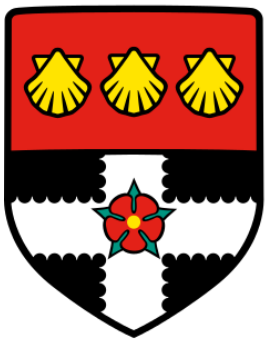


University of Reading

School of Mathematical, Physical and Computational Sciences

Using urban observations in numerical weather prediction: mathematical techniques for multi-scale data assimilation

Zackary Bell

A thesis submitted for the degree of Doctor of Philosophy

July 2021

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Publications

Chapters 5, 6 and 7 of this thesis are reproduced from the following publications respectively.

1. Bell, Z., Dance, S. L., Waller, J. A., and O'Boyle, K. (2021). Quality-control of vehicle-based temperature observations and future recommendations. *Forecasting Research Technical Report 644, Met Office, United Kingdom.*

2. Bell, Z., Dance, S. L. and Waller, J. A., 2021. Exploring the characteristics of a vehicle-based temperature dataset for convection-permitting numerical weather prediction. In review *Meteorological Applications.* `https://arxiv.org/abs/2105.12526`

3. Bell, Z., Dance, S. L. and 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), pp.1-21.

All work undertaken in these publications was carried out by Zackary Bell with co-authors providing guidance and review.

Data Statement

The Met Office trial dataset used in this thesis is not available in order to protect the privacy of the volunteer participants. We would like to thank Highways England for granting permission to use the roadside weather information station data. The MIDAS and UKV datasets used in this thesis are available from the British Atmospheric Data Centre.

Abstract

The skill of weather forecasts depends on the initial conditions obtained through data assimilation. Generation of accurate initial conditions for convection-permitting numerical weather prediction (NWP) requires assimilation of a large number of near-surface observations of high resolution. To optimally extract information from high-resolution observations, methods of data assimilation must compensate for the uncertainty due to unresolved scales. Here, we examine uncertainty due to unresolved scales from two perspectives: crowdsourced vehicle-based observations and data assimilation. To investigate the potential of crowdsourced observations for convection-permitting NWP, we examine a novel vehicle-based temperature dataset. A new quality-control procedure is developed for the vehicle-based temperature dataset. Approximately 75% of the dataset fails quality-control primarily due to missing or inaccurate metadata. The characteristics of quality-controlled data are explored through comparison with other meteorological datasets. We find that the uncertainty of vehicle-based observation-model comparisons is likely weather-dependent and possibly vehicle-dependent. We investigate two different methods to account for observation uncertainty due to unresolved scales in data assimilation. The standard approach includes the uncertainty due to unresolved scales in the observation error covariance matrix. The alternative approach, used by the Schmidt-Kalman filter (SKF), considers the variability of the small-scale processes to estimate the large-scale state. It is shown that the SKF is most suitable in regimes of high uncertainty due to unresolved scales and low instrument uncertainty. The SKF is extended to a novel ensemble transform formulation suitable for nonlinear models and shown to be most beneficial when the uncertainty due to unresolved scales is greater than the instrument uncertainty. We conclude that crowdsourced observations can help fill the gap in near-surface observations for convection-perimitting NWP. Our new ensemble transform SKF has the potential to account for the associated uncertainty due to unresolved scales in their assimilation.

Contents

Chapter 1

Introduction

For urban areas, where the population number is projected to reach 6.3 billion by 2050 (World Meteorological Organisation, 2021), weather forecasts to mitigate the effects of high-impact weather are more important than ever. Using these forecasts, the public can be alerted to the dangers of high-impact weather and appropriate countermeasures may be implemented in advance. For example, when a heatwave is forecast there will be increased media alerts about keeping cool, support will be given to organisations to reduce unnecessary travel, and community and voluntary support will be mobilised (Public Health England, 2021). The grid spacing of numerical weather prediction (NWP) models used in current operational centres is currently 1-4km (Gustafsson et al., 2018). As a result of this small grid spacing, operational NWP can focus on local weather conditions that effect the public such as rain, cloud, visibility, air temperature and wind.

To produce accurate weather forecasts of local conditions requires spatially-dense and temporally-frequent observations across the forecast area (Sun et al., 2014; Gustafsson et al., 2018; Dance et al., 2019). This is currently an issue for operational weather prediction as the horizontal coverage of near-surface observations over land is below the recommended minimum requirements (World Meteorological Organization, 2021). To fill this observation gap, there is a growing interest in the use of opportunistic data sources such as smartphones and vehicles (Mahoney III and O'Sullivan, 2013; Madaus et al., 2014; Muller et al., 2015; Waller, 2020). Such unconventional observations may be obtained for a low cost from the public through crowdsourcing methods or from partnerships with other organisations (Muller et al., 2015; Hintz et al., 2019a). Crowdsourced observations have the additional benefit that they will be densest in urban areas where there are few scientific surface observation stations. Issues regarding the use of crowdsourced observations include their uncertainty, the need for suitable metadata (e.g., location, time), and data privacy (Hintz et al., 2019a). Furthermore, crowdsourced observations will observe meteorological features that may not be represented in NWP models (Muller et al.,

2015). When crowdsourced observations are used to produce weather forecasts, it is important that the uncertainty caused by this representation difference is taken into account. In this thesis we investigate the characteristics of a crowdsourced dataset of vehicle-based observations obtained from a Met Office proof-of-concept trial.

Data assimilation is a mathematical technique that compares observations with their modelled prediction to find the most likely state of some system (Asch et al., 2016). The observations assimilated in NWP can be either in-situ (e.g. measurements from vehicle sensors) or remotely sensed (e.g., satellite observations). The uncertainty of an observation is quantified through the observation error statistics. The two contributing sources of observation error are the instrument error and the representation error. While the terminology has varied between authors, the representation error usually refers to the error that occurs whenever the prediction model is unable to perfectly represent an observed process. Using the definition of Janjić et al. (2018), the representation error is formed of three distinct errors: the pre-processing error, the observation operator error and the error due to unresolved scales. The pre-processing error is defined as any errors arising from the preparation of an observation for assimilation. The observation operator error is caused by an incorrect or approximate conversion of model quantities to observation quantities. The error due to unresolved scales is the result of any mismatch in the size or period of the processes observed and modelled.

To obtain the best initial conditions from assimilation of crowdsourced observations, the error due to unresolved scales must be accounted for by the method of data assimilation. The treatment of error due to unresolved scales has received substantial attention from the data assimilation community this decade (Janjić et al., 2018). The standard approach to compensate for error due to unresolved scales is to increase the uncertainty associated with the observation (Hodyss and Satterfield, 2016; Fielding and Stiller, 2019). An alternative, and relatively unexplored, approach is to use the statistics of the small-scale processes to increase the uncertainty associated with the observation–model comparison (Janjić and Cohn, 2006). A single observation will have a reduced influence on the initial conditions calculated through data assimilation algorithm in both of these approaches which will avoid overfitting and improve forecast skill. In this thesis we investigate the ability of the Schmidt-Kalman filter, which explicitly accounts for the statistics of the large- and small-scale processes, to compensate for error due to unresolved scales. We also develop an ensemble form of the SKF to accommodate more applications. Considering this alternative approach may also provide new insight into observation uncertainty due to unresolved scales and why it is compensated for.

1.1 Thesis aims

The objective of this thesis is to improve high-impact weather prediction via a better treatment of observation uncertainty in data assimilation. This is particularly important for low-cost crowdsourced observations that are often obtained in urban areas.

We investigate observation uncertainty due to unresolved scales from two perspectives. We first examine observation uncertainty for crowdsourced observations. The aim of this part of the thesis is to answer the following research questions:

1. **How can we quality-control the vehicle-based air-temperature dataset obtained from a Met Office proof-of-concept trial?**
   In 2018 the Met Office ran a proof-of-concept trial to collect observations of air temperature recorded by private vehicles. We investigate:
   - Which quality-control tests would be suitable for a crowdsourced vehicle-based air-temperature dataset?
   - Which improvements can be made to the data collection procedure for vehicle-based observations used by the Met Office?
2. **What are the characteristics of vehicle-based observations of air temperature?**
   The vehicle-based observations from the Met Office proof-of-concept trial are quality-controlled in chapter 5. Using the data that passed quality-control, we investigate:
   - What is the uncertainty associated with vehicle-based observation–model comparisons?
   - Does the uncertainty of vehicle-based observation–model comparisons change with weather conditions or vehicle type?

We next examine how observation uncertainty due to unresolved scales is accounted for in data assimilation. The aim of this part of the thesis is to answer the following research questions:

3. **Can the Schmidt-Kalman filter effectively treat observation error and bias due to unresolved scales?**
   The Schmidt-Kalman filter is a method of data assimilation that explicitly accounts for the statistics of unresolved processes in the filter algorithm. We investigate:
   - Which regimes of observation uncertainty is the Schmidt-Kalman filter most suitable for?

- How can biases due to unresolved scales be treated with the Schmidt-Kalman filter?

4. **How can the Schmidt-Kalman filter be adapted for nonlinear models?**
   To use the Schmidt-Kalman filter for atmospheric data assimilation, it must be extended to an ensemble formulation. We investigate:
   - How can the small-scale variability utilised in the Schmidt-Kalman filter be represented in an ensemble formulation?
   - How does the ensemble formulation of the Schmidt-Kalman filter perform in comparison to standard ensemble Kalman filters?

1.2 Principal new results

The principal new results of this thesis (numbered according to the corresponding research questions) are:

1. We developed a novel quality-control procedure for the vehicle-based dataset. The quality of the available metadata is key for the quality-control of vehicle-based observations.
2. For this dataset, we found that the standard deviation of the observation-model comparisons is approximately between 1.2°C and 1.6°C. The uncertainty of vehicle-based observation-model comparisons is likely weather-dependent and possibly vehicle-dependent.
3. The Schmidt-Kalman filter can effectively treat observation error and bias due to unresolved scales. The Schmidt-Kalman filter is most suitable in regimes of high representation uncertainty and low instrument uncertainty. Our novel Schmidt-Kalman filter with bias correction scheme is capable of correcting observation biases and accounting for observation uncertainty due to unresolved scales simultaneously.
4. We developed a novel ensemble transform Schmidt-Kalman filter that can compensate for errors due to unresolved scales in nonlinear systems. The ensemble transform Schmidt-Kalman filter is most beneficial when the uncertainty due to unresolved scales is greater than the instrument uncertainty.

1.3 Outline

The structure of this thesis is as follows.

- **Chapter 2** introduces atmospheric data assimilation and discusses methods that have been used in operational NWP centres. The Kalman filter and the ensemble transform Kalman filter are discussed in detail as they are relevant to chapters 7 and 8. The chapter is concluded by discussing different metrics that will be used to evaluate the performance of the ensemble transform Kalman filter in chapter 8.

- **Chapter 3** defines the different types of observation error that occur in atmospheric data assimilation with a focus on the scale-mismatch component of the representation error. Methods to include the representation uncertainty as part of the observation error covariance are discussed. Literature on using approximate observation error covariances to compensate for representation uncertainty is reviewed. The Schmidt-Kalman filter, a key filter researched in chapters 7 and 8, is introduced as an alternative approach to compensate for observation uncertainty due to unresolved scales.

- **Chapter 4** introduces crowdsourced observations as a new potential data source for NWP. Examples of crowdsourced observations, their quality-control and their application in meteorological studies is reviewed.

- **Chapter 5** describes a novel low-precision vehicle-based air-temperature dataset obtained from a trial run by the Met Office. The new quality-control procedure developed for this dataset is detailed. Approximately 75% of the vehicle-based temperature observations are flagged during quality-control. It is shown that the method of data collection for this dataset is currently unsuitable due to missing or inaccurate metadata for many observations. Future recommendations on the method of data collection and choice of quality-control tests are given. This chapter has been published as Bell et al. (2021b).

- **Chapter 6** contains an exploration of the characteristics of the vehicle-based observations obtained in the Met Office trial through comparison with NWP model data and other observation datasets. The possible sources of instrument and representation uncertainty for vehicle-based observations of air temperature are discussed. The uncertainty of vehicle-based observation–model comparisons is examined. It is shown that the vehicle-based observation uncertainty is likely weather-dependent and possibly vehicle-dependent. This chapter has been submitted to Meteorological Applications for publication as Bell et al. (2021a).

- **Chapter 7** examines the Schmidt-Kalman filter's ability to compensate for observation uncertainty and bias due to unresolved scales. The Schmidt-Kalman filter is shown to be most suitable in regimes of high uncertainty due to unresolved scales and low instrument uncertainty. A novel Schmidt-Kalman filter

with bias-correction scheme is derived as a method to correct observation biases and compensate for uncertainty due to unresolved scales. This chapter has been published as Bell et al. (2020).

- **Chapter 8** details the derivation of the novel ensemble transform Schmidt-Kalman filter suitable for nonlinear dynamical systems such as those encountered in NWP. Numerical experiments using the swinging spring (elastic pendulum) model show that the ensemble transform Schmidt-Kalman filter is capable of compensating for uncertainty due to unresolved scales in nonlinear dynamical systems. The ensemble transform Schmidt-Kalman filter is shown to be most beneficial when the uncertainty due to unresolved scales is greater than the instrument uncertainty and performs similarly to the ensemble transform Kalman filter that includes the representation uncertainty in the observation error covariance matrix.

- **Chapter 9** summarises the results of this thesis for the four research questions presented in chapter 1. The main implications of this work are:

  - The quality-control procedure developed in this thesis presents a possible starting point for future quality-control of vehicle-based observations of air temperature.

  - Vehicle-based observations of air temperature are a promising source of data for convection-permitting NWP, but their uncertainty is likely weather-dependent and possibly vehicle-dependent.

  - The ensemble transform Schmidt-Kalman filter can compensate for uncertainty due to unresolved scales in nonlinear dynamical systems and may be suitable for weather prediction.

  Finally, we suggest potential avenues of future work for vehicle-based observations and the ensemble transform Schmidt-Kalman filter.

Chapter 2

Introduction to atmospheric data assimilation

In atmospheric data assimilation, observations of the atmospheric state are combined with prior information, weighted by their respective error statistics, to provide a best guess of the current atmospheric state (Talagrand, 1997). Data assimilation is widely used by operational numerical weather prediction centres to obtain initial conditions to use in their forecast models (Navon, 2009).

2.1 Notation

In numerical weather prediction, the evolution of the atmospheric state is described by the nonlinear dynamical system

$$\boldsymbol{x}_k^t = \mathcal{M}(\boldsymbol{x}_{k-1}^t) - \boldsymbol{\eta}_k, \tag{2.1.1}$$

where $\boldsymbol{x}_k^t \in \mathbb{R}^{N_t}$ is the true model state vector, $\mathcal{M} : \mathbb{R}^{N_t} \to \mathbb{R}^{N_t}$ is the imperfect nonlinear forecast model used to evolve the state, $\boldsymbol{\eta}_k \in \mathbb{R}^{N_t}$ is the model error, and the subscripts k and $k-1$ indicate time. The model errors are assumed to be Gaussian with zero-mean and covariance given by the expectation of the outer product of the model errors, $\mathbf{Q}_k \equiv E[\boldsymbol{\eta}_k(\boldsymbol{\eta}_k)^T]$. If the model is assumed perfect then $\boldsymbol{\eta}_k = \mathbf{0}$ for all k.

Meteorological observations $\boldsymbol{y}_k \in \mathbb{R}^p$ can be categorised as either in-situ, where a specific point is measured, or remotely sensed, where the measurement is made by a sensor that is not in direct contact with the area being observed. Observations are related to the true atmospheric state $\boldsymbol{x}_k^t$ by

$$\boldsymbol{y}_k = \mathbf{h}_k(\boldsymbol{x}_k^t) + \mathbf{e}_k^o. \tag{2.1.2}$$

Here, $\mathbf{h}_k : \mathbb{R}^{N_t} \to \mathbb{R}^p$ is the nonlinear observation operator that maps the true state from model space to observation space and $\mathbf{e}_k^o \in \mathbb{R}^p$ is the total observation error defined as the observation departure from the truth. The observation errors are assumed to be Gaussian with zero-mean and covariance given by the expectation of the outer product of the observation errors, $\mathbf{R}_k \equiv E[\mathbf{e}_k^o(\mathbf{e}_k^o)^T]$. The errors associated with observations are the instrument error, which is determined by the type of instrument used to make the measurement, and the representation error, which is determined by any misrepresentation of an observation by the model (Hodyss and Nichols, 2015; Janjić et al., 2018). The representation error will be discussed further in chapter 3.

In addition to observations, prior information on the atmospheric state is provided by a previous forecast. This state, $\boldsymbol{x}_k^b \in \mathbb{R}^{N_t}$, is referred to as the background and is related to the true state by

$$\boldsymbol{x}_k^b = \boldsymbol{x}_k^t + \mathbf{e}_k^b, \tag{2.1.3}$$

where $\mathbf{e}_k^b \in \mathbb{R}^{N_t}$ is the background error, assumed to be Gaussian with zero-mean and covariance $\mathbf{B}_k \equiv E[\mathbf{e}_k^b(\mathbf{e}_k^b)^T]$.

2.2 Data Assimilation

Data assimilation aims to find the most likely state through application of Bayes theorem:

$$p(\boldsymbol{x}_k|\boldsymbol{y}_k) \propto p(\boldsymbol{x}_k)p(\boldsymbol{y}_k|\boldsymbol{x}_k). \tag{2.2.1}$$

Here, $p(\boldsymbol{x}_k)$ is the prior knowledge of our model state provided by the previous forecast, $p(\boldsymbol{y}_k|\boldsymbol{x}_k)$ is the likelihood of the observations given the model state, and $p(\boldsymbol{x}_k|\boldsymbol{y}_k)$ is the posterior probability of the model state given the observations. Assuming that the probability density functions are Gaussian, the prior distribution is given by

$$p(\boldsymbol{x}_k) = \frac{1}{\sqrt{(2\pi)^{N_t} det(\mathbf{B}_k)}} \exp\left[-\frac{1}{2}\left(\boldsymbol{x}_k - \boldsymbol{x}_k^b\right)^T \mathbf{B}_k^{-1}\left(\boldsymbol{x}_k - \boldsymbol{x}_k^b\right)\right], \tag{2.2.2}$$

and the likelihood is given by

$$p(\boldsymbol{y}_k|\boldsymbol{x}_k) = \frac{1}{\sqrt{(2\pi)^{p} det(\mathbf{R}_k)}} \exp\left[-\frac{1}{2}\left(\boldsymbol{y}_k - \mathbf{h}_k(\boldsymbol{x}_k)\right)^T \mathbf{R}_k^{-1}\left(\boldsymbol{y}_k - \mathbf{h}_k(\boldsymbol{x}_k)\right)\right]. \tag{2.2.3}$$

Hence, the posterior distribution may be expressed as

$$p(\boldsymbol{x}_k|\boldsymbol{y}_k) \propto \exp\left[-\frac{1}{2}\left(\boldsymbol{x}_k - \boldsymbol{x}_k^b\right)^T \mathbf{B}_k^{-1}\left(\boldsymbol{x}_k - \boldsymbol{x}_k^b\right) - \frac{1}{2}\left(\boldsymbol{y}_k - \mathbf{h}_k(\boldsymbol{x}_k)\right)^T \mathbf{R}_k^{-1}\left(\boldsymbol{y}_k - \mathbf{h}_k(\boldsymbol{x}_k)\right)\right]. \tag{2.2.4}$$

The *maximum a posteriori* (MAP) estimate is obtained by maximising the posterior probability, which may be calculated by minimising the 3D-Var cost function,

$$\mathcal{J}(\boldsymbol{x}_k) = \frac{1}{2}\left(\boldsymbol{x}_k - \boldsymbol{x}_k^b\right)^T \mathbf{B}_k^{-1}\left(\boldsymbol{x}_k - \boldsymbol{x}_k^b\right) + \frac{1}{2}\left(\boldsymbol{y}_k - \mathbf{h}_k(\boldsymbol{x}_k)\right)^T \mathbf{R}_k^{-1}\left(\boldsymbol{y}_k - \mathbf{h}_k(\boldsymbol{x}_k)\right), \tag{2.2.5}$$

with respect to $\boldsymbol{x}_k$ (Talagrand, 2010). The first term in the 3D-Var cost function is a measure of the deviation of $\boldsymbol{x}_k$ from the background $\boldsymbol{x}_k^b$ whereas the second term is a measure of the deviation of $\mathbf{h}_k(\boldsymbol{x}_k)$ from the observation $\boldsymbol{y}_k$. For linear models and observation operators, this is equivalent to the minimum variance solution when the error statistics are Gaussian as the mean and mode are equivalent (Nichols, 2010). The state that minimises equation (2.2.5) is denoted the analysis, $\boldsymbol{x}_k^a \in \mathbb{R}^{N_t}$, and is used as initial conditions for the next forecast. The analysis is related to the true state by

$$\boldsymbol{x}_k^a = \boldsymbol{x}_k^t + \mathbf{e}_k^a, \tag{2.2.6}$$

where $\mathbf{e}_k^a \in \mathbb{R}^{N_t}$ is the analysis error assumed to be Gaussian with zero-mean and covariance $\mathbf{P}_k^a \equiv \mathbf{E}\left[\mathbf{e}_k^a(\mathbf{e}_k^a)^T\right]$.

Numerical weather prediction is a large and computationally expensive procedure where $N_t = \mathcal{O}(10^9)$ and $p = \mathcal{O}(10^7)$. As such, the methods of data assimilation used for weather prediction are constrained to those that can handle a large number of degrees of freedom in a computationally feasible manner. In the following subsections, we discuss several methods of data assimilation that have been previously, or are currently, used in operational NWP.

2.3 The best linear unbiased estimate

The simplest sequential method of data assimilation can be derived through finding the minimum of equation (2.2.5) explicitly. The analysis obtained is then given by

$$\boldsymbol{x}_k^a = \boldsymbol{x}_k^b + \mathbf{B}_k\mathbf{H}_k^T\left(\mathbf{H}_k\mathbf{B}_k\mathbf{H}_k^T + \mathbf{R}_k\right)^{-1}\left(\mathbf{y}_k - \mathbf{H}_k\boldsymbol{x}_k^b\right), \tag{2.3.1}$$

where the observation operator $\mathbf{H}_k$ is linearised around the background state and $\mathbf{K}_k = \mathbf{B}_k\mathbf{H}_k^T(\mathbf{H}_k^T\mathbf{B}_k\mathbf{H}_k^T + \mathbf{R}_k)^{-1}$ is known as the Kalman gain matrix (Nichols,

2010). Assimilating data through equation (2.3.1) provides the best linear unbiased estimate (BLUE). The BLUE will be used in chapter 7 to compare the performance of two methods of data assimilation. While approximate versions of the BLUE have been used in operational weather prediction centres (e.g., Lorenc, 1981; Lorenc et al., 1991), they have mostly been abandoned in favour of more complex methods of data assimilation that do not require as many simplifying assumptions (e.g., Kalnay, 2003, pp 172–174).

