

The ensemble transform Schmidt-Kalman filter: a novel method to compensate for observation uncertainty due to unresolved scales

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Bell, Z., Dance, S. L. ORCID: <https://orcid.org/0000-0003-1690-3338> and Waller, J. A. (2025) The ensemble transform Schmidt-Kalman filter: a novel method to compensate for observation uncertainty due to unresolved scales. *Atmospheric Science Letters*, 26 (5). e1296. ISSN 1530-261X doi: 10.1002/ASL.1296 Available at <https://centaur.reading.ac.uk/121688/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1002/ASL.1296>

Publisher: John Wiley & Sons

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

The ensemble transform Schmidt–Kalman filter: A novel method to compensate for observation uncertainty due to unresolved scales

Zackary Bell¹  | Sarah L. Dance^{1,2}  | Joanne A. Waller³

¹University of Reading, Reading, UK

²National Centre for Earth Observation, Reading, UK

³Met Office, Reading, UK

Correspondence

Zackary Bell, University of Reading, Reading, UK.

Email: zak.n.bell@gmail.com

Funding information

University of Reading; National Centre for Earth Observation, University of Reading, UK; Engineering and Physical Sciences Research Council, Grant/Award Number: EP/P002331/1

Abstract

Data assimilation is a mathematical technique that uses observations to improve model predictions through consideration of their respective uncertainties. Observation error due to unresolved scales occurs when there is a difference in scales observed and modeled. To obtain an optimal estimate through data assimilation, the error due to unresolved scales must be accounted for in the algorithm. In this work, we derive a novel ensemble transform formulation of the Schmidt–Kalman filter (ETSKF) to compensate for observation uncertainty due to unresolved scales in nonlinear dynamical systems. The ETSKF represents the small-scale variability through an ensemble sampled from the representation error covariance. This small-scale ensemble is added to the large-scale forecast ensemble to obtain an ensemble representative of all scales resolved by the observations. We illustrate our new method using a simple nonlinear system of ordinary differential equations with two timescales known as the swinging spring (or elastic pendulum). In this simple system, our novel method performs similarly to another method of compensating for uncertainty due to unresolved scales. Indeed, the use of small-scale ensemble statistics has potential as a new approach to compensate for uncertainty due to unresolved scales in nonlinear dynamical systems but will need further testing using more complicated systems.

KEYWORDS

data assimilation, ensemble transform Kalman filter, error due to unresolved scales, observation uncertainty, representation uncertainty, Schmidt–Kalman filter

1 | INTRODUCTION

In ensemble data assimilation (DA), the state of a dynamical system is estimated using observational data and an ensemble of forecasts obtained from numerical models (e.g., Houtekamer & Zhang, 2016; Vetra-Carvalho

et al., 2018), taking into account their uncertainties. We focus on accounting for uncertainty due to unresolved scales that are caused by a mismatch in scales observed and modeled. Uncertainty due to unresolved scales has been shown to be state-dependent and correlated in idealized systems (Janjić & Cohn, 2006), numerical weather

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 Crown copyright and The Author(s). *Atmospheric Science Letters* published by John Wiley & Sons Ltd on behalf of Royal Meteorological Society. This article is published with the permission of the Controller of HMSO and the King's Printer for Scotland.

prediction (NWP) modeling studies (Waller et al., 2014; Waller et al., 2021), and observational studies (Bell et al., 2022; Waller et al., 2016; Waller et al., 2019).

Methods to account for uncertainty due to unresolved scales will depend on the size of the observation footprint relative to the model grid (Janjić et al., 2018). When the observation footprint is larger than the model grid length, the scale mismatch may be compensated for by averaging the model state over the observed area. When the reverse is true (e.g., using high-resolution observations in convection-permitting NWP, Dance et al., 2019; Gustafsson et al., 2018; Hu et al., 2023), the scale-mismatch uncertainty must be accounted for in the DA algorithm (Bell et al., 2020; Simonin et al., 2019; Stewart et al., 2013). The standard approach is to include the representation uncertainty in the observation error covariance (e.g., Hodyss & Satterfield, 2016). Representation uncertainty may be estimated through DA diagnostic methods (e.g., Waller et al., 2016; Waller et al., 2019) or by using high-resolution model data (e.g., Daley, 1993; Liu & Rabier, 2002; Schutgens et al., 2016; Waller et al., 2021). The error due to unresolved scales may instead be treated as a model error and compensated for with inflation techniques (e.g., Carrassi & Vannitsem, 2011). However, none of these approaches are fully utilized in operational systems due to a lack of feasible computational methods (Hu & Dance, 2021; Janjić et al., 2018).

Multiscale approaches have been shown to be successful for both variational (e.g., Bannister, 2007; Buehner & Shlyueva, 2015) and ensemble (e.g., Wang et al., 2021) assimilation. Typically, these approaches treat both small and large resolved scales through scale decomposition of the error covariance matrices. The Multiscale Local Gain Form Ensemble Transform Kalman Filter, for instance, performs explicit scale decomposition of the resolved scales in the background ensemble perturbations to improve the assimilation of multiscale processes (Wang et al., 2021). There is limited literature that explicitly treats unresolved scales (e.g., Grooms et al., 2014). We consider an approach based on the Schmidt–Kalman filter (SKF) (Schmidt, 1966), where the large-scale state resolved by a model is estimated and the statistics of the unresolved small-scale processes are considered; this has been shown to be effective in linear systems (Bell et al., 2020; Janjić & Cohn, 2006).

To use the SKF for atmospheric DA, it must be adapted to be compatible with nonlinear models. While a stochastic ensemble formulation of the SKF has been derived (Lou et al., 2018), observation perturbations are required to obtain the desired analysis error statistics. Deterministic filters, such as the ensemble transform

Kalman filter (ETKF), have a smaller computational cost of implementation (Vetra-Carvalho et al., 2018) and eliminate the sampling error associated with observation perturbations. Our aim is to develop a novel ensemble transform formulation of the SKF and provide proof of concept in a simple model.

2 | THEORETICAL FRAMEWORK

Filtering in data assimilation involves using recent observations to forecast the atmospheric state forward in time. The forecast state is an estimate of the true atmospheric state conditioned on previous observations. We first define the model used for forecasting in Section 2.1 and then the observations of this state in Section 2.2. Filtering with the Schmidt–Kalman filter is introduced in Section 3.

2.1 | Model configuration

We follow the framework given in Bell et al. (2020). We first introduce the perfect and forecast models. The large-scale dynamics are assumed to reside in a subspace of the phase space for the full system. The phase space for the small-scale dynamics will be the complement of the large-scale subspace. We therefore express the true state $\mathbf{x}^t \in \mathbb{R}^{N_t}$ in the partitioned form

$$\mathbf{x}^t = \begin{pmatrix} \mathbf{x}^{l,t} \\ \mathbf{x}^{s,t} \end{pmatrix}, \quad (2.1)$$

where $\mathbf{x}^{l,t} \in \mathbb{R}^{N_l}$ is the true large-scale state, $\mathbf{x}^{s,t} \in \mathbb{R}^{N_s}$ is the true small-scale state, and $N_t = N_l + N_s$. Throughout this paper, any component with a t -superscript indicates it is a true variable. The l - and s -superscripts correspond to the large- and small-scale processes within the complete system dynamics. We note that it is our intent that \mathbf{x}^l is the actual large-scale model state and not the state obtained from an additional spectral decomposition of model variables. We also note that, while $\mathbf{x}^{s,t}$ is meaningful in its own right, its significance stems from its interaction with the large-scale state.

Throughout this section, the subscripts on brackets denote the time of the arguments inside the bracket. For nonlinear dynamical systems, the perfect model evolving the true state is given by the coupled system,

$$\begin{pmatrix} \mathbf{x}^{l,t} \\ \mathbf{x}^{s,t} \end{pmatrix}_k = \begin{pmatrix} \mathcal{M}^{l,t}(\mathbf{x}^{l,t}, \mathbf{x}^{s,t}) \\ \mathcal{M}^{s,t}(\mathbf{x}^{l,t}, \mathbf{x}^{s,t}) \end{pmatrix}_{k-1}, \quad (2.2)$$

where $\mathcal{M}^{l,t} : \mathbb{R}^{N_l} \times \mathbb{R}^{N_s} \rightarrow \mathbb{R}^{N_l}$ and $\mathcal{M}^{s,t} : \mathbb{R}^{N_l} \times \mathbb{R}^{N_s} \rightarrow \mathbb{R}^{N_s}$ are the true nonlinear models that map the true large- and small-scale states forward in time, respectively.

In NWP, subgrid-scale parameterizations are often used to represent small-scale processes (Janjić et al., 2018; Janjić & Cohn, 2006). We therefore assume the imperfect forecast model has the partitioned form:

$$\begin{pmatrix} \mathbf{x}^{l,t} \\ \mathbf{x}^{s,t} \end{pmatrix}_k = \begin{pmatrix} \mathcal{M}^l(\mathbf{x}^{l,t}) \\ \mathbf{x}^{s,t} \end{pmatrix}_{k-1} - \begin{pmatrix} \boldsymbol{\eta}^l \\ \boldsymbol{\eta}^s \end{pmatrix}_k, \quad (2.3)$$

where $\mathcal{M}^l : \mathbb{R}^{N_l} \rightarrow \mathbb{R}^{N_l}$ is the imperfect nonlinear forecast model that maps the large-scale state forward in time and $\boldsymbol{\eta}^l \in \mathbb{R}^{N_l}$ and $\boldsymbol{\eta}^s \in \mathbb{R}^{N_s}$ are the large- and small-scale model errors assumed to have zero-mean and static covariance:

$$\begin{pmatrix} \mathbf{Q}^{ll} & \mathbf{Q}^{ls} \\ \mathbf{Q}^{sl} & \mathbf{Q}^{ss} \end{pmatrix}. \quad (2.4)$$

Here, $\mathbf{Q}^{ll} \equiv \langle \boldsymbol{\eta}^l (\boldsymbol{\eta}^l)^T \rangle \in \mathbb{R}^{N_l \times N_l}$ and $\mathbf{Q}^{ss} \equiv \langle \boldsymbol{\eta}^s (\boldsymbol{\eta}^s)^T \rangle \in \mathbb{R}^{N_s \times N_s}$ are the model error covariances for the large- and small-scale processes, and $\mathbf{Q}^{ls} \equiv \langle \boldsymbol{\eta}^l (\boldsymbol{\eta}^s)^T \rangle \in \mathbb{R}^{N_l \times N_s}$ (with $\mathbf{Q}^{sl} = (\mathbf{Q}^{ls})^T$) is the cross-covariance between the large- and small-scale model errors. The assumption of a persistence small-scale forecast model is related to the implementation of our novel filter which we discuss in Section 3.3.

