to multiple models as

$$\mathcal{J}[\mathbf{x}] = \sum_{m=1}^{M} \|\mathbf{G}_m\mathbf{x} - \mathbf{x}^{\mathrm{f}}_m\|^2_{(\mathbf{P}^{\mathrm{f}}_m)^{-1}} + \|\mathbf{H}\mathbf{x} - \mathbf{y}\|^2_{\mathbf{R}^{-1}}. \tag{2}$$

Note that this generalization implicitly assumes that the forecast errors are mutually uncorrelated; see Section 5.2.1 for more details.

1.2.1.2. The Direct Solution

Minimizing the multi-model cost function $\mathcal{J}[\mathbf{x}]$ above gives the analysis solution $\mathbf{x}^{\mathrm{a}}$ and its corresponding covariance matrix $\mathbf{P}^{\mathrm{a}}$ for the multi-model Kalman filter (Narayan et al., 2012):

$$\mathbf{x}^{\mathrm{a}} = \mathbf{P}^{\mathrm{a}}\left(\sum_{m=1}^{M} \mathbf{G}_m^T \left(\mathbf{P}_m^{\mathrm{f}}\right)^{-1} \mathbf{x}_m^{\mathrm{f}} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{y}\right), \tag{3}$$

where $(\cdot)^T$ is the transposition operator and

$$\mathbf{P}^{\mathrm{a}} = \left(\sum_{m=1}^{M} \mathbf{G}_m^T \left(\mathbf{P}_m^{\mathrm{f}}\right)^{-1} \mathbf{G}_m + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}\right)^{-1}. \tag{4}$$

The solution is thus a weighted mean, where the weights for each model m are inversely proportional to $\mathbf{P}_m^{\mathrm{f}}$, and the weight of the observations is inversely proportional to $\mathbf{R}$. Note that if we set $M = 1$, we recover the regular Kalman filter equations.

The analysis $\mathbf{x}^{\mathrm{a}}$ will be in the reference model space. The analysis in the model space for $m \neq m_r$ can then be obtained by computing

$$\mathbf{x}_m^{\mathrm{a}} = \mathbf{G}_m \mathbf{x}^{\mathrm{a}}, \tag{5}$$

and the analysis covariance matrix as

$$\mathbf{P}_m^{\mathrm{a}} = \mathbf{G}_m \mathbf{P}^{\mathrm{a}} \mathbf{G}_m^T. \tag{6}$$

For the next forecast cycle, each model is applied to the analysis state:

$$\mathbf{x}_m^{\mathrm{f}}(t_{i+1}) = \mathbf{M}_m(t_i)\mathbf{x}_m^{\mathrm{a}}(t_i), \tag{7}$$

and the covariance is propagated according to

$$\mathbf{P}_m^{\mathrm{f}}(t_{i+1}) = \mathbf{M}_m(t_i)\mathbf{P}_m^{\mathrm{a}}(t_i)\mathbf{M}_m(t_i)^T + \mathbf{Q}_m(t_i), \tag{8}$$

where $\mathbf{Q}_m$ is the model error covariance matrix for model m. We have introduced explicit time dependence here for clarity. In Figure 1, we show a schematic diagram of single- and multi-model assimilation–forecast cycles. The model error is discussed in greater detail in Section 2.1.1.

Although here we considered the variational formulation, the same equations for the multi-model Kalman filter can also be derived from the Bayesian formulation of the problem (Logutov & Robinson, 2005; Narayan et al., 2012).

The single-model Kalman filter is the optimal linear filter in the sense of being the minimum variance unbiased estimator. It has not previously been shown, though, that the multi-model Kalman filter is optimal in terms of minimizing variance, and we prove this in Appendix A.

1.2.2. Iterative Solution

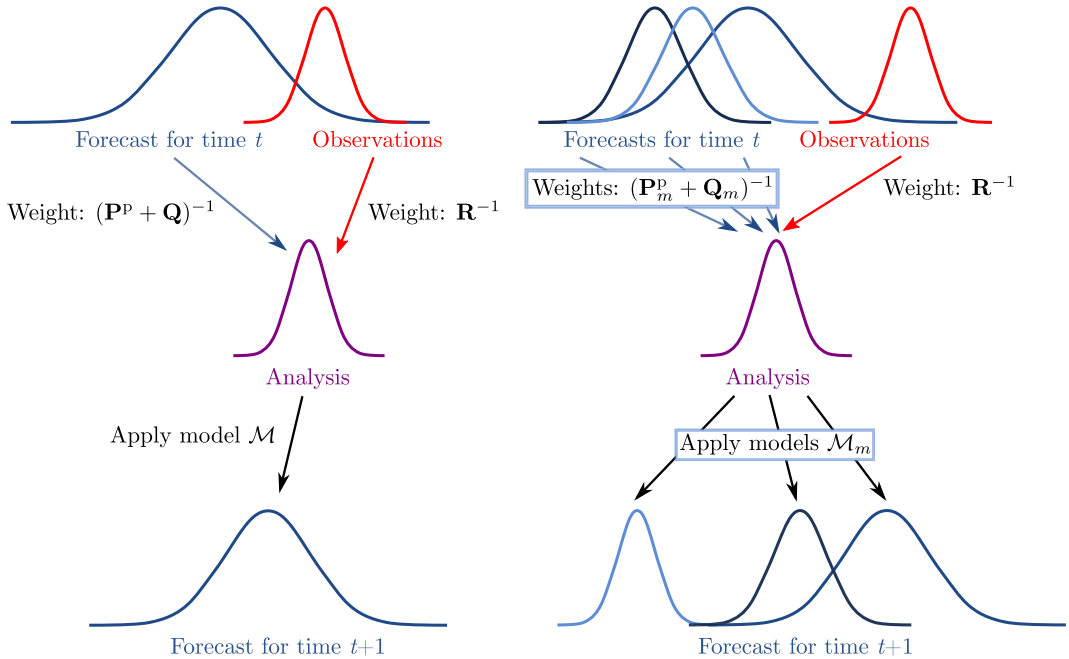
In some cases, it may be possible to directly compute the right-hand side of Equation 3 or to minimize Equation 2 using an approach similar to the three-dimensional variational algorithm (3D-Var) (Asch et al., 2016; Kalnay, 2002). Narayan et al. (2012) show, instead, how to solve the problem iteratively.

In this iteration, the analysis of the previous model $m - 1$ is considered as the forecast for the subsequent model m, and the forecast of model m is considered as an observation.

$$\mathbf{K}_m = \mathbf{P}^{\mathrm{f}'}_{1:m-1} \mathbf{G}_m^T \left(\mathbf{G}_m \mathbf{P}^{\mathrm{f}'}_{1:m-1} \mathbf{G}_m^T + \mathbf{P}_m^{\mathrm{f}}\right)^{\dagger}, \tag{9a}$$

$$\mathbf{x}^{\mathrm{f}'}_{1:m} = \mathbf{x}^{\mathrm{f}'}_{1:m-1} + \mathbf{K}_m \left(\mathbf{x}_m^{\mathrm{f}} - \mathbf{G}_m \mathbf{x}^{\mathrm{f}'}_{1:m-1}\right), \tag{9b}$$

$$\mathbf{P}^{\mathrm{f}'}_{1:m} = (\mathbf{I} - \mathbf{K}_m \mathbf{G}_m)\mathbf{P}^{\mathrm{f}'}_{1:m-1}; \tag{9c}$$



(a) Single-model data assimilation (b) Multi-model data assimilation

Figure 1. Schematics of a forecast–assimilation cycle for single- and multi-model data assimilation. For the purposes of this figure, we assume $\mathbf{H} = \mathbf{G}_m = \mathbf{I}$.

here † indicates the Moore–Penrose pseudoinverse, $\mathbf{x}^{\mathrm{f}'}_{1:m}$ indicates the combined forecast of models 1 to m, and $\mathbf{P}^{\mathrm{f}'}_{1:m}$ indicates the forecast covariance matrix of $\mathbf{x}^{\mathrm{f}'}_{1:m}$. Once done with the M models, one assimilates the actual observations.

$$\mathbf{K} = \mathbf{P}^{\mathrm{f}'}_{1:M}\mathbf{H}\left(\mathbf{H}\mathbf{P}^{\mathrm{f}'}_{1:M}\mathbf{H}^T + \mathbf{R}\right)^{\dagger}, \tag{10a}$$

$$\mathbf{x}^{\mathrm{a}} = \mathbf{x}^{\mathrm{f}'}_{1:M} + \mathbf{K}\left(\mathbf{y} - \mathbf{H}\mathbf{x}^{\mathrm{f}'}_{1:M}\right), \tag{10b}$$

$$\mathbf{P}^{\mathrm{a}} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^{\mathrm{f}'}_{1:M}. \tag{10c}$$

When the covariance matrices $\mathbf{P}^{\mathrm{f}}_m$ and $\mathbf{R}$ are positive definite, the iterative solution is equivalent to the direct solution. However, unlike the direct solution, the iterative solution allows the covariance matrices $\mathbf{P}^{\mathrm{f}}_m$ and $\mathbf{R}$ to be singular. The iterative solution can be shown to be independent of the order in which the models and observations are assimilated, as long as there are no inconsistent zero-variance components (Narayan et al., 2012).

Importantly, the iterative procedure suggests a way to use single-model DA methods to estimate a solution to the MM-DA problem. Notice that Equations 9a–9c and 10a–10c are the assimilation step of a single-model Kalman filter, and thus they can be replaced by any single-model DA method.

Assume now that we have a DA method that takes as input the forecast state $\mathbf{x}^{\mathrm{f}}$, forecast error covariance $\mathbf{P}^{\mathrm{f}}$, observation vector $\mathbf{y}$, observation error covariance $\mathbf{R}$, and observation operator $\mathcal{H}$, and returns as output the analysis state $\mathbf{x}^{\mathrm{a}}$ and analysis error covariance $\mathbf{P}^{\mathrm{a}}$. Denote this function by $(\mathbf{x}^{\mathrm{a}}, \mathbf{P}^{\mathrm{a}}) = \mathcal{F}_{\mathrm{DA}}\left(\mathbf{x}^{\mathrm{f}}, \mathbf{P}^{\mathrm{f}}, \mathbf{y}, \mathbf{R}, \mathcal{H}\right)$. Then, for $m = 2, \ldots, M$:

$$\left(\mathbf{x}^{\mathrm{f}'}_{1:m}, \mathbf{P}^{\mathrm{f}'}_{1:m}\right) = \mathcal{F}_{\mathrm{DA}}\left(\mathbf{x}^{\mathrm{f}'}_{1:m-1}, \mathbf{P}^{\mathrm{f}'}_{1:m-1}, \mathbf{x}^{\mathrm{f}}_m, \mathbf{P}^{\mathrm{f}}_m, \mathcal{G}_m\right), \tag{11}$$

and $\mathbf{x}^{\mathrm{f}'}_{1:1} = \mathbf{x}^{\mathrm{f}}_1$, $\mathbf{P}^{\mathrm{f}'}_{1:1} = \mathbf{P}^{\mathrm{f}}_1$. Finally,

$$(\mathbf{x}^{\mathrm{a}}, \mathbf{P}^{\mathrm{a}}) = \mathcal{F}_{\mathrm{DA}}\left(\mathbf{x}^{\mathrm{f}'}_{1:M}, \mathbf{P}^{\mathrm{f}'}_{1:M}, \mathbf{y}, \mathbf{R}, \mathcal{H}\right). \tag{12}$$

Note that we allow in Equations 11 and 12 for possibly nonlinear operators $\mathcal{G}_m$ and $\mathcal{H}$, thus relaxing the linearity assumption on the operators $\mathbf{G}_m$ and $\mathbf{H}$, since many DA methods can deal with nonlinear observation operators. However, unless these operators are linear and the $\mathcal{F}_{\text{DA}}$ function is the Kalman filter assimilation step, the solution Equation 12 is only an approximation to the direct solution. Furthermore, order-independence is no longer guaranteed. A related issue occurs in serial EnKFs, wherein observations are assimilated one at a time, and localization generally introduces order dependence. Kotsuki et al. (2017) investigated the use of different ordering rules in this setting, and a similar investigation could be carried out for MM-DA. In our results in Section 4, we briefly explore empirically the role of the order in which the iterative solution is computed.

2. A Multi-Model Ensemble Kalman Filter (MM-EnKF)

As discussed in the previous section, MM-DA can potentially be used with any DA method. In this section, we describe the development and implementation of a multi-model ensemble Kalman filter (MM-EnKF). One of the advantages of EnKFs in general is that they dynamically estimate the forecast error covariance matrices, and are thus able to adapt to current conditions, or "errors of the day" (Kalnay, 2002). In the MM-EnKF, this flow dependence is then reflected in the weights assigned to each model and the observations in the state estimate.

For each m, we take its ensemble to have N_m members and denote the forecast and analysis ensembles as $\mathbf{E}_m^{\text{f}} = \left[\left(\mathbf{x}_m^{\text{f}}\right)_i\right]_{i=1}^{N_m}$ and $\mathbf{E}_m^{\text{a}} = \left[(\mathbf{x}_m^{\text{a}})_i\right]_{i=1}^{N_m}$, respectively; here $\left(\mathbf{x}_m^{\text{f}}\right)_i$ and $(\mathbf{x}_m^{\text{a}})_i$ denote the ith member in the forecast or analysis ensemble. We denote the means of the forecast and analysis ensemble by $\overline{\mathbf{x}}_m^{\text{f}}$ and $\overline{\mathbf{x}}_m^{\text{a}}$, respectively.

2.1. Incorporation of Model Error

2.1.1. The Model Error

Narayan et al. (2012) did not explicitly address model error covariances as part of the multi-model Kalman filter. Yang et al. (2017) did include model errors in their multi-model filter equations, but did not discuss methods to estimate them. We stress here that considering model errors is critical for the MM-DA's correctly weighting models, and that the multi-model filter must therefore be supplemented by a model error estimation method.

We assume that the true state evolution of the system can be expressed, for each model $\mathbf{M}_m$, as

$$\mathbf{G}_m\mathbf{x}^{\text{t}}(t_i) = \mathbf{M}_m(t_{i-1})\mathbf{G}_m\mathbf{x}^{\text{t}}(t_{i-1}) + \boldsymbol{\eta}_m(t_{i-1}), \tag{13}$$

where $\mathbf{x}^{\text{t}}(t_i)$ is the true state at time t_i and $\boldsymbol{\eta}_m$ is a model error with mean $\mathbf{0}$ and covariance $\mathbf{Q}_m$.

For model m, the forecast error covariance $\mathbf{P}_m^{\text{f}}$ at time t_{i+1} can then be estimated by Equation 8. This equation holds exactly only for a linear model (Tandeo et al., 2020). Thus $\mathbf{P}_m^{\text{f}}$ can be written as a sum of two terms,

$$\mathbf{P}_m^{\text{f}}(t_{i+1}) = \mathbf{P}_m^{\text{p}}(t_{i+1}) + \mathbf{Q}_m(t_i). \tag{14}$$

The term $\mathbf{P}^{\text{p}}$ is sometimes called the predictability error (Berry & Sauer, 2013), and is due to the effect of the system's dynamics on the uncertainty in the initial conditions. Therefore Kalman filters, without incorporating $\mathbf{Q}$, are prone to underestimate $\mathbf{P}^{\text{f}}$.

Note that the assumption that the total forecast error can be decomposed as a sum of an initial-condition error and a model error becomes less justified at longer lead times, due to the correlations between the initial condition and model errors (Carrassi, Vannitsem, & Nicolis, 2008; Mitchell & Carrassi, 2015).

Besides the underestimation problem, the consideration of model error in MM-DA is critical, since estimating the forecast error covariance from the ensemble spread as in EnKFs may give similar weights to models of different accuracy. For example, in Li et al. (2009), the perfect model was found to have similar spread to an imperfect model. Another issue is that a systematically overconfident model would be given higher weight if only the spread is accounted for.