2.4 Variational data assimilation

An alternative class of methods to obtain the analysis involve minimising the cost function numerically. This type of data assimilation is referred to as variational data assimilation (Talagrand, 2010). Minimising the 3D-Var cost function given by equation (2.2.5) is known as 3-dimensional variational data assimilation. The observation time-dependence can be accounted for using 4D-Var; an extension to 3D-Var that uses the dynamical model to assimilate observations at the correct time (Sasaki, 1970). The 4D-Var cost function is given by

$$\mathcal{J}(\boldsymbol{x}_0) = \frac{1}{2}\left(\boldsymbol{x}_0 - \boldsymbol{x}_0^b\right)^T \mathbf{B}_0^{-1}\left(\boldsymbol{x}_0 - \boldsymbol{x}_0^b\right) + \frac{1}{2}\sum_{k=0}^{n}\left(\boldsymbol{y}_k - \mathbf{h}_k(\boldsymbol{x}_k)\right)^T \mathbf{R}_k^{-1}\left(\boldsymbol{y}_k - \mathbf{h}_k(\boldsymbol{x}_k)\right), \tag{2.4.1}$$

subject to the constraint

$$\boldsymbol{x}_k = \mathcal{M}(\boldsymbol{x}_{k-1}), \tag{2.4.2}$$

for $k = 0, \ldots, n$. Note that for $n = 0$, equation (2.4.1) reduces to equation (2.2.5). Variational data assimilation has been successfully been used for the large systems encountered in operational numerical weather prediction for many years (Bannister, 2017).

The 3D-Var and 4D-Var cost functions are nonlinear least-squares problems which are commonly solved using Gauss-Newton methods (Courtier et al., 1994; Lawless et al., 2005). In this approach, the cost function is transformed into an incremental method where a progression of linear least-squares cost functions are solved. One key benefit of variational methods is their ability to assimilate a wide range of different observation types. They are especially appealing for remotely sensed observations, which are currently the dominant source of meteorological information for global forecasts (Eyre et al., 2020), as they can accommodate nonlinear and nonlocal observation operators. A disadvantage of variational methods is their treatment of the spatio-temporal structure of the background and model errors. We note that 4D-Var has some flow dependence as the background errors are implic-

itly evolved in the assimilation window. In addition, model error can be accounted for using a weak-constraint formulation. However, each assimilation window starts with a climatological background error covariance. The flow-dependence of the background errors may be accounted for by using hybrid methods that represent the background error covariance as a weighted sum of a climatological background error covariance and a background error covariance estimated through an ensemble of forecasts (Bannister, 2017). Hybrid 4D-Var methods are currently used in the Met Office operational global data assimilation system (Clayton et al., 2013; Lorenc and Jardak, 2018) and the ECMWF EDA system (Bonavita et al., 2012).

2.5 The Kalman filter

The Kalman filter (Kalman, 1960) is a sequential method of data assimilation for linear models and observation operators where observations are assimilated at the time they occur. In addition to minimising the trace of the analysis error covariance, the Kalman filter accounts for the spatio-temporal structure of the background errors through calculation of "flow-dependent" background error statistics (Nichols, 2010). The Kalman filter is used extensively in chapter 7.

The Kalman filter algorithm can be summarised in two main steps: a correction step and a prediction step. The correction step is achieved through assimilation of an observation and is identical to equation (2.3.1) where, by convention, the background state $\boldsymbol{x}^b$ and error covariance $\mathbf{B}$ are referred to as the forecast state $\boldsymbol{x}^f$ and the forecast error covariance $\mathbf{P}^f$. The analysis error covariance is given by

$$\mathbf{P}_k^a = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\,\mathbf{P}_k^f, \tag{2.5.1}$$

where $\mathbf{K}_k = \mathbf{P}_k^f\mathbf{H}_k^T(\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^T + \mathbf{R}_k)^{-1}$ is the Kalman gain matrix and $\mathbf{H}_k$ is the linear observation operator. Equation (2.5.1) is obtained through minimising the trace of $\mathbf{E}\left[\mathbf{e}_k^a(\mathbf{e}_k^a)^T\right]$ (e.g., Asch et al., 2016, pp 92–95).

The prediction step is achieved by evolving the analysis to the time of the next observation through the dynamical system

$$\boldsymbol{x}_{k+1}^f = \mathbf{M}\boldsymbol{x}_k^a, \tag{2.5.2}$$

where the forecast model $\mathbf{M}$ is linear and $\boldsymbol{x}_{k+1}^f$ is the BLUE of the forecast. (Note that equation (2.5.2) is obtained by taking the mean of (2.1.1) for a linear $\mathbf{M}$). The uncertainty associated with the forecast state is then given by

$$\mathbf{P}_{k+1}^f = \mathbf{M}\mathbf{P}_k^a\mathbf{M}^T + \mathbf{Q}_{k+1}. \tag{2.5.3}$$

Under the assumption of a perfect model, $\mathbf{Q}_{k+1}$ should be omitted from the forecast

error covariance update.

The Kalman filter is only applicable to systems with linear models and observation operators which makes it an unsuitable method of data assimilation for numerical weather prediction. A way to accommodate nonlinearity is to linearise the model and observation operator around the forecast or analysis states respectively, which leads to a method known as the extended Kalman filter (e.g. Gelb, 1974, pp 182 – 200). This method, however, is unsuitable for operational numerical weather prediction due to the unbounded instability resulting from the linearization of the uncertainty propagation equations (Evensen, 1992) and the computational expense for high dimensional systems. However, the extended Kalman filter is used for some applications of land-surface data assimilation (e.g., De Rosnay et al., 2013).

2.6 The ensemble Kalman filter

The ensemble Kalman filter is an extension to the Kalman filter that uses ensemble statistics to obtain approximations of the forecast uncertainty (Evensen, 1994). Use of Monte Carlo methods to approximate the flow-dependent forecast error statistics allows for the ensemble Kalman filter to be a computationally tractable method of data assimilation for the highly-nonlinear models and observation operators found in numerical weather prediction. Implementation issues resulting from small ensemble sizes are discussed in section 2.7.

2.6.1 Ensemble notation

At time t_k, the forecast is represented by an ensemble of m state estimates $\boldsymbol{x}_k^{(i)}$ for $i = 1, \ldots, m$. The ensemble mean is denoted with an over-bar and given by

$$\overline{\boldsymbol{x}}_k = \frac{1}{m} \sum_{i=1}^{m} \boldsymbol{x}_k^{(i)}. \tag{2.6.1}$$

The ensemble is stored in the ensemble state matrix given by

$$\mathcal{X}_k = \begin{pmatrix} \boldsymbol{x}_k^{(1)} & \boldsymbol{x}_k^{(2)} & \ldots & \boldsymbol{x}_k^{(m)} \end{pmatrix}, \tag{2.6.2}$$

where each column corresponds to an ensemble member. We now define the ensemble perturbation matrix, $\mathbf{X}_k \in \mathbb{R}^{N_t \times m}$, as

$$\mathbf{X}_k = \frac{1}{\sqrt{m-1}} \begin{pmatrix} \boldsymbol{x}_k^{(1)} - \overline{\boldsymbol{x}}_k & \boldsymbol{x}_k^{(2)} - \overline{\boldsymbol{x}}_k & \ldots & \boldsymbol{x}_k^{(m)} - \overline{\boldsymbol{x}}_k \end{pmatrix}. \tag{2.6.3}$$

Here, we divide by $\sqrt{m-1}$ so that the outer product of an ensemble perturbation matrix with its transpose produces the unbiased ensemble covariance (e.g., Brandt,

1999, pp 111-115),

$$\mathbf{P}_k = \mathbf{X}_k \left(\mathbf{X}_k\right)^T. \tag{2.6.4}$$

There are two main categories of ensemble Kalman filters which are determined by the method used to calculate the analysis ensemble (Houtekamer and Zhang, 2016; Vetra-Carvalho et al., 2018). Ensemble Kalman filters that generate additional random sample noise for the calculation of the analysis ensemble are referred to as stochastic. An example of a stochastic ensemble Kalman filter is the perturbed observation ensemble Kalman filter (Burgers et al., 1998) where noise vectors drawn from a Gaussian distribution with zero-mean and covariance $\mathbf{R}_k$ are added to the model equivalent of the observations. If no additional random sampling noise is generated, the ensemble Kalman filter is referred to as deterministic. Instead, the analysis ensemble is generated by updating the forecast perturbation matrix into an analysis perturbation matrix. Examples of deterministic ensemble Kalman filters include the ensemble adjustment Kalman filter (Anderson, 2001) and the ensemble transform Kalman filter. We now describe the ensemble transform Kalman filter (Bishop et al., 2001; Wang and Bishop, 2003) which is used in chapter 8.

2.6.2 The ensemble transform Kalman filter

The ensemble transform Kalman filter (ETKF) consists of a prediction step, known as the forecast, and a correction step, known as the analysis. Given an analysis ensemble at time t_{k-1}, we forecast each ensemble member to time t_k where the next observation is to be assimilated,

$$\boldsymbol{x}_k^{f,(i)} = \mathcal{M}(\boldsymbol{x}_{k-1}^{a,(i)}). \tag{2.6.5}$$

The forecast mean, $\overline{\boldsymbol{x}}_k^f$, can be obtained using equation (2.6.1) and the forecast error covariance, $\mathbf{P}_k^f$, can be calculated using equations (2.6.2)-(2.6.4).

The ETKF (Bishop et al., 2001; Wang and Bishop, 2003) uses a mean analysis state update given by

$$\overline{\boldsymbol{x}}_k^a = \overline{\boldsymbol{x}}_k^f + \mathbf{K}_k \left(\boldsymbol{y}_k - \overline{\mathbf{h}_k(\boldsymbol{x}_k^f)}\right). \tag{2.6.6}$$

The analysis ensemble at time t_k is calculated using the statistics obtained from the forecast ensemble at time t_k. We note that it is not necessary to calculate the analysis error covariance to implement the ensemble Kalman filter, which substantially reduces the computation cost of this method. To obtain the analysis ensemble, we calculate the analysis perturbation matrix, $\mathbf{X}_k^a$, scale it by $\sqrt{m-1}$, and add the mean analysis state, $\overline{\boldsymbol{x}}_k^a$, to each of its columns. The analysis perturbation matrix

$\mathbf{X}_k^a \in \mathbb{R}^{N_t \times m}$ is given by

$$\mathbf{X}_k^a = \mathbf{X}_k^f \mathbf{T}_k, \tag{2.6.7}$$

where $\mathbf{T}_k \in \mathbb{R}^{m \times m}$. To derive $\mathbf{T}_k$, we express the analysis error covariance update given by equation (2.5.1) in terms of perturbation matrices to obtain

$$\mathbf{X}_k^a \left(\mathbf{X}_k^a\right)^T = \mathbf{X}_k^f \left(\mathbf{X}_k^f\right)^T - \mathbf{X}_k^f \left(\mathbf{Y}_k^f\right)^T \mathbf{D}^{-1} \mathbf{Y}_k^f \left(\mathbf{X}_k^f\right)^T \tag{2.6.8}$$

$$= \mathbf{X}_k^f \left(\mathbf{I}_m - \left(\mathbf{Y}_k^f\right)^T \mathbf{D}_k^{-1} \mathbf{Y}_k^f\right) \left(\mathbf{X}_k^f\right)^T, \tag{2.6.9}$$

where $\mathbf{Y}_k^f \in \mathbb{R}^{p \times m}$ is given by

$$\mathbf{Y}_k^f = \frac{1}{\sqrt{m-1}} \left(\mathbf{h}_k(\boldsymbol{x}_k^{f,(1)}) - \overline{\mathbf{h}_k(\boldsymbol{x}_k^f)} \quad \dots \quad \mathbf{h}_k(\boldsymbol{x}_k^{f,(m)}) - \overline{\mathbf{h}_k(\boldsymbol{x}_k^f)}\right), \tag{2.6.10}$$

and $\mathbf{D}_k \in \mathbb{R}^{p \times p}$ is the innovation covariance given by

$$\mathbf{D}_k = \mathbf{Y}_k^f \left(\mathbf{Y}_k^f\right)^T + \mathbf{R}_k. \tag{2.6.11}$$

Hence, we find that

$$\mathbf{T}_k \mathbf{T}_k^T = \mathbf{I}_m - \left(\mathbf{Y}_k^f\right)^T \mathbf{D}_k^{-1} \mathbf{Y}_k^f. \tag{2.6.12}$$

Exploiting the identity

$$\mathbf{I}_m - \left(\mathbf{Y}_k^f\right)^T \mathbf{D}_k^{-1} \mathbf{Y}_k^f = \left(\mathbf{I}_m + \left(\mathbf{Y}_k^f\right)^T \mathbf{R}_k^{-1} \mathbf{Y}_k^f\right)^{-1} \tag{2.6.13}$$

(Bishop et al., 2001) and using the singular value decomposition on $(\mathbf{Y}_k^f)^T \mathbf{R}_k^{-1/2}$, the perturbation update matrix $\mathbf{T}_k$ may be expressed as

$$\mathbf{T}_k = \mathbf{U}_k \left(\mathbf{I}_m + \boldsymbol{\Sigma}_k \boldsymbol{\Sigma}_k^T\right)^{-1} \mathbf{U}_k^T \tag{2.6.14}$$

(Livings, 2005), where $\mathbf{U}_k$ is an orthogonal matrix containing the left-singular vectors of $(\mathbf{Y}_k^f)^T \mathbf{R}_k^{-1/2}$ and $\boldsymbol{\Sigma}_k$ contains the corresponding non-zero singular values. We note that $\mathbf{T}_k = \mathbf{U}_k(\mathbf{I}_m + \boldsymbol{\Sigma}_k \boldsymbol{\Sigma}_k^T)^{-1}$ would also be a solution to equation (2.6.12), but the symmetric choice for $\mathbf{T}_k$ ensures the ETKF is unbiased (Livings et al., 2008). Once the analysis ensemble is calculated, it is forecast to the time of the next observation using equation (2.6.5) and the cycle repeated.

The ETKF is currently used for convection-permitting data assimilation by the German weather service (DWD) in the kilometre-scale ensemble data assimilation (KENDA) system (Schraff et al., 2016) and is a promising candidate for proba-

bilistic forecasting with the HARMONIE (Hirlam ALADIN Research on Meso-scale Operational NWP in Euromed) convection-permitting ensemble prediction system (Frogner et al., 2019). The flow dependence of the forecast errors result in more realistic error variances, and hence analysis updates, in domains with complex topography are more realistic in comparison with standard variational methods (Pu et al., 2013; Ha and Snyder, 2014). However, in order for the ETKF to be used successfully for large systems, the effects of sampling errors caused by small ensemble sizes must be compensated for. In the following section, we describe additional measures that are taken in order for the ensemble Kalman filter to be implemented for operational NWP.

2.7 Implementation issues with the ensemble Kalman filter

The performance of the ensemble Kalman filter is highly dependent on the ensemble size, which is often several orders of magnitude smaller than the number of model grid points and observations. In subsections 2.7.1 and 2.7.2, we describe these problems and commonly adopted methods to compensate for them.

2.7.1 Inflation

The variability represented by a forecast ensemble will systematically underestimate the true error variance of the mean primarily due to sampling errors (Lorenc, 2003). We note that other causes of underestimated ensemble variability can be due to errors from simplified model dynamics, incorrect parameterizations and systematic observation errors (Furrer and Bengtsson, 2007). Practical solutions to increase the forecast variability include multiplicative covariance inflation (e.g., Anderson and Anderson, 1999; Zheng, 2009) or additive inflation (i.e. additive model error) in the evolution of individual forecast ensemble members (e.g., Mitchell and Houtekamer, 2000; Raanes et al., 2015). We note that use of filters designed to avoid the need for inflation altogether (e.g., Bocquet, 2011; Luo and Hoteit, 2011) or relaxation-to-prior methods (e.g., Zhang et al., 2004; Whitaker and Hamill, 2012) may also be used to increase forecast ensemble variability.

Multiplicative inflation is achieved by multiplying the forecast error covariance by a scalar $\tau > 1$ where τ is determined through consideration of the dynamical system and filter used (Hamill et al., 2001). In simple models the inflation factor τ is close to 1 and multiplicative inflation tends to work well. In complex NWP models, higher values of τ are needed which can lead to error growth in regions with sparse observations. To account for this, adaptive algorithms to determine τ have been developed (e.g., Anderson, 2009; El Gharamti, 2018). Multiplicative inflation is

currently used in KENDA (Schraff et al., 2016).

Additive inflation (i.e. model error) can be incorporated into the ensemble Kalman filter by adding random noise to an ensemble member every time-step (Houtekamer et al., 2009). It is common to take the covariance of this random noise to be proportional to a climatological background error covariance used in variational assimilation systems (Mitchell and Houtekamer, 2000; Houtekamer and Zhang, 2016). Therefore, the covariance of the additive noise can be viewed as a tuning parameter used to increase the spread of the forecast ensemble. However, if the model error covariance is specified in this way, any spatio-temporal structures in the true model errors may not be fully represented.

While either inflation method is suitable to increase the forecast variability, Whitaker and Hamill (2012) found that using a combination of multiplicative inflation and additive inflation produced a better filter performance than using either method independently. This lead to their hypothesis that multiplicative inflation is best suited to compensate for sampling error, whereas additive inflation is best suited to compensate for model error. In this thesis, additive inflation will be used to compensate for model error in chapter 8.

2.7.2 Localisation

Undersampling in the ensemble Kalman filter can lead to spurious correlations between physically unrelated variables or over unrealistically large distances (Houtekamer and Mitchell, 1998; Anderson, 2012). As ensemble sizes increase, the true error statistics will be better approximated through sample statistics and hence these correlations become less of an influencing factor. For NWP, where it is computationally infeasible to have an ensemble large enough to circumvent long-range and spurious correlations, it is common to employ a technique known as localisation where long-range covariances are damped.

Covariance localisation is built on the assumption that correlations between variables of a dynamical system decrease at a fast rate with physical distance (Hamill et al., 2001). There are two primary methods of localisation that are employed in NWP (Greybush et al., 2011). The first method involves taking the Schur product (element-wise multiplication) of the forecast error covariance, $\mathbf{P}^f$, with a localisation matrix to remove any correlations between far away grid points. The localisation matrix will be a square-symmetric positive-definite matrix of the same dimension as the forecast error covariance matrix (e.g., Gaspari and Cohn, 1999). Using $\mathbf{P}^f$ localisation will also increase the rank of the ensemble and allow more degrees of freedom for the analysis update (Vetra-Carvalho et al., 2018). The second method involves multiplying the inverse of the observation error covariance, $\mathbf{R}^{-1}$, by a distance-dependent function such that observations which are far away have

infinite uncertainty. We note that, to obtain a similar effect to $\mathbf{P}^f$ localisation, $\mathbf{R}^{-1}$ localisation is used in conjunction with a technique known as domain localisation (Hunt et al., 2007). In domain localisation, only observations local to the analysis state correction are used (Ott et al., 2004). Using domain localisation can result in significant computational savings as the analysis updates on the smaller domains may be done independently and hence in parallel (Nerger et al., 2006).

The optimal static localisation length-scale may be determined by performing multiple experiments using different localisation length-scales and using the value that produces the smallest root-mean-square error. To avoid such numerical tuning experiments, adaptive localisation methods may be employed (e.g., Bishop and Hodyss, 2009a,b). These methods, however, still require tuning through an observing system simulation experiment which may be more costly than determining the optimal static localisation length-scale (Vetra-Carvalho et al., 2018). In the KENDA system, a simple adaptive horizontal $\mathbf{R}^{-1}$ localisation scheme and a static vertical $\mathbf{R}^{-1}$ localisation scheme are used (Schraff et al., 2016).

2.8 Performance metrics for the ensemble Kalman filter

To gauge the performance of an ensemble Kalman filter it is useful to examine the results of several performance metrics. We now detail three different performance metrics used in this thesis.

2.8.1 Root mean squared error

The root mean squared error (RMSE) may be used to determine the average error size of a time-series (e.g. Jolliffe and Stephenson, 2012, pp 82-84). For a vector of length N_k, the RMSE is given by

$$\text{RMSE} = \sqrt{\frac{\sum_{k=1}^{N_k} (x_k - x_k^t)^2}{N_k}}. \tag{2.8.1}$$

The RMSE is suitable for experiments where the truth is known such as twin experiments where a truth run is recreated using data assimilation. Some limitations of the RMSE are that the direction of the errors are not accounted for and that the RMSE value will be much more sensitive to large errors than small errors.

The RMSE will be used in chapters 7 and 8 to calculate the average forecast and analysis error for Kalman filter methods. For ensemble Kalman filters, the ensemble mean is used to calculate the RMSE. We also use RMSE in chapter 6 to compare forecasts with observations.

2.8.2 Rank histograms

Rank histograms, or Talagrand diagrams, may be used to show how well the true uncertainty of the observations is captured by an ensemble Kalman filter (Hamill, 2001; Candille and Talagrand, 2005). This is achieved by examining where observations lie within the forecast ensemble.

Rank histograms are constructed for scalar observations as follows. At observation time k, the forecast ensemble is mapped into observation space and arranged into ascending order as $\mathrm{H}x^{f,(1)} \leq \mathrm{H}x^{f,(2)} \leq \ldots \leq \mathrm{H}x^{f,(m)}$. If the observation is imperfect, we add random noise distributed by the observation error statistics to each ensemble member before they are arranged into ascending order. Using the sorted ensemble, we define the $m+1$ histogram bins $(-\infty, \mathrm{H}x^{f,(1)}], (\mathrm{H}x^{f,(1)}, \mathrm{H}x^{f,(2)}], \ldots, (\mathrm{H}x^{f,(m)}, \infty)$. If the observation lies in the i-th bin, also called a rank, we save that value in a list. This is repeated for each observation and a histogram plotted using the saved values.

The shape of the rank histogram will provide insight on the dispersion characteristics of the forecast ensemble (e.g. Jolliffe and Stephenson, 2012, pp 146–149). Common shapes for rank histograms are shown in figure 2.1. Flat rank histograms (figure 2.1a) suggest that the forecast ensemble is reliable or consistent. A $\cap$-shaped rank histogram (figure 2.1b) suggests that there is too much variability (i.e. overdispersive) in the ensemble whereas a $\cup$-shaped rank histogram (figure 2.1c) suggests there is too little variability (i.e. underdispersive) in the ensemble. An asymmetric rank histogram (figure 2.1d) suggests that the ensemble is biased. If the left ranks are largest then the bias will be positive and if the right ranks are largest then the bias will be negative.

Rank histograms are a useful method to visually check the dispersion characteristics of ensemble systems. We note that a flat rank histogram does not necessarily indicate a reliable forecast (Hamill, 2001). For example, the forecast ensembles used to generate a flat rank histogram may have different conditional biases of the covariance between model grid points can be misrepresented. We also note that rank histograms are sensitive to the presence of observation errors (Saetra et al., 2004). If the incorrect observation error statistics are used to perturb the ensemble, it is likely that the middle ranks will be over or under populated. To avoid misinterpretation of the forecast dispersion characteristics, it is vital that a good approximation of the true observation error statistics is used to perturb the ensemble. Rank histograms will be used in chapter 8 to diagnose the dispersion characteristics of forecast ensembles.