Analogously, the complete forecast state $\begin{pmatrix} (\mathbf{x}^{lf})^T & (\mathbf{x}^{sf})^T \end{pmatrix}_k \in \mathbb{R}^{N_l}$ satisfies

$$\begin{pmatrix} \mathbf{x}^{lf} \\ \mathbf{x}^{sf} \end{pmatrix}_k = \begin{pmatrix} \mathcal{M}^l(\mathbf{x}^{lf}) \\ \mathbf{x}^{sf} \end{pmatrix}_{k-1}, \quad (2.5)$$

where the f -superscript indicates the forecast. We note that \mathcal{M}^l is meant to be representative of the models used in operational weather prediction systems. We also note that the small-scale forecast \mathbf{x}^{sf} is usually omitted in practice as it is unknown. Similarly to the model error covariance, the forecast state error covariance at time t_k will be in the partitioned form:

$$\begin{pmatrix} \mathbf{P}^{ll,f} & \mathbf{P}^{ls,f} \\ \mathbf{P}^{sl,f} & \mathbf{P}^{ss,f} \end{pmatrix}_k \approx \begin{pmatrix} \mathbf{M}^l & \mathbf{0}_{N_l \times N_s} \\ \mathbf{0}_{N_s \times N_l} & \mathbf{I}_{N_s \times N_s} \end{pmatrix} \begin{pmatrix} \mathbf{P}^{ll,f} & \mathbf{P}^{ls,f} \\ \mathbf{P}^{sl,f} & \mathbf{P}^{ss,f} \end{pmatrix}_{k-1} \times \begin{pmatrix} \mathbf{M}^l & \mathbf{0}_{N_l \times N_s} \\ \mathbf{0}_{N_s \times N_l} & \mathbf{I}_{N_s \times N_s} \end{pmatrix}^T + \begin{pmatrix} \mathbf{Q}^{ll} & \mathbf{Q}^{ls} \\ \mathbf{Q}^{sl} & \mathbf{Q}^{ss} \end{pmatrix}, \quad (2.6)$$

where each block of the forecast error covariance has the same dimensions as the corresponding model error block

and \mathbf{M}^l is the tangent linear approximation to \mathcal{M}^l . We note that (2.6) corresponds to the error covariance at the previous analysis at $k=0$ (see Section 3). Even if the initial large- and small-scale forecast errors are uncorrelated (i.e., $\mathbf{P}_0^{ls,f} = \mathbf{0}_{N_l \times N_s}$ and $\mathbf{P}_0^{sl,f} = \mathbf{0}_{N_s \times N_l}$), the cross-covariances $\mathbf{P}^{ls,f}$ and $\mathbf{P}^{sl,f}$ can still evolve into a non-zero matrix due to model error. Using this formulation of the complete system dynamics, we apply filters that analyse the large-scale state only but account for the small-scale processes in different ways.

2.2 | Observation configuration

In this section, we assume each component is valid at time k , allowing us to drop the time subscript. The observations, $\mathbf{y} \in \mathbb{R}^p$, are assumed to be the sum of the true large- and small-scale states mapped into observation space, expressed as

$$\mathbf{y} = \mathbf{h}^l(\mathbf{x}^{l,t}) + \mathbf{h}^s(\mathbf{x}^{s,t}) + \boldsymbol{\epsilon}, \quad (2.7)$$

where $\mathbf{h}^l : \mathbb{R}^{N_l} \rightarrow \mathbb{R}^p$ is the known imperfect large-scale observation operator, $\mathbf{h}^s : \mathbb{R}^{N_s} \rightarrow \mathbb{R}^p$ is the possibly unknown imperfect small-scale observation operator, and $\boldsymbol{\epsilon} \in \mathbb{R}^p$ is the random and unbiased instrument error. The observation error is given by

$$\mathbf{e}^o = \mathbf{y} - \mathbf{h}^l(\mathbf{x}^{l,t}) = \mathbf{h}^s(\mathbf{x}^{s,t}) + \boldsymbol{\epsilon} \quad (2.8)$$

where $\mathbf{h}^s(\mathbf{x}^{s,t})$ is the error due to unresolved scales, and we have assumed no observation operator error. Therefore, the instrument uncertainty is given by $\mathbf{R}^l \equiv \langle \boldsymbol{\epsilon} \boldsymbol{\epsilon}^T \rangle$ and the representation uncertainty is given by

$$\tilde{\mathbf{R}}^H \equiv \left\langle \left(\mathbf{h}^s(\mathbf{x}^{s,t}) - \overline{\mathbf{h}^s(\mathbf{x}^{s,t})} \right) \left(\mathbf{h}^s(\mathbf{x}^{s,t}) - \overline{\mathbf{h}^s(\mathbf{x}^{s,t})} \right)^T \right\rangle, \quad (2.9)$$

where the angular brackets denote the mathematical expectation over the corresponding error distribution. Here, we have used a tilde (\sim) to indicate the true representation uncertainty. In later sections, approximations to the representation uncertainty will not have a tilde.

3 | THE DETERMINISTIC ENSEMBLE SCHMIDT-KALMAN FILTER

Throughout this section, all components are assumed to be valid at the same time, so we omit the time subscripts.

3.1 | The Schmidt–Kalman filter

The SKF is a linear filter that analyzes the large-scale state only through consideration of the large-scale uncertainty and the variability of the small-scale processes (Bell et al., 2020; Janjić & Cohn, 2006). The analysis state update for the SKF is given by

$$\mathbf{x}^{l,a} = \mathbf{x}^{lf} + \mathbf{K}^l (\mathbf{y} - \mathbf{H}^l \mathbf{x}^{lf}), \quad (3.1)$$

where the large-scale gain matrix \mathbf{K}^l is given by

$$\mathbf{K}^l = (\mathbf{P}^{llf} \quad \mathbf{P}^{lsf}) \begin{pmatrix} (\mathbf{H}^l)^T \\ (\mathbf{H}^s)^T \end{pmatrix} \left((\mathbf{H}^l \quad \mathbf{H}^s) \begin{pmatrix} \mathbf{P}^{llf} & \mathbf{P}^{lsf} \\ \mathbf{P}^{slf} & \mathbf{C}^s \end{pmatrix} \begin{pmatrix} (\mathbf{H}^l)^T \\ (\mathbf{H}^s)^T \end{pmatrix} + \mathbf{R}^l \right)^{-1}. \quad (3.2)$$

Here, the forecast error covariance blocks \mathbf{P}^{llf} , \mathbf{P}^{lsf} , and \mathbf{P}^{slf} are defined in Equation (2.6), $\mathbf{C}^s \in \mathbb{R}^{N_s \times N_s}$ is a climatological approximation to the true small-scale variability, given by the covariance of the true small-scale state $\langle (\mathbf{x}^{s,t} - \overline{\mathbf{x}}^{s,t})(\mathbf{x}^{s,t} - \overline{\mathbf{x}}^{s,t})^T \rangle$, and the observation operator is in partitioned form where $\mathbf{H}^l \in \mathbb{R}^{p \times N_l}$ is the large-scale observation operator and $\mathbf{H}^s \in \mathbb{R}^{p \times N_s}$ the small-scale observation operator.

From Equation (2.9), an approximation of the representation error covariance is given by

$$\mathbf{R}^H = \mathbf{H}^s \mathbf{C}^s (\mathbf{H}^s)^T \quad (3.3)$$

(Hodyss & Satterfield, 2016). The large-scale forecast error covariance \mathbf{P}^{llf} and forecast cross-covariance \mathbf{P}^{lsf} are analysed by

$$\mathbf{P}^{ll,a} = (\mathbf{I}_{N_l} - \mathbf{K}^l \mathbf{H}^l) \mathbf{P}^{llf} - \mathbf{K}^l \mathbf{H}^s \mathbf{P}^{slf} \quad (3.4)$$

and

$$\mathbf{P}^{ls,a} = (\mathbf{I} - \mathbf{K}^l \mathbf{H}^l) \mathbf{P}^{lsf} - \mathbf{K}^l \mathbf{H}^s \mathbf{C}^s. \quad (3.5)$$

In the forecast step, the SKF evolves the large-scale analysis state $\mathbf{x}^{l,a}$, the large-scale analysis error covariance $\mathbf{P}^{ll,a}$, and the analysis cross-covariance $\mathbf{P}^{ls,a}$ (Equations (3.15)–(3.18), Bell et al. (2020)). In practice, subgrid-scale parameterizations are used to approximate contributions from the small-scale processes. For example, Clark et al. (2021) developed a stochastic boundary layer perturbation scheme to approximate the effects of

unresolved processes, enhancing the representation of boundary layer variability. Hence, only knowledge of the large-scale model is required to evolve $\mathbf{x}^{l,a}$ and $\mathbf{P}^{ll,a}$. To evolve the cross-covariance $\mathbf{P}^{ls,a}$, additional assumptions would be required such as assuming the small-scale state is modelled by persistence (as is done in (Equation 2.3) with addition of model error). For a detailed description of the SKF equations, we refer the reader to Bell et al. (2020).