Common ways to handle model error include: estimating the model error covariance matrix $\mathbf{Q}$ and using it to inflate the forecast covariance (additive inflation); inflating the forecast covariance with scalars (multiplicative inflation); or attempting to directly correct model error (bias correction). Gharamti (2018) discusses several additional methods.

Additive inflation generally works better than simple multiplicative inflation in accounting for model errors (Hamill & Whitaker, 2005; Li et al., 2009; Raanes et al., 2015; Whitaker & Hamill, 2012), since the latter assumes that model errors will have the same structure as errors due to initial conditions, which is not generally the case. Estimating scalar inflation factors, however, is more feasible in high-dimensional and data-scarce settings than estimating the matrix $\mathbf{Q}$. Moreover, there are methods for multiplicative covariance inflation that allow the inflation to vary in space and time (Anderson, 2009; Gharamti, 2018; Tandeo et al., 2020). Such methods are likely to narrow the performance gap or surpass temporally fixed additive inflation.

Several sophisticated state-dependent bias correction schemes have been developed and used in DA (Farchi et al., 2021; Li et al., 2009). The best results are usually obtained by a combination of bias correction and inflation (Baek et al., 2006; Li et al., 2009), with the latter accounting for the model error remaining after the bias correction.

In this paper, we use additive inflation to account for model error. Future work could apply bias correction to each model in addition to inflation. In the algorithms that follow, we use $\mathbf{b}_m$ to refer to the bias of model m, when bias estimation is employed; otherwise, $\mathbf{b}_m = \mathbf{0}$.

2.1.2. Estimation Method and Use in Filtering

In this paper, we use a simple, innovation-based estimation method for model error covariance, which we describe in Appendix B. However, there are a variety of methods for estimating $\mathbf{Q}$, often simultaneously with estimating $\mathbf{R}$; see the reviews of Duník et al. (2017) and Tandeo et al. (2020). When estimating $\mathbf{Q}$ is not computationally feasible, many methods for adaptive estimation of multiplicative covariance inflation are available, as described in the last section.

Several methods to estimate $\mathbf{Q}$, including the one we use, rely on the statistics of the innovations, that is, of the differences between observations and forecasts. In order to compute innovations for our MM-EnKF, we must define an additional observation operator $\mathbf{H}_m : \mathbb{R}^{n_m} \to \mathbb{R}^p$ for each model, which maps the model space to the observation space. For the reference model $m = m_r$, $\mathbf{H}_{m_r} = \mathbf{H}$. When $\mathbf{G}_m$ is injective, $\mathbf{H}_m$ is given by

$$\mathbf{H}_m = \mathbf{G}_m^{\dagger}\mathbf{H}. \tag{15}$$

In case $\mathbf{G}_m$ is not injective, $\mathbf{H}_m$ would have to be specified for every model. The innovations for model m are given by $\mathbf{d}_m = \mathbf{y} - \mathbf{H}_m\mathbf{x}_m^{\mathrm{f}}$.

Given an estimate of $\mathbf{Q}_m$, in order to account for it in the ensemble, samples drawn from the multivariate Gaussian distribution $\mathcal{N}\left(\mathbf{0}, \tilde{\mathbf{Q}}_m\right)$ can be added to the mth forecast ensemble (Asch et al., 2016; Mitchell & Carrassi, 2015), as done herein. Mitchell and Carrassi (2015) found this stochastic method to perform better than directly inflating the covariance matrix. Raanes et al. (2015) showed, however, that some methods work better for additive inflation in square-root filters than random sampling.

Here, we estimate $\mathbf{Q}_m$ for each model independently using the method described in Appendix B. Logutov and Robinson (2005), though, showed that an error estimation method—in their case, the direct estimation of the forecast error covariance matrices $\mathbf{P}_m^{\mathrm{f}}$—using all the models simultaneously can be more effective, especially when there is a small number of verifying observations. While not taken here, the latter approach could prove useful in the future.

2.2. Ensemble Perturbations

By applying the MM-DA framework directly to an EnKF, the iterative procedure results, prior to assimilating observations, in a combined multi-model forecast ensemble $\mathbf{E}_{1:M}^{\mathrm{f}'}$. This ensemble lives in the reference model space, and has N_{m_r} ensemble members. A disadvantage of this approach is that, already at the beginning of the forecast cycle, it reduces the number of ensemble members from $\sum_m N_m$ to N_{m_r}. Even though the information from these members is included in $\mathbf{E}_{1:M}^{\mathrm{f}'}$, a larger ensemble helps reduce sampling error.

Furthermore, once observations are assimilated, we obtain $\mathbf{E}^{\mathrm{a}}$, an analysis ensemble in the reference model space. How does one then obtain the analysis ensemble $\mathbf{E}_m^{\mathrm{a}}$ in each model space m, in order to use it as initial conditions for the next forecast cycle?

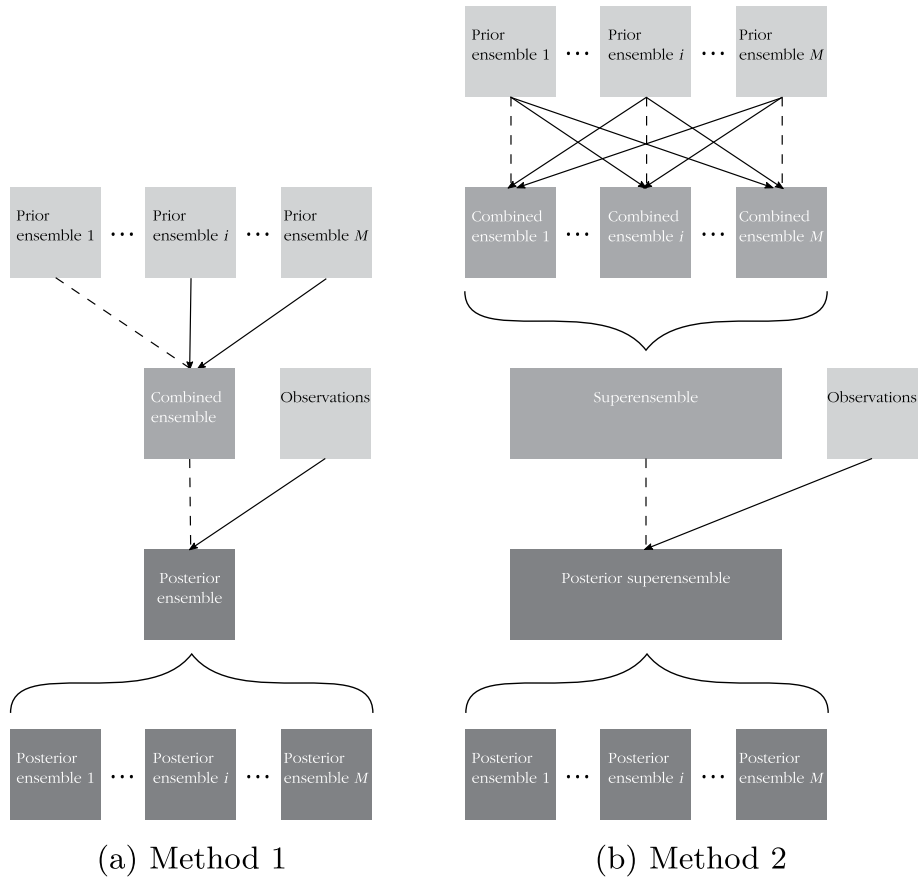


Figure 2. Diagram of the proposed algorithms. In Method 1, one model is treated as the reference model. In Method 2, each model is treated as the reference model in turn, and the combined forecast ensembles are concatenated into a superensemble before assimilating observations. Arrows indicate information treated as observations in data assimilation, dotted lines indicate those treated as background. Braces indicate concatenation for Method 2, and application of Equation 16 for Method 1.

Previous work on MM-DA did not address these questions dealing with ensemble perturbations in an MM-EnKF. Here, we discuss three ways of doing so.

2.2.1. Method 1

This method is a straightforward application of the iterative methodology described above: we simply compute $\mathbf{E}^{f'}_{1:M}$. After assimilating observations, we then take

$$\mathbf{E}^{\mathrm{a}}_{m} = \mathbf{G}_{m}\mathbf{E}^{\mathrm{a}} \tag{16}$$

as the analysis ensemble in model space m. A schematic diagram is shown in Figure 2a.

This method has a disadvantage in terms of sampling error, as described above. Moreover, Equation 16 implies that each new model ensemble will now have the same number of ensemble members as that of the reference model: $N_m = N_{m_r}$. If $N_{m_r} \geq N_m$ for all $m \neq m_r$, though, a random choice of N_m ensemble perturbations out of the N_{m_r} could still be made for each $m \neq m_r$.

A related issue is that, for Method 1, each model's posterior ensemble has the same perturbations but transformed into the model space m, since $\mathbf{E}^{\mathrm{a}} = \overline{\mathbf{x}}^{\mathrm{a}}\mathbf{1}^{T} + (N-1)^{1/2}\mathbf{X}^{\mathrm{a}}$ (Asch et al., 2016) implies that $\mathbf{G}_m\mathbf{E}^{\mathrm{a}} = \overline{\mathbf{x}}^{\mathrm{a}}_{m}\mathbf{1}^{T} + (N-1)^{1/2}\mathbf{G}_m\mathbf{X}^{\mathrm{a}}$, where $\mathbf{X}^{\mathrm{a}}$ are the reference model's analysis ensemble perturbations and $\mathbf{1}$ is a vector of ones. This may reduce the effective number of ensemble members in the multi-model forecast ensemble, when the models are similar enough.

2.2.2. Method 2

We propose an alternative method for handling the ensemble perturbations. Here, we repeat the iterative procedure of Equation 9 m times at each assimilation step, changing the reference model to $m_r = m$ each time. Then, we have m model ensembles, each in their own model space. We then map all these ensembles into a single model space, considering them as a single "superensemble." The observations are then assimilated into this superensemble, and the analysis ensemble members can be mapped back into their respective model spaces. A schematic diagram appears in Figure 2b.

This method uses all the ensemble members in assimilating the observations, and will thus suffer from lower sampling error than Method 1. Furthermore, one obtains an analysis ensemble for each model which retains the number of ensemble members N_m, and has distinct ensemble perturbations for each model.

This method, though, has a larger computational cost than Method 1: the multi-model forecast combination cost in terms of operation count will increase by a factor of about M, although each of the M assimilation steps can be done in parallel. Likewise, if the ensemble sizes are equal, the memory requirement for the analysis ensembles will increase by a factor of M. Furthermore, this method requires mappings from every model space m to every other model space m': $\mathcal{G}_{m\to m'}$. This is only possible if the mappings are invertible. Hence, this method is easiest to implement when all the models are in the same space or when there is a simple mapping between them, but it will not work with different dimensions of model space.

2.2.3. Other Approaches

Lastly, the analysis ensemble could be regenerated for each $m \neq m_r$ by matching the known moments of the analysis distribution $\left(\overline{\mathbf{x}}^{\mathrm{a}}_m, \mathbf{P}^{\mathrm{a}}_m\right)$, as obtained from Equations 5 and 6. There is no unique set of ensemble members that possess these moments, but some that do can be generated by sampling from the multivariate normal $\mathcal{N}\left(\overline{\mathbf{x}}^{\mathrm{a}}_m, \mathbf{P}^{\mathrm{a}}_m\right)$.

Another way to generate appropriate ensemble members is to use sigma points as in the unscented Kalman filter (Julier & Uhlmann, 2004). Doing so, however, requires at least $2n_m$ sigma points, which is not feasible for high-dimensional, computationally expensive models.

In the numerical experiments that follow, we compare only Methods 1 and 2.

2.3. Computational Considerations

For the iterative form of MM-DA, the Kalman gain Equation 9a can be written as

$$\mathbf{K}_m\left(\mathbf{G}_m\mathbf{P}^{\mathrm{f}'}_{1:m-1}\mathbf{G}_m^T + \mathbf{P}^{\mathrm{f}}_m\right) = \mathbf{P}^{\mathrm{f}'}_{1:m-1}\mathbf{G}_m^T, \tag{17}$$

where the linear system is solved for $\mathbf{K}_m$ in order to avoid explicit matrix inversion. Note that here, since the forecast of model m is treated as an observation, we are required to solve a system in the model space $\mathbb{R}^{n_m}$. This can be too computationally expensive for high-dimensional models.

2.3.1. Taking Advantage of Lower-Dimensional Models

If only one of the models has very high dimension, that model can be chosen to be the reference model m_r. Then, in the iterative procedure, the inversions will only have to be done in the lower-dimensional model spaces, thus facilitating the computations.

Alternatively, if several models have high dimensions but only large-scale features are of interest, their forecasts could be mapped to a lower-dimensional space prior to assimilation, and the $\mathbf{G}_m$ modified accordingly. In the case of weather or climate models, this could consist in mapping the forecasts to a coarser grid. Possible solutions for multiple high-dimensional models with high-dimensional features that are relevant will be discussed below.

2.3.2. Taking Advantage of Low Rank

When using an EnKF, both $\mathbf{P}^{\mathrm{f}'}_{1:m-1}$ and $\mathbf{P}^{\mathrm{f}}_m$ in Equation 17 will be sample covariance matrices and they will be rank-deficient if the ensemble sizes are smaller than the model dimensions, as is typically the case. Localization generally increases the rank; when not applied, the low rank of these covariance matrices can be exploited to obtain a least-squares solution in $\mathcal{O}\left(n_m N'^2_m\right)$ operations, where N'_m is the rank of the matrix $\mathbf{G}_m\mathbf{P}^{\mathrm{f}'}_{1:m-1}\mathbf{G}_m^T + \mathbf{P}^{\mathrm{f}}_m$ (Mandel, 2006).

2.3.3. Right-Multiplied ESRFs

In the following approaches to efficient MM-EnKF implementation, an important role is played by square-root Kalman filters (SRFs) (Bellantoni & Dodge, 1967; Bierman, 1977), and in particular ensemble SRFs (ESRFs) (Tippett et al., 2003). In their historical account, Grewal and Andrews (2010) state that the SRF is an "improvement […] over conventional Kalman filtering [achieving] 'the same accuracy with half as many bits' of precision."

An alternative form of the gain is obtained by applying the Sherman–Morrison–Woodbury formula (Hager, 1989) to Equation 17:

$$\mathbf{K}_m = \left(\left(\mathbf{P}^{\mathrm{f}'}_{1:m-1}\right)^{-1} + \mathbf{G}_m^T\left(\mathbf{P}^{\mathrm{f}}_m\right)^{-1}\mathbf{G}_m\right)^{-1}\mathbf{G}_m^T\left(\mathbf{P}^{\mathrm{f}}_m\right)^{-1} \tag{18}$$

Some ensemble Kalman filter variants use gains of the form Equation 18, but express the analysis in the ensemble subspace (Asch et al., 2016). These are known as right-multiplied ESRFs (Sakov & Bertino, 2011).

The ensemble transform Kalman filter (ETKF: Bishop et al., 2001) is an important form of right-multiplied ESRF. The gain can be computed by solving a linear system without explicitly inverting $\mathbf{P}^{\mathrm{f}}_m$. If $\mathbf{P}^{\mathrm{f}}_m$ is assumed to have a block-diagonal structure with relatively small blocks, the computation becomes feasible. This block-diagonal structure is intrinsic to analyses being done locally, as in the local ETKF (LETKF; Hunt et al., 2007).

Similar to the low-rank case discussed above, when $\mathbf{P}^{\mathrm{f}}_m$ is rank-deficient with rank N_m, its pseudoinverse can be computed in $\mathcal{O}\left(n_m N_m^2\right)$ operations.