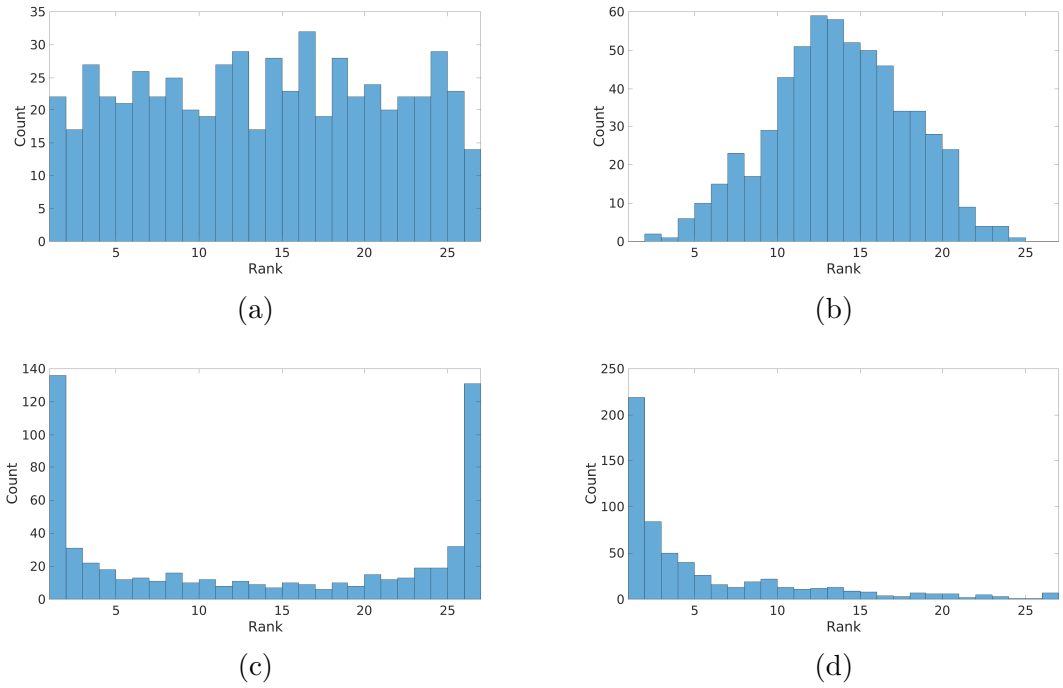


Figure 2.1: Common rank histogram shapes. An ensemble which captures the true variability of the observations well will have a uniform distribution as in (a). An ensemble that is too variable will have a ∩-shape as in (b). An ensemble that has too little variability will have a ∪-shape as in (c). A positively biased ensemble will have their ranks skewed to the left side as in (d).

2.8.3 Continuous ranked probability score

The continuous ranked probability score (CRPS) is an attractive performance metric as it summarises forecast verification information on reliability, resolution, and uncertainty into a single score (Hersbach, 2000). Furthermore, it can be used to compare the accuracy of probabilistic forecasts with deterministic forecasts (Casati et al., 2008). Denoting the probability density function for the forecast ensemble as $f(x^f)$, we define

$$\text{CRPS} = \int_{-\infty}^{\infty} \left[F(x^f) - \Theta(x^f - x^t)\right]^2 dx^f, \tag{2.8.2}$$

where $F(x^f)$ and $\Theta(x^f)$ are the cumulative distribution functions given by

$$F(x) = \int_{-\infty}^{x} f(y)dy, \tag{2.8.3}$$

$$\Theta(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}. \tag{2.8.4}$$

Here, $\Theta(x)$ is known as the Heaviside function and is used in forecast verification for point values (Casati et al., 2008). We also note that the CRPS is defined for scalar quantities and so is evaluated for each variable in the state vector. The best

$0 < i < m$	α_i	β_i
$x^t > x^{f,(i+1)}$	$x^{f,(i+1)} - x^{f,(i)}$	0
$x^{f,(i+1)} > x^t > x^{f,(i)}$	$x^t - x^{f,(i)}$	$x^{f,(i+1)} - x^t$
$x^{f,(i)} > x^t$	0	$x^{f,(i+1)} - x^{f,(i)}$

Table 2.1: Values for α_i and β_i in equation (2.8.8).

possible value of the CRPS is 0 and the worst possible value is 1. For a deterministic forecast, the CRPS is equivalent to the mean absolute error.

The application of CRPS for ensemble forecasts has been derived by Hersbach (2000). The (scalar) forecast ensemble members are assumed to be equally probable and arranged in ascending order such that their cumulative distribution function is given by

$$F(x) = \frac{1}{m} \sum_{i=1}^{m} \Theta(x - x^{f,(i)}). \tag{2.8.5}$$

Defining $x^{f,(0)} = -\infty$ and $x^{f,(m+1)} = \infty$, we have for $x^{f,(i)} < x < x^{f,(i+1)}$ that

$$c_i = \int_{x^{f,(i)}}^{x^{f(i+1)}} \left(\frac{i}{m} - \Theta(x - x^t) \right)^2 dx \tag{2.8.6}$$

where

$$CRPS = \sum_{i=0}^{m} c_i. \tag{2.8.7}$$

This may be rewritten as

$$c_i = \alpha_i \left(\frac{i}{m} \right)^2 + \beta_i \left(1 - \frac{i}{m} \right)^2 \tag{2.8.8}$$

where the values of α_i and β_i are given in table 2.1. A pictorial example of the CRPS is shown in 2.2. It should be noted that outliers will increase the CRPS substantially. The units for the CRPS are the same as for the forecast ensemble members. The CRPS will be used to verify the accuracy of forecast ensembles in chapter 8.

2.9 Summary

In this chapter we have introduced data assimilation for numerical weather prediction. A brief overview of optimal interpolation and variational data assimilation has been given. A more detailed explanation of the Kalman filter and its extension to the ensemble Kalman filter have been provided as they are both pertinent to

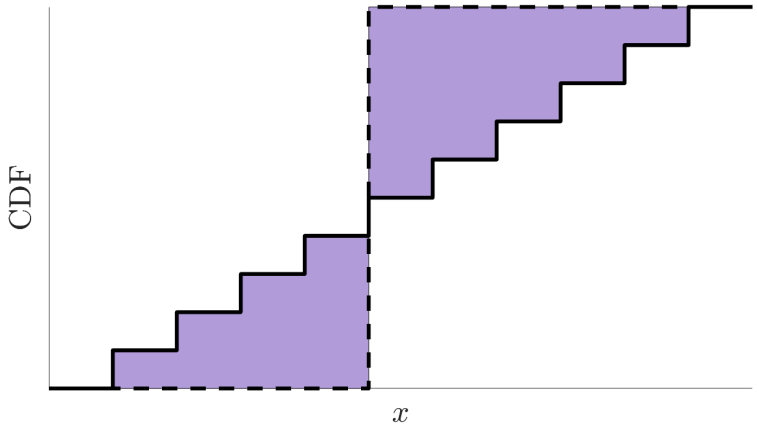


Figure 2.2: The true cumulative distribution function (dashed line) and a cumulative distribution function of a forecast ensemble with 10 members (solid line). The continuous ranked probability score is the total shaded area.

chapters 7 and 8. Performance metrics have been described. These are used in chapter 8. In the next chapter, we discuss observation errors in data assimilation and how we may compensate for the scale-mismatch component of the representation uncertainty.

Chapter 3

Observation uncertainty due to unresolved scales

To produce the optimal estimate of a state through data assimilation requires correctly specified error covariances. While it is impossible to know the exact statistics in practice, a good state estimate may still be obtained using approximations of the true error statistics. In this chapter we define each component of the observation error and discuss methods to approximate the representation error covariance. We also introduce an alternative approach to compensate for error due to unresolved scales that is used in this thesis.

3.1 Partitioned formulation of the complete system dynamics and observations

In this section, we establish notation to mathematically define the components of the observation error. In particular, we introduce notation for the complete system dynamics and observations in a partitioned form such that the large-scales resolved by an NWP model and small-scales unresolved by an NWP model are separated. This is because in NWP, the numerical models used to predict the weather will be unable to represent all of the observed spatio-temporal scales (Janjić et al., 2018). This partitioned form of the complete system dynamics will be used in chapters 7 and 8.

To express the complete system dynamics in partitioned form, we assume that the true state $\boldsymbol{x}_k^t$ may be partitioned as $(\boldsymbol{x}_k^{l,t} \;\; \boldsymbol{x}_k^{s,t})^T$ where $\boldsymbol{x}_k^{l,t} \in \mathbb{R}^{N_l}$ is the true large-scale state, $\boldsymbol{x}_k^{s,t} \in \mathbb{R}^{N_s}$ is the true small-scale state and $N_t = N_l + N_s$. The dynamical system describing the large- and small-scale dynamics is assumed to have the

form

$$\begin{pmatrix} \boldsymbol{x}^{l,t} \\ \boldsymbol{x}^{s,t} \end{pmatrix}_k = \begin{pmatrix} \mathbf{M}^l & \mathbf{0}_{N_l \times N_s} \\ \mathbf{M}^{sl} & \mathbf{M}^s \end{pmatrix} \begin{pmatrix} \boldsymbol{x}^{l,t} \\ \boldsymbol{x}^{s,t} \end{pmatrix}_{k-1} - \begin{pmatrix} \boldsymbol{\eta}^l \\ \boldsymbol{\eta}^s \end{pmatrix}_k , \tag{3.1.1}$$

where $\mathbf{M}^l \in \mathbb{R}^{N_l \times N_l}$ is the large-scale model assumed to contain any subgrid-scale parameterizations used to mimic the influence of the small-scale processes (Janjić and Cohn, 2006), $\mathbf{M}^{sl} \in \mathbb{R}^{N_s \times N_l}$ is the contribution from the large-scale processes to the small-scale state, $\mathbf{M}^s \in \mathbb{R}^{N_s \times N_s}$ is the small-scale model, and $\boldsymbol{\eta}_k^l \in \mathbb{R}^{N_l}$ and $\boldsymbol{\eta}_k^s \in \mathbb{R}^{N_s}$ are the large- and small-scale model errors respectively assumed to have zero mean. We note that $\mathbf{M}^l$ is meant to be representative of the prediction models used in NWP (Janjić and Cohn, 2006).

The state error covariance $\mathbf{P}_k \equiv \left\langle (\boldsymbol{x}_k - \boldsymbol{x}_k^t)(\boldsymbol{x}_k - \boldsymbol{x}_k^t)^T \right\rangle \in \mathbb{R}^{N_t \times N_t}$ will be in the partitioned form

$$\mathbf{P}_k = \begin{pmatrix} \mathbf{P}_k^{ll} & \mathbf{P}_k^{ls} \\ \mathbf{P}_k^{sl} & \mathbf{P}_k^{ss} \end{pmatrix}, \tag{3.1.2}$$

where $\mathbf{P}_k^{ll} \in \mathbb{R}^{N_l \times N_l}$ and $\mathbf{P}_k^{ss} \in \mathbb{R}^{N_s \times N_s}$ are the large- and small-scale state error covariances, and $\mathbf{P}_k^{ls} \in \mathbb{R}^{N_l \times N_s}$ is the cross-covariance between the large- and small-scale state errors and satisfies $\mathbf{P}_k^{sl} = \left(\mathbf{P}_k^{ls}\right)^T$.

Similarly to $\mathbf{P}_k$, the model error covariance matrix $\mathbf{Q}_k$ will also be in block form comprised of the model error covariances for the large- and small-scales along the diagonal and the cross-covariance between the large- and small-scales on the off-diagonal:

$$\mathbf{Q}_k = \begin{pmatrix} \mathbf{Q}_k^{ll} & \mathbf{Q}_k^{ls} \\ \mathbf{Q}_k^{sl} & \mathbf{Q}_k^{ss} \end{pmatrix}. \tag{3.1.3}$$

Each model error block has the same dimension as the corresponding block of the state error covariance given by equation (3.1.2). Now that we have established the notation, we next consider the components of the observation error.

3.2 Observation errors in data assimilation

The observations assimilated into NWP models can be classified as remotely sensed or in-situ. Remotely sensed observations are obtained from instruments not in direct contact with the area they are observing. Satellite observations are an example of remotely sensed observations and are currently the dominant source of observations for global NWP (Eyre et al., 2020). In-situ observations have the property that they observe a single point in space. Such observations are the main source of information

in the surface-layer and are mostly comprised of meteorological stations, aircraft, and radiosondes (Candy et al., 2021). In chapter 4, we discuss unconventional in-situ observations which are not currently widely used in operational NWP.

In data assimilation, the uncertainty of observations can be attributed to two sources: the instrument uncertainty and the representation uncertainty. The instrument uncertainty corresponds to any errors made by the instrument in the observation process. The representation uncertainty corresponds to errors caused by the misrepresentation of an observation by the model. Following the definition of Janjić et al. (2018), the representation error consists of the pre-processing error, the observation operator error and the error due to unresolved scales. In this thesis, we assume that an observation at time t_k may be expressed as the sum of the true large- and small-scale states mapped into observation space,

$$\boldsymbol{y}_k = \mathbf{h}_k^{l,t}(\boldsymbol{x}_k^{l,t}) + \mathbf{h}_k^{s,t}(\boldsymbol{x}_k^{s,t}) + \boldsymbol{\epsilon}_k + \boldsymbol{\nu}_k, \tag{3.2.1}$$

where $\mathbf{h}_k^{l,t} : \mathbb{R}^{N_l} \to \mathbb{R}^p$ and $\mathbf{h}_k^{s,t} : \mathbb{R}^{N_s} \to \mathbb{R}^p$ are the true large- and small-scale observation operators respectively, $\boldsymbol{\epsilon}_k \in \mathbb{R}^p$ is the instrument error, and $\boldsymbol{\nu}_k \in \mathbb{R}^p$ is the pre-processing error. We now discuss each component of the representation error in greater detail.

The pre-processing error, $\boldsymbol{\nu}_k$, is caused by the incorrect preparation of an observation for assimilation. For an observation to be assimilated, there must be adequate knowledge of the observation metadata such as location and time the observation was taken. If these are incorrectly specified, the data assimilation algorithm will not compare the observation with the corresponding prediction resulting in a pre-processing error. In practice, quality-control procedures are used to reject any observations that cannot be modelled correctly or are in gross error (Zahumenskỳ, 2004). Observations that are incorrectly accepted by the quality-control procedure and assimilated will also exhibit pre-processing error. The quality-control of observations for NWP will be discussed in chapter 4. We also note that observations processed or corrected before assimilation can exhibit pre-processing error. For instance, surface observations of air temperature are corrected to model-level height using environmental lapse rates (e.g., Dutra et al., 2020). The correction applied to these observations will be approximate and hence result in a pre-processing error.

The observation operator error is caused by an incorrect or approximate observation operator used to map the state variables to observation space. Therefore, using equation (3.2.1), the observation operator error is given by

$$\boldsymbol{\gamma}_k^l = \mathbf{h}_k^{l,t}(\boldsymbol{x}_k^{l,t}) - \mathbf{h}_k^{l}(\boldsymbol{x}_k^{l,t}), \tag{3.2.2}$$

where $\mathbf{h}_k^l : \mathbb{R}^{N_l} \to \mathbb{R}^p$ is the large-scale observation operator used by the filter. This

error will usually occur for remotely sensed observations where the observed quantities are related to the model state variables through complicated nonlinear equations. For example, Doppler radial winds exhibit observation operator uncertainty at far ranges if the observation operator does not account for the beam broadening and bending (Waller et al., 2016c).

The error due to unresolved scales occurs when there is a difference in scales and processes observed and modelled. The error due to unresolved scales is state-dependent and correlated in time resulting in a potential representation error bias (Janjić and Cohn, 2006; Waller et al., 2014b). Methods to compensate for error due to unresolved scales will depend on the observation footprint, which may be larger or smaller than the model grid (Janjić et al., 2018). For in-situ observations, the spatio-temporal scales observed will usually be smaller than those represented in the model. However, for remotely sensed observations, the reverse is possible where scales represented in the model are smaller than observed (Janjić et al., 2018). In this thesis, we consider observations that contain information on scales smaller than those represented by an NWP model. Hence, as we will explain in chapter 7, the error due to unresolved scales is given by $\mathbf{h}_k^{s,t}(\boldsymbol{x}_k^{s,t})$. We note that as the error due to unresolved scales arises when the forecast model is a truncated representation of the truth (Hodyss and Nichols, 2015), the scale-mismatch could also be treated as a model error as there is an improper comparison of the observation and the model state (e.g., Carrassi and Vannitsem, 2011; Mitchell and Carrassi, 2015). Such an approach would still increase the uncertainty associated with the observation-model comparison in the analysis step.

The remainder of this chapter will discuss how observation error due to unresolved scales may be accounted for in data assimilation. Investigation on how the error due to unresolved scales changes for an opportunistic observation type will be presented in chapter 6. Chapters 7 and 8 will present new research on how to compensate for the error using an approach described in section 3.4.

3.3 The observation error covariance

As discussed in chapter 2, the observation error covariance, $\mathbf{R}_k$, is used to penalise the deviations of $\mathbf{h}_k(\boldsymbol{x}_k)$ from the observations $\boldsymbol{y}_k$ in the 3D-Var and 4D-Var cost functions given by equations (2.2.5) or (2.4.1) respectively. Therefore, to produce a good analysis, it is vital that the contributions to the observation error are understood so that a good approximation of $\mathbf{R}_k$ may be calculated.

In operational centres, it is typical to compensate for the error due to unresolved scales by including its statistics in the total observation error covariance. Mathematically, the total observation error covariance may be expressed as $\mathbf{R}_k = \mathbf{R}_k^I + \mathbf{R}_k^H$,

where $\mathbf{R}_k^I$ is the instrument error covariance and $\mathbf{R}_k^H$ is the representation error covariance. The instrument error covariance, $\mathbf{R}_k^I$, will depend on the instrument used to measure the observation. It is commonly assumed that the instrument error is uncorrelated such that $\mathbf{R}_k^I$ is a diagonal matrix. The representation error covariance, $\mathbf{R}_k^H$, has a more complicated structure. In particular, the error due to unresolved scales is state-dependent and correlated (Janjić and Cohn, 2006; Waller et al., 2014b). Throughout this thesis, accounting for representation uncertainty by including it in observation error covariance matrix will be referred to as the standard approach and is used in chapters 7 and 8.

3.3.1 Residual-based diagnosis

In data assimilation, the representation error statistics are unknown and methods to approximate observation error statistics are employed. Statistical diagnostic methods such as the Hollingsworth-Lönnberg method (Hollingsworth and Lönnberg, 1986; Garand et al., 2007) or the Desroziers et al diagnostic (Desroziers et al., 2005) are common methods to obtain an approximation of the entire observation error covariance.

The Hollingsworth-Lönnberg method estimates the complete innovation error covariance using the relationship

$$E\left[\left(\boldsymbol{y}_k - \mathbf{h}_k(\boldsymbol{x}_k^b)\right)\left(\boldsymbol{y}_k - \mathbf{h}_k(\boldsymbol{x}_k^b)\right)^T\right] = \mathbf{H}_k\mathbf{B}_k\mathbf{H}_k^T + \mathbf{R}_k. \tag{3.3.1}$$

To separate the covariances, it was originally assumed that the background errors are spatially correlated whereas the observations are not. Therefore, any spatial observation error correlations caused by the unresolved scales are neglected. This method was adapted to account for observation error correlations by (Garand et al., 2007) by fitting correlation functions to the background uncertainties. This modified version of the Hollingsworth-Lönnberg method has been used to show significant interchannel correlations for various satellite instruments (Garand et al., 2007; Bormann and Bauer, 2010; Bormann et al., 2010). However, the size of the observation error correlations is dependent on the fitted correlation function. Determining the best correlation function to use is difficult and makes the results of the modified Hollingsworth-Lönnberg method subjective.

The Desroziers et al diagnostic estimates the complete observation error covariance using the relationship

$$E\left[\left(\boldsymbol{y}_k - \mathbf{h}_k(\boldsymbol{x}_k^a)\right)\left(\boldsymbol{y}_k - \mathbf{h}_k(\boldsymbol{x}_k^b)\right)^T\right] = \mathbf{R}_k. \tag{3.3.2}$$

This diagnostic has been widely used to estimate satellite interchannel error correlations (e.g., Bormann and Bauer, 2010; Bormann et al., 2010; Weston, 2011; Weston

et al., 2014; Stewart et al., 2014; Campbell et al., 2017; Gauthier et al., 2018) and spatial correlations (e.g., Waller et al., 2016a,c; Cordoba et al., 2017; Michel, 2018). However, it is still common for uncorrelated observation error covariances to be used in operational data assimilation systems. For example, the Desroziers et al diagnostic has been used to obtain diagonal observation error covariances in the COSMO-KENDA system by neglecting the off-diagonal components (Schraff et al., 2016; Lange and Janjić, 2016; Zeng et al., 2021). We note that the results of the diagnostic will only be correct if the true statistics are used in the assimilation. Since approximate error statistics will be used in operational data assimilation, the Desroziers et al diagnostic will produce an inexact approximation of the complete observation error covariance for operational systems. Theoretical interpretation of the results obtained from the Desroziers et al diagnostic when incorrect error statistics are used are given in Waller et al. (2016b) and Ménard (2016). Incorporation of the Desroziers et al diagnostic into the ensemble Kalman filter algorithm has also been researched (Li et al., 2009; Miyoshi et al., 2013; Waller et al., 2014a, 2017).

Both the Hollingsworth-Lönnberg method and the Desroziers et al diagnostic may be computed offline such that no extra computational cost is incurred in operational systems. A large number of observation samples are required for these methods which are usually taken over an assimilation window. Hence, when using either method it is implicitly, and incorrectly, assumed that the errors are not changing over time such that $\mathbf{R}$ is time-independent.

3.3.2 The metrological approach

An alternative approach to approximate the total observation uncertainty is to estimate each component of the observation error covariance matrix individually and combine them. In this approach, the instrument error covariance is usually supplied by the manufacturer and the representation error covariance is approximated from first principles. A common method to estimate the uncertainty due to unresolved scales is to use high-resolution data as a proxy for the truth. This dataset can be high-resolution observations (e.g., Oke and Sakov, 2008) or high-resolution model data (Daley, 1993; Liu and Rabier, 2002; Schutgens et al., 2016).

Oke and Sakov (2008) presented two methods to approximate the uncertainty due to unresolved scales using high-resolution observations for ocean data assimilation. The first method averages the observations to the resolution of the model and then interpolates this average back to the high-resolution grid. The difference between the high-resolution observations and the interpolated averages is taken as the error due to unresolved scales. If the high-resolution observations do not resolve all scales, the uncertainty due to unresolved scales will be underestimated. We also note that the interpolation can cause an additional source of error. The second method

approximates the uncertainty due to unresolved scales for a grid box by calculating the variance of the high-resolution observations contained within it. For this method to produce acceptable approximations, a grid box must contain a sufficient number of high-resolution observations that adequately span it. Therefore, the estimates that are obtained from this method can exhibit substantial sampling error.