The SKF accounts for the small-scale variability in state space, which allows for consideration of flow-dependent cross-covariances between the large-scale

uncertainty and the small-scale variability. As a result, the observation error covariance for the SKF will consist of instrument uncertainty only. In contrast, the standard Kalman filter may account for the small-scale variability in observation space (i.e., $\mathbf{R} = \mathbf{R}^l + \mathbf{R}^H$) or completely disregard it altogether (i.e., $\mathbf{R} = \mathbf{R}^l$) such that only the large-scale filter and model components are needed for its implementation. We note that in the limit of zero cross-covariances, the SKF and standard Kalman Filter are equivalent in their compensation of representation uncertainty (Hodyss & Nichols, 2015). The disadvantage of using the SKF over the standard Kalman filter is a higher computational cost associated with the augmentation of the state error covariances (Equation (3.2)). The computational cost of the standard Kalman filter can be reduced by adopting an ensemble formulation such as the ensemble transform Kalman filter, ETKF (Vetra-Carvalho et al., 2018, described in the Supplementary Material). The computational cost of our new ensemble transform SKF is discussed in Section 3.4.

3.2 | The ensemble transform Schmidt–Kalman filter (ETSKF)

In the novel ensemble transform SKF (ETSKF), the small-scale covariance \mathbf{C}^s must be approximated through ensemble statistics in the analysis step. The uncertainty associated with the large-scale forecast ensemble mean $\overline{\mathbf{x}}^{lf} \in \mathbb{R}^{N_l}$ is given by

$$\mathbf{P}^{llf} = \mathbf{X}^{lf} (\mathbf{X}^{lf})^T, \quad (3.6)$$

where $\mathbf{X}^{lf} \in \mathbb{R}^{N_t \times m}$ is the large-scale forecast perturbation matrix given by

$$\mathbf{X}^{lf} = \frac{1}{\sqrt{m-1}} (\mathbf{x}^{lf,(1)} - \bar{\mathbf{x}}^{lf} \dots \mathbf{x}^{lf,(m)} - \bar{\mathbf{x}}^{lf}). \quad (3.7)$$

Here, m is the ensemble size and $\mathbf{x}^{lf,(i)}$ is the i -th ensemble member. As $\mathbf{x}^{s,t}$ is not estimated in the SKF equations, the small-scale perturbation matrix, $\mathbf{X}^s \in \mathbb{R}^{N_s \times m}$, may be obtained by sampling an ensemble of size $N_s \times m$ from the climatological approximation of the small-scale variability, \mathbf{C}^s , and dividing by $\sqrt{m-1}$. Hence, the forecast cross-covariance is given by

$$\mathbf{P}^{ls,f} = \mathbf{X}^{lf} (\mathbf{X}^s)^T. \quad (3.8)$$

Further discussion regarding the construction of \mathbf{X}^s is given in Section 3.3.

We next describe the ETSKF analysis update equations. Expressing the error covariances in Equation (3.2) in terms of perturbation matrices, the gain matrix for the ensemble formulation of the Schmidt–Kalman filter is given by

$$\mathbf{K}^l = \left(\mathbf{X}^{lf} (\mathbf{Y}^{lf})^T + \mathbf{X}^{lf} (\mathbf{Y}^s)^T \right) \mathbf{D}^{-1}, \quad (3.9)$$

where $\mathbf{Y}^{lf} = \mathbf{H}^l \mathbf{X}^{lf}$, $\mathbf{Y}^s = \mathbf{H}^s \mathbf{X}^s$ and \mathbf{D} is the innovation covariance given by

$$\mathbf{D} = (\mathbf{Y}^{lf} + \mathbf{Y}^s) (\mathbf{Y}^{lf} + \mathbf{Y}^s)^T + \mathbf{R}^l. \quad (3.10)$$

For the same reasons as discussed in Section 3.1, the observation error covariance assumed by the ETSKF will not contain a component corresponding to the uncertainty due to unresolved scales. The ETSKF mean analysis state update is obtained by replacing the large-scale state variable in Equation (3.1) with the ensemble mean.

To derive the analysis perturbation matrix, we start by writing the SKF large-scale analysis error covariance update (Equation 3.4) in terms of perturbation matrices to obtain

$$\begin{aligned} \mathbf{X}^{l,a} (\mathbf{X}^{l,a})^T &= \left(\mathbf{I}_{N_t} - \left(\mathbf{X}^{lf} (\mathbf{Y}^{lf})^T + \mathbf{X}^{lf} (\mathbf{Y}^s)^T \right) \mathbf{D}^{-1} \mathbf{H}^l \right) \mathbf{X}^{lf} (\mathbf{X}^{lf})^T \\ &\quad - \left(\mathbf{X}^{lf} (\mathbf{Y}^{lf})^T + \mathbf{X}^{lf} (\mathbf{Y}^s)^T \right) \mathbf{D}^{-1} \mathbf{H}^s \mathbf{X}^s (\mathbf{X}^{lf})^T \\ &= \mathbf{X}^{lf} \left(\mathbf{I}_m - [\mathbf{Y}^{lf} + \mathbf{Y}^s]^T \mathbf{D}^{-1} [\mathbf{Y}^{lf} + \mathbf{Y}^s] \right) (\mathbf{X}^{lf})^T. \end{aligned} \quad (3.11)$$

We now define $\mathbf{T} \in \mathbb{R}^{m \times m}$ such that

$$\mathbf{T} \mathbf{T}^T = \mathbf{I}_m - [\mathbf{Y}^{lf} + \mathbf{Y}^s]^T \mathbf{D}^{-1} [\mathbf{Y}^{lf} + \mathbf{Y}^s]. \quad (3.12)$$

We next apply the Sherman–Morrison–Woodbury identity on \mathbf{D}^{-1} and then use the singular value decomposition on $[\mathbf{Y}^{lf} + \mathbf{Y}^s]^T (\mathbf{R}^l)^{-1/2}$ to obtain

$$\mathbf{T} = \mathbf{U} (\mathbf{I}_m + \mathbf{\Sigma} \mathbf{\Sigma}^T)^{-1/2} \mathbf{U}^T, \quad (3.13)$$

where $\mathbf{U} \in \mathbb{R}^{m \times m}$ is an orthogonal matrix containing the left singular vectors of $(\mathbf{Y}^{lf} + \mathbf{Y}^s)^T (\mathbf{R}^l)^{-1/2}$ and $\mathbf{\Sigma} \in \mathbb{R}^{m \times p}$ is a rectangular diagonal matrix containing the corresponding non-zero singular values. Finally, the analysis perturbation can be implemented in the filter using

$$\mathbf{X}^{l,a} = \mathbf{X}^{lf} \mathbf{T}. \quad (3.14)$$

The ETSKF forecast step is identical to the ETKF forecast step given in the Supplementary Material. An algorithm block for the ETSKF is given in Section 3.3.

3.3 | Discussion of the small-scale perturbation matrix

The ETSKF compensates for uncertainty due to unresolved scales through the use of the small-scale perturbation matrix \mathbf{X}^s . Since the small-scale perturbation matrix is mapped into observation space, we may construct the small-scale ensemble by sampling an ensemble of size $p \times m$ from an approximation of the uncertainty due to unresolved scales given by Equation (3.3). Sampling in observation space circumvents the need for a small-scale observation operator \mathbf{H}^s . An approximation to the climatological uncertainty due to unresolved scales may be obtained in several ways (Daley, 1993; Liu & Rabier, 2002; Schutgens et al., 2016; Waller et al., 2021).

To determine a strategy to sample the small-scale ensemble, we examine how the cross-covariance between the large-scale uncertainty and the small-scale variability, \mathbf{P}^{ls} , is treated by the SKF. While the small-scale covariance, \mathbf{C}^s , is taken to be time-independent by the SKF, \mathbf{P}^{ls} is explicitly evolved. The cross-covariance is analysed by

$$\mathbf{P}^{ls,a} = \mathbf{X}^{lf} \mathbf{T} \mathbf{T}^T (\mathbf{X}^s)^T, \quad (3.15)$$

which is obtained by expressing $\mathbf{P}^{ls,f}$ and \mathbf{C}^s in Equation (3.5) in terms of perturbation matrices. To account for the evolution of cross-covariances, the partitioned perturbation matrix $\left((\mathbf{Y}^{l,a})^T (\mathbf{Y}^s)^T \right)^T$ is sampled from $\mathcal{N}(\mathbf{0}, \mathbf{\Psi})$ where

$$\begin{aligned} \Psi &= \begin{pmatrix} \mathbf{Y}^{lf} \mathbf{T} \mathbf{T}^T (\mathbf{Y}^{lf})^T & \mathbf{Y}^{lf} \mathbf{T} \mathbf{T}^T (\mathbf{Y}^s)^T \\ \mathbf{Y}^s \mathbf{T} \mathbf{T}^T (\mathbf{Y}^{lf})^T & \mathbf{R}^H \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{H}^l \mathbf{P}^{ll,a} (\mathbf{H}^l)^T & \mathbf{H}^l \mathbf{P}^{ls,a} (\mathbf{H}^s)^T \\ \mathbf{H}^s \mathbf{P}^{sl,a} (\mathbf{H}^l)^T & \mathbf{H}^s \mathbf{C}^s (\mathbf{H}^s)^T \end{pmatrix}. \end{aligned} \quad (3.16)$$

Here, the second equality is obtained using Equations (3.3), (3.11), and (3.15). We note that since \mathbf{C}^s is a climatological approximation to the small-scale variability, we treat it constant in time so that the cross-covariances are correctly forecast by the ETSKF. We also note that \mathbf{Y}^s is sampled after $\mathbf{X}^{l,a}$ is calculated and is used in the analysis step at the next observation time. The entire ETSKF implementation is given in Algorithm 1.