2.3.4. Structured Covariance Matrices

Another way of making the matrix operations less expensive is to take either $\left(\mathbf{G}_m\mathbf{P}^{\mathrm{f}'}_{1:m-1}\mathbf{G}_m^T + \mathbf{P}^{\mathrm{f}}_m\right)$ in gains of the form of Equation 17 or $\mathbf{P}^{\mathrm{f}}_m$ in gains of the form of Equation 18 to have a simplified structure. As discussed above, block-diagonal structure is one such possibility. Block-diagonality also enables the use of sequential EnKFs (Houtekamer & Mitchell, 2001). Several simplified structures were considered for observation error covariance matrices in Stewart et al. (2013).

The simplest structure, but a rather restrictive one, is assuming the matrices to be diagonal; then the inverse is trivial to compute and store. The diagonality assumption is often made for the covariance matrices of observation errors. Note that if all the $\mathbf{P}^{\mathrm{f}}_m$ are treated as diagonal in computing Equations 3 and 4 with the direct method, the solution corresponds to the minimum variance estimator when the weights for each $\mathbf{x}^{\mathrm{f}}_m$ are vectors instead of matrices; see Corollary 2 in Sun (2004).

For right-multiplied ESRFs, if we only impose the simplified structure when inverting $\mathbf{P}^{\mathrm{f}}_m$, the simplified-structure assumption is not made for the forecast covariance of the reference model, $\mathbf{P}^{\mathrm{f}}_{m_r}$.

In the experiments below, since the dimensionality is relatively low, we first apply localization to each $\mathbf{P}^{\mathrm{f}}_m$ and then invert directly.

2.4. Multi-Model Forecasting

MM-DA can be used for real-time forecasting with multiple models by carrying out the iterative procedure for the available models and not assimilating any observations (Logutov & Robinson, 2005; Narayan et al., 2012). Doing so corresponds simply to the use of Equation 9 to combine the multiple models.

We can let $\mathbf{R}^{-1} \to \mathbf{0}$ in Equation 2, since this limit of infinite variance simply corresponds to no observations being available. Then, Equations 3 and 4 of MM-DA for the analysis state and covariance become.

$$\mathbf{x}^{\mathrm{a}} = \mathbf{P}^{\mathrm{a}}\left(\sum_{m=1}^{M}\mathbf{G}_m^T\left(\mathbf{P}^{\mathrm{f}}_m\right)^{-1}\mathbf{x}^{\mathrm{f}}_m\right), \tag{19a}$$

$$\mathbf{P}^{\mathrm{a}} = \left(\sum_{m=1}^{M}\mathbf{G}_m^T\left(\mathbf{P}^{\mathrm{f}}_m\right)^{-1}\mathbf{G}_m\right)^{-1}. \tag{19b}$$

Thus, MM-DA neatly handles multi-model forecasting in addition to DA. Note that, when $\mathbf{G}_1 = \mathbf{G}_2 = \cdots = \mathbf{G}_M = \mathbf{I}$ and $\mathbf{P}_1^{\mathrm{f}} = \mathbf{P}_2^{\mathrm{f}} = \cdots = \mathbf{P}_M^{\mathrm{f}}$, Equation 19a reduces simply to the unweighted multi-model average.

When forecasting at long lead times, it can be helpful to apply MM-DA recursively at intermediate leads. The set of model error covariance matrices should be specified for each lead time; it is known in the seasonal climate prediction context, for example, that the "best model" can depend on the lead time (Hagedorn et al., 2005).

Assume that we have estimated the model error covariance matrix for each model m at different intermediate lead times $k\tau$, which we denote by $\mathbf{Q}_m^{k\tau}$. Then, if the desired forecast horizon is $T = K\tau$, MM-DA can be applied first at lead time τ with model error covariance matrices $\mathbf{Q}_m^{\tau}$. The analysis for this horizon is then used as an initial condition for the forecasts out to time 2τ, whereupon MM-DA is applied with $\mathbf{Q}_m^{2\tau}$, etc. This recursive method tends to perform better than directly applying MM-DA at horizon T, since the trajectory is repeatedly corrected.

One may wonder whether, at long lead times, when the error growth of a nonlinear forecast model ceases to obey linearized dynamics, Equation 14 for the forecast error covariance is still a good approximation. Here, it is more useful to think of $\mathbf{Q}$ as an additive inflation that compensates for overconfidence in the prediction.

2.5. Filter Algorithm

Localization is critical for EnKFs (Carrassi et al., 2018). Here, we apply localization at each step of the iterative procedure, and also when observations are assimilated.

We use the left-multiplied form of the ESRF, as described in Sakov and Bertino (2011), for both the multi-model combination and the assimilation of observations. This EnKF is a deterministic filter for which it is particularly simple to express covariance localization.

The left-multiplied ESRF equations are given by.

$$\mathbf{X} = (N-1)^{-1/2}\left(\mathbf{E}^{\mathrm{f}} - \overline{\mathbf{x}}^{\mathrm{f}}\mathbf{1}^T\right), \tag{20a}$$

$$\mathbf{P}^{\mathrm{f}} = \boldsymbol{\rho} \circ \left(\mathbf{X}\mathbf{X}^T\right), \tag{20b}$$

$$\mathbf{K} = \mathbf{P}^{\mathrm{f}}\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^{\mathrm{f}}\mathbf{H}^T + \mathbf{R}\right)^{-1}, \tag{20c}$$

$$\overline{\mathbf{x}}^{\mathrm{a}} = \overline{\mathbf{x}}^{\mathrm{f}} + \mathbf{K}\left(\mathbf{y} - \mathbf{H}\overline{\mathbf{x}}^{\mathrm{f}}\right), \tag{20d}$$

$$\mathbf{E}^{\mathrm{a}} = \overline{\mathbf{x}}^{\mathrm{a}}\mathbf{1}^T + (N-1)^{1/2}(\mathbf{I} - \mathbf{K}\mathbf{H})^{1/2}\mathbf{X}, \tag{20e}$$

where $\boldsymbol{\rho}$ is the localization matrix; $\circ$ is the Hadamard, or element-wise, product; and $\mathbf{X}$ are the normalized ensemble perturbations.

In the iterative procedure, we use the ensemble mean $\overline{\mathbf{x}}_m^{\mathrm{f}}$ of model m as the observation for the MME $\mathbf{E}_{1:m-1}^{\mathrm{f}\prime}$.This ESRF form is not efficient for high-dimensional systems, since the update is done in the state space instead of the ensemble space. For high-dimensional systems, right-multiplied ESRFs are more practical. See Section 2.3 for more details on computational issues.

2.6. Inflation

EnKFs generally underestimate the forecast covariance due to model and sampling errors, thus imposing the need for inflation (Carrassi et al., 2018). While we attempted to account for the model error in each individual model, we found that the multi-model forecast covariance is usually still underestimated, and the underestimation increases with M.

This underestimation is due to the assumption that the models are unbiased, and that the errors for distinct models are independent of one another. That is, if the model forecasts were unbiased and independent, one would expect the error in a multi-model average to decrease as $M^{-1/2}$, but this does not happen. See Knutti et al. (2010) and Christiansen (2020) for an explanation of this phenomenon in MMEs. Furthermore, in assimilating forecast states

of one model into another one, which has a different attractor, one inherently encounters representation error (Hodyss & Nichols, 2015). Hence, we also need to apply inflation to the multi-model forecast.

Here, we use a simple multiplicative covariance inflation scheme, with the inflation factor $\hat{\lambda}$ estimated as in Tandeo et al. (2020):

$$\hat{\lambda} = \frac{\mathbf{d}^T\mathbf{d} - \mathrm{tr}(\mathbf{R})}{\mathrm{tr}(\mathbf{H}\mathbf{P}^{\mathrm{f}}\mathbf{H}^T)}. \tag{21}$$

Since the inflation is applied to the multi-model forecast, we take $\mathbf{d} = \mathbf{y} - \mathbf{H}\mathbf{x}^{\mathrm{f}'}_{1:M}$ and $\mathbf{P}^{\mathrm{f}} = \mathbf{P}^{\mathrm{f}'}_{1:M}$ and then apply a temporal smoothing, as in Equation B3, which yields

$$\tilde{\lambda}(k+1) = \gamma\hat{\lambda}(k) + (1-\gamma)\tilde{\lambda}(k), \tag{22}$$

for some $0 < \gamma < 1$. Note that the numerator of Equation 21 is not guaranteed to be positive, although its expected value is. However, negativity of $\hat{\lambda}$ does not pose a problem as long as the smoothed estimate $\tilde{\lambda}$ is positive. Encountering a negative $\tilde{\lambda}$ suggests either a misspecification of the error covariance matrices or a γ-value that is too large, allowing for rapid fluctuations in $\tilde{\lambda}$.

Due to the MM-DA–specific reasons above, the resulting values of λ are higher than typically encountered with regular covariance inflation: in the experiments below, for instance, we have encountered $\tilde{\lambda}$-values as large as 4.

2.7. Algorithms

We are ready now to summarize in pseudo-code the two proposed versions of the MM-EnKF, as Algorithms 1 and 2. To maintain generality, we define the following `DA_step` function, which represents the analysis step for any EnKF, and in which the observation operator $\mathcal{H}$ is kept as possibly nonlinear, since ensemble Kalman filters allow for nonlinear observation operators.

function *DA_step*

Input:

- $\mathbf{E}^{\mathrm{f}}$, the prior ensemble
- $\mathbf{y}$, the observation vector
- $\mathbf{R}$, the observation error covariance
- $\mathcal{H}$, the observation operator

Output: $\mathbf{E}^{\mathrm{a}}$, the posterior ensemble

Note that, for real-time forecasting, line 12 in the pseudocode for Method 1 or line 16 in the pseudocode for Method 2 is removed.

3. Relation to Other Methods

3.1. MM-EnKF Properties

The MM-EnKF has the following properties, compared to other methods for MM-DA and forecasting:

- The method is a natural generalization of the standard Kalman filter to multiple models, and can be derived from both the variational and Bayesian viewpoints (Narayan et al., 2012), as well as from linear minimum variance estimation (see Appendix A). This fact allows for the use of well-understood DA methods, and the theoretical apparatus of optimal state estimation and Kalman filters (e.g., Jazwinski, 1970; Simon, 2006).
- The methods reviewed in Section 1 mostly involve scalar weights. Here, the weights are matrices, which allows for variables to be weighted differently. In the case of spatiotemporal models, this allows the weights

Algorithm 1. Multi-Model Ensemble Kalman Filter Step (Method 1)

```
/* Inflate ensemble members from estimated model error distribution */
for m in (1, ..., M) do
    for i in (1, ..., N_m) do
        η_i ~ N(−b_m, Q_m)
        x_i^f = x_i^f + η_i
    end
end
/* Assimilate the other model forecasts into the reference model ensemble */
E_{1:1}^{f'} = E_1^f
for ℓ in (2, ..., M) do
    E_{1:ℓ}^{f'} = DA_step(E_{1:ℓ−1}^{f'}, x̄_ℓ^f, P_ℓ^f, G_ℓ)
end
E^f = x̄ + λ^{1/2}(E_{1:M}^{f'} − x̄)
/* Assimilate observations */
E^a = DA_step(E^f, y, R, H)
/* Integrate each posterior ensemble to the next time */
for m in (1, ..., M) do
    E_m^a = G_m((E^a)_m)
    E_m^f(t + Δt) = M_{t→t+Δt}(E_m^a(t))
end
```

assigned to each model to vary in space; this is important in the case of atmospheric models, where model skill can be highly spatially inhomogeneous (Du & Smith, 2017).

- Each model can have its own model space. Most of the other reviewed methods do not allow for this, instead assuming a common model space. Distinct model spaces allow for the combination of models of different resolutions, those that predict different variables, or those that are restricted to different spatial domains. Some examples of such scenarios are provided in Section 4.
- A common problem of adaptive multi-model methods is the weight of useful models converging to 0 (Smith et al., 2020). With MM-DA, this problem does not occur as long as the filter is stable, since this would require $\left(\mathbf{P}_m^{\mathrm{f}}\right)^{-1} \to \mathbf{0}$. This feature may have its downside when a model is consistently detrimental.

Algorithm 2. Multi-Model Ensemble Kalman Filter Step (Method 2)

```
    /* Inflate ensemble members from estimated model error distribution
       */
1–6 (Same as lines 1–6 in Algorithm 1)
    /* For each model ensemble, assimilate the other model forecasts
       into it                                                        */
7   for m in (1, ..., M) do
8       O = (m, 1, ..., m − 1, m + 1, ..., M) // Order of assimilation
9       E^{f'}_{m,1:1} = E^f_m
10      for ℓ in (2, ..., M) do
11          E^{f'}_{m,1:ℓ} = DA_step(E^{f'}_{m,1:ℓ−1}, x̄^f_{O_ℓ}, P^f_{O_ℓ}, G_{m→O_ℓ})
12      end
13  end
    /* Form a "superensemble" from all the ensembles                  */
14  E^f = [G_{1→m_r}(E^{f'}_{1,1:M}) ··· G_{M→m_r}(E^{f'}_{M,1:M})]
15  E^f = x̄ + λ^{1/2}(E^f − x̄)
    /* Assimilate observations                                        */
16  E^a = DA_step(E^f, y, R, H)
    /* Integrate each posterior ensemble to the next time             */
17–20 (Same as lines 13–16 in Algorithm 1)
```

- If all models are biased in one direction, Bayesian model averaging will result in a forecast worse than the best model. This is not the case with MM-DA (Narayan et al., 2012).
- The MM-EnKF methodology provides probabilistic analyses and forecasts, using ensembles. Many of the methods for forecast combination reviewed herein assume a single deterministic forecast for each model, and do not account for uncertainty.
- DA is designed for forecast problems, and MM-DA is shown in Section 4 to improve forecast skill. However, some multi-model methods target instead improving climatology, that is, the system's long-term statistics. In MM-DA, the $\mathbf{Q}_m$ are specified for a specific lead time; it is not clear how—or whether—these $\mathbf{Q}_m$'s can be adequately adapted to capture climatological error instead. It is often the case, though, that long-term systematic errors are similar to those at short timescales (Martin et al., 2010; Rodwell & Palmer, 2007).
- Several authors (Bach et al., 2021; Chattopadhyay et al., 2022; Chen & Stechmann, 2019; Counillon et al., 2022; Du & Smith, 2017; Ojeda et al., 2013; Potthast et al., 2022) explored the assimilation of forecasts as pseudo-observations. In many of these works, however, the error covariance assigned to the

pseudo-observations was not defined in a consistent way, or the generalization to more than two models was not clear. MM-DA also assimilates forecasts as if they were observations, but in a consistent mathematical framework.

- Rainwater and Hunt (2013) formulated an EnKF that uses ensembles at two different resolutions to compute the background covariance matrix, with a parameter that sets the weights given to each one. The low-resolution state forecast was not used. Hoel et al. (2016, 2020) combined forecasts at different resolutions in an EnKF, but they did not weight them differently. Popov et al. (2021) combined models of different fidelities in an EnKF with a control variate approach. In Section 4, we will show how the MM-EnKF can effectively incorporate forecasts at different resolutions and fidelities.
- In the terminology of Mallet et al. (2009), MM-DA is a convex sequential aggregation rule.