Approximation of the uncertainty due to unresolved scales using high-resolution model data was first presented by Daley (1993) and later used by Liu and Rabier (2002). This method assumes that the observations can be expressed as the mapping of a high-resolution state into observation space, the model state is a truncation of this high-resolution state, the observation operator is linear, and the domain periodic. We note that the approximation obtained from this method will be a time average. Additionally, by assuming the model state is a truncation of the high-resolution state, the approximation will likely be an underestimation. Waller et al. (2014b) used this method in conjunction with convective-scale NWP model data (1.5km grid spacing) in order to diagnose the spatial observation error correlations arising from the error due to unresolved scales for meso-scale NWP (12km grid spacing). Their results showed significant spatial observation error correlations are caused by the error due to unresolved scales and that the natural logarithm of specific humidity has larger error due to unresolved scales than air temperature.

An alternative method to approximate the error due to unresolved scales is to use high-resolution model data as perfect error-free observations of the model state variables. The low-resolution model equivalent to these observations are calculated by spatial and/or temporal averaging the high-resolution model data. Schutgens et al. (2016) first proposed this method to examine the differences between perfect observations and perfect models due to spatial sampling for global aerosol models. Their results showed significant root-mean-square differences for a range of variables. This framework provides a simple way to estimate the error due to unresolved scales explicitly from which statistics may be calculated. However, as the high-resolution model data will not resolve all scales, the uncertainty due to unresolved scales may be underestimated. Using this approach with data from the Met Office's experimental London model (approximately 300m grid spacing), Waller et al. (2021) examined the uncertainty and bias due to unresolved scale for temperature, specific humidity, zonal and meridional wind. Their results showed that the bias and variability is more substantial in lower model levels and significant vertical correlations for each variable. They also found insignificant horizontal correlations which contradicts the results of Waller et al. (2014b). This, however, may be due to averaging over a larger case study period or because of the difference in methods used.

3.3.3 Assimilation using estimated observation error correlations

Until recently, diagonal approximations of the observation error covariance were assumed for all observation types. This was due to the difficulty of estimating observation error statistics and the computational cost of inverting large correlated matrices. However, incorrectly specifying the observation error covariance with diagonal approximations leads to unrealistically smooth and very suboptimal analyses (Rainwater et al., 2015). Using observation error correlations has been shown to improve analysis accuracy (Healy and White, 2005; Stewart et al., 2013) and include more observation information content (Stewart et al., 2008; Rainwater et al., 2015; Fowler et al., 2018; Simonin et al., 2019).

Using approximate observation error covariances obtained through the Desroziers et al diagnostic has improved the assimilation in simple model experiments (e.g., Li et al., 2009; Miyoshi et al., 2013; Waller et al., 2014a) and operational systems (e.g., Weston et al., 2014; Bormann et al., 2016; Campbell et al., 2017; Simonin et al., 2019). In addition, reconditioning correlated observation error covariance matrices has been shown to be useful in overcoming convergence issues (e.g., Weston et al., 2014; Tabeart et al., 2020a,b). As a result, operational centres such as the ECMWF, Met Office and NCEP are increasingly investigating the use of observation error correlations to improve their assimilation systems (e.g., Geer et al., 2019; Bennitt et al., 2017; Bathmann and Collard, 2021).

3.4 Accounting for unresolved scales with the Schmidt-Kalman filter

The Schmidt-Kalman filter (SKF) is an alternative approach to compensate for error due to unresolved scales that was originally developed for navigation problems (Schmidt, 1966; Jazwinski, 1970). In contrast to the standard approach, the SKF uses the statistics of the small-scale processes in state space, without ever evaluating the unresolved state itself. However, by analysing the state resolved by the model only, the SKF is deemed suboptimal as it does not minimise the mean-square error of its estimated states. Janjić and Cohn (2006) applied the SKF within a geophysical context and demonstrated small analysis uncertainty despite the approximations and assumptions required for implementation. The SKF remains largely unexplored within the meteorology community, but has been adapted to a stochastic ensemble form (Lou et al., 2018). The SKF will be used in chapters 7 and 8. In the following subsections, we derive the Schmidt-Kalman filter and explain how it may be applied to compensate for observation uncertainty due to unresolved scales. A more detailed derivation and discussion of the SKF is given in chapter 7.

3.4.1 The Kalman filter in partitioned form

Before we derive the SKF equations, it is necessary to state the Kalman filter equations for the partitioned dynamical system, given by (3.1.1), that analyse all scales (i.e. both the large- and small-scale state). The Kalman filter equations were introduced in a non-partitioned model in chapter 2. For equations relating to the analysis update we drop the time subscripts as all components are valid at the same time. For equations relating to the forecast updates, we retain the subscript as they will differ between components. The Kalman filter for the partitioned model dynamics will be used in chapter 7.

When all scales are analysed, the Kalman gain, given by (2.3.1), in partitioned form will have a large-scale component, $\mathbf{K}^l \in \mathbb{R}^{N_l \times p}$, and small-scale component, $\mathbf{K}^s \in \mathbb{R}^{N_s \times p}$, given by

$$\begin{pmatrix} \mathbf{K}^l \\ \mathbf{K}^s \end{pmatrix} = \begin{pmatrix} \mathbf{P}^{ll,f} & \mathbf{P}^{ls,f} \\ \mathbf{P}^{sl,f} & \mathbf{P}^{ss,f} \end{pmatrix} \begin{pmatrix} (\mathbf{H}^l)^T \\ (\mathbf{H}^s)^T \end{pmatrix} \left[\begin{pmatrix} \mathbf{H}^l & \mathbf{H}^s \end{pmatrix} \begin{pmatrix} \mathbf{P}^{ll,f} & \mathbf{P}^{ls,f} \\ \mathbf{P}^{sl,f} & \mathbf{P}^{ss,f} \end{pmatrix} \begin{pmatrix} (\mathbf{H}^l)^T \\ (\mathbf{H}^s)^T \end{pmatrix} + \mathbf{R} \right]^{-1}, \tag{3.4.1}$$

where $\mathbf{H}^l \in \mathbb{R}^{p \times N_l}$ and $\mathbf{H}^s \in \mathbb{R}^{p \times N_s}$ are the imperfect linear large- and small-scale filter observation operators respectively and the innovation covariance $\mathbf{D} \in \mathbb{R}^{p \times p}$ is given by the term in square brackets. We note that the observation error covariance $\mathbf{R}$ will be representative of the instrument uncertainty, the observation operator uncertainty, and the pre-processing uncertainty. It will not contain the observation uncertainty due to unresolved scales as all scales are analysed. The analysis state update equation (2.3.1) in partitioned form is given by

$$\begin{pmatrix} \boldsymbol{x}^{l,a} \\ \boldsymbol{x}^{s,a} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}^{l,f} \\ \boldsymbol{x}^{s,f} \end{pmatrix} + \begin{pmatrix} \mathbf{K}^l \\ \mathbf{K}^s \end{pmatrix} \left[\boldsymbol{y} - \begin{pmatrix} \mathbf{H}^l & \mathbf{H}^s \end{pmatrix} \begin{pmatrix} \boldsymbol{x}^{l,f} \\ \boldsymbol{x}^{s,f} \end{pmatrix} \right], \tag{3.4.2}$$

where the term in square brackets is the innovation. The partitioned form of the analysis error covariance equation (2.5.1) is given by

$$\begin{pmatrix} \mathbf{P}^{ll,a} & \mathbf{P}^{ls,a} \\ \mathbf{P}^{sl,a} & \mathbf{P}^{ss,a} \end{pmatrix} = \left(\mathbf{I} - \begin{pmatrix} \mathbf{K}^l \\ \mathbf{K}^s \end{pmatrix} \begin{pmatrix} \mathbf{H}^l & \mathbf{H}^s \end{pmatrix} \right) \begin{pmatrix} \mathbf{P}^{ll,f} & \mathbf{P}^{ls,f} \\ \mathbf{P}^{sl,f} & \mathbf{P}^{ss,f} \end{pmatrix}. \tag{3.4.3}$$

The forecast state update equation (2.5.2) for the augmented system is

$$\begin{pmatrix} \boldsymbol{x}^{l,f} \\ \boldsymbol{x}^{s,f} \end{pmatrix}_{k+1} = \begin{pmatrix} \mathbf{M}^l & \mathbf{0}_{N_l \times N_s} \\ \mathbf{M}^{sl} & \mathbf{M}^s \end{pmatrix} \begin{pmatrix} \boldsymbol{x}^{l,a} \\ \boldsymbol{x}^{s,a} \end{pmatrix}_k. \tag{3.4.4}$$

The forecast error covariance update equation (2.5.3) in partitioned form is

$$\begin{pmatrix} \mathbf{P}^{ll,f} & \mathbf{P}^{ls,f} \\ \mathbf{P}^{sl,f} & \mathbf{P}^{ss,f} \end{pmatrix}_{k+1} = \begin{pmatrix} \mathbf{M}^{l} & \mathbf{0}_{N_l \times N_s} \\ \mathbf{M}^{sl} & \mathbf{M}^{s} \end{pmatrix} \begin{pmatrix} \mathbf{P}^{ll,a} & \mathbf{P}^{ls,a} \\ \mathbf{P}^{sl,a} & \mathbf{P}^{ss,a} \end{pmatrix}_{k} \begin{pmatrix} \mathbf{M}^{l} & \mathbf{0}_{N_l \times N_s} \\ \mathbf{M}^{sl} & \mathbf{M}^{s} \end{pmatrix}^{T} + \begin{pmatrix} \mathbf{Q}^{ll} & \mathbf{Q}^{ls} \\ \mathbf{Q}^{sl} & \mathbf{Q}^{ss} \end{pmatrix}_{k+1}. \tag{3.4.5}$$

3.4.2 The Schmidt-Kalman filter equations

To derive the Schmidt-Kalman filter (SKF) from the Kalman filter equations shown in section 3.4.1, we first set $\mathbf{K}^s = \mathbf{0}_{N_s \times p}$ so that only the large-scale state is analysed. Additionally, the SKF will use the innovation $\boldsymbol{y} - \mathbf{H}^l \boldsymbol{x}^{l,f}$ so that the large-scale state may be analysed without knowledge of the small-scale state value. The analysis update for the SKF is then

$$\boldsymbol{x}^{l,a} = \boldsymbol{x}^{l,f} + \mathbf{K}^l (\boldsymbol{y} - \mathbf{H}^l \boldsymbol{x}^{l,f}), \tag{3.4.6}$$

where

$$\mathbf{K}^l = \left(\mathbf{P}^{ll,f} \left(\mathbf{H}^l\right)^T + \mathbf{P}^{ls,f} \left(\mathbf{H}^s\right)^T \right) \left[\begin{pmatrix} \mathbf{H}^l & \mathbf{H}^s \end{pmatrix} \begin{pmatrix} \mathbf{P}^{ll,f} & \mathbf{P}^{ls,f} \\ \mathbf{P}^{sl,f} & \left\langle \boldsymbol{x}^{s,t} \left(\boldsymbol{x}^{s,t}\right)^T \right\rangle \end{pmatrix} \begin{pmatrix} (\mathbf{H}^l)^T \\ (\mathbf{H}^s)^T \end{pmatrix} + \mathbf{R} \right]^{-1}. \tag{3.4.7}$$

We note that since the small-scale state isn't analysed, $\mathbf{P}^{ss,f}$ is replaced by $< \boldsymbol{x}^{s,t} (\boldsymbol{x}^{s,t})^T >$ in the SKF innovation covariance. Hence, $\mathbf{P}^{ls,f}$ and $\mathbf{P}^{sl,f}$ are now the flow-dependent cross-covariances between the large-scale forecast errors and the small-scale variability. Similarly to the partitioned Kalman filter, the observation error covariance will not contain a component corresponding to the error due to unresolved scales.

While the SKF analyses the large-scale state only, it evolves the uncertainty of both the large- and small-scales. The analysis error covariance update is obtained by setting $\mathbf{K}^s = \mathbf{0}_{N_s \times p}$ within the short-form update equation (3.4.3). We note that while the full update of the analysis error covariance update is symmetric, the short form update for the SKF is not symmetric, and so we set $\mathbf{P}^{sl,a} = (\mathbf{P}^{ls,a})^T$. As the small-scale state isn't analysed, $< \boldsymbol{x}^{s,t} \left(\boldsymbol{x}^{s,t}\right)^T >$ is not updated in the analysis update.

The forecast state update for the SKF is given by

$$\boldsymbol{x}^{l,f}_{k+1} = \mathbf{M}^l \boldsymbol{x}^{l,a}_k. \tag{3.4.8}$$

The forecast error covariance update for the SKF is given by equation (3.4.5) except that $\mathbf{P}^{ss,f}$ is replaced by $< \boldsymbol{x}^{s,t} \left(\boldsymbol{x}^{s,t}\right)^T >$.

The SKF provides a methodology to compensate for uncertainty due to unresolved

scales explicitly within the filter algorithm. It is essentially a generalized form of the filter proposed by Mitchell and Daley (1997a,b) to compensate for unresolved scales. The benefit of the SKF is the flow-dependent cross-covariances between the large-scale state errors and the small-scale variability. However, while computationally cheaper than the optimal Kalman filter, the SKF is still an expensive filtering strategy due to the augmentation of the state error covariance.

3.5 Summary

In this chapter we have described the observation errors that occur in data assimilation for NWP. Following the definition of Janjić et al. (2018), each component of the representation error was discussed. In particular, the error due to unresolved scales, and the implications arising from its state-dependence and correlations, was discussed in detail. The standard approach to compensate for representation uncertainty is to include it in the observation error covariance matrix. This may be achieved by estimating the entire observation error covariance through residual-based diagnosis to obtain a time-independent approximation. Alternatively, a metrological approach may be used where each contribution to the observation uncertainty is estimated individually from first principles and suitably combined. The benefits of including the representation uncertainty in the observation error covariance matrix for assimilation were reviewed. This chapter concluded with an introduction to the Schmidt-Kalman filter, which represents an alternative, and relatively unexplored, approach to account for the error due to unresolved scales. The Schmidt-Kalman filter, which is researched in chapters 7 and 8, uses the statistics of the small-scale processes in state-space, without ever evaluating the small-scale state, to estimate the large-scale state. We next review crowdsourced observations and their application to NWP. Crowdsourced observations are likely to be in-situ observations taken in urban areas and will therefore have uncertainty due to unresolved scales.

Chapter 4

Crowdsourced observations for numerical weather prediction

Convection-permitting weather prediction requires a large number of spatially-dense and temporally-frequent observations of high resolution across the forecast area to constrain short-term forecasts (Sun et al., 2014; Gustafsson et al., 2018; Dance et al., 2019). However, it may be impractical to extend current scientific surface observation networks due to the cost of installing, managing and maintaining observing instruments. A potential data source that can help fill this observation gap is the use of non-traditional observing instruments such as personal weather stations, smartphones, and vehicles (e.g., Bell et al., 2013; Mahoney III and O'Sullivan, 2013; Mass and Madaus, 2014; de Vos et al., 2017). Such near-surface land-domain observations can help achieve the horizontal coverage recommended by the World Meteorological Organisation (World Meteorological Organization, 2021). In addition, these observations will likely have high spatio-temporal resolution and may be generated for a low cost through crowdsourcing (Muller et al., 2015). In this chapter, we discuss crowdsourcing of meteorological observations and review recent literature on the subject.

4.1 Crowdsourced observations

The collaboration between the public and researchers to address real-world problems is known as citizen science (Wiggins and Crowston, 2011). The area of citizen science that outsources the generation, storage, and utilisation of data to the public is known as crowdsourcing. In the context of numerical weather prediction, crowdsourced data collectively refers to meteorological reports and data generated by the public through use of privately owned equipment (Hintz et al., 2019a). We note that there are other opportunistic data sources that are currently used in operational weather prediction. For example, the Met Office currently assimilate roadside

weather information station data obtained from highways agencies into the United Kingdom variable-resolution model (Gustafsson et al., 2018). In this thesis, we are primarily interested in crowdsourced observations which will be researched in chapters 5 and 6.

Crowdsourced observations have several advantages and disadvantages when compared to traditional scientific observations (Hintz et al., 2019a). The advantages of crowdsourced observations include that they may be obtained for a low cost, they are in-situ observations with high spatio-temporal resolution, and they can help fill in the horizontal coverage of near-surface observations over land. In particular, urban areas with large populations, where there are few traditional surface observations networks, are likely to be densely populated by crowdsourced observations. Hence, as convection-permitting weather prediction requires a large number of observations of high spatio-temporal resolution (Sun et al., 2014; Gustafsson et al., 2018; Dance et al., 2019), crowdsourced observations are ideal for improving convection-permitting forecasts of urban areas. The disadvantage of crowdsourced observations include that they will be inaccurate in comparison to traditional observations. Other issues regarding the use of crowdsourced data in NWP include data ownership, intermittency, heterogeneity, data provenance and large data volumes (Hintz et al., 2019a).

The uncertainty of crowdsourced observations can be attributed to five areas (Bell et al., 2015); calibration issues, communication and software issues, inaccurate or missing metadata, design flaws, and error due to unresolved scales. We note that the metadata available for crowdsourced datasets will depend on the constraints imposed by data privacy laws and will contribute to the pre-processing component of the representation uncertainty. The error due to unresolved scales will depend on the influence of the environment local to the in-situ crowdsourced observation (Janjić et al., 2018). As crowdsourced observations are likely to be taken in urban areas, they may have substantial uncertainty due to unresolved scales because of the dynamics induced by the surrounding urban environment. For example, a measurement in a sheltered street will give a different reading to one made on the top of a skyscraper, but can still be situated in the same model grid box. In chapter 6, the uncertainty of crowdsourced vehicle-based observations of air temperature is discussed.

4.2 Examples of crowdsourced datasets

In this section we review the use of crowdsourced observations from personal weather stations, smartphones and vehicles in meteorology. Quality-control of crowdsourced observations will be discussed in section 4.3.

4.2.1 Personal weather stations

Personal weather stations (PWSs) are defined by Bell et al. (2015) as "a weather station set up by a member of the public for whom the terms weather enthusiast, volunteer, hobbyist and amateur observer are fitting descriptions." Observations from PWSs have received much academic interest in part due to the direct observation of meteorological variables. Furthermore, it is anticipated that the bias correction of stationary instruments will be simpler than the bias correction of instruments that change location (Hintz et al., 2019a). However, rules and standards adhered to by the World Meteorological Organization (WMO) may not be followed by PWS observations due to the locations they are installed, the measuring instruments they use and their maintenance (Nipen et al., 2020).

PWSs are typically situated in the gardens of amateur weather enthusiasts where they can properly observe atmospheric conditions (Bell et al., 2013). As such, PWSs will be dense in urban and suburban areas making them ideal sources of air-temperature data in urban heat island studies (e.g., Steeneveld et al., 2011; Wolters and Brandsma, 2012; Chapman et al., 2017; Meier et al., 2017). Other uses of PWS observations include analysing severe hailstorms (Clark et al., 2018), hydrometeorological monitoring in metropolitan areas (de Vos et al., 2017, 2018, 2019a) and diagnosing the distribution of urban wind speeds (Droste et al., 2020). PWS observations are not currently assimilated in operational weather prediction, but they are used in post-processing and verification (e.g., Nipen et al., 2020; Kirk et al., 2021; Hintz et al., 2021).

4.2.2 Smartphone observations

Smartphones are commonly used throughout the world and have the potential to produce more crowdsourced observations than any other source (Mass and Madaus, 2014). The most researched observation obtainable from smartphones is surface pressure from built-in barometers (e.g., Madaus et al., 2014; Madaus and Mass, 2017; Price et al., 2018; McNicholas and Mass, 2018a,b; Hintz et al., 2019b, 2021). Other observations that have been obtained from smartphones include air temperature derived from smartphone batteries (Overeem et al., 2013; Droste et al., 2017) and wind speed from anemometers mounted on smartphones (Hintz et al., 2020).

For data assimilation, surface pressure observations (SPOs) are a desirable observation that can provide useful information on the three-dimensional structure of the atmosphere. Additionally, due to the spatial homogeneity of surface pressure fields, the uncertainty due to unresolved scales associated with SPOs will be small (Madaus et al., 2014). Hanson (2016) performed a series of observing system simulation experiments to assess the potential value of SPOs from smartphones for high-resolution weather prediction. They found that, in the absence of any other observation type,

SPOs from smartphones have the potential to produce forecasts with similar skill to forecasts produced using conventional SPOs. Additionally, the benefits of the simulated smartphone observations extended to other surface fields such as 2-metre air temperature and 10-metre wind. Assimilation of SPOs from smartphones into a mesoscale model was initially conducted by Madaus and Mass (2017). Their results showed that systematic sensor biases and elevation uncertainty in the smartphone data caused by poor data collection and quality-control, inaccurate metadata, sensor bias and user behaviour hindered the impact of the observations. As a result of these issues, only marginal improvements were attained. In a follow-up study that compensated for sensor bias and elevation uncertainty, assimilation of smartphone observations produced more substantial forecast improvements (McNicholas and Mass, 2018a,b). The assimilation of bias-corrected SPOs from smartphones into the HARMONIE numerical weather prediction system, where synoptic surface pressure observations were removed, has been conducted by Hintz et al. (2019b). Their results showed a decrease in the surface-pressure model bias and an improvement in the forecast of accumulated precipitation.

4.2.3 Vehicle-based observations

Vehicle-based observations are a source of potentially millions of directly observed or derived surface-based observations that may be utilised in NWP (Mahoney III and O'Sullivan, 2013). Such observations may be obtained through connected vehicle initiatives (e.g., Mahoney III and O'Sullivan, 2013), from built-in sensors of vehicle fleets via the controller area network (CAN) (e.g., Mercelis et al., 2020), or using sensors that have been externally mounted on the vehicle (e.g., Anderson et al., 2012). In chapters 5 and 6, we examine crowdsourced vehicle-based observations of air temperature from built-in sensors obtained from on-board diagnostic (OBD) dongles.

Research into the application of vehicle-based observations of air temperature has primarily been for road weather applications such as modelling (e.g., Hu et al., 2019) and forecasting (e.g., Siems-Anderson et al., 2019; Mercelis et al., 2020). For example, vehicle-based observations of the atmospheric state are a primary information source for the open-source Pikalert system, which combines weather information from disparate sources to provide improved road weather forecasts and hazard assessments (Siems-Anderson et al., 2019). Additionally, assimilation of vehicle-based observations of road-surface temperature into the Finnish Meteorological Institute's road weather model has been shown to improve forecasts in areas with few roadside weather information stations (Karsisto and Lovén, 2019). Research into vehicle-based observations for convection-permitting NWP is still in its infancy, but improvements from assimilation of vehicle-based observations of air temperature, dewpoint temperature, wind speed and wind direction in observation system simu-

lation experiments have been obtained (Siems-Anderson et al., 2020). We note that the vehicle-based observations were simulated using gridded analysis data, surface station data and roadside weather information station data so the error due to unresolved scales caused by the environment local to a vehicle was not fully considered in the study. The use of vehicle-based observations has also been considered for nowcasting by the German Weather service (Hintz et al., 2019a).

4.3 Quality control

Before the meteorological information available from crowdsourced observations can be examined, it is important to remove any observations that are likely in gross error. This is achieved through quality-control (QC); a vital process performed prior to data assimilation. Due to the high spatial-temporal variability of meteorological processes, detection of errors can be a challenging process (Gandin, 1988). Therefore, QC procedures usually include several simple tests that assess different aspects of the observed quantities (Zahumenskỳ, 2004; Fiebrich et al., 2010). We now summarise common tests used in the QC of meteorological observations and their use in the QC of crowdsourced datasets. In chapter 5, we QC crowdsourced vehicle-based observations of air temperature.