3.4 | Discussion of the computational expense

We compare the computational expense of the ETSKF with the ETKF. In principle, to provide a fair comparison, we should consider an ETKF implemented with a fully correlated observation error covariance matrix, $\mathbf{R}^I + \mathbf{R}^H$. However, common implementations of the ETKF typically assume that the observation error covariance matrix is diagonal, and in this case the general leading order of operation count for the ETKF is $\mathcal{O}(N_l m^2 + p m^2 + m^3)$ (Vetra-Carvalho et al., 2018).

In the forecast step, the ETSKF evolves the large-scale state only, which is identical to the ETKF forecast step (see Supplementary Material).

Algorithm 1 Ensemble Transform Schmidt–Kalman Filter (ETSKF) algorithm

- 1: At time $k = 1$, initialize ensemble $\mathbf{x}_k^{lf,(i)}$ for $i = 1 \dots m$, observation operator \mathbf{H}^l , small-scale ensemble \mathbf{Y}_0^s , instrument error covariance \mathbf{R}^I , and representation error covariance \mathbf{R}^H
- 2: **for** $k = 1$ to t_{end} **do**
- 3: // Calculate forecast statistics
- 4: $\bar{\mathbf{x}}_k^{lf} = \frac{1}{m} \sum_{i=1}^m \mathbf{x}_k^{lf,(i)}$
- 5: $\mathbf{X}_k^{lf} = \frac{1}{\sqrt{m-1}} \left(\mathbf{x}_k^{lf,(1)} - \bar{\mathbf{x}}_k^{lf} \quad \dots \quad \mathbf{x}_k^{lf,(m)} - \bar{\mathbf{x}}_k^{lf} \right)$
- 6: $\mathbf{Y}_k^{lf} = \mathbf{H}^l \mathbf{X}_k^{lf}$
- 7: // Compute transform matrix
- 8: $\Sigma_k = \left(\mathbf{Y}_k^{lf} + \mathbf{Y}_k^s \right) \left(\mathbf{R}^I \right)^{-1/2}$
- 9: $\mathbf{T} = \mathbf{U}_k \left(\mathbf{I}_m + \Sigma_k \Sigma_k^T \right)^{-1/2} \mathbf{U}_k^T$
- 10: // Sample \mathbf{Y}_k^s
- 11: Sample $\left(\left(\mathbf{Y}_k^{l,a} \right)^T \quad \left(\mathbf{Y}_k^s \right)^T \right)^T$ from $\Psi_k = \begin{pmatrix} \mathbf{Y}_k^{lf} \mathbf{T}_k \mathbf{T}_k^T \left(\mathbf{Y}_k^{lf} \right)^T & \mathbf{Y}_k^{lf} \mathbf{T}_k \mathbf{T}_k^T \left(\mathbf{Y}_{k-1}^s \right)^T \\ \mathbf{Y}_{k-1}^s \mathbf{T}_k \mathbf{T}_k^T \left(\mathbf{Y}_k^{lf} \right)^T & \mathbf{R}^H \end{pmatrix}$
- 12: // Update analysis ensemble
- 13: $\mathbf{D}_k = \left(\mathbf{Y}_k^{lf} + \mathbf{Y}_k^s \right) \left(\mathbf{Y}_k^{lf} + \mathbf{Y}_k^s \right)^T + \mathbf{R}^I$
- 14: $\mathbf{K}_k^l = \left(\mathbf{X}_k^{lf} \left(\mathbf{Y}_k^{lf} \right)^T + \mathbf{X}_k^{lf} \left(\mathbf{Y}_k^s \right)^T \right) \mathbf{D}_k^{-1}$
- 15: $\bar{\mathbf{x}}_k^{l,a} = \bar{\mathbf{x}}_k^{lf} + \mathbf{K}_k^l \left(\mathbf{y}_k - \mathbf{H}^l \bar{\mathbf{x}}_k^{lf} \right)$
- 16: $\mathbf{X}_k^{l,a} = \mathbf{X}_k^{lf} \mathbf{T}_k$
- 17: **for** $i = 1$ to m **do**
- 18: $\mathbf{x}_k^{l,a,(i)} = \bar{\mathbf{x}}_k^{l,a} + \mathbf{X}_k^{l,a} [i, :]$
- 19: **end for**
- 20: // Forecast analysis ensemble to time $k + 1$
- 21: **for** $i = 1$ to m **do**
- 22: $\mathbf{x}_{k+1}^{lf,(i)} = \mathcal{M}^l \left(\mathbf{x}_k^{l,a,(i)} \right) - \eta_{k+1}^{l,(i)}$
- 23: **end for**
- 24: **end for**

In the analysis step, the ETSKF equations are obtained by replacing \mathbf{Y}^{lf} with $\mathbf{Y}^{lf} + \mathbf{Y}^s$ and \mathbf{R} with \mathbf{R}^l in the ETKF equations. Hence, the two main differences in computational expense between the two methods are (1) the ETKF requires inversion of a fully correlated matrix \mathbf{R} compared with inversion of the usually diagonal \mathbf{R}^l in the ETSKF and (2) the ETSKF samples the small-scale ensemble \mathbf{Y}^s and adds it to \mathbf{Y}^{lf} each analysis step. Regarding (1), the use of fully correlated observation error covariance matrices may be feasible for some observation types (Simonin et al., 2019). However, for large datasets with long correlation lengths this is not currently possible within the constraints of operational schedules (Hu & Dance, 2021). Regarding (2), sampling from the representation error covariance has computational complexity $\mathcal{O}(m\lambda^2)$, where m is the number of samples drawn and λ is the number of rows in the covariance matrix (Gentle, 2010, p. 315–316). Hence, the general leading order of operation count for the ETSKF is $\mathcal{O}(N_l m^2 + pm^2 + m^3 + 4mp^2)$.

4 | ILLUSTRATIVE NUMERICAL EXAMPLE

4.1 | The swinging spring model

We use the swinging spring (elastic pendulum) model (Lynch, 2002) shown in Figure 1 to illustrate the behaviour of the ETSKF. This has been studied in atmospheric science as a simple model with two scales (Vanneste, 2013).

A mass, m , is suspended from a fixed point by a spring with unstretched length ℓ_0 and elasticity k in a uniform gravitational field g . The spring is unable to bend and can stretch lengthways. Using polar coordinates, the nonlinear system of differential equations describing the motion of the swinging spring is given by

$$\dot{\theta} = \frac{p_\theta}{mr^2}, \quad (4.1)$$

$$\dot{p}_\theta = -mgr \sin \theta, \quad (4.2)$$

$$\dot{r} = \frac{p_r}{m}, \quad (4.3)$$

$$\dot{p}_r = \frac{p_\theta^2}{mr^3} - k(r - \ell_0) + mg \cos \theta, \quad (4.4)$$

where θ is the angle measured from the downward vertical with momentum p_θ and r is the length of the spring

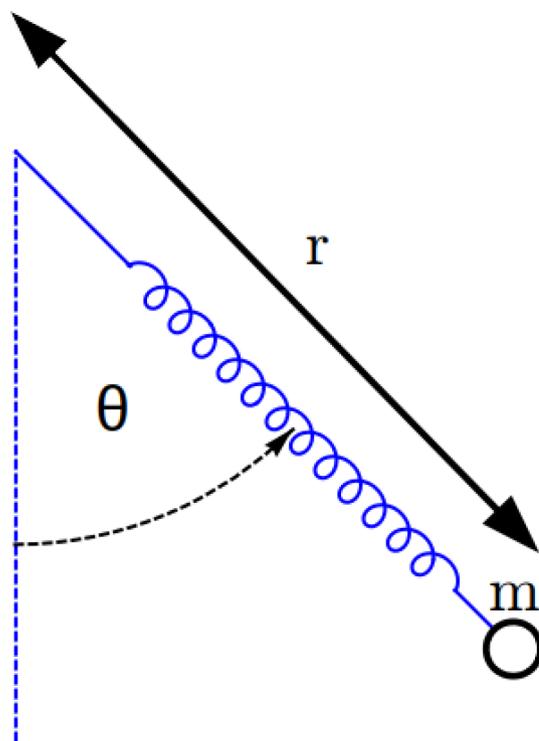


FIGURE 1 Schematic of the swinging spring. The angle θ is measured from the downward vertical and the length of the spring r is measured from the point of suspension to the mass.

measured from the point of suspension with momentum p_r .

The model has two equilibrium points: a stable equilibrium with the spring stretched vertically below the point of suspension and an unstable equilibrium where the spring is compressed vertically above the point of suspension. We consider only the stable equilibrium. The equilibrium length of the spring is given by

$$\ell = \ell_0 + \frac{mg}{k}. \quad (4.5)$$

We may linearize the swinging spring system for small amplitude motions, obtaining two independent oscillations in θ and r (Lynch, 2002). The frequency of the rotational motions is

$$\omega_\theta = \sqrt{\frac{g}{\ell}}, \quad (4.6)$$

and the frequency of the elastic motions is

$$\omega_r = \sqrt{\frac{k}{m}}. \quad (4.7)$$

In the fully nonlinear system, (Equations 4.1–4.4), the rotational and elastic motions are coupled. Nevertheless, provided $\omega_\theta/\omega_r < 1$, the rotational motions will correspond to the large-scale processes and the elastic motions will correspond to the small-scale processes. For our illustration, we set $\ell = 1$ m, $m = 1$ kg, $g = \pi^2$ m/s², and $k = 3\pi^2$ N/m, so that there is a separation in scales and substantial small-scale variability.