3.2. Connection to Synchronization

To combine forecasts of two models we have, from Equations 9b and 9a.

$$\mathbf{x}_1 = \mathbf{x}_1^{\mathrm{f}} + \mathbf{K}_2\left(\mathbf{x}_2^{\mathrm{f}} - \mathbf{G}_{1\to 2}\mathbf{x}_1^{\mathrm{f}}\right), \tag{23a}$$

$$\mathbf{x}_2 = \mathbf{x}_2^{\mathrm{f}} + \mathbf{K}_1\left(\mathbf{x}_1^{\mathrm{f}} - \mathbf{G}_{2\to 1}\mathbf{x}_2^{\mathrm{f}}\right), \tag{23b}$$

where

$$\mathbf{K}_1 = \mathbf{P}_2^{\mathrm{f}}\mathbf{G}_{2\to 1}^T\left(\mathbf{G}_{2\to 1}\mathbf{P}_2^{\mathrm{f}}\mathbf{G}_{2\to 1}^T + \mathbf{P}_1^{\mathrm{f}}\right)^{-1}, \tag{24a}$$

$$\mathbf{K}_2 = \mathbf{P}_1^{\mathrm{f}}\mathbf{G}_{1\to 2}^T\left(\mathbf{G}_{1\to 2}\mathbf{P}_1^{\mathrm{f}}\mathbf{G}_{1\to 2}^T + \mathbf{P}_2^{\mathrm{f}}\right)^{-1}. \tag{24b}$$

Here, $\mathbf{G}_{1\to 2}$ is the matrix mapping from a state in model space 1 to the corresponding state in model space 2, and vice-versa for $\mathbf{G}_{2\to 1}$. Thus, each model is being nudged toward the forecast of the other. This mutual nudging connects MM-DA to the synchronization view of DA (Abarbanel et al., 2017; Carrassi, Ghil, et al., 2008; Penny, 2017; Penny et al., 2019): the multi-model combination step can be considered a form of impulsive synchronization between the models.

In the connected supermodeling approach of Selten et al. (2017), connection terms between model states are introduced into the model equations. The connection coefficients are gathered into matrices $\mathbf{C}$, which can be identified with the gain matrices $\mathbf{K}_i$ in Equation 23. We note, however, that MM-DA differs from the approach of Selten et al. (2017), as the latter directly estimates the connection coefficients by minimizing a cost function with training data. Additionally, the supermodeling approach uses static and diagonal $\mathbf{C}$, does not allow for different model spaces, and does not consider ensembles of each model. Future work could compare the connection coefficients obtained by connected supermodeling with the gains $\mathbf{K}_i$ obtained by MM-DA. Since supermodels are typically formulated in continuous time, determining the exact relationship between MM-DA and supermodels necessitates the derivation of the continuous-time analog of the multi-model Kalman filter, namely a multi-model Kalman–Bucy filter.

A similar connection can be made between MM-DA and weighted supermodeling: in the latter, the supermodel tendency is a weighted average of the individual model tendencies (Schevenhoven et al., 2019), while in MM-DA the analysis is a weighted average of the model forecasts (Equation 19a). Wiegerinck et al. (2013) showed that a connected supermodel becomes a weighted supermodel in the limit of large couplings.

4. Numerical Experiments

Previously, MM-DA was only tested on very low-dimensional models with non-chaotic behavior (Narayan et al., 2012; Yang et al., 2017), and recursive multi-step forecasts were not tested. Methods for multi-model forecasting have often been tested with perfect observations for calibration, single forecasts for each model rather than ensembles, and models that all share the same space (Schevenhoven & Selten, 2017; Schevenhoven et al., 2019); several papers, though, have extended this work to noisy observations (Du & Smith, 2017; Schevenhoven & Carrassi, 2022). Here, we conduct twin experiments of the proposed method for both DA and forecasting in

various settings, including models of different dimensionality and different-sized ensembles. Noisy observations are used for the model error estimation in all cases.

4.1. Experimental Set-Up

In the following numerical experiments, we use the Lorenz96 (Lorenz, 1996) model, except that we allow a different forcing F_i for each site:

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = -x_{i-1}(x_{i-2} + x_{i+1}) - x_i + F_i; \tag{25}$$

here the indices i range from 1 to D and are cyclical. We use $D = 40$ variables in the experiments that follow.

The true model here has $F_i = 8$ for $1 \le i \le 10$, $F_i = 10$ for $11 \le i \le 20$, $F_i = 12$ for $21 \le i \le 30$, and $F_i = 14$ for $31 \le i \le 40$, similar to Du and Smith (2017). We then define four imperfect forecast models to be used in the experiments, having fixed $F \equiv 8$, 10, 12, and 14 for all i.

We also use the two-scale version of model (Lorenz, 1996).

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = -x_{i-1}(x_{i-2} + x_{i+1}) - x_i + F_i - \frac{hc}{b}\sum_{j=1}^{n} y_{j,i}, \tag{26a}$$

$$\frac{\mathrm{d}y_{j,i}}{\mathrm{d}t} = -cby_{j+1,i}(y_{j+2,i} - y_{j-1,i}) - cy_{j,i} + \frac{hc}{b}x_i, \tag{26b}$$

where the indices i range from 1 to D, the indices j range from 1 to d, $y_{d+1,i} = y_{1,i+1}$, and $y_{0,i} = y_{d,i-1}$. The y_{j_i} variables represent smaller-scale dynamics, which interact with the larger-scale x_i's. We set $D = 20$, $d = 10$, $h = 1$, $b = 10$, $c = 10$. With these parameters, the timescale is about 10 times as fast for the y_{j_i}'s as for the x_i's. Given the full state vector containing both the x and y variables,

$$\mathrm{vec}\begin{pmatrix} x_1 & y_{1,1} & y_{2,1} & \cdots & y_{d,1} \\ x_2 & y_{1,2} & y_{2,2} & \cdots & y_{d,2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_D & y_{1,D} & y_{2,D} & \cdots & y_{d,D} \end{pmatrix}, \tag{27}$$

where vec is the vectorization operator which stacks the columns of the matrix on top of one another to obtain a column vector, the corresponding $\mathbf{G}_2$ is the $(d+1)D \times D$ matrix

$$(\mathbf{G}_2)_{i,j} = \begin{cases} 1, & \text{if } i = 1 \\ 0, & \text{otherwise.} \end{cases} \tag{28}$$

The time integrations used the fourth-order Runge–Kutta scheme. For the single-scale Lorenz96 model, we use a timestep of $\Delta t = 0.05$, and for the two-scale one we use $\Delta t = 0.005$.

For localization, we use the Gaspari–Cohn correlation function (Gaspari & Cohn, 1999). For experiments with the single-scale model, we use a localization radius of 4. For experiments with the two-scale model, we apply a localization radius of 4 to the x_i variables and a radius of 40 to the $y_{j,i}$ ones. Cross-scale interactions are localized such that observations of the $y_{j,i}$ are allowed to influence the corresponding x_i, and vice versa.

In the following experiments, we compare the MM-EnKF to the unweighted MME, wherein the multiple single-model ensembles are treated identically as a single ensemble, except that each model ensemble is inflated using the appropriate $\mathbf{Q}_m$. We also compare the results to each of the individual single-model ensembles, again inflated by their respective $\mathbf{Q}_m$. Moreover, scalar inflation is applied for both the MME and the MM-EnKF, as described in Section 2.6.

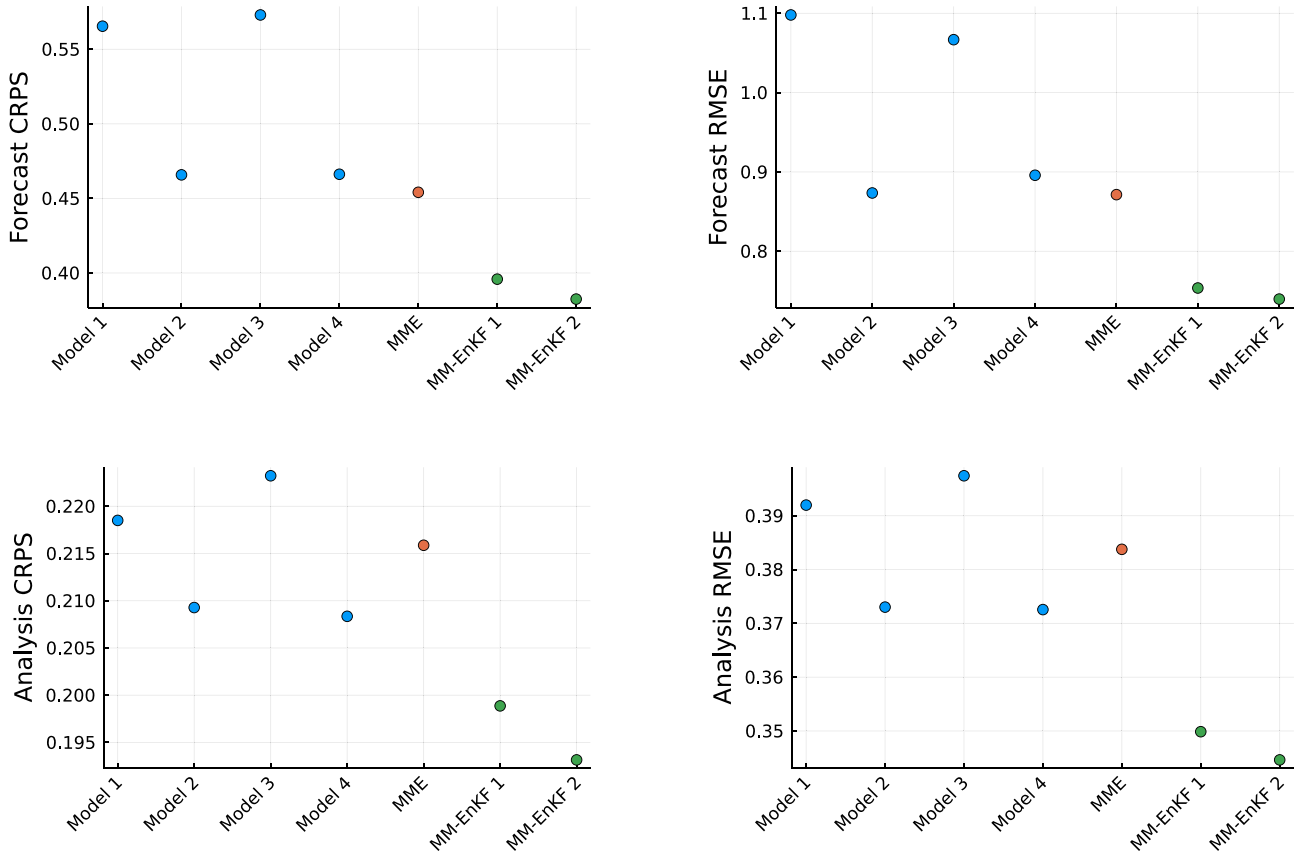


Figure 3. Overall performance of the multi-model ensemble Kalman filter, in terms of both root-mean-square error and the continuous ranked probability score. Here and in the subsequent experiments, we use suitably defined versions of the Lorenz96 model (Lorenz, 1996). The error bars are too small to be visible and hence none are plotted.

4.2. Experiments With Parametric Model Error

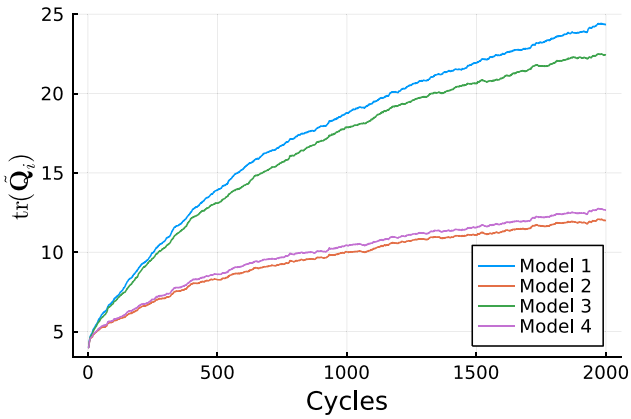
4.2.1. Multi-Model DA

We test out Methods 1 and 2 of the MM-EnKF with the four imperfect models. We use an analysis window, or time interval over which observations are assimilated, of 0.2; $\mathbf{R} = 0.25\mathbf{I}_{40}$; $\delta = 10^{-3}$ for the model error estimation; and $\gamma = 10^{-2}$ for the inflation estimation. Here we fully observe the state, but test partial observations in Section 4.2.1.3.

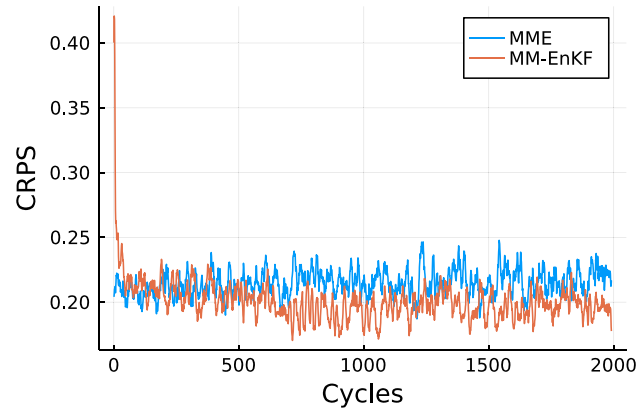
For the MME and the MM-EnKF, we use 20 ensemble members for each model. In order to have a fair comparison, 80 ensemble members are used for each of the individual model experiments. We carry out 10,000 DA cycles, and average the error over the last 2,000.

Figure 3 shows the results for the forecast and analysis errors. The forecast errors are for forecasts initialized from the analyses obtained by the filter, for a lead time equal to the analysis window. In addition to the root-mean-square error (RMSE), we use the continuous ranked probability score (CRPS: Hersbach, 2000), a probabilistic error metric, to measure the discrepancy between the ensemble and the true probability distributions. We apply the univariate CRPS along each dimension, and then take the mean. A strength of the CRPS is that it is a strictly proper scoring rule (Wilks, 2019).

The regular MME performs slightly better than the best model in terms of forecast error, and worse than the best model in terms of analysis error. The MM-EnKF, though, performs better than the MME and any individual model, in both forecast and analysis errors, and Method 2 has a slight edge over Method 1. The latter fact is likely due to Method 2 using a larger ensemble than Method 1 when assimilating the observations; see Section 2.2.



(a) Evolution of the estimated model error covariance for each model, as given by the trace.



(b) Evolution of the analysis CRPS over time. A 10-timestep moving average is applied.

Figure 4. (a) Evolution of estimated model error covariance for the four models. Simultaneously in time, (b) shows a comparison of the evolution of the continuous ranked probability score error metric for the unweighted multi-model ensemble and the multi-model ensemble Kalman filter.

4.2.1.1. Impact of Model Error Estimation

To see the effect of model error estimation on the performance of the MM-EnKF methods, we run it simultaneously with the DA itself, and consider the time evolution of the weights and analysis error. In Figure 4a we see the model weights evolving: initially assigned the same model error covariance, the model error estimation procedure estimates a higher error for models 1 and 3, and they are thus weighted less in the DA. Note that we show here only the trace, but in reality the weights are not the same for all variables.

In Figure 4b, the analysis error is shown over the same time interval. Initially, with the same weight for each model, the MM-EnKF performs worse than an unweighted MME. However, as the model error estimation becomes more accurate, the MM-EnKF reaches a lower asymptotic error than the MME.

4.2.1.2. Impact of Assimilation Order

To test the effect of the order in which the models are assimilated, we repeated the experiment with the $24 = 4!$ possible permutations of the model orders. The results are shown in Figure 5. In this case, model order is not very significant, and all the orders result in errors smaller than the best model and the MME. Furthermore, the standard deviation of the CRPS over all the permutations is about an order of magnitude smaller than the improvement of the MM-EnKF compared to either the individual models or the MME.