4.3.1 Range-validity tests

Range-validity tests involve checking that some observed value is within a predetermined validity range. Such tests are useful to QC crowdsourced datasets as no information on the data source is required and so data privacy issues are avoided. Validity ranges obtained from sensors are determined by the hardware specifications and have been used in the QC of vehicle-based observations using externally-mounted sensors (Chapman et al., 2010; Anderson et al., 2012; Boyce et al., 2017). However, for vehicle-based observations from built-in sensors, whose designs may be kept secret by the manufacturer, sensor-range tests will be unsuitable. Alternatives to sensor ranges include statistically-derived ranges and climatological ranges. For example, Meier et al. (2017) defined a validity range of five standard deviations around a reference minimum temperature in order to determine whether PWSs were located inside buildings. Climatological ranges are usually determined from historical datasets and have been used to QC vehicle-based observations (Chapman et al., 2010; Boyce et al., 2017) and smartphone observations (Hintz et al., 2019b).

4.3.2 Temporal checks

Temporal checks involve examination of a time-series of measurements obtained from a single instrument. The plausibility of the data is usually assessed through

either step or persistence checks. These tests require some sort of identification which may be unavailable for crowdsourced datasets due to privacy concerns (Hintz et al., 2019a). Step tests examine the difference between sequential observations and require some sort of numerical calculation to be satisfied. For instance, the Pikalert system (Boyce et al., 2017; Siems-Anderson et al., 2019) tests for unrealistically large differences between consecutive vehicle-based observations and Meier et al. (2017) tests for the diurnal variability in PWS observations using step checks. Persistence checks are used to determine whether a sensor is stuck on a value by checking that measurements change over a specified time period. Such checks have also been used by the Pikalert system (Boyce et al., 2017).

4.3.3 Consistency tests

Consistency tests identify observations that are inconsistent with other nearby observations. The criteria for inconsistency will depend on the observed variable, the location of the observations being compared and the distance between them. "Buddy checks" are a type of consistency test that compare observations of the same type in close spatio-temporal proximity. These tests have been used to QC PWS observations (Meier et al., 2017; de Vos et al., 2019b; Droste et al., 2020; Nipen et al., 2020), smartphone observations (Hintz et al., 2019b) and vehicle-based observations (Anderson et al., 2012; Boyce et al., 2017). Such tests are useful for the QC of crowdsourced observations as they can take advantage of the large observation density in urban areas and they do not require instrument identification (Nipen et al., 2020). Crowdsourced observations may also be compared with traditional surface observation networks provided they are sufficiently close to one another. Such tests have been used to QC smartphone observations (Madaus and Mass, 2017; McNicholas and Mass, 2018b; Hintz et al., 2019b) and vehicle-based observations (Chapman et al., 2010; Anderson et al., 2012; Boyce et al., 2017).

4.3.4 Background checks

Background checks involve calculating the difference between an observation and the background from an operational NWP system. The size of the background departure will indicate the plausibility of an observation. Background checks have been used to flag SPOs that have a background departure larger than 1hPa, which approximately corresponds to an 8m difference in altitude, as the standard deviation of surface pressure in the NWP model is of the order of 0.3–0.4hPa (Hintz et al., 2019b). Background checks are suitable for crowdsourced observations as the forecast can be interpolated to the location of each crowdsourced observation. However, if the forecast for an area is poor, this test may remove good observations in that area.

4.4 Summary

In this chapter we have introduced crowdsourced observations as a dataset of opportunity for weather prediction. Crowdsourced observations are typically non-traditional observations generated from the public through use of their privately owned equipment. Such observations are receiving increased interest for weather prediction as they can help achieve the horizontal coverage recommended by the World Meteorological Organization (2021) for near-surface land-domain observations for a low cost. Examples of crowdsourced datasets that are currently researched by the meteorological community have been discussed including vehicle-based observations which are investigated in chapters 5 and 6. This chapter concluded with discussion of range-validity tests, temporal checks, consistency tests and background checks for the quality-control of meteorological datasets. Applications of these tests to crowdsourced datasets in published works, as well as any limitations, have been discussed. A new quality-control procedure will be developed for crowdsourced vehicle-based observations in chapter 5. In the next chapter we begin our investigation into vehicle-based observations obtained from a Met Office proof-of-concept trial.

Chapter 5

Quality-control of vehicle-based temperature observations and future recommendations

The first aim of this thesis is to investigate a crowdsourced vehicle-based temperature dataset as a low-cost source of urban observations. In this chapter we answer the first research question given in chapter 1: *how can we quality-control the vehicle-based air-temperature dataset obtained from a Met Office proof-of-concept trial?* In addition, we wish to determine:

- Which quality-control tests are suitable for crowdsourced vehicle-based observations of air temperature?
- Which improvements can be made to the data collection procedure for vehicle-based observations used by the Met Office?

The remainder of this chapter, except for the chapter summary in section 5.7, is strongly based on the paper Bell et al. (2021b).

Abstract:

In numerical weather prediction, *datasets of opportunity* is a collective term used for meteorological observations obtained from unconventional data sources. This report presents the quality-control process designed for a vehicle-based dataset of opportunity containing 67959 observations obtained from a proof-of-concept trial run from 20th February until 30th April 2018 by the Met Office. In this trial, on-board diagnostic (OBD) dongles were used to transmit low precision dry bulb temperature measurements from a vehicle to the driver's phone which were subsequently uploaded to the Met Office cloud servers with time, location, and vehicle identification metadata using a Met Office phone app. The raw data from the trial

was first filtered to remove observations with missing fields or invalid measurements and metadata. The resultant filtered dataset contained 32179 observations (47.4% of original dataset) which underwent further testing. The quality-control tests applied to the filtered dataset included a climatological range test, a stuck instrument test, and a GPS test that checks whether an observation location is physically consistent with the vehicle's previous location. A substantial number of observations were flagged by the GPS test due to the accuracy of smartphone GPS measurements, GPS location update app settings, and poor GPS signal, while the majority of observations passed the climatological range and stuck instrument test. The 19094 observations which passed all previous quality-control tests were put through a final sensor ventilation test to determine if the vehicle drove at a sufficient speed for the temperature sensor to be adequately ventilated. This test flagged 1669 observations with speeds below a predetermined sensor ventilation threshold. In total, the quality-controlled dataset consists of 17425 observations (25.6% of original dataset). The results of the quality-control process have shown that the observation location metadata can be inaccurate due to unsuitable app settings and poor GPS signal. Additionally, inadequate sensor ventilation can result in observations with a warm bias. Recommendations on future data collection include revising OBD dongle and app settings/features to correct observation GPS, methods to circumvent the need for vehicle identification in quality-control, and the use of higher precision instruments.

5.1 Introduction

The advancement of convection-allowing data assimilation requires a large number of observations of high spatio-temporal resolution relevant to the weather processes being modelled (Sun et al., 2014; Gustafsson et al., 2018; Dance et al., 2019). Due to the enormous cost of installation, management, and maintenance, it may not be practical to extend traditional scientific observing networks. A potential alternative source of inexpensive, high-resolution meteorological observations to constrain convective-scale numerical weather prediction forecasts is from crowdsourced data which are currently receiving increased interest from the numerical weather prediction community (e.g., Nipen et al., 2020).

In the context of numerical weather prediction, crowdsourced data collectively refers to reports and data generated by the public through use of privately owned equipment (Hintz et al., 2019a). These observations will be inaccurate in comparison to those obtained from scientific observing networks but have the potential to far exceed the number of scientific observations currently produced. Application of crowdsourced data is relatively new and studies into collection methods and observation characteristics are currently active areas of research (e.g., Bell et al., 2015;

McNicholas and Mass, 2018b; Hintz et al., 2019a; Nipen et al., 2020).

Crowdsourced observations from citizen observing networks have been shown to successfully observe meteorological phenomena such as urban heat islands directly through digital and car thermometer observations (Knight et al., 2010) and amateur/citizen science weather stations observations (Steeneveld et al., 2011; Wolters and Brandsma, 2012; Chapman et al., 2017; Meier et al., 2017). Additionally, urban heat islands have been shown to be observable through temperature measurements derived from smartphone battery temperatures (Overeem et al., 2013; Droste et al., 2017). To ensure representative measurements from direct observations, several precautionary measures must be taken into consideration (Bell et al., 2015). For instance, citizen weather stations need to be shielded from radiation and temperature sensors need to be located a sufficient distance from buildings and in a naturally ventilated area. Sensor ventilation is especially important to prevent large air temperature errors in circumstances of large radiative forcing (e.g., World Meteorological Organization, 2008; Richardson et al., 1999; Nakamura and Mahrt, 2005). Though guidance on the proper use of meteorological instruments was provided in these studies, the overall credibility of the crowdsourced datasets must be evaluated through quality-control methods before the observed meteorological processes can be examined.

Quality-control is a vital process performed prior to data assimilation to reject observations that are likely to contain gross errors (i.e. observations that are clearly incorrect). The techniques which comprise the quality-control procedure usually include simple checks designed to test different aspects of the observed values (Zahumenskỳ, 2004; Fiebrich et al., 2010). Quality-control of crowdsourced observations is a notably difficult task (Muller et al., 2015). Non-traditional data sources may suffer from numerous issues that traditional scientific observations will not (e.g. calibration, user behaviour such as locations above/below ground, low precision data, sensor-specific measurement errors, etc). As an example, surface pressure observations from smartphones may not be the desired meteorological measurement due to user behaviour and inadequate location and elevation metadata. We note that while smartphone GPS accuracy may be degraded by urban environments, recent studies on horizontal position accuracy in urban areas have shown they can produce acceptable location-based metadata for crowdsourced observations (e.g., Merry and Bettinger, 2019). Due to the previously mentioned issues over half the smartphone observations examined in the work of Madaus and Mass (2017) and Hintz et al. (2019b) were removed through quality-control. However, smartphone observations have been shown to improve forecasts after bias correction and thorough quality-control had been applied to the data (McNicholas and Mass, 2018a,b; Hintz et al., 2019b).

Adverse weather is a leading contributor to severe congestion, large travel time delays and harmful incidents for surface transportation networks (Snelder and Calvert, 2016). To combat this, vehicle-based meteorological observations have been utilised by the Clarus Initiative (Limber et al., 2010) and the Pikalert System (Boyce et al., 2017; Siems-Anderson et al., 2019) to provide improved road and atmospheric hazard products to road maintenance operators and the travelling public. Examples of observations obtainable by vehicles include precipitation (e.g., Haberlandt and Sester, 2010; Rabiei et al., 2013), air quality (e.g., Devarakonda et al., 2013; Rada et al., 2016), atmospheric pressure and temperature (e.g., Drobot et al., 2010; Chapman et al., 2010; Anderson et al., 2012). Studies on the quality-control of vehicle-based temperature and pressure observations have been performed for experiments where a set number of vehicles were driven along predetermined routes using fitted meteorological sensors with known error characteristics in order to generate weather statistics for road segments (Chapman et al., 2010; Drobot et al., 2010; Anderson et al., 2012). This freedom over the experimental design allowed for Anderson et al. (2012) to use spatial consistency checks between neighbouring vehicles and nearby surface-stations as well as sensor range tests to check the meteorological instruments. Chapman et al. (2010) also used a nearby surface-station consistency check and a sensor range test but used a climatological range test instead of a spatial consistency test between neighbouring vehicles. In contrast, the small novel dataset examined in this report allowed for no such freedom in the experimental design and severely limits the quality-control tests applicable.

The objective of this report is to quality-control a vehicle-based observation dataset and to provide recommendations on future data collection methods. The structure of this report is as follows. In section 5.2 we provide an overview of the Met Office proof-of-concept vehicle-based observations trial. Preparation of the data obtained from the trial into a filtered dataset to be quality-controlled is discussed in section 5.3. The detailed description of the implementation and results of the quality-control tests is given in section 5.4. The results of the quality-control process highlight that the observation location metadata can be inaccurate due to poor GPS signal and app settings. Additionally, inadequate sensor ventilation can result in observations with a warm bias. Discussion of the quality-controlled dataset is given in section 5.5. A summary with recommendations is provided in section 5.6.

5.2 Met Office trial

The vehicle-based observations studied in this report are obtained from a trial by the Met Office from 20th February 2018 to 30th April 2018. In the trial, volunteer participants connected an on-board diagnostics (OBD) dongle to their vehicle engine management interface. The OBD dongle used in this trial is an inexpensive adapter

which is inserted into the OBD port of a vehicle which transfers vehicle diagnostic data to a Bluetooth-connected computing device. Meteorological parameters and vehicle speed (given in km/h) selected on the app "Met Office OBD App 0909" are broadcast to the participant's Android phone. This app decodes any data sent by the dongle and appends the corresponding date-time (given by the app as the date and 24 hour clock time) and GPS location metadata derived through the phone. In addition, a unique observation ID and sensor ID are appended to each observation. A sensor ID is used to determine if observations come from the same vehicle. However, the sensor ID is specific to the installed version of the app and so reinstalling the app used to record the observations would result in a new identifier for the participant. The observations along with relevant metadata are then uploaded via 3G or 4G to the Met Office Weather Observations Website (Met Office, 2011). The preset data collection frequency and GPS update period are set to 1 minute while the preset minimum distance for a GPS update is 500 metres. These settings can be changed by the participant through the app interface.

In total, 31 participants were successful in producing vehicle-based observations from journeys they undertook during the trial period. For clarity, we define a journey as a subset of data originating from the same vehicle over a fixed time interval. Throughout this trial, fewer observations were usually obtained during weekends than week-days. A "call for data" was made to obtain data for certain weather conditions on the 9th March (rainy), 22nd March (benign) and 16th April (sunny). In addition, a large number of observations were obtained on 27th March due to a few long journeys.

The observations of interest obtained through this trial include dry bulb temperature (°C), engine intake temperature (°C) and air pressure (hPa). Both temperature observations have low precision (1°C). The air pressure is precise to 10hPa and hence not useful in data assimilation. In this report we examine dry bulb temperature only. Engine intake temperature measures the dry bulb temperature in the vehicle engine which will not reflect the true atmospheric air temperature. However, a fault known to have occurred for some observations during this trial is for engine intake temperature to be recorded as dry bulb temperature.

5.3 The filtered dataset

In order to assess the quality of a dataset, each datum must contain information we are interested in examining. Any data which are obviously in error or do not have the relevant observation field we wish to examine will be discarded. The remaining data will be referred to as the *filtered dataset* and will undergo the quality-control process detailed in section 5.4.

5.3.1 Preparation of the filtered dataset

To obtain the filtered dataset we carry out some gross checks on the dry bulb temperature measurements and accompanying time and speed metadata of the complete dataset. In particular, we discard any datum which exhibit any of the following properties:

1. the vehicle speed is less than 0km/h,
2. there is an invalid date-time in the observation metadata,
3. there is no dry bulb temperature observation.

Data that do not exhibit any of these properties will form the filtered dataset.

In total, 35780 observations were discarded resulting in the filtered dataset containing 32179 observations. No observations were discarded due to an invalid date-time. Many of the discarded observations contain speeds with value -32768km/h which is the value used when the vehicle speed is unable to be recorded by the app. This is the minimum short signed integer for a 2-bit system. A single observation with speed 255km/h, over double the speed limit for dual carriageways and motorways in the UK (UK Government, 2015), is discarded as it is likely incorrect. There were 34681 discarded observations without a dry bulb temperature field.

The data that passed this data filtering test underwent further quality-control tests. We now examine the observations in the filtered dataset before we describe the implementation of the GPS and sensor ventilation tests detailed in sections 5.4.3 and 5.4.4 respectively.

5.3.2 Characteristics of the filtered dataset

To implement the quality-control process described in section 5.4 it is necessary to have knowledge of the characteristics of the filtered dataset. In this section we examine the dry bulb temperature observations (section 5.3.2.1), the temporal frequency of the observations (section 5.3.2.2) and the GPS measurements (section 5.3.2.3) of the filtered dataset.

5.3.2.1 Dry bulb temperature

The number of dry bulb observations for each month as well as their distribution is shown in figure 5.1. In total, 5684 observations were taken in February, 16211 in March, and 10284 in April. We note that, except the two outliers in April that had a dry bulb temperature of -22°C, there are no observations with dry bulb temperature less than -8°C. Any observations with a temperature greater than 20°C occurred in April. Furthermore, the largest dry bulb temperature values (27-34°C) occurred

for vehicle speeds less than 25km/h. As discussed in section 5.1, this is likely to be caused by inadequate sensor ventilation.

	February	March	April	All data
Number of observations	5684	16211	10282	32179
Mean	3.33°C	6.59°C	11.38°C	7.55°C
Standard deviation	3.99°C	3.48°C	4.34°C	4.84°C
Skew	−0.09	−0.44	0.67	0.27
Excess kurtosis	−0.74	0.27	1.13	1.14

Table 5.1: Summary of the descriptive statistics for the filtered dataset and split into each month. The two values of −22°C have been removed from the descriptive statistics calculations of "April" and "All data" as they are clearly outliers.

A summary of the descriptive statistics for the filtered dataset and each month is given in table 5.1. As expected, the distribution of dry bulb temperatures varies significantly with each month with mean temperatures increasing from February to April due to season. The distribution for February is the most concentrated distribution (negative excess kurtosis) with slight asymmetry (negative skew). March is a less concentrated distribution than February with a longer left tail. The left tails for both February and March are likely to be caused by the Beast from the East (22/02/2018 to 05/03/2018 (Met Office, 2020)). Removing the extreme dry bulb temperatures from April reveals it is the least concentrated distribution with a longer right tail that is likely caused by inadequate sensor ventilation during more frequent sunny weather conditions and the April hot spell (18th-22nd April 2018 (Met Office, 2020)). Examining the variability of each month shows that April is the most variable month (highest standard deviation) and March the least variable. As the majority of the observations occurred in March and April, the characteristics of the combined temperature distribution share the most similarities with these months. Namely, the distribution is concentrated around similar temperatures to March and has a longer right tail like April. However, the combined temperature distribution has higher variability than each individual month.

5.3.2.2 Multiple reporting

The observations from this trial were designed to have a 1 minute temporal frequency. However, several observations are reported within a minute of the previous observation but retained a 1 minute temporal frequency with other similar observations. An example of this is shown in figure 5.2 which shows the dry bulb temperature observations for a journey segment from a single vehicle. The blue observations maintain a 1 minute temporal frequency with other blue observations while the red observations occur 7 seconds after a blue observation and maintain a 1 minute temporal frequency with other red observations. An observation reported

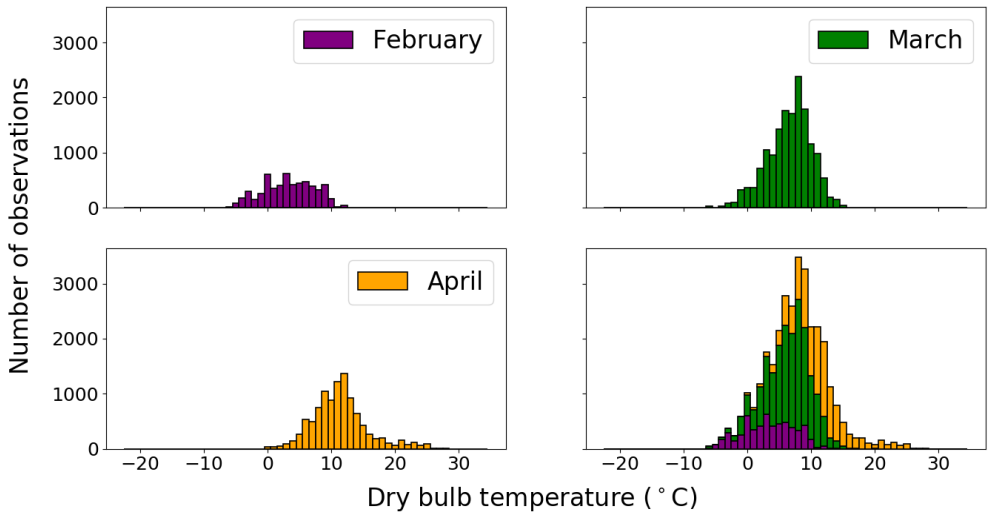


Figure 5.1: Distribution of dry bulb temperature observations for each month of the trial for the filtered dataset. The purple bar segments indicate the number of February 2018 observations, the green segments March 2018, and the orange segments April 2018. The combined distribution is a stacked histogram that shows the contribution from each month to the total number of observations for each dry bulb temperature.

within a minute of the previous observation will be referred to as a *multiple reported observation* (MRO) and is suspected to be caused by phone or dongle hardware technical issues. We note that the GPS metadata may be inaccurate for MROs due to the short time between observations. This will be discussed further in section 5.4.3.

Denoting the time-gap between consecutive observations from the same vehicle on a given day as Δt, a histogram of all Δt binned into one-minute intervals is shown in figure 5.3. Noting the log-scale, the vast majority of Δt have length 0-2 mins (i.e. the first two bins). The first bin contains all $\Delta t \in [0, 1)$ minutes which occurs 15364 times. This corresponds to the number of MROs in the filtered dataset. MROs form a large part of the filtered dataset and will need to be accounted for in the additional quality check tests we impose on this data. The second bin contains all $\Delta t \in [1, 2)$ minutes which occurs 15025 times with 14646 corresponding to the observations which maintain a 1 minute temporal frequency with the previous observation. Any $\Delta t \geq 60$ minutes are placed in the 60+ minute bin. Any $\Delta t \in [2, 60)$ are caused by breaks in a vehicle journey, issues with the collection method such as loss of phone signal, or removal of data without the necessary information needed for quality-control.

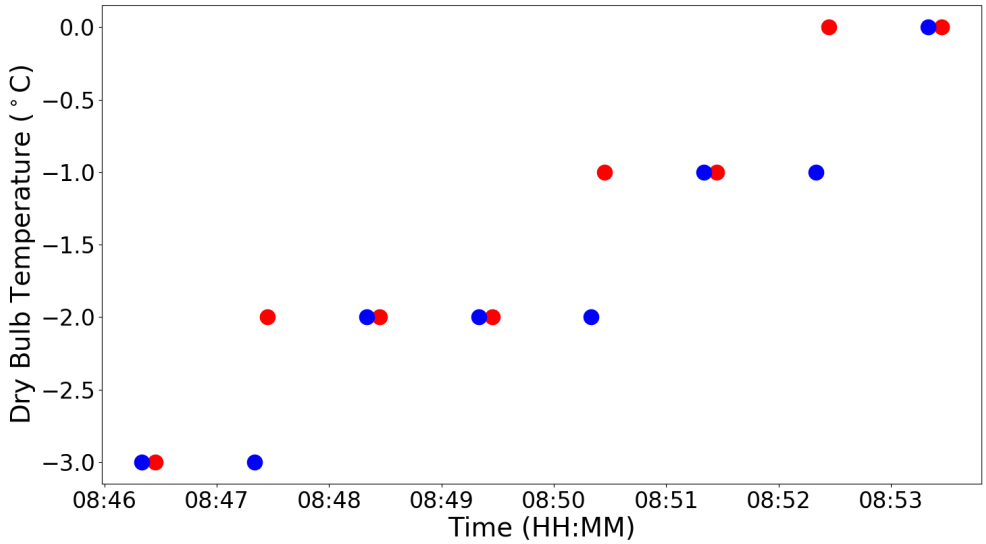


Figure 5.2: Time series segment of dry bulb temperature for a single vehicle taken on the 28th of February 2018. The blue observations retain a 1 minute temporal frequency with other blue observations. The red observations retain a 1 minute temporal frequency with other red observations but always occur 7 seconds after a blue observation.