To obtain the true partitioned model in Equation (2.2), we define $r = \ell + \rho$, where ρ is the displacement of the spring from the stable equilibrium with momentum p_ρ . Since ℓ is fixed, the variability of r is completely associated with ρ . With this transformation, the swinging spring system is given by

$$\dot{\theta} = \frac{p_\theta}{m(\ell + \rho)^2}, \quad (4.8)$$

$$\dot{p}_\theta = -mg(\ell + \rho)\sin\theta, \quad (4.9)$$

$$\dot{\ell} = 0, \quad (4.10)$$

$$\dot{p}_\rho = \frac{p_\rho}{m}, \quad (4.11)$$

$$\dot{p}_\rho = \frac{p_\rho^2}{m(\ell + \rho)^3} - k(\ell + \rho - l_0) + mg\cos\theta. \quad (4.12)$$

Using this form of the equations, the complete state is $(\theta \ p_\theta \ \ell \ \rho \ p_\rho)^T$ which consists of the large-scale state $(\theta \ p_\theta \ \ell)^T$ and the small-scale state $(\rho \ p_\rho)^T$. Typical behaviour of the state variables is shown in Figure 2. The frequency of θ and p_θ is lower than the frequency of ρ and p_ρ and ℓ is constant. Hence, θ , p_θ , and ℓ may be correctly represented with a coarser temporal resolution than ρ and p_ρ . Additionally, the large- and small-scale states are weakly correlated.

To obtain the large-scale forecast model (Equation 2.3), we neglect ρ from Equations (4.8)–(4.10) such that the large- and small-scale processes are uncoupled:

$$\dot{\theta} = \frac{p_\theta}{m\ell^2}, \quad (4.13)$$

$$\dot{p}_\theta = -mg\ell\sin\theta, \quad (4.14)$$

$$\dot{\ell} = 0. \quad (4.15)$$

We note that the forecast model for θ and p_θ will have model error due to unresolved scales. Typical behaviour of the forecast model is shown in Figure 3.

To integrate the true system in Equations (4.8)–(4.12) and the forecast model in Equations (4.13)–(4.15), we use the Octave (version 9.2.0) *ode45* solver with a relative error tolerance of 10^{-3} and an absolute error tolerance of 10^{-6} . The value of the system is recorded every 0.01 model seconds.

4.2 | Observations and their uncertainty

For our illustration, we consider observations of θ^t and $r^t = \ell^t + \rho^t$. Expressing the observations $\mathbf{y} \in \mathbb{R}^2$ in the form of Equation (2.7), we have

$$\mathbf{y} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \theta^t \\ p_\theta^t \\ \ell^t \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \rho^t \\ p_\rho^t \end{pmatrix} + \boldsymbol{\epsilon}, \quad (4.16)$$

where the first matrix–vector product corresponds to $\mathbf{H}^t \mathbf{x}^{l,t}$, the second corresponds to the error due to unresolved scales $\mathbf{H}^s \mathbf{x}^{s,t}$, and $\boldsymbol{\epsilon} \in \mathbb{R}^2$ is the instrument error.

To generate observations, we take the values of θ^t and r^t at specified time-steps and add instrument error $\boldsymbol{\epsilon}$. The instrument error has distribution $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, 0.1^2 \mathbf{I})$ such that the uncertainty due to unresolved motion is substantially greater than the instrument uncertainty. Following Bell et al. (2020), we expect the ETSKF to provide the most benefit in this regime.

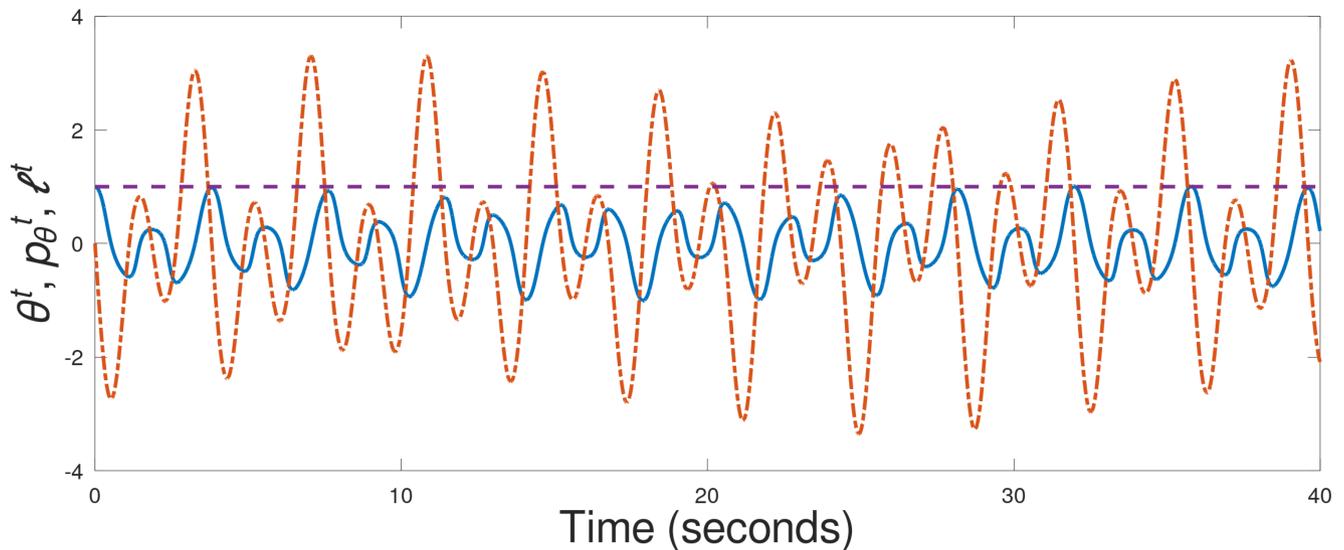
We consider filters that analyse the large-scale state only. Hence, the observations of r^t are treated as observations of ℓ^t with error due to unresolved scales ρ^t . There is a bias due to unresolved scales as the time average of ρ^t is non-zero (see Section 4.3).

4.3 | Twin experiment

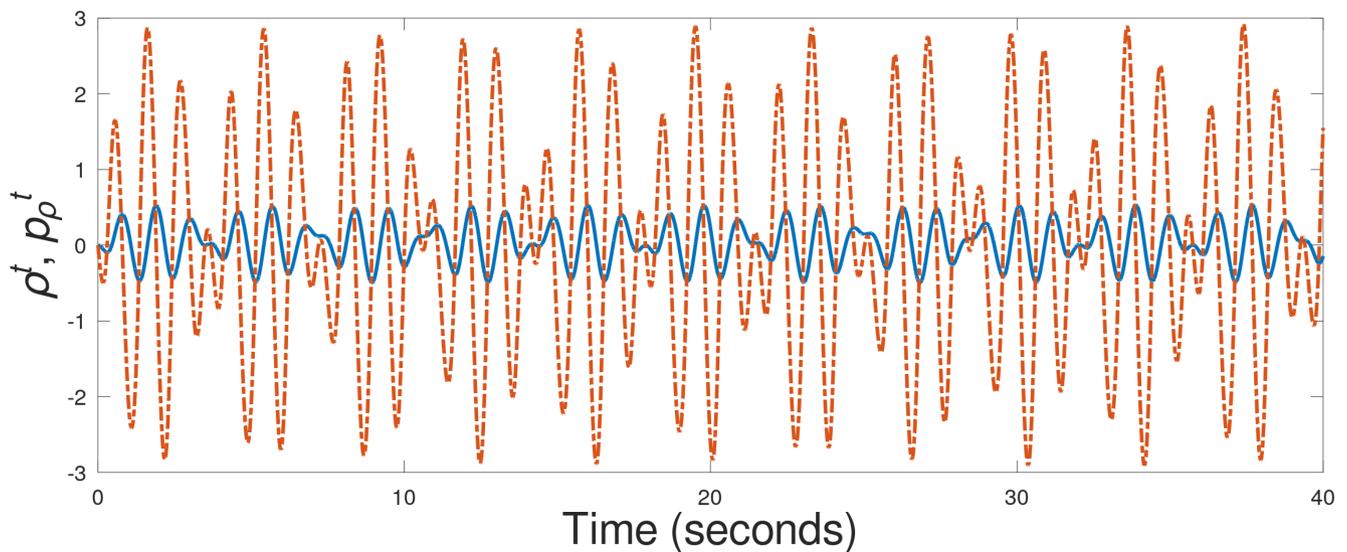
We use three filters: the ETKF that disregards uncertainty due to unresolved scales (i.e., $\mathbf{R} = \mathbf{R}^l$), denoted by ETKF-LS, the ETKF that compensates for uncertainty due to unresolved scales (i.e., $\mathbf{R} = \mathbf{R}^l + \mathbf{R}^H$), denoted by ETKF-RH, and the ETSKF. Each filter uses the same large-scale forecast model, given by Equations (4.13)–(4.15). We note that the ETSKF and ETKF-RH are equivalent in the limit that the cross-covariance terms are zero.

To evaluate the performance of each filter, we perform cycled assimilation experiments and evaluate the forecasts against the corresponding truth. Each experiment consists of the following initialization steps:

1. Using the true model in Equations (4.8)–(4.12), evolve the state $(\theta \ p_\theta \ \ell \ \rho \ p_\rho)^T = (1 \ 0 \ 1 \ 0 \ 0)^T$ for 100 seconds. Using the Octave (version 9.2.0)



(a) Top panel: θ (blue solid line), p_θ (orange dot-dash line) and ℓ (purple dashed line).