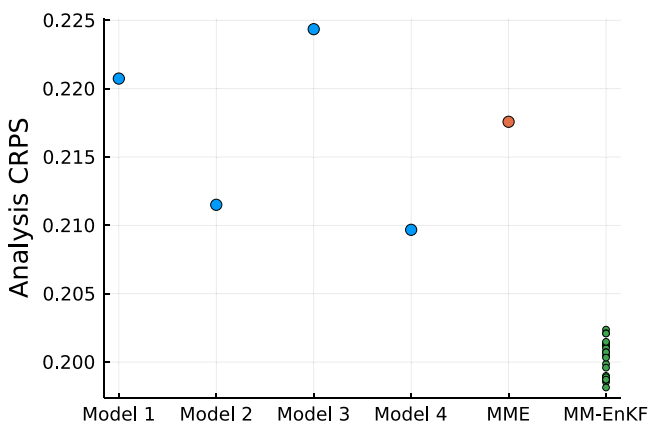


Figure 5. Performance of Method 1 of the multi-model ensemble Kalman filter, with all permutations in the order of assimilating the four models plotted in green. The single-model analyses and unweighted multi-model ensemble are included for reference.

Although for this case the assimilation order has a minimal effect, it will be important to test this sensitivity in other set-ups.

4.2.1.3. Partial Observations

We test a case where we only have partial observations of the system. In particular, here we observe only the odd-numbered x_i.

Figure 6 shows the results for both forecasting and analysis. Note that the forecast step here is not different than for full observations; however, the forecasts are initialized from analyses obtained using the partial observations. Again, the MM-EnKF produces the best forecasts and analyses.

4.2.2. Multi-Model Forecasts

We now test the MM-EnKF for real-time forecasting at different lead times. The experimental set-up is the same as in the previous Section 4.2.1, except

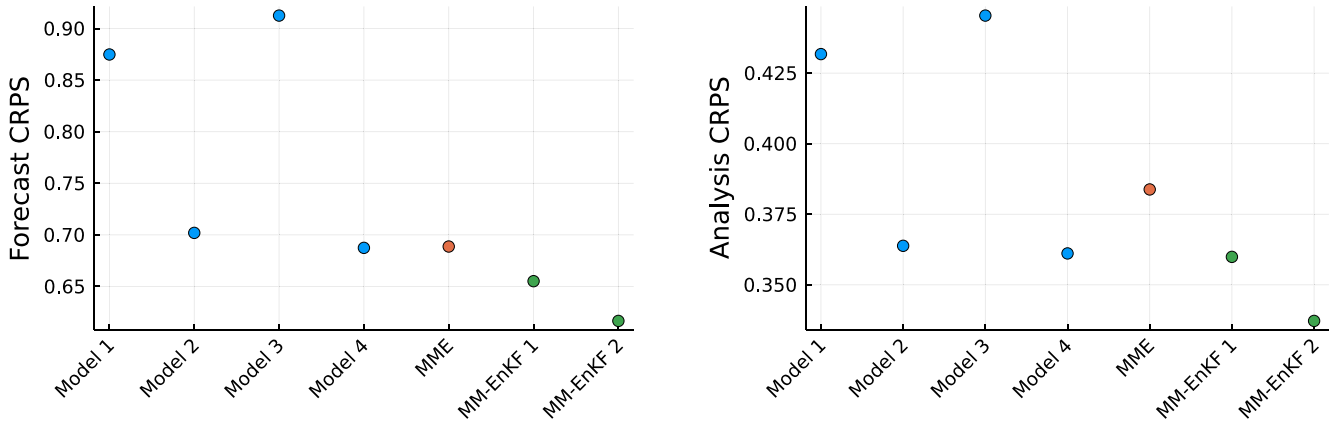


Figure 6. Overall performance of the multi-model ensemble Kalman filter with partial observations, in terms of continuous ranked probability score. The different models and multi-model combinations are indicated on the abscissa.

that for each forecast cycle, we obtain the initial ensembles from a previous analysis with observations having an error of $\mathbf{R} = 0.1\mathbf{I}$. We run 5,000 forecast cycles for each lead time, and compute the error statistics over the last 3,000 cycles.

Figure 7 shows that, for real-time forecasting, the MME error is similar to that of the best model, while the MM-EnKF consistently outperforms the MME and the individual models until the forecast errors start to saturate.

4.2.2.1. Recursive Multi-Step Forecasts

We then try to apply the multi-model forecasting recursively. After an interval of 0.2, we form the multi-model forecast and use it as the initial conditions for the next interval. Figure 8 shows that this results in much greater error reductions, while Method 2 has again a slight advantage.

4.2.3. Impact of Flow Dependence

In order to estimate the impact of flow dependence in the weights, we test a 3D-Var–like version of the filter, wherein instead of using the ensemble-estimated $\mathbf{P}^{\mathrm{f}}_m$ for each model, we use a static $\mathbf{B}_m$. This version is more similar to that of Logutov and Robinson (2005), which uses static forecast covariance matrices; it differs, though, from 3D-Var in that we keep the ensemble for the state update. These $\mathbf{B}_m$, instead of representing an instantaneous estimate of the forecast error covariance, represent the models' long-term statistical properties, and are often referred to as climatological error covariance matrices. We estimate these $\mathbf{B}_m$ by averaging the ensemble-estimated $\mathbf{P}^{\mathrm{f}}_m$ over 100 cycles. We thus remove any flow dependence in the weights attached to the models and observations.

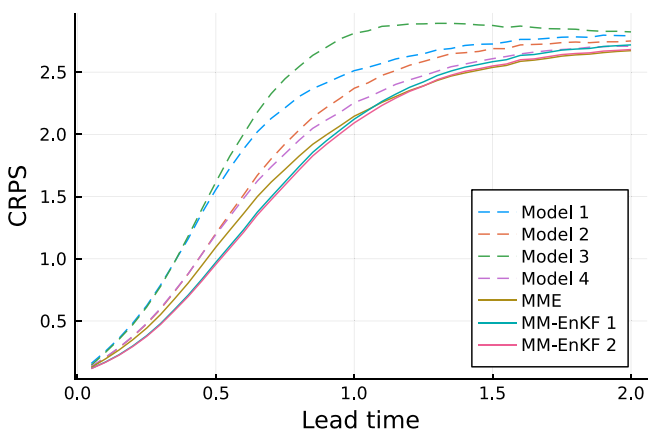


Figure 7. Lead time dependence of single- and multi-model forecast performance.

Comparing this non–flow-dependent version of the MM-EnKF to the flow-dependent one for both DA and forecasting in Table 2, we find that the flow-dependent MM-EnKF outperforms the non–flow-dependent version. Although the EnKF is generally known to outperform DA methods which lack flow dependence, such as 3D-Var, it is notable that the flow dependence also impacts the forecast skill. The flow dependence helps account for the uncertainty in the multi-model forecast, which is reflected in the improved CRPS. However, the flow dependence also improves the mean of the forecast ensemble, as reflected in the improved CRPS and RMSE, the latter depending only on the ensemble mean.

4.3. Experiments With Models of Different Fidelities

Suppose one has two models of different accuracy and computational cost: one is more computationally expensive and more accurate, the other less expensive and less accurate. Then, can a larger ensemble of the cheaper model improve DA or forecasts of the more expensive one? Such scenarios are often encountered in operational prediction where, due to constraints on

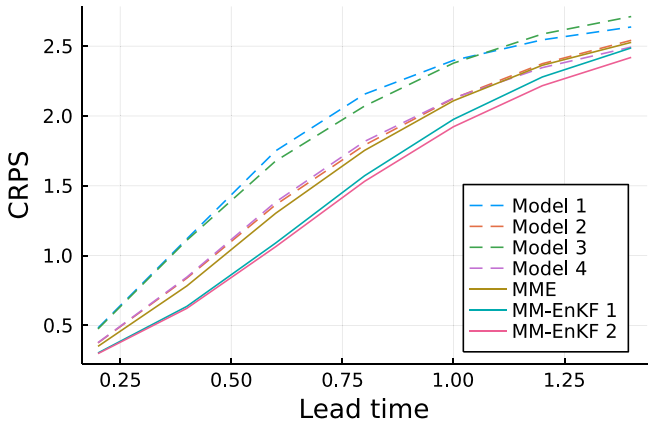


Figure 8. Continuous ranked probability score error metric of recursive single- and multi-model forecasts by lead time.

computational resources, only a small ensemble at a higher resolution can be afforded, but this can be supplemented by large low-resolution ensembles (Gascón et al., 2019). We test this scenario by applying Method 2 with models having different ensemble sizes, and additive model errors of different magnitudes.

We generate a 40×40 banded matrix $\mathbf{B}$ with bandwidth 20; the entries within the nonzero band are drawn from a uniform distribution $\mathcal{U}(0, 1)$. We then prescribe the model error for HF (for "high fidelity") to have covariance $\mathbf{Q}_1 = (1/10)(\mathbf{B} - 0.4\mathbf{J}_{40})(\mathbf{B} - 0.4\mathbf{J}_{40})^T$, and LF (for "low fidelity") to have covariance $\mathbf{Q}_2 = (\mathbf{B} - 0.4\mathbf{J}_{40})(\mathbf{B} - 0.4\mathbf{J}_{40})^T$, where $\mathbf{J}_{40}$ is the 40×40 matrix of ones.

We use a 5-member ensemble for HF and a 40-member ensemble for LF. In this case, the single-model forecasts do not have 45 ensemble members, rather 5 and 40, since a 45-member ensemble of HF would clearly outperform any MME which adds LF members at the expense of HF members. Rather, the question is whether the forecast skill of a small HF ensemble can be improved by adding LF members.

Figure 9 shows the performance of the MM-EnKF for recursive multi-step forecasts in this scenario. Here, the MME has error in between the errors of the HF and LF models, as would be expected from a simple average. On the other hand, the MM-EnKF clearly outperforms the MME and the 5-member ensemble of the more accurate HF model.

4.4. Experiments With Models With Different Resolved Scales

Here, we apply Method 1 with the two-scale Lorenz96 model being labeled HR (for "high-resolution") and the single-scale version being labeled LR (for "low-resolution"). The two-scale Lorenz96 model includes the small-scale dynamics $\{y_j\}$ of Equation 26 affecting the large scales $\{x_i\}$, while the single-scale version only includes the latter large scales of Equation 26a. This experiment thus serves as a test case for having ensembles at two different scales, one at higher resolution than the other.

The true model here is the two-scale Lorenz96 model with forcing as defined in Section 4.1. In this case, we prescribe an imperfect large-scale forcing of $F = 8.5$ for $1 \leq i \leq 10$ and $F = 9.5$ for $11 \leq i \leq 20$ for the higher-resolution model, while the lower-resolution model's forcing is perfect but model error is still present due to the unresolved scales.

Table 3 shows the results in terms of analysis RMSE in the large-scale variables $\{x_i\}$ and small-scale variables $\{y_{j,i}\}$. The results demonstrate that higher resolution is, at least in the present setting, more valuable than accurate forcing for the DA performance, and that the MM-EnKF provides further improvement over the better one of the two models.

We then test forecasting with the same two models. In these experiments, we obtain the ensembles at the beginning of each forecast cycle from a previous analysis with observations having an error of 10% of the climatological variance. For the MM-EnKF, we forecast recursively, combining the forecasts every 0.2 time units. We run 500 cycles and show the results for the last 200 cycles in Figure 10. The MM-EnKF again outperforms both individual models by a substantial margin.

Table 2
The Errors Obtained for the Static and Flow-Dependent Versions of the Filter

	Analysis CRPS	Forecast CRPS	Forecast RMSE
Static	0.235 ± 0.001	0.447 ± 0.003	0.813 ± 0.006
Flow dependent	**0.202** ± 0.001	**0.433** ± 0.003	**0.803** ± 0.007

Note. The ± indicates the standard error in the time mean. Bold indicates the lowest error in a column.

4.5. Implementation

We implemented the method in the Julia language, with the open-source code available (see Data Availability Statement). The code is modular, making it easy to add different DA methods and models. The CRPS error metric was computed with the proper scoring library (The Climate Corporation, 2015).

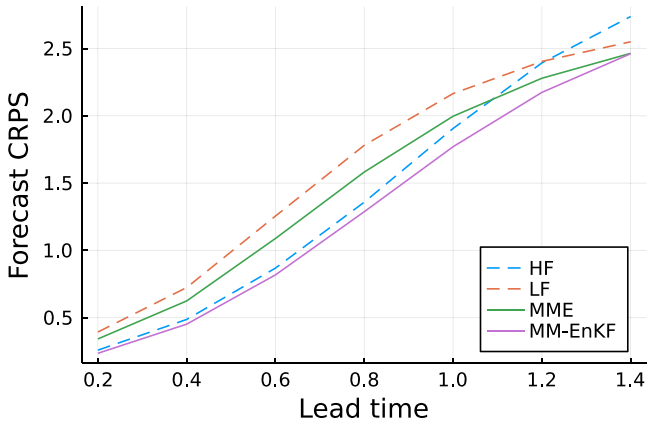


Figure 9. Continuous ranked probability score error metric of recursive multi-model forecasts by lead time.

We used the parasweep library for Python of Bach (2021) to facilitate the running in parallel of multiple experiments at different lead times and with different parameter values.

5. Concluding Remarks

5.1. Summary and Conclusions

In this paper, we proposed and implemented a MM-EnKF, based on the framework of Narayan et al. (2012). We addressed several gaps in previous work on MM-DA, including the formulation of an appropriate EnKF algorithm for high-dimensional systems and incorporation of model error estimation. Using numerical experiments with several versions of a chaotic model (Lorenz, 1996), we showed that the MM-EnKF is a robust and versatile method for making use of multiple imperfect models of a system in DA, as well as forecasting.

5.2. Future Work

In future work, the MM-EnKF could be applied in more high-dimensional and complex models, including operational numerical weather prediction models. Section 2.3 discusses the computational considerations for such high-dimensional systems. Because the MM-EnKF can be implemented by iteratively using an EnKF, it can be adopted in contexts where an EnKF-based assimilation system already exists. It is, moreover, non-obtrusive, meaning that it does not require changes to the model equations.

In our numerical experiments, we did not perform any bias correction. However, in climate contexts, where model biases—as opposed to model error that can be approximated as unbiased noise with covariance $\mathbf{Q}_m$—become increasingly important, one will need to address the model bias issue; see Section 2.1.1. Furthermore, the interpretation of the $\mathbf{Q}_m$ becomes unclear when the ensembles mix on the models' attractors; see Section 2.4. Future work should address whether, and if so, how, the MM-EnKF can be extended to such climate problems.

Of course, the application of the MM-EnKF would also require the availability of multiple model forecasts. Examples of such operational multi-model systems include the North American Multimodel Ensemble (NMME: Kirtman et al., 2014) and the North American Ensemble Forecast System (NAEFS: Candille, 2009). To our knowledge, there is not yet any operational DA systems that use multiple models. Secondarily, one would require the construction of the $\mathcal{G}_m$ operators mapping to a common space. Such operations are already used in multi-model forecast contexts, when the distinct forecasts have to be regridded to a common grid before averaging.

We discuss further avenues for future work below.

5.2.1. Correlated Forecast Errors

The formulation of MM-DA assumes that the forecast errors are uncorrelated from each other (Logutov & Robinson, 2005). This may not be a good assumption for climate models, especially when distinct models have the same historical provenance (Abramowitz et al., 2019; Christiansen, 2020; Knutti et al., 2010). Future work could formulate a multi-model Kalman filter which accounts for correlated forecast errors, following Kalman filters that include correlations between observation and model errors (Berry & Sauer, 2018; Section 7.1 in Simon, 2006). In fact, the derivation in Appendix A is easily modified for correlated forecast error. Although for a different problem, combining correlated state estimates arises in multi-sensor fusion (Kim, 1994; Sun, 2004), and similar ways of incorporating cross-correlation information could be tested with the MM-EnKF.