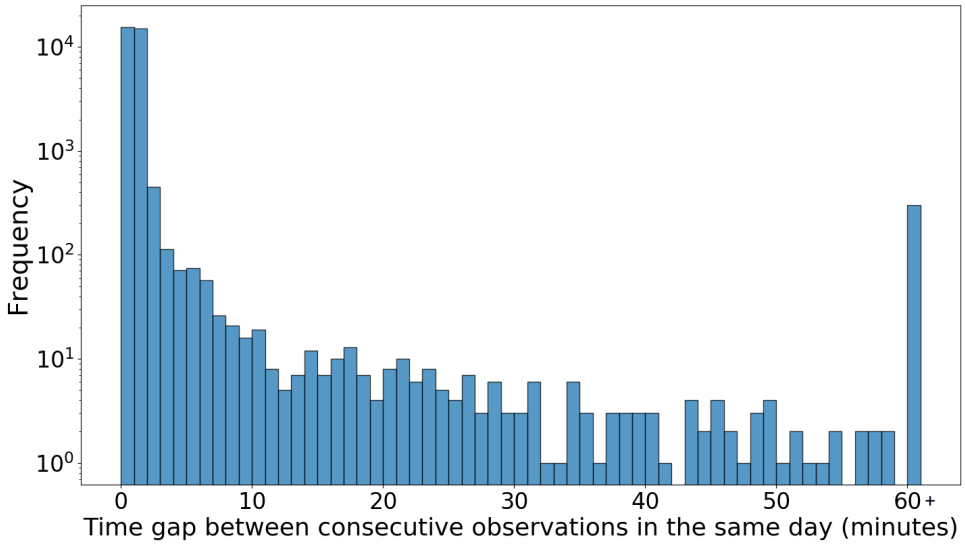


Figure 5.3: The distribution of time-gaps between consecutive observations from the same vehicle and day for the filtered data set binned into minute intervals. Note that a log-scale has been used for the time-gaps frequency. Any time-gaps greater than 60 mins are placed into the 60+ minute interval.

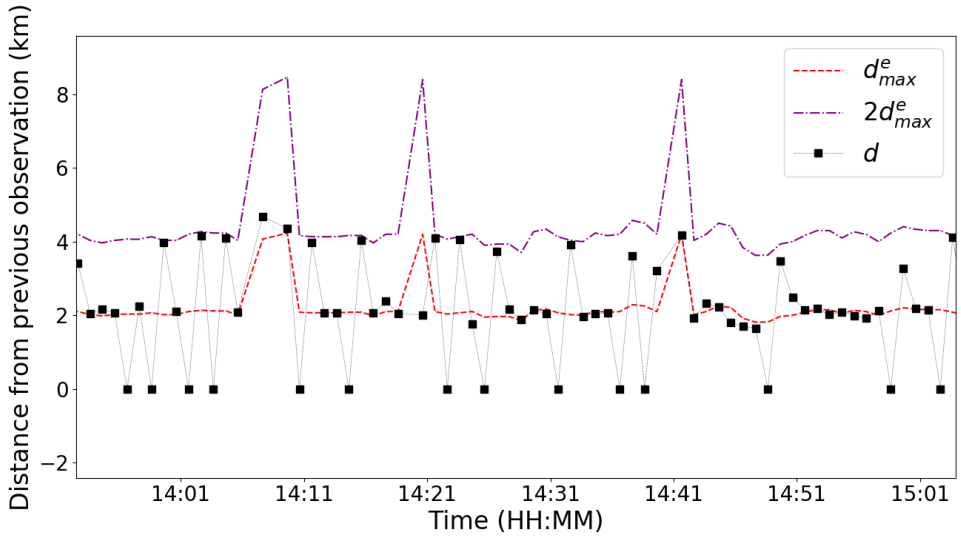


Figure 5.4: Time series of the distance between consecutive observations calculated using the great circle distance (black squares), denoted d, the distance estimated using the time-gap between the observations and the speeds of the observations, denoted d^e_{max} (red-dashed line) and $2d^e_{max}$ (purple dot-dashed line), for the first half of a journey along the M5 motorway during the 25th of March 2018. We have used two estimates to provide a reference for when a distance is realistic (i.e. close to the red line) or unrealistic (i.e. close to the purple line). Instances of $d = 0$km indicate the observation at that time is a GPS-lagged observation. Almost all instances of $d = 0$km are immediately followed by $d^e_{max} < d \leq 2d^e_{max}$ which corresponds to the distance travelled between the observation before and after the GPS-lagged observation.

5.3.2.3 GPS-lagged observations

The method of data collection in the Met Office trial used smartphones to obtain location-based metadata. A common occurrence in the filtered dataset is GPS location not updating due to poor GPS signal, the vehicle has not travelled far enough to trigger a GPS update or insufficient time between observations. The default GPS update distance and period for the app are 500 metres and 60 seconds respectively. This results in some observations having identical GPS location to the previous observation taken by the same vehicle. These observations will be referred to as *GPS-lagged observations*.

Figure 5.4 shows data from the first half of a journey along the M5 motorway during the 25th of March 2018. We have plotted a time series of distance between consecutive observations calculated using the great circle distance,

$$d = 2r \sin^{-1}\left(\sqrt{\sin^2\left(\frac{\Phi_1 - \Phi_2}{2}\right) + \cos(\Phi_1)\cos(\Phi_2)\sin^2\left(\frac{\lambda_1 - \lambda_2}{2}\right)}\right), \quad (5.3.1)$$

where Φ_1 (λ_1), Φ_2 (λ_2), are the latitudes (longitudes) of the two locations and $r = 6371$km is the radius of the Earth. We also show the distance estimated using

the time-gap between the observations and the speeds of the observations,

$$d^e_{max} = \max(v_1, v_2) \times \Delta t, \tag{5.3.2}$$

where v_1 and v_2 are the recorded speeds at the time of the two observations and Δt is the time-gap between them. We have used two estimates to provide reference for when a distance is realistic (i.e. $d \approx d^e_{max}$) or unrealistic (i.e. $d \approx 2d^e_{max}$). Instances of $d = 0$km indicate the observation at that time is a GPS-lagged observation. Almost all instances of $d = 0$km are immediately followed by $d^e_{max} < d \leq 2d^e_{max}$ which corresponds to the distance travelled between the observation before and after the GPS-lagged observation.

5.4 Quality-control tests

In this section we describe the quality-control tests we use on the filtered dataset. A schematic showing the complete quality-control process applied to the trial data is shown in figure 5.5. The climatological range test (section 5.4.1), stuck instrument test (section 5.4.2) and GPS test (section 5.4.3) are applied in parallel to the filtered dataset obtained in section 5.3. Observations that have passed each quality-control test undergo a final sensor ventilation test (section 5.4.4). The final quality-controlled dataset will be comprised of observations passed by every quality-control test. Throughout this section we use the units that the data have been recorded in which are given in section 5.2.

5.4.1 Climatological range test (CRT)

The Climatological range test (CRT) identifies observations which fall outside of location-specific climatological ranges (Limber et al., 2010; Boyce et al., 2017).

5.4.1.1 Test implementation

To implement this test, we use the Met Office integrated data archive system (MIDAS) daily temperature data (Met Office, 2006). This dataset contains observations of the maximum and minimum temperatures over a specified time window (usually 12 to 24 hours) for various locations in the UK. MIDAS temperature data have a resolution of 0.1°C and an uncertainty of 0.2°C.

To create our climatology dataset we use surface stations active during 2018. For each station, we obtain the minimum and maximum dry bulb temperature for February, March, and April using pre-2018 MIDAS data. We note that the climatology of each surface station will depend on when it was made operational and so the climatology length of each station will vary. Due to this, we remove the MIDAS

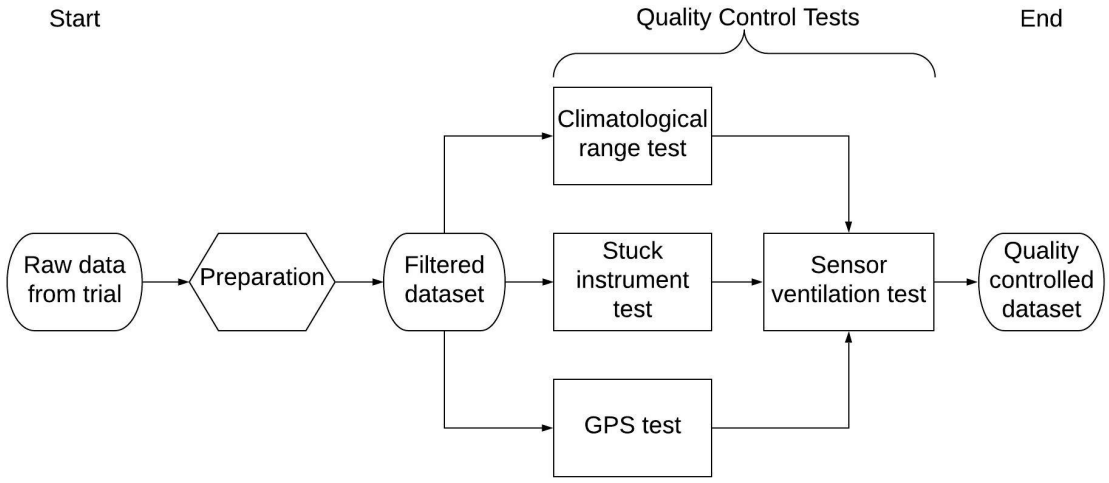


Figure 5.5: The complete quality-control process applied to the complete dataset obtained from the Met Office trial. The complete data is first prepared into the filtered dataset whereby all data without dry bulb temperature and necessary meta-data are discarded. The climatological range test (section 5.4.1), stuck instrument test (section 5.4.2) and GPS test (section 5.4.3) are applied in parallel to the filtered dataset. Observations which have passed each quality-control test undergo a final sensor ventilation test (section 5.4.4). The final quality-controlled dataset consist of observations that pass the sensor ventilation test.

stations with site IDs 62083 and 62119 from our April climatology dataset as they became active in 2016 and 2017 respectively. We also remove the MIDAS stations with site ID 6313 and 15365 from our April climatology dataset as they have implausibly small maximum temperatures despite becoming active in 1914 and 1988 respectively. While it is possible there exist other similarly problematic MIDAS stations, we have not found any further evidence to support removing any other stations from our climatology.

The CRT is performed by checking if the vehicle dry bulb temperature observations are within a predetermined tolerance of the minimum to maximum range for the nearest surface station. The nearest surface station is calculated through the use of the great circle distance (5.3.1). (The vehicle GPS measurement will be addressed in section 5.4.3).

For this test we use a 2°C tolerance to compensate for a number of factors. For example, dry bulb temperature would be expected to change with elevation in the surface layer (e.g. Stull, 1988, pp 9–19). Additionally, dry bulb temperature measurements taken on surfaces with higher albedo, such as grass, can produce noticeably different measurements from those taken on surfaces with lower albedo, such as asphalt (e.g., Huwald et al., 2009). The purpose of the tolerance used in this test is to account for elevation and surface differences between a vehicle and its nearest MIDAS surface station. This tolerance will also partially compensate for extreme events that occurred during the trial such as the snow and low temperatures that occurred during

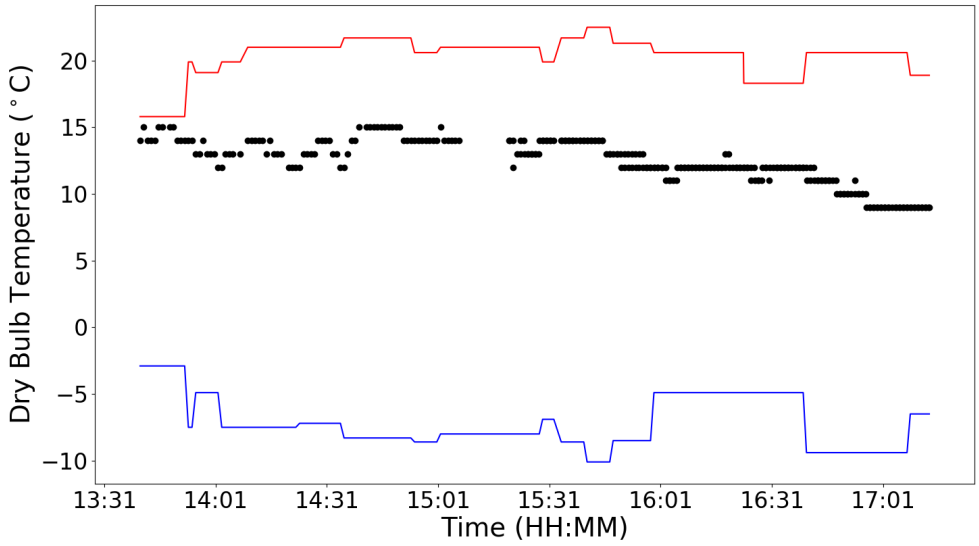


Figure 5.6: The dry bulb temperature observations (black) from one vehicle on the 25th of March 2018 and the climatological minimum (blue) and maximum (red) temperatures of the nearest MIDAS surface station. Any fluctuation in the minimum and maximum lines can be attributed to the tested observation being nearest to a surface station different from the previously tested observation. The two surface stations used to test the observations around 16.30 have the same minimum but different maxima. Each dry bulb temperature observation shown passes the CRT as each lies between the minimum and maximum temperature of the nearest MIDAS surface station.

late February and early March 2018 and a hot spell that occurred from 18th-22nd April 2018 (Met Office, 2020). Operational MIDAS surface stations will likely be shielded from radiation by Stevenson screens but radiation errors may still occur for calm and/or sunny conditions due to poor ventilation (Harrison, 2015). Vehicle dry bulb temperature sensors are not shielded from radiation and will be affected by re-radiated heat from road surfaces (Donegan, 2017).

The CRT applied to a single vehicle on the 25th of March 2018 is shown in figure 5.6. Observations taken during different segments of longer journeys are likely to be nearer to different MIDAS surface stations. This is seen by the jumps in the climatological maximum and minimum lines. Each dry bulb temperature observation plotted here would pass the CRT as each lies between the climatological minimum and maximum air temperature of the nearest MIDAS surface station.

5.4.1.2 Algorithm

The CRT algorithm is shown in algorithm 1. Here, $N_{filtered}$ is the total number of observations in the filtered dataset, y_i is the i-th observation in the filtered dataset with corresponding dry bulb temperature T_i, θ_{max} and θ_{min} are the maximum and minimum climatological air temperature observations of the nearest MIDAS surface

station and *tol* is the tolerance.

Algorithm 1: Climatological range test pseudocode

```
for i = 1 ... N_filtered do
    Determine corresponding MIDAS climatology to use (i.e. February, March or April)
    Find MIDAS station nearest to y_i to obtain θ_max and θ_min
    if θ_min − tol ≤ T_i ≤ θ_max + tol then
        Pass y_i
    else
        Flag y_i
    end if
end for
```

5.4.1.3 CRT results

Using the settings described in section 5.4.1.1 we find that 32129 observations have been passed (over 99%) by the CRT and 50 observations have been flagged. Since few observations were flagged by this test we conclude that a 2°C tolerance is suitable for this dataset but larger tolerances may be more suitable for other vehicle-based observation datasets.

5.4.2 Stuck instrument test (SIT)

Persistence tests are a common quality-control test to determine whether an instrument is stuck and/or if the variability of the measurements of some meteorological field is physically realistic (e.g., Zahumenskỳ, 2004; Drobot et al., 2011). Standard persistence tests will be unsuitable for the filtered dataset due to a large number of short journeys undertaken by participants during the trial. Furthermore, insufficient knowledge on the variability of dry bulb temperature on the scales measured by vehicles, as well as whether this variability would be adequately reflected in low precision measurements, add to the infeasibility of persistence tests. We therefore implement a simplified form of a persistence test which only checks whether an instrument is stuck on some value. This test will be referred to as the stuck instrument test (SIT).

5.4.2.1 Test implementation

To implement the SIT for an observation y_{test} valid at time t_{test} we first create a sample of observations y_{sample} valid at times t_{sample} from the same vehicle such that

$$t_{test} - 15 \text{ mins} \leq t_{sample} \leq t_{test} + 15 \text{ mins}. \tag{5.4.1}$$

The accepted sample time window is chosen to be symmetric in t_{test} so that observations at the start and end of a journey can be tested. We note that a large time-window is chosen for our sample to compensate for the low precision of the dry bulb temperature observations. Provided at least one observation in the test sample has a different dry bulb temperature value to y_{test} then y_{test} is passed.

5.4.2.2 Algorithm

The SIT algorithm is shown in algorithm 2. Here, $N_{filtered}$ corresponds to the number of observations in the filtered dataset, y_{test} is the test observation taken at time t_{test} with dry bulb temperature T_{test} and sensor ID SID_{test} and y_i is the i-th observation taken at time t_i with dry bulb temperature T_i and sensor ID SID_i.

Algorithm 2: Stuck instrument test pseudocode

```
for i ... N_filtered do
    Set y_test = y_i
    for i ... N_filtered do
        if y_test ≠ y_i, t_i ∈ [t_test − 15 mins, t_test − 15 mins] and SID_test = SID_i
        then
            Add T_i to sample S
        end if
    end for
    if the sample is empty then
        Cannot test y_test
    else if at least one dry bulb temperature doesn't equal T_test then
        Pass y_test
    else
        Flag y_test
    end if
end for
```

5.4.2.3 Results

There are 30124 observations passed by the SIT, 2008 observations flagged, and 47 observations which could not be tested as there were no other observations from the same vehicle in the sample window.

5.4.3 GPS test

Due to the high spatial resolution of vehicle-based observations, it is vital that accurate knowledge of the observation location is known. GPS-based positioning has been shown to be acceptable for accurate vehicle navigation, but requires frequent location polling (Menard et al., 2011). The accuracy of GPS measurements is known to be heavily affected by the smartphone and application used (Hess et al.,

2012; Bauer, 2013). The GPS test verifies whether the GPS metadata of individual observations provide physically plausible vehicle locations.

5.4.3.1 Test implementation

To quality check the GPS measurements associated with the vehicle data we examine the location of a test observation against the location of a reference observation. The reference observation will have been taken up to 30 minutes before the test observation and by the same vehicle. We note that the first observation taken by a vehicle and any observations with a time-gap larger than 30 minutes from the previous observation cannot be tested by this method as there will be no suitable reference observation to test against.

The GPS test first calculates the distance between the tested observation and reference observation, d_{test}, through the great circle distance (5.3.1). Next, using the metadata of the two observations, we calculate estimates of the minimum and maximum distances the vehicle could have travelled between the times the reference and test observation were taken. The maximum distance will be estimated by (5.3.2) where v_1 (v_2) is replaced by the speed of the test (reference) observation v_{test} (v_{ref}) and Δt is the time-gap between the test and reference observations. Similarly, the minimum distance will be estimated by

$$d^e_{min} = min\left(v_{test}, v_{ref}\right) \times \Delta t. \tag{5.4.2}$$

The estimates provided by (5.3.2) and (5.4.2) may not be reflective of the true distance traversed by the vehicle due to speed fluctuations and the route travelled between the observations. When estimating the maximum distance we must account for the vehicle having a larger speed than v_{test} and v_{ref} between the times the observations were taken. When estimating the minimum distance we must account for more factors than with the maximum distance. In addition to the vehicle having a lower speed than v_{test} and v_{ref} between the observation times, the journey may occur during heavy traffic congestion, the route travelled may not be a straight line, or the route traversed may be through a residential area. Additionally, MROs must be accounted for as it is possible their GPS will not have been updated because of the minimum time and distance update conditions on the app (see section 5.2).

The tested observation will pass the GPS test if $\Gamma_{min} \times d^e_{min} \leq d_{test} \leq \Gamma_{max} \times d^e_{max}$ where $\Gamma_{min} < 1$ and $\Gamma_{max} > 1$ are tolerance constants used to account for the uncertainty in d^e_{min} and d^e_{max} respectively. When an observation is passed by the GPS test, if $\Delta t \geq 1$ minute and $d_{test} > 0$km between this passed observation and the reference observation it was tested against it becomes the reference observation for the next test observation. Otherwise, the reference observation is unchanged for the next test observation. This is to avoid GPS-lagged observations being reference observations

as they are known to have inaccurate GPS (see section 5.3.2.3). If the next test observation is over 30 minutes from the reference observation, the test observation is unable to be tested and we set it to be the next reference observation.

For our implementation we set the tolerance constants as $\Gamma_{max} = 1.3$ and $\Gamma_{min} = 0.6$. To show the suitability of our choices for Γ_{max} and Γ_{min} in the GPS test we calculate the distances between observations and their respective reference observations denoted d_{test}. Figure 5.7 shows each d_{test} plotted against their corresponding d^e_{max} (black dots) and lines $d_{test} = d^e_{max}$ (red dashed line) and $d_{test} = 1.3 \times d^e_{max}$ (red solid line). The gradients of the two lines represent possible choices for Γ_{max}. By using $\Gamma_{max} = 1.3$, all points above the solid line will be flagged by the GPS test, since for these data, the vehicle appears to have travelled further than physically plausible since the previous reference observation. Points below the solid line are not flagged but must also pass a minimum distance test. Figure 5.8 shows each d_{test} plotted against their corresponding d^e_{min} (black dots) and lines $d_{test} = d^e_{min}$ (blue dashed line) and $d_{test} = 0.6 \times d^e_{min}$ (blue solid line). The gradients of the two lines represent possible choices for Γ_{min}. Similarly to figure 5.7, by using $\Gamma_{min} = 0.6$, most points below the solid line will be flagged by the GPS test, since for these data, the vehicle appears to have travelled less than physically plausible since the previous reference observation. Points above the solid line are not flagged but must also pass a maximum distance test. Additionally, we will set $\Gamma_{min} = 0$ when $v_{ref}, v_{test} < 25$km/h or $\Delta t < 1$ minute as we expect the test observation to be relatively near to the reference observation. (The specific choice of 25km/h is related to the sensor ventilation test which will discussed in section 5.4.4). Therefore, many observations beneath the solid line in figure 5.8 will not be flagged by the GPS test. We note that the horizontal threshold $d_{test} \approx 0.5$km in figures 5.7 and 5.8 is caused by the 500 metre default GPS update distance used by the app.

5.4.3.2 Algorithm

The GPS test algorithm is shown in algorithm 3. Here, N_{IDs} corresponds to the number of unique sensor IDs in the filtered dataset which determine if observations come from the same source (i.e. vehicle), N_{obs} is the number of observations for the i-th sensor ID and y_j is the j-th observation such that y_j occurs before y_{j+1}.