(b) Bottom panel: ρ (blue solid line) and p_ρ (orange).

FIGURE 2 Typical behaviour of the swinging spring governed by Equations (4.8)–(4.12) where $m = 1$, $\ell = 1$, $g = \pi^2$, and $k = 3\pi^2$. For the experiments described in Section 4.3, an assimilation interval of 0.9 s is used.

function randi to determine the starting point t_{start} , a random 10-second segment of this 100-s run is used as the truth for the assimilation.

2. Create observations following Equation (4.16) from the truth trajectory every 0.9 s and remove the bias due to unresolved scales. This observation frequency is chosen such that each filter produced good approximations of ℓ^t , but allowed difference in performance between the various filters. The observation bias is calculated as $\langle r^t \rangle - \ell^t$, where the angular brackets denote the temporal average of the 100-second run of r^t produced in step 1 and $\ell^t = 1$. This is removed from the observations before their assimilation so that we may focus our results on uncertainty due to unresolved scales. We

note that in operational practice, the bias can be removed prior to assimilation (e.g., surface observations) or estimated online through bias correction schemes. We also note that there is no bias in θ .

3. Using the large-scale forecast model in Equations (4.13)–(4.15), evolve the state $(1 \ 0 \ 1 + \zeta)^T$, where $\zeta \sim \mathcal{N}(0, 0.2^2)$, to time t_{start} to obtain the mean forecast state at the beginning of the assimilation window.
4. Generate an initial forecast ensemble of $m = 50$ members by adding random noise $\mathbf{e}_0^{lf} \sim \mathcal{N}(\mathbf{0}, \mathbf{P}_0^{lf})$ to the mean forecast state at time t_{start} obtained in 3. The initial forecast error covariance is set to

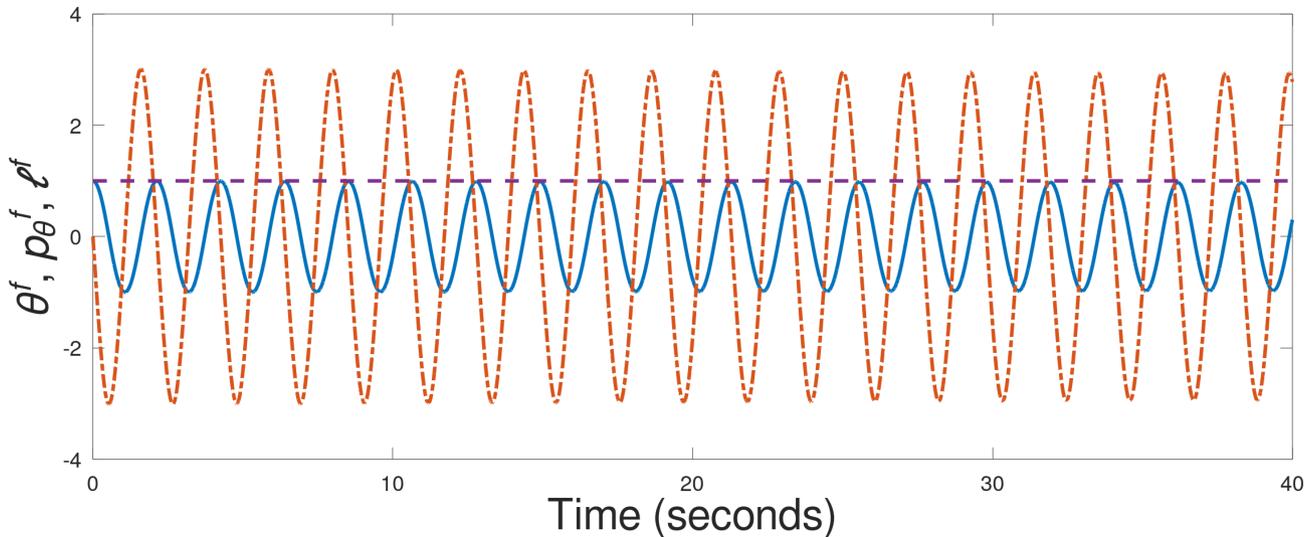


FIGURE 3 Typical behaviour of the large-scale forecast model governed by Equations (4.13)–(4.15) where $m = 1$, $\ell = 1$, $g = \pi^2$, and $k = 3\pi^2$. In this figure, θ is a blue solid line, p_θ is an orange dot-dash line, and ℓ is a dashed purple line. For the experiments described in Section 4.3, an assimilation interval of 0.9 s is used.

$$\mathbf{P}_0^{lf} = \begin{pmatrix} 0.2 & 0 & 0 \\ 0 & 0.6 & 0 \\ 0 & 0 & 0.2 \end{pmatrix}^2, \quad (4.17)$$

such that the standard deviation of the forecast error is approximately 20% the amplitude of θ^t and p_θ^t and size of ℓ^t .

This experiment is repeated 200 times and we calculate statistics from the results of each experiment. For the ETSKF, the cross-covariances \mathbf{P}^{lsf} and \mathbf{P}^{slf} are initially assumed to be zero matrices, which are evolved throughout the assimilation as described in Section 3.3. We remind the reader that the cross-covariances appear in both terms of the ETSKF gain in Equation (3.9).

The ETKF-RH and ETSKF use a climatological approximation of the representation uncertainty, obtained by taking the covariance of a single 100 model second truth run of $(\rho^t \ p_\rho^t)^T$ obtained in step 1 of the initialization procedure and mapping to observation space with \mathbf{H}^s . The climatological approximation of the representation uncertainty is given by

$$\mathbf{R}_{clim}^H = \begin{pmatrix} 0 & 0 \\ 0 & 0.28^2 \end{pmatrix}. \quad (4.18)$$

There are two sources of large-scale model error in our experiments: sampling error due to ensemble size and structural error caused by the imperfect forecast model for θ and p_θ (compare Equations 4.8–4.9 with Equations 4.13 and 4.14). Additionally, the model for ℓ^f has no growing modes (it is modelled by persistence, see Equation 4.15). Thus, practical

computations with moderate ensemble size need additive inflation to avoid ensemble collapse. However, further experiments (not shown) indicate that ℓ^f does not need model error when the ensemble size is sufficiently large. Model noise is added to each ensemble member every 0.01 model seconds. The large-scale model-noise covariance \mathbf{Q}^l is chosen such that rank histograms (Hamill, 2001) produced from the ETSKF forecast ensembles are uniform (see figures 8.3 and 8.4 of Bell (2021)). Consequently, the \mathbf{Q}^l that is used by each filter in our experiments is given by

$$\mathbf{Q}^l = \begin{pmatrix} 0.05 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 0.001 \end{pmatrix}^2. \quad (4.19)$$

4.4 | Performance metrics

To assess the performance of the ETSKF, we use two metrics.

1. The root-mean-square error (RMSE) for a single component of the state vector. For a vector of length N_k , the number of time-steps, the RMSE is given by

$$\text{RMSE} = \sqrt{\frac{\sum_{k=1}^{N_k} (\bar{x}_k^{lf} - x_k^{lt})^2}{N_k}}. \quad (4.20)$$

2. The continuous rank probability score (CRPS) which quantifies the deviation between the true distribution function and the filter's probabilistic forecast for a single component of the state vector. Calculated at time

k , the extra ensemble members $x^{lf,(0)} = -\infty$ and $x^{lf,(m+1)} = \infty$ are defined and ensemble members are sorted in ascending order (Hersbach, 2000). Assuming equal probability for each member, the CRPS is computed as

$$CRPS = \sum_{j=0}^m \left[\alpha_j \left(\frac{j}{m} \right)^2 + \beta_j \left(1 - \frac{j}{m} \right)^2 \right], \quad (4.21)$$

with α_j and β_j given in Table 1. The best possible CRPS value is 0 which corresponds to a perfect deterministic forecast.

To provide a comparison between filters, the RMSE and mean CRPS are expressed in relative percentage form,

$$\frac{A - B}{A} \times 100\% \quad (4.22)$$

where A corresponds to the value for the ETKF-LS and B corresponds to another filter. Here, a negative value indicates worse performance than the ETKF-LS, zero represents the same performance, and a positive value indicates better performance. Further, the quantity $(A - B)/A \times 100\%$ is bounded between $-\infty$ and 100, as A

TABLE 1 Values for α_j and β_j in Equation (4.21).

$0 < j < m$	α_j	β_j
$x^{lt} > x^{lf,(j)}$	$x^{lf,(j+1)} - x^{lf,(j)}$	0
$x^{lf,(j+1)} > x^{lt} > x^{lf,(j)}$	$x^{lt} - x^{lf,(j)}$	$x^{lf,(j+1)} - x^{lt}$
$x^{lf,(j)} > x^{lt}$	0	$x^{lf,(j+1)} - x^{lf,(j)}$

Note: The subscript j indicates the ensemble member in the ensemble arranged into ascending order (i.e., $x^{lf,(j)} < x^{lf,(j+1)}$).

and B can each be only positive or zero for RMSE and CRPS.