Table 3
The Analysis Root-Mean-Square Error Over the Large-Scale Variables $\{x_i\}$ and Small-Scale Variables $\{y_{i,j}\}$ in Data Assimilation Experiments With the Two-Scale Lorenz96 Model

	Analysis RMSE in x_i	Analysis RMSE in $y_{i,j}$
HR	0.511 ± 0.007	0.072 ± 0.002
LR	0.540 ± 0.004	–
MM-EnKF	**0.447** ± 0.005	**0.066** ± 0.001

Note. Here the analysis window is 0.05, with 2,000 cycles and errors averaged over the last 500. Bold indicates the lowest error in a column.

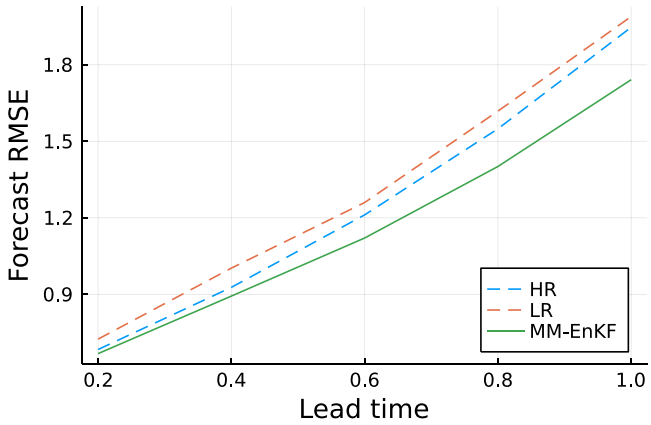


Figure 10. Forecast root-mean-square error in the large-scale variables $\{x_i\}$ by lead time.

5.2.2. Hybrid Forecasting and DA

Hybrid methods combining statistical or machine learning (ML) forecasts with a dynamical model of a system are a promising approach for improving on pure dynamical forecasts. Pathak et al. (2018) demonstrated the advantage of hybrid models in forecasting high-dimensional chaotic systems, showing that a hybrid that combines an ML forecast with a forecast from an imperfect dynamical model can be skillful for longer than either one individually. In Bach et al. (2021), the authors also demonstrated the advantage of combining a data-driven model with a dynamical model in leveraging the predictability of a system's oscillatory modes. In Chattopadhyay et al. (2023), the authors demonstrated that augmenting an atmospheric model ensemble with a large ensemble of deep learning–based forecasts can significantly improve estimation of the forecast covariance in an EnKF.

The MM-EnKF could be tested for hybrid DA and forecasting. As demonstrated in Sections 4.3 and 4.4, MM-EnKF is able to successfully combine models of different accuracy and resolution. This feature could be used for combining physical and data-driven forecasts: namely, one of the model ensembles could be generated by a physical model and the other one by a data-driven model.

5.2.3. Multi-Fidelity and Multi-Resolution Forecasting and DA

Given a limited computational budget, it appears to be advantageous to supplement a small ensemble of expensive, high-fidelity model runs with a large ensemble of cheaper, lower-fidelity runs. Future work could further explore the use of the MM-EnKF for combining ensembles at multiple fidelities or multiple resolutions. The MM-EnKF could also be tested for combining a global atmospheric forecast with multiple higher-resolution limited-area models, as in Kretschmer et al. (2015).

In addition to models that can be numerically simulated at lower resolution, another class of low-fidelity models consists of reduced-order models (ROMs) that approximate a high-fidelity model by dynamics in a lower-dimensional space. With projection-based model order reduction methods, in particular, one can project from the higher-dimensional space to the reduced space (Amsallem & Farhat, 2012; Antoulas, 2005). This idea allows one to combine high-fidelity models and ROMs within the MM-EnKF framework by defining the $\mathbf{G}_m$ operators to map from the full to the reduced space.

5.2.4. Multi-Model Smoothing

While filtering is the problem of optimally estimating the state of a system given all observations prior to the analysis time, smoothing takes into account also observations of the system obtained after the analysis time. Various forms of ensemble Kalman smoothers have been developed (Evensen, 2018), and future work could adapt the MM-EnKF to solve the smoothing problem. For climate applications, this could enable the development of multi-model reanalyses.

Appendix A: Optimality of the Multi-Model Kalman Filter as the Linear Minimum Variance Estimator

We state this result in the form of a theorem and provide its proof herewith.

Theorem 1. *Let* $\{\hat{\mathbf{x}}_\ell\}_{\ell=1}^{L}$ *be unbiased state estimates of the n-dimensional vector* $\mathbf{x}$ *under the linear transformation* $\mathbf{G}_\ell$*, such that*

$$\hat{\mathbf{x}}_\ell = \mathbf{G}_\ell \mathbf{x} + \mathbf{e}_\ell, \tag{A1}$$

where $\mathbb{E}[\mathbf{e}_\ell] = \mathbf{0}$, $\mathbb{E}\left[\mathbf{e}_\ell \mathbf{e}_\ell^T\right] = \mathbf{P}_\ell$, *and* $\mathbb{E}\left[\mathbf{e}_\ell \mathbf{e}_{\ell'}^T\right] = \mathbf{0}$ *when* $\ell \neq \ell'$.
Then, the minimum variance linear unbiased estimator of $\mathbf{x}$ *is given by*

$$\hat{\mathbf{x}} = \sum_{\ell=1}^{L} \mathbf{A}_\ell \hat{\mathbf{x}}_\ell, \tag{A2}$$

where

$$\mathbf{A}_\ell = \left(\sum_{\ell'=1}^{L} \mathbf{G}_{\ell'}^T \mathbf{P}_{\ell'}^{-1} \mathbf{G}_{\ell'} \right)^{-1} \mathbf{G}_\ell^T \mathbf{P}_\ell^{-1}. \tag{A3}$$

In multi-sensor information fusion, a problem of the same form appears, except that all the $\mathbf{G}_\ell = \mathbf{I}$. For that case, minimum-variance optimality has been proven in Corollary 1 of Sun (2004). We largely follow the latter proof, but allow for general $\mathbf{G}_\ell$.

Proof Begin by defining an estimator $\hat{\mathbf{x}}$ of $\mathbf{x}$ as a linear combination of the $\hat{\mathbf{x}}_\ell$:

$$\hat{\mathbf{x}} = \sum_{\ell=1}^{L} \mathbf{A}_\ell \hat{\mathbf{x}}_\ell. \tag{A4}$$

Taking the expectation of $\hat{\mathbf{x}}$ and using the linearity of the expectation operator,

$$\mathbb{E}\left[\hat{\mathbf{x}}\right] = \sum_{\ell=1}^{L} \mathbf{A}_\ell \mathbb{E}\left[\hat{\mathbf{x}}_\ell\right] = \sum_{\ell=1}^{L} \mathbf{A}_\ell \mathbf{G}_\ell \mathbb{E}[\mathbf{x}]. \tag{A5}$$

Then, in order for $\hat{\mathbf{x}}$ to be unbiased—namely $\mathbb{E}\left[\hat{\mathbf{x}}\right] = \mathbb{E}[\mathbf{x}]$—we must have

$$\sum_{\ell=1}^{L} \mathbf{A}_\ell \mathbf{G}_\ell = \mathbf{I}. \tag{A6}$$

The error $\tilde{\mathbf{x}}$ in $\hat{\mathbf{x}}$ can be expressed as

$$\tilde{\mathbf{x}} = \mathbf{x} - \hat{\mathbf{x}} = \sum_{\ell=1}^{L} \mathbf{A}_\ell (\mathbf{G}_\ell \mathbf{x} - \hat{\mathbf{x}}_\ell), \tag{A7}$$

with covariance matrix

$$\mathbf{P} = \mathbb{E}\left[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T\right] = \sum_{\ell=1}^{L} \mathbf{A}_\ell \mathbf{P}_\ell \mathbf{A}_\ell^T. \tag{A8}$$

In order to obtain the minimum variance estimator, we wish to minimize $J \equiv \mathrm{tr}(\mathbf{P})$. By linearity of the trace,

$$J = \sum_{\ell=1}^{L} \mathrm{tr}\left(\mathbf{A}_\ell \mathbf{P}_\ell \mathbf{A}_\ell^T\right). \tag{A9}$$

We minimize J using the method of Lagrange multipliers (see, e.g., Boyd & Vandenberghe, 2004). The Lagrangian $\mathcal{L}$ is defined as follows:

$$\mathcal{L} = J + \sum_{j=1}^{n} \left[\lambda_j^T \left(\sum_{\ell=1}^{L} \mathbf{A}_\ell \mathbf{G}_\ell - \mathbf{I} \right) \mathbf{e}_j \right], \tag{A10}$$

where $\boldsymbol{\lambda}_j = \left[\lambda_{1j}, \ldots, \lambda_{nj}\right]^T$ is the jth vector of Lagrange multipliers and $\mathbf{e}_j$ is a vector with a 1 in the jth coordinate and zeros elsewhere. A necessary condition for $\mathcal{L}$ to have a stationary point is that

$$\frac{\partial \mathcal{L}}{\partial \mathbf{A}_\ell} = \mathbf{A}_\ell \mathbf{P}_\ell + \frac{1}{2} \boldsymbol{\Lambda} \mathbf{G}_\ell^T = \mathbf{0} \tag{A11}$$

for all ℓ. We gather Equations A6 and A11 into a block matrix equation:

$$\begin{pmatrix} \boldsymbol{\Sigma} & \overline{\mathbf{G}} \\ \overline{\mathbf{G}}^T & \mathbf{0} \end{pmatrix} \begin{pmatrix} \overline{\mathbf{A}} \\ \frac{1}{2}\boldsymbol{\Lambda}^T \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{I} \end{pmatrix}, \tag{A12}$$

in which $\boldsymbol{\Sigma}$, $\overline{\mathbf{A}}$, and $\overline{\mathbf{G}}$ are the block matrices

$$\boldsymbol{\Sigma} = \begin{pmatrix} \mathbf{P}_1 & & \\ & \ddots & \\ & & \mathbf{P}_M \end{pmatrix}, \overline{\mathbf{A}} = \begin{pmatrix} \mathbf{A}_1^T \\ \vdots \\ \mathbf{A}_M^T \end{pmatrix}, \overline{\mathbf{G}} = \begin{pmatrix} \mathbf{G}_1 \\ \vdots \\ \mathbf{G}_M \end{pmatrix}. \tag{A13}$$

Using a block matrix inversion identity on Equation A12, we obtain

$$\overline{\mathbf{A}} = \boldsymbol{\Sigma}^{-1}\overline{\mathbf{G}}\left(\overline{\mathbf{G}}^T\boldsymbol{\Sigma}^{-1}\overline{\mathbf{G}}\right)^{-1}, \tag{A14}$$

which implies

$$\mathbf{A}_\ell = \left(\sum_{\ell'=1}^{L} \mathbf{G}_{\ell'}^T\mathbf{P}_{\ell'}^{-1}\mathbf{G}_{\ell'}\right)^{-1} \mathbf{G}_\ell^T\mathbf{P}_\ell^{-1}. \tag{A15}$$

Note that the $\mathbf{A}_\ell\mathbf{G}_\ell$ are positive semidefinite matrices. Thus, in the scalar case, Equation A6 is a convex linear combination, that is, the weights $\mathbf{A}_\ell\mathbf{G}_\ell$ are nonnegative and sum to 1. The multivariate case generalizes this property by having the weights be positive semidefinite matrices that sum to the identity matrix.

To apply this theorem to the assimilation step of the multi-model Kalman filter, take $L = M + 1$. Then, for $m = 1, \ldots, M$, take $\hat{\mathbf{x}}_m = \mathbf{x}_m^{\mathrm{f}}$ and $\mathbf{P}_m = \mathbf{P}_m^{\mathrm{f}}$. Finally, take $\hat{\mathbf{x}}_{M+1} = \mathbf{y}$, $\mathbf{P}_{M+1} = \mathbf{R}$, and $\mathbf{G}_{M+1} = \mathbf{H}$. At this point, identifying $\hat{\mathbf{x}}$ with $\mathbf{x}^{\mathrm{a}}$, we recover Equation 3 of Section 1.2.1.

Appendix B: Model Error Estimation Method

Here, we suggest a method for estimating $\mathbf{Q}$ that is closely related to the one of Berry and Sauer (2013) and Hamilton et al. (2016), but we assume that the observation noise covariance $\mathbf{R}$ is known. This assumption allows us to derive a simple estimate for $\mathbf{Q}$ that does not require either lagged innovations or the gain matrix. Nor is model linearization required in the case of an EnKF applied to a nonlinear forward model.

The method for estimating $\mathbf{Q}$ relies on the statistics of the innovations $\mathbf{d}(t_i) = \mathbf{y}(t_i) - \mathbf{H}\mathbf{x}^{\mathrm{f}}(t_i)$, which equal the difference between observations and forecasts. A standard result for the Kalman filter states that

$$\mathbb{E}\left[\mathbf{d}(t_i)\mathbf{d}(t_i)^T\right] = \mathbf{H}\mathbf{P}^{\mathrm{f}}(t_i)\mathbf{H}^T + \mathbf{R}; \tag{B1}$$

see, for instance, Desroziers et al. (2005) or Section 10.1 of Simon (2006).

If the state is not fully observed, as is usually the case in DA problems, then $\mathbf{H}$ is not invertible. However, for idealized cases when $\mathbf{H}$ is invertible, we can obtain an estimate $\hat{\mathbf{Q}}$ of $\mathbf{Q}$ by substituting Equation 14 into Equation B1. and rearranging:

$$\hat{\mathbf{Q}}(t_{i-1}) = \mathbf{H}^{-1}\left(\mathbb{E}\left[\mathbf{d}(t_i)\mathbf{d}(t_i)^T\right] - \mathbf{R} - \mathbf{H}\mathbf{P}^{\mathrm{p}}(t_i)\mathbf{H}^T\right)\mathbf{H}^{-T}. \tag{B2}$$

See Section B2 below for the general case in which $\mathbf{H}$ is not invertible.

In order to avoid abrupt changes in $\hat{\mathbf{Q}}$ over time, and to preserve positive semidefiniteness (see below), a temporal smoothing needs to be applied:

$$\tilde{\mathbf{Q}}(t_{i+1}) = \delta\hat{\mathbf{Q}}(t_i) + (1 - \delta)\tilde{\mathbf{Q}}(t_i), \tag{B3}$$

where $0 < \delta < 1$ is a tunable parameter (Berry & Sauer, 2013; Tandeo et al., 2020), and $\tilde{\mathbf{Q}}$ is the smoothed estimate. Then, $\mathbf{P}^{\mathrm{f}}(t_{i+1})$ is estimated by adding $\tilde{\mathbf{Q}}(t_i)$ to the $\mathbf{P}^{\mathrm{p}}$ estimated by the filter. In what follows, we drop the time indices for simplicity.

Covariance matrices must be positive semidefinite: in other words, their eigenvalues are real and nonnegative, that is, $\lambda_{\min} \geq 0$. Due to the observation noise entering the $\mathbb{E}\left[\mathbf{d}\mathbf{d}^T\right]$ term in Equation B2, the estimate $\tilde{\mathbf{Q}}$ can often lack this property. To avoid this problem, a small enough δ must be chosen, and the "initial guess" $\tilde{\mathbf{Q}}(t_0)$ should be positive semidefinite. When forecasting at multiple lead times, we initialize at lead $k\tau$ by $\tilde{\mathbf{Q}}(t_0)k^2$, inspired by the quadratic growth of model error described in Carrassi, Vannitsem, and Nicolis (2008).

In general, the larger the observation noise relative to the model error, the smaller δ must be. However, if the estimated $\tilde{\mathbf{Q}}$ does become indefinite at some t_j, definiteness can be restored. The matrix satisfying $\lambda_{\min} \geq \epsilon$ that is nearest in the Frobenius norm $\|\cdot\|_F$ (Horn & Johnson, 2013) to the problematic one at $t = t_j$ can be computed by using the spectral decomposition and setting all $\lambda_i < \epsilon$ to ϵ (Cheng & Higham, 1998).