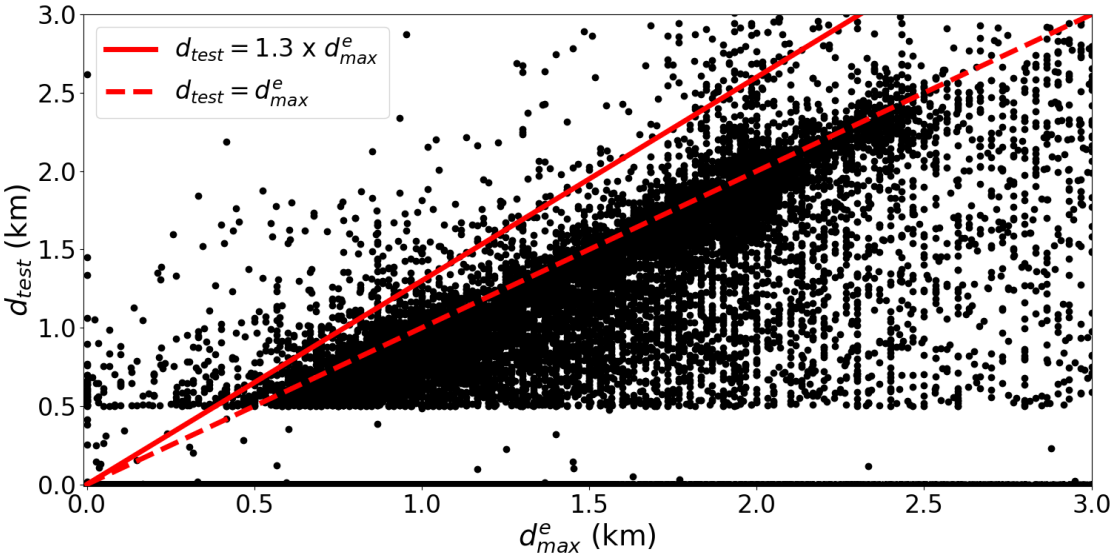


Figure 5.7: Scatter plot to show the distances between observations and their respective reference observation, denoted d_{test}, against their corresponding d^e_{max} (black dots) and the lines $d_{test} = d^e_{max}$ (red dashed line) and $d_{test} = 1.3 \times d^e_{max}$ (red solid line). The gradients of the two lines represent possible choices for Γ_{max}. We note that the threshold $d_{test} \approx 0.5$km is due to the 500 metre default GPS update distance of the app used in this trial. By using $\Gamma_{max} = 1.3$, all points above the solid line will be flagged by the GPS test, since for these data, the vehicle appears to have travelled further than physically plausible since the previous reference observation.

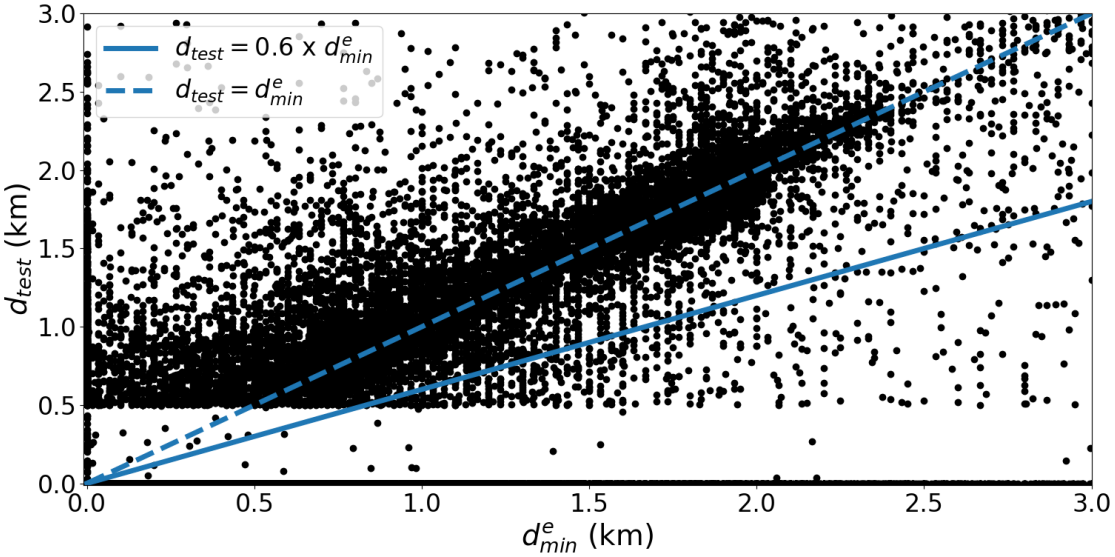


Figure 5.8: Scatter plot to show the distances between observations and their respective reference observation, denoted d_{test}, against their corresponding d^e_{min} (black dots) and the lines $d_{test} = d^e_{min}$ (blue dashed line) and $d_{test} = 0.6 \times d^e_{min}$ (blue solid line). The gradients of the two lines represent possible choices for Γ_{min}. We note that the threshold $d_{test} \approx 0.5$km is due to the 500 metre default GPS update distance of the app used in this trial. By using $\Gamma_{min} = 0.6$, all points below the solid line will be flagged by the GPS test, since for these data, the vehicle appears to have travelled less than physically plausible since the previous reference observation.

Algorithm 3: GPS test pseudocode

for i ... N_{IDs} **do**
  Collect all observations y belonging to the i-th sensor ID into a test dataset
  Set $y_{ref} = y_1$
  for $j = 2$ ... N_{obs} **do**
    Calculate the time-gap Δt between y_{ref} and y_j
    if $\Delta t \geq 30$ mins **then**
      Set $y_{ref} = y_j$
    else
      if $\Delta t < 1$ minute or $v_{test}, v_{ref} < 25$km/h **then**
        $\Gamma_{min} = 0$
      else
        $\Gamma_{min} = 0.6$
      end if
      Calculate d^e_{min}, d^e_{max} and d_{test}
      if $\Gamma_{min} \times d^e_{min} \leq d_{test} \leq \Gamma_{max} \times d^e_{max}$ **then**
        Pass y_j
        if $\Delta t > 1$ minute and $d_{test} > 0$km **then**
          Set $y_{ref} = y_j$
        end if
      else
        Flag y_j
      end if
    end if
  end for
end for

5.4.3.3 Results

There are 20162 observations that pass the GPS test of which 9939 are MROs. There are 11181 flagged observations and 836 observations which could not be tested. The results of the GPS test are summarised in table 5.2. The majority of the observations that couldn't be tested were due to large time-gaps between the test and reference observations.

For the flagged dataset, 1331 (1429) observations were flagged for having $d_{test} > d^e_{max}$ ($d_{test} < d^e_{min}$). We note that it is possible many of these observations may have accurate location metadata but have inaccurate speed and time metadata resulting in the disagreement between d_{test} and d^e_{max} or d^e_{min}. The remaining 8421 flagged observations were GPS-lagged observations. The GPS-lagged observations are primarily due to the 500 metre default GPS update distance of the app but also because of poor GPS signal resulting in more time being needed for a location update.

Determining the uncertainty of the GPS for this dataset would require examination of the GPS uncertainty caused by the smartphone app and each smartphone make used by the participants during the trial. Merry and Bettinger (2019) found an average horizontal position accuracy in urban areas of 7-13m for an iPhone 6. For convection-permitting NWP this would likely be an acceptable GPS uncertainty for vehicle-based observations.

A greater concern is the minimum GPS update distance used by the app (500m) which resulted in a substantial portion of the filtered dataset being flagged by the GPS test. For all crowdsourced observations accurate spatial location metadata are needed due to the scales of the atmospheric processes being observed. This is especially true for vehicles as their locations are non-stationary.

5.4.4 Sensor ventilation test (SVT)

The final quality-control test we apply in this report is the sensor ventilation test (SVT). As discussed in section 5.1, radiative forcing can result in large errors in measurements of air temperature. In order to produce realistic temperature measurements it is necessary for sensors to be adequately ventilated. For vehicle-based observations, sensor ventilation will be determined by the speed the vehicle is travelling at. From our examination of the dry bulb temperatures of the filtered dataset in section 5.3.2.1, we define the sensor ventilation threshold for the filtered dataset as $v_{sensor} = 25\text{km/h}$.

The SVT is used to check for adequate temperature sensor ventilation by removing any observations with speed less than the sensor ventilation threshold v_{sensor}. This test is implemented last as each observation can be tested individually and observations flagged by this test are still useful for the SIT or GPS test. We also note that $\Gamma_{min} = 0$ was used in the GPS test for observations with speed less than v_{sensor} as those observations will be flagged by the SVT regardless of the GPS test result.

5.4.4.1 Algorithm

The algorithm for the SVT is shown in algorithm 4. Here, N is the number of observations that have passed all previous quality-control tests and y_i is the i-th observation with corresponding speed u_i.

Algorithm 4: Sensor ventilation test pseudocode

```
for i = 1 to N do
    if u_i ≥ 25 then
        Pass y_i
    else
        Flag y_i
    end if
end for
```

5.4.4.2 Results

The SVT is applied to the 19094 observations that have passed all other quality-control tests. In total, the SVT flags 1669 observations (8.7%). We note that a large number of observations relative to the number tested are flagged by this test due to the large number of observations that occurred when the vehicle was stationary (i.e. 0km/h speed).

5.5 The quality-controlled dataset

The dataset resulting from the quality-control process described in section 5.4 consists of 17425 observations (25.6% of the complete dataset). A summary of the quality-control test results is given in table 5.2. The number of observations that passed all tests, were flagged by at least one test, and could not be tested by all tests for each day is shown in figure 5.9.

	Number of tested observations	Number of passed observations	Number of flagged observations	Number of untested observations
Climatological range test	32179	32129	50	0
Stuck instrument test	32179	30124	2008	47
GPS test	32179	20162	11181	836
Sensor ventilation test	19094	17425	1669	0

Table 5.2: Summary table containing the results from all quality checking tests.

A summary of the descriptive statistics for the quality-controlled dataset and each month is given in table 5.3. In total, 2833 observations were taken in February, 9012 in March, and 5580 in April. The ratio between the number of observations that occurred in each month in the quality-controlled dataset is relatively unchanged from that of the filtered dataset. The distributions of the dry bulb temperatures for each month, shown in figure 5.10, retain the majority of the characteristics of their filtered dataset counterparts discussed in section 5.3.2.1. There has been a slight reduction in skew and kurtosis for February and April as higher temperatures have been removed by the SVT. Additionally, extreme temperature anomalies have been removed from April by the CRT.

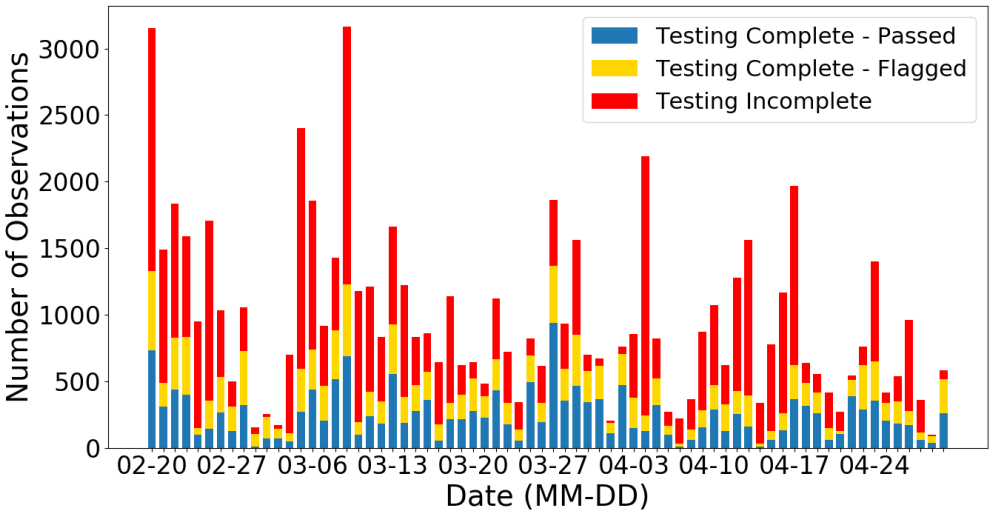


Figure 5.9: The observations that passed all tests (blue bar segments totalling 17425), were flagged by at least one test (yellow bar segments totalling 13897) and could not be tested by all tests (red bar segments totalling 36637). We note that the red bar segments include all observations removed in the initial filtering stage and those untested by any QC test.

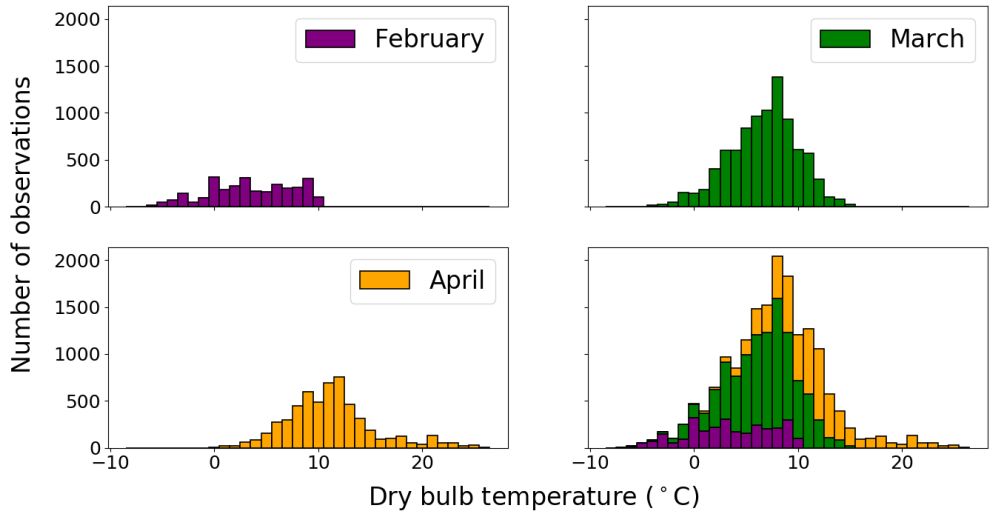


Figure 5.10: Distribution of dry bulb temperature observations for each month of the trial for the quality-controlled dataset. The purple bar segments indicate the number of February 2018 observations, the green segments March 2018, and the orange segments April 2018. The combined distribution is a stacked histogram that shows the contribution from each month to the total number of observations for each dry bulb temperature.

	February	March	April	All data
Number of observations	2833	9012	5580	17425
Mean	3.52°C	6.67°C	11.29°C	7.63°C
Standard deviation	4.04°C	3.30°C	4.26°C	4.64°C
Skew	−0.22	−0.38	0.72	0.26
Excess kurtosis	−0.88	0.13	0.92	1.07

Table 5.3: Summary of the descriptive statistics for the quality-controlled dataset and split into each month.

5.6 Summary and recommendations

The use of crowdsourced observations in numerical weather prediction is a new research area that is quickly receiving much attention from the meteorology community. Indeed, the high spatio-temporal resolution of the observations is particularly attractive for convection-permitting data assimilation where expansion and management of conventional scientific surface observing networks are too costly.

This report details the quality-control process applied to a novel low-precision vehicle-based observation dataset. In order to quality-control the raw dataset, we first needed to filter the dataset in order to determine which data had the necessary information. The climatological range test (CRT), stuck instrument test (SIT), and GPS test were applied in parallel to the filtered dataset. The sensor ventilation test (SVT) was applied to data that had passed the three previous tests. We note that as the CRT can be applied to each observation individually it may be more suitable to apply this test after the SIT or GPS test. The quality-control dataset consists of the 17425 observations (25.6% of original dataset) which have passed all quality-control tests. This is in stark contrast to the quality-control of scientific surface observing networks where most observations are retained. For example, in the Met Office system approximately 10% of dry bulb temperature observations are flagged or discarded (Waller, 2019). Recommendations for future quality-control and data collection are given at the end of this section.

The CRT was used as a range validity test on the dry bulb temperatures. This type of quality-control test has been successfully used for smartphone observations (Hintz et al., 2019b) and vehicle-based observations (Chapman et al., 2010; Limber et al., 2010; Boyce et al., 2017). To implement this test we used monthly climatologies of the MIDAS surface stations active during the trial. Hence, this test is limited by the climatology of each MIDAS station. The CRT flagged the lowest number of observations out of any quality-control test and would be suitable for operational quality-control of vehicle-based observations.

The SIT was used as a simplified persistence test to determine if the sensor was stuck on some value. In order to implement this test, a vehicle identifier was required to

determine if observations came from the same source. For operational persistence tests, it will be important to consider how observations taken from short finite journeys can be tested for persistence. We note that, despite the simplicity of this test, there is still a small amount of data in the filtered dataset unable to be tested. We also note that, because of the low precision of the data, it is also possible that valid observations have been flagged by this test. However, as the majority of observations able to be tested were passed, it is unlikely that our sample time-span was an issue for the implementation of this test on the filtered dataset.

The GPS test was performed to verify the plausibility of the GPS metadata. GPS accuracy has been an issue for crowdsourced smartphone observations where it is important to know the elevation of pressure observations (Madaus and Mass, 2017; Hintz et al., 2019b) but inaccuracies in the horizontal have not been a reported concern. Similarly to the SIT, vehicle identification is required to implement the GPS test. The GPS test involves comparing the distance between GPS coordinates of two observations with minimum and maximum estimates calculated through their metadata. As such, the results of this test are also dependent on the accuracy of the speed and time metadata of the observations. We also note that the starting observation of a journey or after a large time-gap was used as a reference observation in this test despite their GPS being untested and possibly inaccurate. The number of flagged observations is substantially greater than the number of observations flagged by the CRT, SIT and SVT combined. This is primarily due to the existence of GPS-lagged observations in the filtered dataset which are caused by poor GPS signal, insufficient distance travelled, or time between observations to trigger a GPS location update. As vehicles are able to traverse a greater distance in a short time-span than some current operational weather prediction model grid spacing it is important that lags in GPS be accounted for.

The SVT is the final quality-control test and is another filtering test where observations with speed less than $v_{sensor} = 25$km/h are flagged. A similar precaution was used by Knight et al. (2010) who found vehicle-based temperature observations to be reasonably accurate provided the vehicle was moving for a number of minutes prior to the observation time. Despite being applied to fewer observations than any other test, a large number of observations were flagged by the SVT. It is expected that a large amount of vehicle-based observations will be flagged due to traffic congestion or driving through residential areas.

Unfortunately, we were unable to perform a spatial consistency test on the filtered dataset due to a lack of observations occurring at similar locations and times. Additionally, due to the high spatial resolution of each observation in this dataset, we were unable to obtain an independent dataset which most observations from the filtered dataset could be tested against. It is possible in the future that vehicle-based

observations form vast observation networks that are densest in urban areas making spatial consistency tests more feasible. Additionally, the examination of any biases present in the observations was also infeasible due to a lack of data. Vehicle-based observations are likely to be biased due to the sensing instrument and heat from the car engine and road surface.

Recommendations for future data collection and quality-control of vehicle-based observations are as follows.

- A substantial amount of data was found without valid speed metadata. Correcting the OBD dongle or app settings/features causing this will result in fewer observations discarded prior to the quality-control process.
- Reduce the GPS update distance to reduce the number of GPS-lagged observations.
- The climatology datasets used in the CRT each contained around 400 MIDAS stations situated within the UK making this test a computationally expensive matching procedure. In operational settings, it may be more appropriate to use reduced datasets such as monthly regional climatologies. Additionally, the comparison of vehicle-based observations with WOW surface station data may provide another suitable quality-control test.
- The temperature observations obtained from the Met Office trial are all low-precision measurements. Because of this, a simplified persistence test was used as a quality-control test. If precision was increased to 0.1°C then more standard persistence tests can be used as a quality-control test. This will allow for testing of whether the sensor is stuck and whether the variability of the observed fields is physically plausible simultaneously.
- Both the SIT and GPS test could not have been implemented without a sensor ID. Due to data privacy, it may be unfeasible to have sensor IDs with potential vehicle-based observation sources. Using appropriate encryption techniques on vehicle sensor IDs may allow for the use of vehicle time series in quality-control (e.g., Verheul et al., 2019). Alternatively, if a phone app is used in the data collection process it may be possible to quality-control locally on a smartphone before uploading to WOW servers. Such methods have been used for the quality-control and bias correction of smartphone observations (Hintz et al., 2019b).
- To check if a vehicle sensor is stuck without vehicle observation time series it may be suitable to record the amount of time since the vehicle observed a different value as metadata for that observation.
- To check GPS accuracy without vehicle location time series it would be use-

ful to record GPS signal strength between a satellite and a phone, as well as how many satellites are involved in the GPS polling, as metadata for an observation. Having access to this information will also assist in determining the GPS measurement uncertainty. Additionally, provided enough observations are available, spatial consistency tests may be capable of flagging some observations with incorrect GPS.

The Met Office proof-of-concept trial has shown it is possible to obtain vehicle-based observations from in-built vehicle sensors using smartphones and OBD dongles. However, the observations obtained through this trial leave much to be desired. While the quality-control procedure presented in this report may be a suitable reference point for other crowdsourced datasets, there is much improvement to be made in both data collection and quality-control for such observations to be utilisable. It is therefore necessary to conduct further trials possibly with alternative data collection methods which address the issues raised in this report.

5.7 Chapter summary

In this chapter, we addressed the first research question given in chapter 1: *How can we quality-control the vehicle-based air-temperature dataset obtained from a Met Office proof-of-concept trial?* A novel quality-control procedure was implemented on crowdsourced vehicle-based observations of air temperature obtained from a Met Office proof-of-concept trial. The quality-control procedure consisted of range-validity tests, a new GPS test and a time-series check. We note that the GPS test and time-series check may be unsuitable more generally due to data privacy. In total, only 25.6% of the initial dataset passed quality-control. We found that the majority of observations that were unable to be tested or failed quality-control were due to missing or inaccurate metadata. Hence, the quality of the available metadata is key for the quality-control of vehicle-based observations. In the next chapter, we will explore the characteristics of the vehicle-based observations that passed quality-control.

Chapter 6

Exploring the characteristics of a vehicle-based temperature dataset for convection-permitting numerical weather prediction

In this chapter we use the quality-controlled dataset obtained in chapter 5 to answer the second research question given in chapter 1: *what are the characteristics of vehicle-based observations of air temperature?* In particular, we wish to determine if the uncertainty of vehicle-based observations of air temperature is weather-dependent or vehicle-dependent. The remainder of this chapter, except for the chapter summary in section 6.6, is strongly based on the paper submitted to Meteorological Applications for publication (Bell et al., 2021a).

Abstract:

Crowdsourced vehicle-based observations have the potential to improve forecast skill in convection-permitting numerical weather prediction (NWP). The aim of this paper is to explore the characteristics of vehicle-based observations of air temperature. We describe a novel low-precision vehicle-based observation dataset obtained from a Met Office proof-of-concept trial. In this trial, observations of air temperature were obtained from built-in vehicle air-temperature sensors, broadcast to an application on the participant's smartphone and uploaded, with relevant metadata, to the Met Office servers. We discuss the instrument and representation uncertainties associated with vehicle-based observations and present a new quality-control procedure. It is shown that, for some observations, location metadata may be inaccurate due to unsuitable smartphone application settings. The characteristics of the data that passed quality-control are examined through comparison with United Kingdom

variable-resolution model data, roadside weather information station observations, and Met Office integrated data archive system observations. Our results show that the uncertainty associated with vehicle-based observation-minus-model comparisons is likely to be weather-dependent and possibly vehicle-dependent. Despite the low precision of the data, vehicle-based observations of air temperature could be a useful source of spatially-dense and temporally-frequent observations for NWP.