4.5 | Numerical illustration

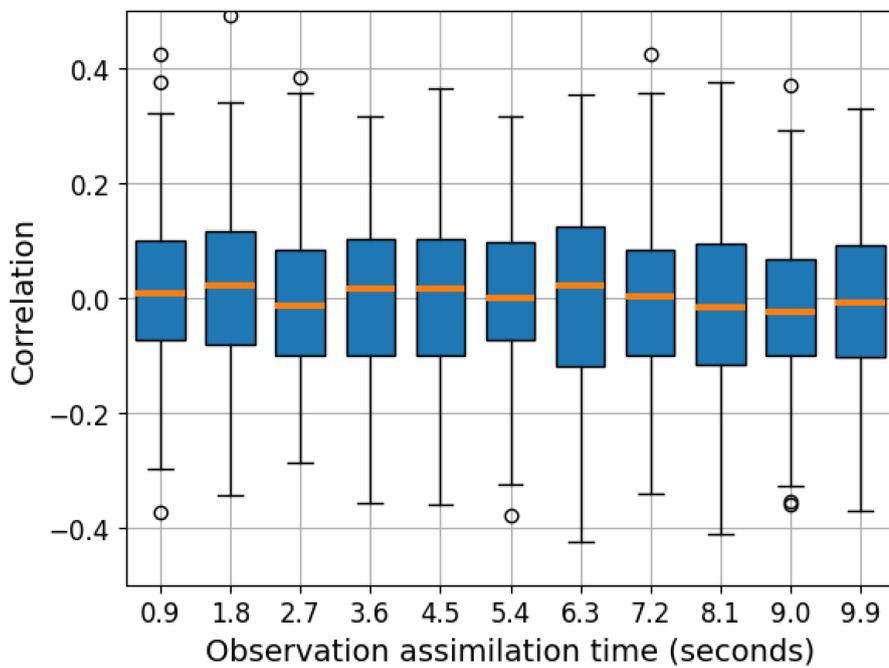
We illustrate the performance of the three filters in a regime of high representation uncertainty and low instrument uncertainty. The results for our illustration are shown in Table 2. We evaluate the filters using both forecast and analysis metrics. For the forecast evaluation, we generate forecast ensembles from each analysis. The forecast RMSE and mean CRPS are computed by comparing these predictions against the true state at each time-step between and $t = 5$ s and $t = 10$ s. For the analysis evaluation, we compare the analysis against the truth at the 6 observation times between $t = 5$ s and $t = 10$ s. The metrics calculated are then averaged over 200 experiments.

To determine the statistical significance of the results, we use the t -test. The difference in RMSE and mean CRPS for θ^a , p_θ^a , θ^f and p_θ^f between the ETKF-LS and each of the ETKF-RH and ETSKF were not found to be statistically significant. However, the differences in RMSE and mean CRPS for ℓ^a and ℓ^f between the ETKF-LS and each other filter are statistically significant. From Table 2, we see that the ETKF-RH and ETSKF have a large, positive percentage improvement in RMSE and mean CRPS for ℓ^a and ℓ^f compared to the ETKF-LS. Hence, the ETKF-RH and ETSKF on average produce a more accurate probabilistic forecast with a smaller average forecast-mean error than the ETKF-LS. Additionally, the difference in RMSE and mean CRPS between the ETKF-RH and ETSKF is not statistically significant. Therefore, the ETSKF performs similarly to the ETKF-RH in a high representation uncertainty and low instrument uncertainty regime. This can be attributed to the correlations between the large-scale forecast ensembles

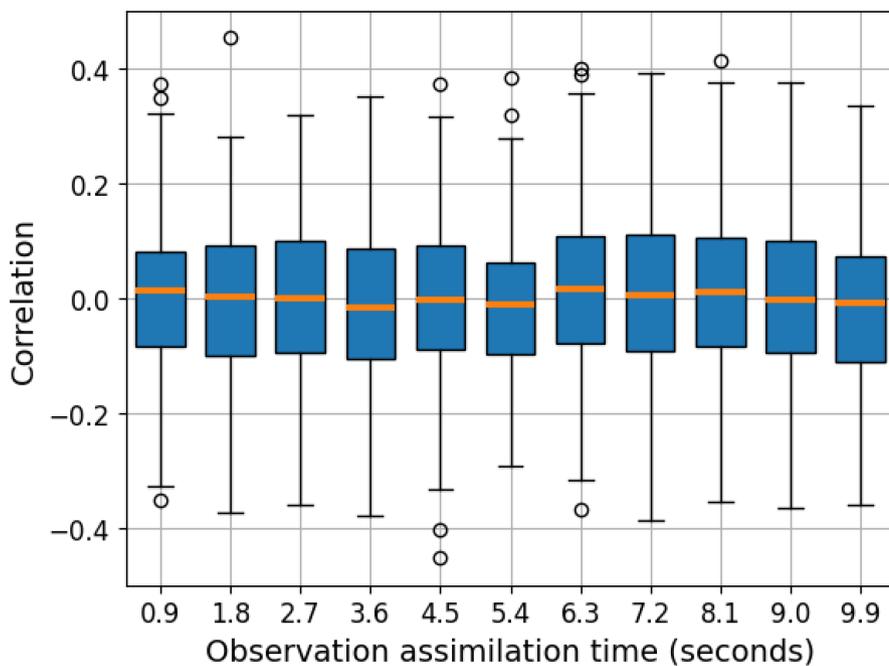
TABLE 2 Summary of the results for the RMSE and mean CRPS obtained for each filter averaged across 200 experiments.

R^I	State	ETKF-LS		ETKF-RH		ETSKF	
		RMSE	Mean CRPS	RMSE (%)	Mean CRPS (%)	RMSE (%)	Mean CRPS (%)
$0.1^2 \mathbf{I}$	θ^a	0.098	0.002	0.20	-2.84	0.21	6.76
	θ^f	0.050	0.006	1.95	2.47	-1.66	0.17
	p_θ^a	1.570	0.023	1.67	3.01	1.74	10.71
	p_θ^f	0.184	0.017	-0.17	4.20	-1.41	7.05
	ℓ^a	0.121	0.016	16.92	87.45	17.04	81.26
	ℓ^f	0.127	0.017	17.23	87.29	15.87	81.85

Note: The same model error covariance was used for each filter (see Section 4.3). The results are given in relative percentage form (see Equation (4.22)) for the ETKF-RH and ETSKF. Relative percentages in bold indicate when the improvement of the ETKF-RH or ETSKF over the ETKF-LS is statistically significant.



(a) Correlation coefficients between θ^f ensemble and sampled small-scale variability ρ .



(b) Correlation coefficients between ℓ^f ensemble and sampled small-scale variability ρ .

for the directly observed states, θ^f and ℓ^f , and the sampled small-scale variability, ρ , as shown in Figure 4. Since the correlations are weak, the cross-covariances have little effect on the ETSKF gain matrix. Additional experiments with increased instrument uncertainty and analysis of the cross-covariance terms in the ETSKF can be found in Bell (2021).

5 | CONCLUSION

We introduced a novel ensemble transform formulation of the Schmidt–Kalman filter (ETSKF) to compensate for the error due to unresolved scales when observations contain information on scales smaller than those represented by the numerical model. In the ETSKF, the small-

FIGURE 4 The correlation coefficients at each observation assimilation time over the 200 experiments. The lower and upper edges of the blue boxes represent the 25th percentile ($Q1$) and 75th percentile ($Q3$), respectively; the orange line indicates the median, and the whiskers extend from $Q1 - 1.5 \times (Q3 - Q1)$ to $Q3 + 1.5 \times (Q3 - Q1)$. Circles represent outliers outside the whisker range. The majority of the correlation coefficients are weak at each observation assimilation time, indicating they have little effect on the ETSKF performance.

scale variability is represented through an ensemble sampled from the representation error covariance.

We carried out “proof-of-concept” numerical experiments using an idealized two-scale model and three filters: an ensemble transform Kalman filter that disregards the uncertainty due to unresolved scales (ETKF-LS), an ETKF that accounts for the uncertainty due to unresolved scales by incorporating it in the observation error covariance matrix (ETKF-RH), and the ETSKF. The three filters were implemented without localization as the swinging spring system is low dimensional, and we used reasonable 50-member ensemble sizes with additive inflation (see Equation (4.19)) to prevent ensemble collapse.

We examined the improvement of the ETKF-RH and ETSKF relative to the ETKF-LS in a regime of high representation uncertainty and low instrument uncertainty. The ETKF-RH and ETSKF showed statistically significant improvements over the ETKF-LS for one of the three large-scale variables. We also found that the performance of the ETSKF was as good as the ETKF-RH. The ETSKF will be most appropriate when the large- and small-scale processes are strongly correlated. Further research is needed to determine the potential benefits of the ETSKF over the ETKF-RH in such cases.

The ETSKF has been shown to be capable of compensating for uncertainty due to unresolved scales in nonlinear systems. However, further experiments are required to understand its robustness. Specifically, the limitations highlighted by this study include the consideration of only a single observation uncertainty regime, the focus on scenarios with a larger number of ensemble members compared to state variables, and the consideration of a single observing system. Addressing these limitations will be critical to evaluating the importance of cross-covariances and determining the suitability of the ETSKF for operational DA.

AUTHOR CONTRIBUTIONS

Zackary Bell: Software; methodology; validation; visualization; writing – review and editing; writing – original draft; investigation; conceptualization; formal analysis.

Sarah L. Dance: Supervision; writing – review and editing; formal analysis; funding acquisition; writing – original draft; conceptualization; investigation; methodology; validation. **Joanne A. Waller:** Supervision; writing – review and editing; formal analysis; writing – original draft; conceptualization; investigation; methodology; validation.

FUNDING INFORMATION

Zackary Bell was supported by a University of Reading PhD studentship. Sarah L. Dance and Joanne A. Waller were supported in part by Engineering and Physical

Sciences Research Council grant EP/P002331/1 (DARE). Sarah L. Dance was also supported in part by the National Centre for Earth Observation, University of Reading, UK National Centre for Earth Observation.