B1. Ensemble Filters

In the case of an ensemble Kalman filter, we estimate $\mathbb{E}\left[\mathbf{d}\mathbf{d}^T\right] \simeq \left(\mathbf{y} - \mathbf{H}\bar{\mathbf{x}}^{\mathrm{f}}\right)\left(\mathbf{y} - \mathbf{H}\bar{\mathbf{x}}^{\mathrm{f}}\right)^T$, where $\bar{\mathbf{x}}^{\mathrm{f}}$ is the mean of the forecast ensemble.

In ensemble filters, $\mathbf{P}^{\mathrm{p}}$ is estimated as

$$\mathbf{P}^{\mathrm{p}} = \frac{1}{m-1}\sum_{i=1}^{m}\left(\mathbf{x}_i^{\mathrm{f}} - \bar{\mathbf{x}}^{\mathrm{f}}\right)\left(\mathbf{x}_i^{\mathrm{f}} - \bar{\mathbf{x}}^{\mathrm{f}}\right)^T, \tag{B4}$$

where $\mathbf{x}_i^{\mathrm{f}}$ is the ith ensemble member and m is the ensemble size. We use this $\mathbf{P}^{\mathrm{p}}$ directly in Equation B2, thus avoiding the need for a tangent linear model, as in Equation 8, when $\mathcal{M}$ is nonlinear.

B2. Rank-Deficient Observations

When $\mathbf{H}$ is not invertible, we can find a solution that minimizes the Frobenius norm, as in Berry and Sauer (2013). We let $\hat{\mathbf{Q}}$ in Equation B3. be a linear combination of fixed matrices, $\hat{\mathbf{Q}} = \sum_p q_p \mathbf{Q}_p$. This formulation can be used to specify a simplified structure, such as a diagonal matrix or a block-constant one.

Let $\mathbf{q}$ be the vector of coefficients $\{q_p\}$. Then,

$$\mathbf{q} = \underset{\{q_p\}}{\arg\min}\left\|\mathbf{C} - \sum_p q_p \mathbf{H}\mathbf{Q}_p\mathbf{H}^T\right\|_F, \tag{B5}$$

where

$$\mathbf{C} = \mathbb{E}\left[\mathbf{d}\mathbf{d}^T\right] - \mathbf{R} - \mathbf{H}\mathbf{P}^{\mathrm{p}}\mathbf{H}^T. \tag{B6}$$

The minimization in Equation B5. is carried out by finding the least-squares solution of

$$\mathbf{A}\mathbf{q} \simeq \mathrm{vec}(\mathbf{C}), \tag{B7}$$

where the pth column of $\mathbf{A}$ is $\mathrm{vec}(\mathbf{H}\mathbf{Q}_p\mathbf{H}^T)$.

Data Availability Statement

Version 2022-12 of the Julia code implementing the MM-EnKF used in this manuscript is preserved at Bach (2022), available via the MIT License and developed openly at https://github.com/eviatarbach/mmda. No data was used in this study. Scripts for numerical experiments are available in the MM-EnKF repository.

Acknowledgments
We thank Marc Bocquet for several helpful suggestions, V. Balaji for discussions on correlated model error, Safa Mote for helpful discussions regarding hybrid methods, Tapio Schneider for discussions on smoothing applications, and four anonymous referees for additional suggestions. E.B. was funded by the Make Our Planet Great Again (MOPGA) postdoctoral program of the French Ministry for Europe and Foreign Affairs. The present work is TiPES contribution #142; the TiPES (Tipping Points in the Earth System) project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No. 820970. M.G. acknowledges support by the EIT Climate-KIC; EIT Climate-KIC is supported by the European Institute of Innovation & Technology (EIT), a body of the European Union.

References

Abarbanel, H. D. I., Shirman, S., Breen, D., Kadakia, N., Rey, D., Armstrong, E., & Margoliash, D. (2017). A unifying view of synchronization for data assimilation in complex nonlinear networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *27*(12), 126802. https://doi.org/10.1063/1.5001816

Abramowitz, G., Herger, N., Gutmann, E., Hammerling, D., Knutti, R., Leduc, M., et al. (2019). ESD Reviews: Model dependence in multi-model climate ensembles: Weighting, sub-selection and out-of-sample testing. *Earth System Dynamics*, *10*(1), 91–105. https://doi.org/10.5194/esd-10-91-2019

Akca, A., & Efe, M. Ö. (2019). Multiple model Kalman and particle filters and applications: A survey. *IFAC-PapersOnLin*e, *52*(3), 73–78. https://doi.org/10.1016/j.ifacol.2019.06.013

Amsallem, D., & Farhat, C. (2012). Stabilization of projection-based reduced-order models. *International Journal for Numerical Methods in Engineering*, *91*(4), 358–377. https://doi.org/10.1002/nme.4274

Anandalingam, G., & Chen, L. (1989). Linear combination of forecasts: A general Bayesian model. *Journal of Forecasting*, *8*(3), 199–214. https://doi.org/10.1002/for.3980080306

Anderson, J. (2009). Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus A: Dynamic Meteorology and Oceanography*, *61*(1), 72–83. https://doi.org/10.1111/j.1600-0870.2007.00361.x

Antoulas, A. C. (2005). *Approximation of large-scale dynamical systems*. Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9780898718713

Asch, M., Bocquet, M., & Nodet, M. (2016). *Data assimilation: Methods, algorithms, and applications*. Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611974546

Bach, E. (2021). parasweep: A template-based utility for generating, dispatching, and post-processing of parameter sweeps. *SoftwareX*, *13*, 100631. https://doi.org/10.1016/j.softx.2020.100631

Bach, E. (2022). eviatarbach/mm-enkf (Version 2022-12) [Software]. Zenodo. https://doi.org/10.5281/zenodo.7450931

Bach, E., Mote, S., Krishnamurthy, V., Sharma, A. S., Ghil, M., & Kalnay, E. (2021). Ensemble oscillation correction (EnOC): Leveraging oscillatory modes to improve forecasts of chaotic systems. *Journal of Climate*, *34*(14), 5673–5686. https://doi.org/10.1175/JCLI-D-20-0624.1

Baek, S.-J., Hunt, B. R., Kalnay, E., Ott, E., & Szunyogh, I. (2006). Local ensemble Kalman filtering in the presence of model bias. *Tellus A: Dynamic Meteorology and Oceanography*, *58*(3), 293–306. https://doi.org/10.1111/j.1600-0870.2006.00178.x

Bar-Shalom, Y., Li, X.-R., & Kirubarajan, T. (2001). *Estimation with applications to tracking and navigation: Theory, algorithms and software*. John Wiley & Sons, Ltd. https://doi.org/10.1002/0471221279

Bates, J. M., & Granger, C. W. J. (1969). The combination of forecasts. *Operational Research Quarterly*, *20*(4), 451–468. https://doi.org/10.2307/3008764

Bellantoni, J. F., & Dodge, K. W. (1967). A square root formulation of the Kalman-Schmidt filter. *AIAA Journal*, *5*(7), 1309–1314. https://doi.org/10.2514/3.4189

Berry, T., & Sauer, T. (2013). Adaptive ensemble Kalman filtering of non-linear systems. *Tellus A: Dynamic Meteorology and Oceanography*, *65*(1), 20331. https://doi.org/10.3402/tellusa.v65i0.20331

Berry, T., & Sauer, T. (2018). Correlation between system and observation errors in data assimilation. *Monthly Weather Review*, *146*(9), 2913–2931. https://doi.org/10.1175/MWR-D-17-0331.1

Bierman, G. J. (1977). *Factorization methods for discrete sequential estimation*. Academic Press.

Bishop, C. H., Etherton, B. J., & Majumdar, S. J. (2001). Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Monthly Weather Review*, *129*(3), 420–436. https://doi.org/10.1175/1520-0493(2001)129⟨0420:ASWTET⟩2.0.CO;2

Boyd, S., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge University Press. https://doi.org/10.1017/CBO9780511804441

Candille, G. (2009). The multiensemble approach: The NAEFS example. *Monthly Weather Review*, *137*(5), 1655–1665. https://doi.org/10.1175/2008MWR2682.1

Carrassi, A., Bocquet, M., Bertino, L., & Evensen, G. (2018). Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *Wiley Interdisciplinary Reviews: Climate Change*, *9*(5), e535. https://doi.org/10.1002/wcc.535

Carrassi, A., Ghil, M., Trevisan, A., & Uboldi, F. (2008). Data assimilation as a nonlinear dynamical systems problem: Stability and convergence of the prediction-assimilation system. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *18*(2), 023112. https://doi.org/10.1063/1.2909862

Carrassi, A., Vannitsem, S., & Nicolis, C. (2008). Model error and sequential data assimilation: A deterministic formulation. *Quarterly Journal of the Royal Meteorological Society*, *134*(634), 1297–1313. https://doi.org/10.1002/qj.284

Chattopadhyay, A., Mustafa, M., Hassanzadeh, P., Bach, E., & Kashinath, K. (2022). Towards physics-inspired data-driven weather forecasting: Integrating data assimilation with a deep spatial-transformer-based U-NET in a case study with ERA5. *Geoscientific Model Development*, *15*(5), 2221–2237. https://doi.org/10.5194/gmd-15-2221-2022

Chattopadhyay, A., Nabizadeh, E., Bach, E., & Hassanzadeh, P. (2023). Deep learning-enhanced ensemble-based data assimilation for high-dimensional nonlinear dynamical systems. *Journal of Computational Physics*, 111918. https://doi.org/10.1016/j.jcp.2023.111918

Chen, Y., & Stechmann, S. N. (2019). Multi-model communication and data assimilation for mitigating model error and improving forecasts. *Chinese Annals of Mathematics, Series B*, *40*(5), 689–720. https://doi.org/10.1007/s11401-019-0157-1

Cheng, S. H., & Higham, N. J. (1998). A modified Cholesky algorithm based on a symmetric indefinite factorization. *SIAM Journal on Matrix Analysis and Applications*, *19*(4), 1097–1110. https://doi.org/10.1137/S0895479896302898

Christiansen, B. (2020). Understanding the distribution of multimodel ensembles. *Journal of Climat*e, *33*(21), 9447–9465. https://doi.org/10.1175/JCLI-D-20-0186.1

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, *5*(4), 559–583. https://doi.org/10.1016/0169-2070(89)90012-5

The Climate Corporation. (2015). Properscoring [Software]. GitHub. Retrieved from https://github.com/TheClimateCorporation/properscoring

Coelho, E. F., Hogan, P., Jacobs, G., Thoppil, P., Huntley, H. S., Haus, B. K., et al. (2015). Ocean current estimation using a multi-model ensemble Kalman filter during the grand Lagrangian deployment experiment (GLAD). *Ocean Modelling*, *87*, 86–106. https://doi.org/10.1016/j.ocemod.2014.11.001

Counillon, F., Keenlyside, N. S., Wang, S., Devilliers, M., Gupta, A. K., Koseki, S., & Shen, M.-L. (2022). Framework for an ocean-connected supermodel of the Earth System. *Earth and Space Science Open Archive*. https://doi.org/10.1002/essoar.10512643.1

Cramer, E. Y., Ray, E. L., Lopez, V. K., Bracher, J., Brennen, A., Rivadeneira, A. J. C., et al. (2022). Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the United States. *Proceedings of the National Academy of Sciences*, *119*(15), e2113561119. https://doi.org/10.1073/pnas.2113561119

Desroziers, G., Berre, L., Chapnik, B., & Poli, P. (2005). Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society*, *131*(613), 3385–3396. https://doi.org/10.1256/qj.05.108

Du, H., & Smith, L. A. (2017). Multi-model cross-pollination in time. *Physica D: Nonlinear Phenomena*, *353–354*, 31–38. https://doi.org/10.1016/j.physd.2017.06.001

Duane, G. S., Grabow, C., Selten, F., & Ghil, M. (2017). Introduction to focus issue: Synchronization in large networks and continuous media—Data, models, and supermodels. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *27*(12), 126601. https://doi.org/10.1063/1.5018728

Dumont Le Brazidec, J., Bocquet, M., Saunier, O., & Roustan, Y. (2021). Quantification of uncertainties in the assessment of an atmospheric release source applied to the autumn 2017 106Ru event. *Atmospheric Chemistry and Physics*, *21*(17), 13247–13267. https://doi.org/10.5194/acp-21-13247-2021

Duník, J., Straka, O., Kost, O., & Havlík, J. (2017). Noise covariance matrices in state-space models: A survey and comparison of estimation methods—Part I. *International Journal of Adaptive Control and Signal Processing*, *31*(11), 1505–1543. https://doi.org/10.1002/acs.2783

Evensen, G. (2003). The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, *53*(4), 343–367. https://doi.org/10.1007/s10236-003-0036-9

Evensen, G. (2018). Analysis of iterative ensemble smoothers for solving inverse problems. *Computational Geosciences*, *22*(3), 885–908. https://doi.org/10.1007/s10596-018-9731-y

Farchi, A., Bocquet, M., Laloyaux, P., Bonavita, M., & Malartic, Q. (2021). A comparison of combined data assimilation and machine learning methods for offline and online model error correction. *Journal of Computational Science*, *55*, 101468. https://doi.org/10.1016/j.jocs.2021.101468

Fragoso, T. M., Bertoli, W., & Louzada, F. (2018). Bayesian model averaging: A systematic review and conceptual classification. *International Statistical Review*, *86*(1), 1–28. https://doi.org/10.1111/insr.12243

Gascón, E., Lavers, D., Hamill, T. M., Richardson, D. S., Ben Bouallègue, Z., Leutbecher, M., & Pappenberger, F. (2019). Statistical postprocessing of dual-resolution ensemble precipitation forecasts across Europe. *Quarterly Journal of the Royal Meteorological Society*, *145*(724), 3218–3235. https://doi.org/10.1002/qj.3615

Gaspari, G., & Cohn, S. E. (1999). Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, *125*(554), 723–757. https://doi.org/10.1002/qj.49712555417

Gerding, S., & Myers, B. (2003). *Adaptive data fusion of meteorological forecast modules*. Paper presented at 3rd Conference on Artificial Intelligence Applications, American Meteorological Society, Long Beach, CA. https://ams.confex.com/ams/annual2003/techprogram/paper_55795.htm

Gharamti, M. E. (2018). Enhanced adaptive inflation algorithm for ensemble filters. *Monthly Weather Review*, *146*(2), 623–640. https://doi.org/10.1175/MWR-D-17-0187.1

Ghil, M., Cohn, S., Tavantzis, J., Bube, K., & Isaacson, E. (1981). Applications of estimation theory to numerical weather prediction. In L. Bengtsson, M. Ghil, & E. Källén (Eds.), *Dynamic meteorology: Data assimilation methods* (pp. 139–224). Springer. https://doi.org/10.1007/978-1-4612-5970-1_5

Gonzalez, P. L. M., Brayshaw, D. J., & Ziel, F. (2021). A new approach to extended-range multimodel forecasting: Sequential learning algorithms. *Quarterly Journal of the Royal Meteorological Society*, *147*(741), 4269–4282. https://doi.org/10.1002/qj.4177

Grewal, M. S., & Andrews, A. P. (2010). Applications of Kalman filtering in aerospace 1960 to the present [historical perspectives]. *IEEE Control Systems Magazine*, *30*(3), 69–78. https://doi.org/10.1109/MCS.2010.936465

Hagedorn, R., Doblas-Reyes, F. J., & Palmer, T. N. (2005). The rationale behind the success of multi-model ensembles in seasonal forecasting—I. Basic concept. *Tellus*, *57*(3), 219–233. https://doi.org/10.1111/j.1600-0870.2005.00103.x

Hager, W. W. (1989). Updating the inverse of a matrix. *SIAM Review*, *31*(2), 221–239. https://doi.org/10.1137/1031049