CONFLICT OF INTEREST STATEMENT

No potential conflict of interests was reported by the authors.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Zackary Bell  <https://orcid.org/0000-0002-8427-1400>

Sarah L. Dance  <https://orcid.org/0000-0003-1690-3338>

REFERENCES

- Bannister, R.N. (2007) Can wavelets improve the representation of forecast error covariances in variational data assimilation? *Monthly Weather Review*, 135(2), 387–408. Available from: <https://doi.org/10.1175/MWR3279.1>
- Bell, Z. (2021) Using urban observations in numerical weather prediction: mathematical techniques for multi-scale data assimilation. PhD thesis, University of Reading. <https://doi.org/10.48683/1926.00109792>
- Bell, Z., Dance, S.L. & Waller, J.A. (2020) Accounting for observation uncertainty and bias due to unresolved scales with the Schmidt–Kalman filter. *Tellus A: Dynamic Meteorology and Oceanography*, 72(1), 1–21. Available from: <https://doi.org/10.1080/16000870.2020.1831830>
- Bell, Z., Dance, S.L. & Waller, J.A. (2022) Exploring the characteristics of a vehicle-based temperature dataset for kilometre-scale data assimilation. *Meteorological Applications*, 29(3), e2058. Available from: <https://doi.org/10.1002/met.2058>
- Buehner, M. & Shlyayeva, A. (2015) Scale-dependent background-error covariance localisation. *Tellus A: Dynamic Meteorology and Oceanography*, 67(1), 28027. Available from: <https://doi.org/10.3402/tellusa.v67.28027>
- Carrassi, A. & Vannitsem, S. (2011) Treatment of the error due to unresolved scales in sequential data assimilation. *International Journal of Bifurcation and Chaos*, 21(12), 3619–3626. Available from: <https://doi.org/10.1142/S0218127411030775>
- Clark, P., Halliwell, C. & Flack, D. (2021) A physically based stochastic boundary layer perturbation scheme. Part I: formulation and evaluation in a convection-permitting model. *Journal of the Atmospheric Sciences*, 78(3), 727–746. Available from: <https://doi.org/10.1175/JAS-D-19-0291.1>
- Daley, R. (1993) Estimating observation error statistics for atmospheric data assimilation. *Annales Geophysicae*, 11, 634–647.
- Dance, S.L., Ballard, S.P., Bannister, R.N., Clark, P., Cloke, H.L., Darlington, T. et al. (2019) Improvements in forecasting intense rainfall: results from the FRANC (forecasting rainfall exploiting new data assimilation techniques and novel observations of convection) project. *Atmosphere*, 10(3), 125. Available from: <https://doi.org/10.3390/atmos10030125>

- Gentle, J.E. (2010) *Computational statistics*. New York: Springer.
- Grooms, I., Lee, Y. & Majda, A.J. (2014) Ensemble Kalman filters for dynamical systems with unresolved turbulence. *Journal of Computational Physics*, 273, 435–452. Available from: <https://doi.org/10.1016/j.jcp.2014.05.037>
- Gustafsson, N., Janjić, T., Schraff, C., Leuenberger, D., Weissmann, M., Reich, H. et al. (2018) Survey of data assimilation methods for convective-scale numerical weather prediction at operational centres. *Quarterly Journal of the Royal Meteorological Society*, 144(713), 1218–1256. Available from: <https://doi.org/10.1002/qj.3179>
- Hamill, T.M. (2001) Interpretation of rank histograms for verifying ensemble forecasts. *Monthly Weather Review*, 129(3), 550–560. Available from: [https://doi.org/10.1175/1520-0493\(2001\)129<0550:IORHFV>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2)
- Hersbach, H. (2000) Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, 15(5), 559–570. Available from: [https://doi.org/10.1175/1520-0434\(2000\)015<0559:DOTCRP>2.0.CO;2](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. Available from: <https://doi.org/10.3402/tellusa.v67.24822>
- Hodyss, D. & Satterfield, E. (2016) The treatment, estimation, and issues with representation error modelling. In: Park, S.K. & Xu, L. (Eds.) *Data assimilation for atmospheric, oceanic and hydrologic applications*, Vol. 3. Berlin: Springer, pp. 177–194. Available from: <https://doi.org/10.1007/978-3-319-43415-5>
- Houtekamer, P.L. & Zhang, F. (2016) Review of the ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 144(12), 4489–4532. Available from: <https://doi.org/10.1175/MWR-D-15-0440.1>
- Hu, G. & Dance, S.L. (2021) Efficient computation of matrix–vector products with full observation weighting matrices in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 147(741), 4101–4121.
- Hu, G., Dance, S.L., Bannister, R.N., Chipilski, H.G., Guillet, O., Macpherson, B. et al. (2023) Progress, challenges, and future steps in data assimilation for convection-permitting numerical weather prediction: report on the virtual meeting held on 10 and 12 november 2021. *Atmospheric Science Letters*, 24(1), e1130.
- Janjić, T., Bormann, N., Bocquet, M., Carton, J., Cohn, S., Dance, S. et al. (2018) On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 144(713), 1257–1278. Available from: <https://doi.org/10.1002/qj.3130>
- Janjić, T. & Cohn, S.E. (2006) Treatment of observation error due to unresolved scales in atmospheric data assimilation. *Monthly Weather Review*, 134(10), 2900–2915. Available from: <https://doi.org/10.1175/MWR3229.1>
- Liu, Z.-Q. & Rabier, F. (2002) The interaction between model resolution, observation resolution and observation density in data assimilation: a one-dimensional study. *Quarterly Journal of the Royal Meteorological Society*, 128(582), 1367–1386. Available from: <https://doi.org/10.1256/003590002320373337>
- Lou, T.-s., Chen, N.-h., Xiong, H., Li, Y.-x. & Wang, L. (2018) Ensemble consider Kalman filtering. In: *2018 IEEE CSAA guidance, navigation and control conference (CGNCC)*. Xiamen: IEEE. Available from: <https://doi.org/10.1109/GNCC42960.2018.9018907>
- Lynch, P. (2002) The swinging spring: a simple model of atmospheric balance. In: Norbury, J. & Roulstone, I. (Eds.) *Large-scale atmosphere-ocean dynamics II*. Cambridge: Cambridge University Press, pp. 64–108.
- Schmidt, S.F. (1966) Application of state-space methods to navigation problems. In: Leondes, C.T. (Ed.) *Advances in control systems*, Vol. 3. New York: Academic Press, pp. 293–340. Available from: <https://doi.org/10.1016/B978-1-4831-6716-9.50011-4>
- Schutgens, N.A., Gryspeerdt, E., Weigum, N., Tsyro, S., Goto, D., Schulz, M. et al. (2016) Will a perfect model agree with perfect observations? The impact of spatial sampling. *Atmospheric Chemistry and Physics*, 16(10), 6335–6353. Available from: <https://doi.org/10.5194/acp-16-6335-2016>
- Simonin, D., Waller, J.A., Ballard, S.P., Dance, S.L. & Nichols, N.K. (2019) A pragmatic strategy for implementing spatially correlated observation errors in an operational system: an application to Doppler radial winds. *Quarterly Journal of the Royal Meteorological Society*, 145(723), 2772–2790. Available from: <https://doi.org/10.1002/qj.3592>
- 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. Available from: <https://doi.org/10.3402/tellusa.v65i0.19546>
- Vanneste, J. (2013) Balance and spontaneous wave generation in geophysical flows. *Annual Review of Fluid Mechanics*, 45, 147–172.
- Vetra-Carvalho S. Leeuwen P. J. Nerger L. Barth A. Altaf M. U. Brasseur P. Kirchgessner P. Beckers J.-M. 2018 State-of-the-art stochastic data assimilation methods for high-dimensional non-Gaussian problems *Tellus A: Dynamic Meteorology and Oceanography* 70(1), 1–43. <https://doi.org/10.1080/16000870.2018.1445364>
- Waller, J.A., Bauernschubert, E., Dance, S.L., Nichols, N.K., Potthast, R. & Simonin, D. (2019) Observation error statistics for Doppler radar radial wind superobservations assimilated into the DWD COSMO-KENDA system. *Monthly Weather Review*, 147(9), 3351–3364. Available from: <https://doi.org/10.1175/MWR-D-19-0104.1>
- Waller, J.A., Dance, S.L., Lawless, A.S., Nichols, N.K. & Eyre, J. (2014) Representativity error for temperature and humidity using the Met Office high-resolution model. *Quarterly Journal of the Royal Meteorological Society*, 140(681), 1189–1197. Available from: <https://doi.org/10.1002/qj.2207>
- Waller, J.A., Dance, S.L. & Lean, H.W. (2021) Evaluating errors due to unresolved scales in convection-permitting numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 147(738), 2657–2669. Available from: <https://doi.org/10.1002/qj.4043>
- Waller, J.A., Simonin, D., Dance, S.L., Nichols, N.K. & Ballard, S.P. (2016) Diagnosing observation error correlations for Doppler radar radial winds in the Met Office UKV model using observation-minus-background and observation-minus-analysis statistics. *Monthly Weather Review*, 144(10), 3533–3551. Available from: <https://doi.org/10.1175/MWR-D-15-0340.1>
- Wang, X., Chipilski, H.G., Bishop, C.H., Satterfield, E., Baker, N. & Whitaker, J.S. (2021) A multiscale local gain form ensemble

transform Kalman filter (MLGETKF). *Monthly Weather Review*, 149(3), 605–622. Available from: <https://doi.org/10.1175/MWR-D-20-0290.1>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Bell, Z., Dance, S. L., & Waller, J. A. (2025). The ensemble transform Schmidt–Kalman filter: A novel method to compensate for observation uncertainty due to unresolved scales. *Atmospheric Science Letters*, 26(5), e1296. <https://doi.org/10.1002/asl.1296>