Hamill, T. M., & Whitaker, J. S. (2005). Accounting for the error due to unresolved scales in ensemble data assimilation: A comparison of different approaches. *Monthly Weather Review*, *133*(11), 3132–3147. https://doi.org/10.1175/MWR3020.1

Hamilton, F., Berry, T., & Sauer, T. (2016). Ensemble Kalman filtering without a model. *Physical Review X*, *6*(1), 011021. https://doi.org/10.1103/PhysRevX.6.011021

Hersbach, H. (2000). Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, *15*(5), 559–570. https://doi.org/10.1175/1520-0434(2000)015⟨0559:DOTCRP⟩2.0.CO;2

Hodyss, D., & Nichols, N. (2015). The error of representation: Basic understanding. *Tellus A: Dynamic Meteorology and Oceanography*, *67*(1), 24822. https://doi.org/10.3402/tellusa.v67.24822

Hoel, H., Law, K. J. H., & Tempone, R. (2016). Multilevel ensemble Kalman filtering. *SIAM Journal on Numerical Analysis*, *54*(3), 1813–1839. https://doi.org/10.1137/15M100955X

Hoel, H., Shaimerdenova, G., & Tempone, R. (2020). Multilevel Ensemble Kalman Filtering based on a sample average of independent EnKF estimators. *Foundations of Data Science*, *2*(4), 351–390. https://doi.org/10.3934/fods.2020017

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, *14*(4), 382–417. https://doi.org/10.1214/ss/1009212519

Horn, R. A., & Johnson, C. R. (2013). *Matrix analysis* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/9781139020411

Houtekamer, P. L., & Mitchell, H. L. (2001). A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, *129*(1), 123–137. https://doi.org/10.1175/1520-0493(2001)129⟨0123:ASEKFF⟩2.0.CO;2

Houtekamer, P. L., & Zhang, F. (2016). Review of the ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, *144*(12), 4489–4532. https://doi.org/10.1175/MWR-D-15-0440.1

Hunt, B. R., Kostelich, E. J., & Szunyogh, I. (2007). Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena*, *230*(1), 112–126. https://doi.org/10.1016/j.physd.2006.11.008

Ide, K., Courtier, P., Ghil, M., & Lorenc, A. C. (1997). Unified notation for data assimilation: Operational, sequential and variational. *Journal of the Meteorological Society of Japan. Ser. II*, *75*(1B), 181–189. https://doi.org/10.2151/jmsj1965.75.1B_181

Jazwinski, A. H. (1970). *Stochastic processes and filtering theory*. Academic Press, Inc.

Julier, S. J., & Uhlmann, J. K. (2004). Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, *92*(3), 401–422. https://doi.org/10.1109/JPROC.2003.823141

Kalnay, E. (2002). *Atmospheric modeling, data assimilation and predictability*. Cambridge University Press.

Kim, K. (1994). Development of track to track fusion algorithms. In *Proceedings of 1994 American Control Conference—ACC '94* (Vol. 1, pp. 1037–1041). https://doi.org/10.1109/ACC.1994.751905

Kirtman, B. P., Min, D., Infanti, J. M., Kinter, J. L., Paolino, D. A., Zhang, Q., et al. (2014). The North American multimodel ensemble: Phase-1 seasonal-to-interannual prediction; Phase-2 toward developing intraseasonal prediction. *Bulletin of the American Meteorological Society*, *95*(4), 585–601. https://doi.org/10.1175/BAMS-D-12-00050.1

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., & Meehl, G. A. (2010). Challenges in combining projections from multiple climate models. *Journal of Climate*, *23*(10), 2739–2758. https://doi.org/10.1175/2009JCLI3361.1

Kotsuki, S., Greybush, S. J., & Miyoshi, T. (2017). Can we optimize the assimilation order in the serial ensemble Kalman filter? A study with the Lorenz-96 model. *Monthly Weather Review*, *145*(12), 4977–4995. https://doi.org/10.1175/MWR-D-17-0094.1

Kretschmer, M., Hunt, B. R., Ott, E., Bishop, C. H., Rainwater, S., & Szunyogh, I. (2015). A composite state method for ensemble data assimilation with multiple limited-area models. *Tellus A: Dynamic Meteorology and Oceanography*, *67*(1), 26495. https://doi.org/10.3402/tellusa.v67.26495

Krishnamurti, T. N., Kumar, V., Simon, A., Bhardwaj, A., Ghosh, T., & Ross, R. (2016). A review of multimodel superensemble forecasting for weather, seasonal climate, and hurricanes. *Reviews of Geophysics*, *54*(2), 336–377. https://doi.org/10.1002/2015RG000513

Li, H., Kalnay, E., Miyoshi, T., & Danforth, C. M. (2009). Accounting for model errors in ensemble data assimilation. *Monthly Weather Review*, *137*(10), 3407–3419. https://doi.org/10.1175/2009MWR2766.1

Logutov, O. G., & Robinson, A. R. (2005). Multi-model fusion and error parameter estimation. *Quarterly Journal of the Royal Meteorological Society*, *131*(613), 3397–3408. https://doi.org/10.1256/qj.05.99

Lorenz, E. N. (1996). Predictability: A problem partly solved. In *Proceedings of a seminar held at ECMWF on predictability, 4–8 September 1995* (Vol. 1, pp. 1–18). ECMWF. https://www.ecmwf.int/en/elibrary/75462-predictability-problem-partly-solved

Mallet, V. (2010). Ensemble forecast of analyses: Coupling data assimilation and sequential aggregation. *Journal of Geophysical Research*, *115*(D24), D24303. https://doi.org/10.1029/2010JD014259

Mallet, V., Stoltz, G., & Mauricette, B. (2009). Ozone ensemble forecast with machine learning algorithms. *Journal of Geophysical Research*, *114*(D5), D05307. https://doi.org/10.1029/2008JD009978

Mandel, J. (2006). *Efficient implementation of the ensemble Kalman filter* (Center for Computational Mathematics Reports No. 231). University of Colorado at Denver and Health Sciences Center. http://ccm.ucdenver.edu/reports/rep231.pdf

Martin, G. M., Milton, S. F., Senior, C. A., Brooks, M. E., Ineson, S., Reichler, T., & Kim, J. (2010). Analysis and reduction of systematic errors through a seamless approach to modeling weather and climate. *Journal of Climate*, *23*(22), 5933–5957. https://doi.org/10.1175/2010JCLI3541.1

Mitchell, L., & Carrassi, A. (2015). Accounting for model error due to unresolved scales within ensemble Kalman filtering. *Quarterly Journal of the Royal Meteorological Society*, *141*(689), 1417–1428. https://doi.org/10.1002/qj.2451

Myers, W., Wiener, G., Linden, S., & Haupt, S. (2011). *A consensus forecasting approach for improved turbine hub height wind speed predictions*. Paper presented at AWEA Windpower Conference & Exhibition 2011, American Wind Energy Association (AWEA), Anaheim, CA. http://n2t.net/ark:/85065/d7j105rc

Narayan, A., Marzouk, Y., & Xiu, D. (2012). Sequential data assimilation with multiple models. *Journal of Computational Physics*, *231*(19), 6401–6418. https://doi.org/10.1016/j.jcp.2012.06.002

Ojeda, L. L., Kibangou, A. Y., & de Wit, C. C. (2013). Adaptive Kalman filtering for multi-step ahead traffic flow prediction. In *2013 American Control Conference* (pp. 4724–4729). https://doi.org/10.1109/ACC.2013.6580568

Okuno, S., Aihara, K., & Hirata, Y. (2019). Combining multiple forecasts for multivariate time series via state-dependent weighting. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *29*(3), 033128. https://doi.org/10.1063/1.5057379

Otsuka, S., & Miyoshi, T. (2015). A Bayesian optimization approach to multimodel ensemble Kalman filter with a low-order model. *Monthly Weather Review*, *143*(6), 2001–2012. https://doi.org/10.1175/MWR-D-14-00148.1

Panofsky, H. A. (1949). Objective weather-map analysis. *Journal of Meteorology*, *6*(6), 386–392. https://doi.org/10.1175/1520-0469(1949)006⟨0386:OWMA⟩2.0.CO;2

Pathak, J., Wikner, A., Fussell, R., Chandra, S., Hunt, B. R., Girvan, M., & Ott, E. (2018). Hybrid forecasting of chaotic processes: Using machine learning in conjunction with a knowledge-based model. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *28*(4), 041101. https://doi.org/10.1063/1.5028373

Penny, S. G. (2017). Mathematical foundations of hybrid data assimilation from a synchronization perspective. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *27*(12), 126801. https://doi.org/10.1063/1.5001819

Penny, S. G., Bach, E., Bhargava, K., Chang, C.-C., Da, C., Sun, L., & Yoshida, T. (2019). Strongly coupled data assimilation in multiscale media: Experiments using a quasi-geostrophic coupled model. *Journal of Advances in Modeling Earth Systems*, *11*(6), 1803–1829. https://doi.org/10.1029/2019MS001652

Popov, A. A., Mou, C., Sandu, A., & Iliescu, T. (2021). A multifidelity ensemble Kalman filter with reduced order control variates. *SIAM Journal on Scientific Computing*, *43*(2), A1134–A1162. https://doi.org/10.1137/20M1349965

Potthast, R., Vobig, K., Blahak, U., & Simmer, C. (2022). Data assimilation of nowcasted observations. *Monthly Weather Review*, *150*(5), 969–980. https://doi.org/10.1175/MWR-D-21-0017.1

Raanes, P. N., Carrassi, A., & Bertino, L. (2015). Extending the square root method to account for additive forecast noise in ensemble methods. *Monthly Weather Review*, *143*(10), 3857–3873. https://doi.org/10.1175/MWR-D-14-00375.1

Rainwater, S., & Hunt, B. (2013). Mixed-resolution ensemble data assimilation. *Monthly Weather Review*, *141*(9), 3007–3021. https://doi.org/10.1175/MWR-D-12-00234.1

Rodwell, M. J., & Palmer, T. N. (2007). Using numerical weather prediction to assess climate models. *Quarterly Journal of the Royal Meteorological Society*, *133*(622), 129–146. https://doi.org/10.1002/qj.23

Sakov, P., & Bertino, L. (2011). Relation between two common localisation methods for the EnKF. *Computational Geosciences*, *15*(2), 225–237. https://doi.org/10.1007/s10596-010-9202-6

Schevenhoven, F. J., & Carrassi, A. (2022). Training a supermodel with noisy and sparse observations: A case study with CPT and the synch rule on speedo—V.1. *Geoscientific Model Development*, *15*(9), 3831–3844. https://doi.org/10.5194/gmd-15-3831-2022

Schevenhoven, F. J., & Selten, F. M. (2017). An efficient training scheme for supermodels. *Earth System Dynamics*, *8*(2), 429–438. https://doi.org/10.5194/esd-8-429-2017

Schevenhoven, F. J., Selten, F. M., Carrassi, A., & Keenlyside, N. (2019). Improving weather and climate predictions by training of supermodels. *Earth System Dynamics*, *10*(4), 789–807. https://doi.org/10.5194/esd-10-789-2019

Schunk, R. W., Scherliess, L., Eccles, V., Gardner, L. C., Sojka, J. J., Zhu, L., et al. (2016). Space weather forecasting with a multimodel ensemble prediction system (MEPS). *Radio Science*, *51*(7), 1157–1165. https://doi.org/10.1002/2015RS005888

Selten, F. M., Schevenhoven, F. J., & Duane, G. S. (2017). Simulating climate with a synchronization-based supermodel. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *27*(12), 126903. https://doi.org/10.1063/1.4990721

Sengupta, U., Amos, M., Hosking, S., Rasmussen, C. E., Juniper, M., & Young, P. (2020). Ensembling geophysical models with Bayesian neural networks. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, & H. Lin (Eds.), *Advances in Neural Information Processing Systems* (Vol. 33, pp. 1205–1217). Curran Associates, Inc. https://proceedings.neurips.cc/paper/2020/file/0d5501edb21a59a43435efa67f200828-Paper.pdf

Simon, D. (2006). *Optimal state estimation*. John Wiley & Sons, Ltd. https://doi.org/10.1002/0470045345

Smith, L. A., Du, H., & Higgins, S. (2020). Designing multimodel applications with surrogate forecast systems. *Monthly Weather Review*, *148*(6), 2233–2249. https://doi.org/10.1175/MWR-D-19-0061.1

Stewart, L. M., Dance, S. L., & Nichols, N. K. (2013). Data assimilation with correlated observation errors: Experiments with a 1-D shallow water model. *Tellus A: Dynamic Meteorology and Oceanography*, *65*(1), 19546. https://doi.org/10.3402/tellusa.v65i0.19546

Sun, S.-l. (2004). Multi-sensor optimal information fusion Kalman filters with applications. *Aerospace Science and Technology*, *8*(1), 57–62. https://doi.org/10.1016/j.ast.2003.08.003

Tandeo, P., Ailliot, P., Bocquet, M., Carrassi, A., Miyoshi, T., Pulido, M., & Zhen, Y. (2020). A review of innovation-based methods to jointly estimate model and observation error covariance matrices in ensemble data assimilation. *Monthly Weather Review*, *148*(10), 3973–3994. https://doi.org/10.1175/MWR-D-19-0240.1

Thorey, J., Mallet, V., & Baudin, P. (2017). Online learning with the continuous ranked probability score for ensemble forecasting. *Quarterly Journal of the Royal Meteorological Society*, *143*(702), 521–529. https://doi.org/10.1002/qj.2940

Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M., & Whitaker, J. S. (2003). Ensemble square root filters. *Monthly Weather Review*, *131*(7), 1485–1490. https://doi.org/10.1175/1520-0493(2003)131⟨1485:ESRF⟩2.0.CO;2

Trenkler, G., & Gotu, B. (1998). *Combination of forecasts: A bibliography*. Sonderforschungsbereich 475, Universität Dortmund. http://doi.org/10.17877/DE290R-7024

Whitaker, J. S., & Hamill, T. M. (2012). Evaluating methods to account for system errors in ensemble data assimilation. *Monthly Weather Review*, *140*(9), 3078–3089. https://doi.org/10.1175/MWR-D-11-00276.1

Wiegerinck, W., Burgers, W., & Selten, F. (2013). On the limit of large couplings and weighted averaged dynamics. In L. Kocarev (Ed.), *Consensus and synchronization in complex networks* (pp. 257–275). Springer. https://doi.org/10.1007/978-3-642-33359-0_10

Wilks, D. S. (2019). *Statistical methods in the atmospheric sciences* (4th ed.). Elsevier. https://doi.org/10.1016/C2017-0-03921-6

Wu, L., Mallet, V., Bocquet, M., & Sportisse, B. (2008). A comparison study of data assimilation algorithms for ozone forecasts. *Journal of Geophysical Research*, *113*(D20), D20310. https://doi.org/10.1029/2008JD009991

Xue, L., & Zhang, D. (2014). A multimodel data assimilation framework via the ensemble Kalman filter. *Water Resources Research*, *50*(5), 4197–4219. https://doi.org/10.1002/2013WR014525

Yang, L., Narayan, A., & Wang, P. (2017). Sequential data assimilation with multiple nonlinear models and applications to subsurface flow. *Journal of Computational Physics*, *346*, 356–368. https://doi.org/10.1016/j.jcp.2017.06.026

Young, G. S. (2002). *Combining forecasts for superior prediction*. Paper presented at 16th Conference on Probability and Statistics in Atmospheric Science, American Meteorological Society, Orlando, FL. https://ams.confex.com/ams/annual2002/techprogram/paper_28061.htm

Erratum

In the originally published version of this article, a few symbols in the first column of Table 1 were not boldfaced as intended. The boldface has been added to bm, B, E{f,a}, K, ρ, P{f,a}, Q, R, x{t,f,a} and X{f,a}, and this may be considered the authoritative version of record.