6.1 Introduction

Convection-permitting numerical weather prediction (NWP) requires a large number of observations of high spatio-temporal resolution to constrain short-term forecasts (Sun et al., 2014; Gustafsson et al., 2018; Dance et al., 2019). However, due to the cost of installation, management, and maintenance of observing instrumentation, it may be impractical to extend traditional scientific observing networks to provide sufficient additional relevant observations. A potential alternative source of inexpensive observations is from opportunistic data generated by the public or other organisations (Waller, 2020; Blair et al., 2021).

The application of opportunistic datasets in NWP has been a popular area of research in recent years (Hintz et al., 2019a). Observations from personal weather stations (PWSs) (Steeneveld et al., 2011; Wolters and Brandsma, 2012; Chapman et al., 2017; Meier et al., 2017; Nipen et al., 2020) and smartphones (Overeem et al., 2013; Droste et al., 2017; Madaus and Mass, 2017; Hintz et al., 2019b, 2020, 2021) are commonly obtained through crowdsourcing. Such observations may be inaccurate when compared with traditional scientific observations. However, the number of crowdsourced observations available has the potential to far exceed the number of scientific surface observations currently produced (Muller et al., 2015). Opportunistic datasets can also be obtained from partnerships with other organisations. For example, roadside weather information station (RWIS) data obtained from highways agencies are currently assimilated into the Met Office United Kingdom variable-resolution (UKV) model (Gustafsson et al., 2018).

Observations obtained from vehicles are another dataset of opportunity (Mahoney III and O'Sullivan, 2013). Similarly to PWS and smartphone observations, vehicle-based observations can be obtained through crowdsourcing and will therefore be most densely distributed in urban areas and on major transport networks. Vehicle-based observations can also be obtained through several non-crowdsourcing methods. For example, the data can be obtained directly from vehicle manufacturers through connected vehicle initiatives (e.g., Mahoney III and O'Sullivan, 2013), from built-in sensors of vehicle fleets via the controller area network (CAN) (e.g., Mercelis et al., 2020), or through externally mounted sensors (e.g., Anderson et al., 2012). In this paper, vehicle-based observations of air temperature are obtained from built-

in vehicle sensors through on-board diagnostic (OBD) dongles. This method of data collection, which is described in section 6.3, could be used for crowdsourcing vehicle-based observations.

Vehicle-based observations are currently used to improve road weather modelling (e.g., Hu et al., 2019) and forecasts to combat adverse road weather conditions on transportation networks (e.g., Karsisto and Nurmi, 2016; Siems-Anderson et al., 2019). Karsisto and Lovén (2019) showed that assimilation of vehicle-based observations into the Finnish Meteorological Institute's road weather model had the greatest forecast impact factor when RWISs were sparse. The use of vehicle-based observations in NWP is still in its infancy, but their use for nowcasting has been investigated by the German weather service (DWD) (Hintz et al., 2019a). Additionally, an observing simulation system experiment (OSSE) conducted by Siems-Anderson et al. (2020) showed a modest but appreciable impact from assimilating simulated vehicle-based observations.

Before opportunistic datasets can be assimilated, they must undergo thorough quality-control (QC) and the contributions to their observation uncertainty identified and investigated. Bell et al. (2015) attributed the total uncertainty of crowdsourced PWS observations to five sources; calibration issues, communication and software issues, inaccurate metadata, design flaws, and error due to unresolved scales. As a result of these issues, which also apply to other opportunistic datasets, the implementation of QC procedures can become substantially more difficult than the QC for traditional observations. In some studies over half the crowdsourced data were removed by the QC procedure (e.g. Meier et al. (2017); Madaus and Mass (2017); Hintz et al. (2019b)). Siems-Anderson et al. (2019) developed QC for vehicle-based observations from disparate sources for use in road weather forecasting systems. However, these QC tests required a large number of observations to be in close spatio-temporal proximity such that spatial comparisons be used. Due to a lack of observations occurring at similar locations and times for the dataset examined in this study, a new QC procedure was developed.

Understanding the characteristics of opportunistic observations is key to their effective use in NWP (Waller, 2020). For data assimilation, an understanding of the instrument and representation errors that contribute to the total observation uncertainty is required. Important meteorological features such as sharp discontinuities caused by precipitation processes can be observed by opportunistic observations but will likely be misrepresented by a NWP model (Mahoney III and O'Sullivan, 2013). Hence, it is likely that there will be significant representation error caused by the mis-match in scales observed and modelled (Janjić et al., 2018). The instrument and representation components of the vehicle-based observation uncertainty are discussed in section 6.2. For the vehicle-based observations of air temperature examined

in this study, an important physical feature misrepresented by a NWP model will be the underlying road surface. The influence of roads on the air temperature measured by vehicles will be complex as the road-surface energy balance at a given location is substantially affected by the availability of water, the quantity of visible sky, and the amount of traffic (e.g., Anandakumar, 1999; Chapman and Thornes, 2011; Oke et al., 2017; Karsisto and Lovén, 2019). To properly understand the discrepancy between what is observed and modelled, it is necessary to examine the characteristics of the differences between the model and the observations. The objective of this paper is to explore the characteristics of a vehicle-based temperature dataset through comparison with other datasets.

The format of this paper is as follows. In section 6.2 the uncertainties associated with vehicle-based observations of air temperature are discussed. The Met Office trial used to obtain the vehicle-based observations in this study, the datasets used for comparison, and the novel quality-control procedure applied to the vehicle-based observations are detailed in section 6.3. The results of the new quality-control process highlight that the observation location metadata can be inaccurate due to poor GPS signal and application settings. A comparison between vehicle-based observations and other datasets is given in section 6.4. Our novel results show that the uncertainty of vehicle-based observations is likely weather-dependent and possibly vehicle-dependent. In section 6.5 our results are summarised and we conclude that vehicle-based observations are a promising opportunistic dataset for convection-permitting data assimilation.

6.2 Uncertainties in vehicle-based observations of air temperature

6.2.1 Vehicle-based observations of air temperature from built-in sensors

Most modern vehicles are equipped with a sensor to measure the air temperature of the surrounding atmosphere. Throughout this paper, these sensors will be referred to as external air-temperature sensors. Measurements obtained from external air-temperature sensors are used by vehicle air conditioning systems to adjust cabin air temperature (Abdelhamid et al., 2014) and alert the driver to safety hazards such as the possible presence of ice on the roads (Padarthy and Heyns, 2019). External air-temperature sensors are commonly negative temperature coefficient thermistors (FierceElectronics, 2014). The location of external air-temperature sensors will vary with vehicle make and manufacturer. Common placements are usually in the airflow at the front of the vehicle, such as behind the grill near the bottom of the vehicle or in the wing mirror (Tchir, 2016). We note that most vehicles also have a sensor

that measures the air temperature inside the vehicle engine commonly referred to as the intake air-temperature sensor. These measurements, however, are contaminated with heat from the vehicle engine and hence will not be representative of the true atmospheric conditions.

6.2.2 Instrument error

Built-in vehicle sensors are not intended to give high-quality meteorological information. As such, observations of air temperature from external air-temperature sensors are likely to have substantial instrument uncertainty. There are several sources of instrument uncertainty for vehicle-based observations of air temperature:

1. The observations may be affected by extraneous influences (Mahoney III and O'Sullivan, 2013).
2. The sensing instrument may not be as accurate or precise as required for meteorological applications (Mahoney III and O'Sullivan, 2013).
3. The ventilation of the sensing instrument may be inadequate (Harrison, 2015).

We now discuss these issues in more detail.

The extraneous influences that vehicle-based observations of air temperature are subject to include heating from the vehicle engine or the underlying road surface. The degree of vehicle influence on the observations will be determined by the sensors proximity to the vehicle engine. Mercelis et al. (2020) found that observations of air temperature from external air-temperature sensors situated far away from the vehicle engine were consistent with reliable observations obtained from road weather information stations. In contrast, observations obtained from sensors near the vehicle engine had to be discarded due to sensor biases. While external air-temperature sensor placement is usually chosen to mitigate the influence of engine heat (Tchir, 2016), radiation reflected from the road surface can be incident on the sensor. Observations of air temperature from external air-temperature sensors in such circumstances may be warmer than the true ambient conditions.

The precision of an observation will depend on the number of significant figures available for the digital representation of the measured value. (The concept we have called precision is known in metrology as resolution (BIPM et al., 2012)). The difference between a continuous variable and its imprecise digital representation is known as the quantization error (Widrow et al., 1996). As the sensing instruments used for opportunistic datasets are not intended to give high-quality meteorological information, quantization uncertainty will likely be part of the instrument uncertainty (e.g., Mirza et al., 2016).

Adequate sensor ventilation is necessary to ensure accurate observations of air tem-

perature (Harrison and Burt, 2020). Sensor ventilation for external air-temperature sensors is determined by how fast the vehicle is moving (e.g., Knight et al., 2010).

6.2.3 Representation error

Representation error is defined as the difference between a perfect observation and a model's representation of that observation (Janjić et al., 2018; Bell et al., 2020). The model's representation of an observation is calculated using an observation operator. An observation operator is function that maps the model state into observation space. According to Janjić et al. (2018), the representation error consists of three components:

1. The pre-processing error caused by the incorrect preparation of an observation.
2. The observation operator error due to any incorrect or approximate observation operators used in the assimilation of an observation.
3. The error due to unresolved scales and processes when there is a mis-match in scales and processes observed and modelled.

We now discuss these errors in more detail.

The pre-processing error for vehicle-based observations of air temperature can be caused by the data collection and quality-control procedures. The height of external air-temperature sensors will vary with vehicle type and sensor-height metadata will likely be unavailable in the collection of crowdsourced datasets. Hence, the observations must be assigned a height, which may differ from the true height, resulting in a height assignment error. The quality-control for the vehicle-based observations is discussed in section 6.3.3.

Since air temperature is usually an NWP variable, the observation operator for vehicle-based observations may be a simple interpolation operator. An observation operator error may result from the misrepresentation of the vehicle-based observation height by the NWP model. The resolution of NWP models is likely to be too coarse to represent the elevation of the vehicle-based observations properly (Waller et al., 2021). This mismatch in elevation between a surface observation and a NWP model field is normally accounted for by correcting the observation to be at the same height as the model field. We note that, for surface observations, the observation operator will not interpolate the model field to the observation height to prevent the interpolation of the model field beneath the model surface. As air temperature is expected to change with altitude in the surface layer (e.g., Stull, 1988, pp 9–19), the model height selected by the observation operator will influence the value of the model-equivalent observation.

As discussed in section 6.1, vehicle-based observations of air temperature will mea-

sure the air temperature above the road surface, which is likely to be influenced by radiative effects (Chapman et al., 2001). The effect of the radiation absorbed by the road surface on vehicle-based observations may be understood through the road-surface energy balance (RSEB) equation

$$I_{net} + TR = G + H + LE, \tag{6.2.1}$$

where I_{net} is the net radiation into the surface, TR is the heating caused by traffic, G is the ground heat flux density, H is the sensible heat flux density and LE is the latent heat flux density (Karsisto and Lovén, 2019). A illustration of the road-surface energy balance is given in figure 6.1. The materials used for road surfaces have a large heat capacity such that the majority of the radiation absorbed by the surface, $I_{net}+TR$, is converted into ground heat flux, G (Anandakumar, 1999). The remaining turbulent heat fluxes, H and LE, are determined by the amount of water available at the road surface (Oke et al., 2017). If water is unavailable at the road surface, the remaining energy, $I_{net}+TR-G$, is entirely converted to sensible heat, H, which will result in a strong vertical air-temperature gradient near the road surface. If water is available at the road surface some, but not all, of the remaining energy, $I_{net}+TR-G$, is converted into latent heat, LE. Hence, the vertical air-temperature gradient near the road surface will reduce and so the air-temperature profile will be more uniform. A common approach for modelling surface fluxes in NWP is through tile schemes (e.g. Essery et al., 2003). Using this approach, the surface flux of a grid box is the weighted average of several different surface fluxes and hence may differ from the RSEB substantially. Therefore, we expect the greatest representation uncertainty to be during dry and sunny conditions when the sensible heat flux, H, emitted by the road surface will be largest.

The amount of radiation absorbed by the road, $I_{net} + TR$, and hence the effect of the road surface on the air temperature above the surface, will vary across the road (Chapman and Thornes, 2011). This may be attributed to the sky-view factor and the complex effect of traffic on road-surface temperature. The sky-view factor corresponds to how much of the sky is visible. Therefore, the sky-view factor determines how much solar radiation is absorbed during the day and how much cooling there is at night. The sky-view factor may be highly spatially variable in rural areas with trees near the road or in urban canyons. Chapman and Thornes (2011) showed a rural example where the sky-view factor caused road-surface temperature to vary by almost 3°C. Road-surface temperature is affected by traffic through the generation of turbulence by vehicles, friction heat dissipation from tyres, sensible heat flux from vehicle engines, heat and moisture from exhaust fumes, and the blocking of incoming solar radiation and outgoing longwave radiation from the road surface. The most influential of these processes on the temperature profile across the road surface is

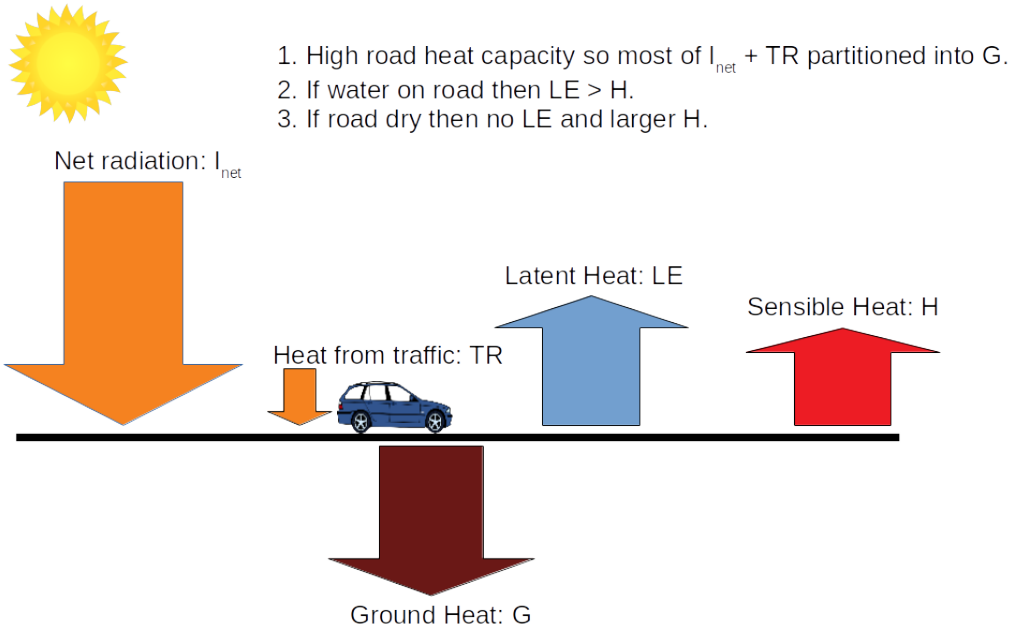


Figure 6.1: Illustration of the road-surface energy balance. The net radiation into the surface, I_{net}, and heat from traffic, TR, are converted into ground heat flux G, sensible heat flux H, and latent heat flux LE. In dry and sunny conditions, there will be no latent heat LE and so the energy $I_{net} + TR - G$ is converted entirely into sensible heat H. When there is water available at the road surface, more, but not all, of the energy $I_{net} + TR - G$ will be converted into latent heat LE than sensible heat H. As the vertical air-temperature gradient above the road surface will be determined by the magnitude of H, we expect the representation uncertainty to be greatest in dry and sunny weather conditions.

the heat from vehicle engines and the frictional heat dissipation from tyres (Prusa et al., 2002; Chapman and Thornes, 2005). Gustavsson et al. (2001) found that, during morning commuting hours in urban areas, traffic caused the road surface temperature to increase by approximately 2°C.

A common approach for modelling surface fluxes in NWP is through tile schemes (e.g., Essery et al., 2003). Using this approach, the surface flux of a grid box is the weighted average of several different surface fluxes and hence may differ from the local RSEB substantially.

For road forecasting applications, outputs from NWP are post-processed in order to take better account of the road physics (e.g., Clark, 1998; Coulson et al., 2012). For this initial study, we do not use these post-processing techniques for simplicity. However, in principle, a more sophisticated observation operator could use a similar approach to road forecasting models and reduce the uncertainty due to unresolved scales.

6.3 Methodology

6.3.1 The Met Office trial

From 20th February 2018 to 30th April 2018 the Met Office ran a proof-of-concept trial to collect vehicle-based observations of air temperature. The measuring instruments used in this trial are those built-in by the manufacturer of the vehicle. In this trial, on-board diagnostic (OBD) dongles were used to broadcast reports from the vehicle engine management interface to an application (app) installed on a participant's smartphone via Bluetooth. Additional metadata derived from the smartphone was appended to the report, and uploaded to the Met Office Weather Observations Website (Kirk et al., 2021) using the smartphone's connection to the mobile network (3G etc). A complete description of this trial can be found in Bell et al. (2021b). We now note some of the important aspects of the trial below.

The data collection frequency and GPS update period was set to 1 minute while the minimum distance for a GPS update was set to 500 metres. We also note that a known fault that occurred during this trial was for engine-intake temperature (i.e. the air temperature inside the vehicle engine) to be recorded as air temperature for some observations. Observations were collected throughout the United Kingdom with their locations corresponding to journeys undertaken by the participants.

The dataset obtained through this trial consists of 67959 reports obtained from 31 Met Office volunteers. Each report contains some combination of observations of air temperature, engine-intake temperature, and air pressure from built-in vehicle sensors. The observations of temperature have a precision of 1°C while the observations of air pressure have a precision of 10hPa. We limit the scope of this study to air temperature only as engine-intake temperature will not reflect the true atmospheric-air temperature and air pressure has too low precision to be useful in NWP. The metadata for each report include vehicle speed (km/h), date-time (given by the application as date and 24 hour clock time), GPS location, vehicle ID, and an unique observation ID. With the exception of vehicle speed which was obtained by the OBD dongle, all metadata was derived by the smartphone app.

6.3.2 Additional datasets used in this study

6.3.2.1 Met Office Integrated Data Archive System data

Met Office Integrated Data Archive System (MIDAS) temperature data consists of observations of 1.25m-air temperature which have a precision of 0.1°C and an uncertainty of 0.2°C for various locations in the UK (Met Office, 2006). We use MIDAS daily maximum and minimum temperature data in our quality-control procedure described in section 6.3.3. We also use MIDAS hourly temperature data to provide

a comparison with vehicle-based observations of air temperature that are within 1.5km of a MIDAS station (see section 6.4.2). These data are linearly interpolated to the time of a vehicle-based observation.

6.3.2.2 NWP model data

To explore the characteristics of the vehicle-based observations that pass quality-control we use Met Office 10-minute UK variable-resolution (UKV) model data (Met Office, 2016). The UKV is a variable resolution configuration of the Unified Model whose domain covers the United Kingdom and Ireland (Lean et al., 2008). The inner domain has grid boxes of size 1.5km × 1.5km and fully covers the United Kingdom (Milan et al., 2020). Surrounding this is a variable-resolution grid with boxes whose edges steadily increase in zonal and/or meridional directions to 4km in size.

The UKV model fields we use in this study are 1.5m-air temperature and surface-air temperature defined as the air temperature at the boundary with the surface. The UKV model data are interpolated to the time and horizontal location of a vehicle-based observation so that we can construct two observation-minus-background (OMB) datasets (i.e. one OMB dataset using surface-air data for the background and another OMB dataset using 1.5m-air temperature for the background). Since a vehicle-based observation and the horizontally interpolated background are both estimates of the true air temperature, their difference is equal to the difference of their errors. If their errors are independent, the variance of their differences will be equal to the sum of their individual error variances. Therefore, examining the statistics of the two observation-minus-background datasets will provide insight into the uncertainty of the vehicle-based observations. As the height of the external air-temperature sensor for each vehicle is unknown, we are unable to interpolate the model data to the height of a vehicle-based observation or correct the vehicle-based observation to be at the height of either UKV model field. It is likely that the vehicle-based observations are between the two model heights and are closer to the surface than the 1.5m height.

The surface flux for each grid box is determined by expressing the percentage of land use as a combination of 5 vegetation and 4 non-vegetation tiles (Essery et al., 2003; Porson et al., 2010). For each grid box, the surface flux is obtained by calculating the sum of the weighted average of the fluxes from each tile (where instantaneous interaction between tiles is neglected). The UKV uses the urban canopy model MORUSES (Met Office-Reading Urban Surface Exchange Scheme) as the urban tile. MORUSES represents the impervious urban surface through a roof tile and a canyon tile (Hertwig et al., 2020). However, observations taken on motorways and major routes will often be surrounded by rural areas, and so the road fraction of the UKV grid box will be small. For example, a typical UK motorway traversing a

rural grid box occupies less than 2% of the total area (Bremner, 2019).

6.3.2.3 Roadside weather information station observations

Vehicle-based observations of air temperature from built-in sensors are known to be consistent with reliable observations obtained from roadside weather information stations (RWISs), provided the external air-temperature sensor is located away from the engine block (Mercelis et al., 2020). We therefore use RWIS data provided by Highways England (2018) to provide a comparison with similar point observations for different weather conditions. There are over 250 RWISs in England located along major roads and major routes providing various roadside meteorological information with a temporal frequency of 10 minutes (Buttell et al., 2020). In this study, we use RWIS observations of air temperature that have precision of at least 0.1°C. To give an indication of the total uncertainty of these observations, we note that the Met Office currently assimilate RWIS observations of air temperature into the UKV with an uncertainty of 1°C. We note that the height that RWISs measure air temperature is estimated to be between 2 and 3 metres, but can be outside of this range if the site is located on a bank (Highways England, 2020). Road-state classifiers (i.e. dry, trace amounts of water, wet) provided by RWISs are used to indicate the availability of water at the road surface. The RWIS observations are linearly interpolated to the time that a vehicle passed a station.

6.3.3 Quality-control

In this section, we briefly describe the quality-control (QC) process applied to the vehicle-based dataset. Further details are given by Bell et al. (2021b). We note that, due to the size and spatio-temporal sparsity of this dataset, we were unable to use spatial consistency QC tests.

Before the QC process was implemented, an initial filtering of the raw data from the trial was performed to ensure each observation had an air temperature observation and the relevant metadata needed for each test. This filtering removed 35780 observations due to either a missing air temperature observation or an invalid speed. The resultant dataset will be referred to as the filtered dataset.

The QC process applied to the vehicle-based dataset began with three tests applied in parallel: the climatological range test (CRT), the stuck instrument test (SIT), and the global positioning system (GPS) test. Lastly, observations that passed each of these tests were then put through a sensor ventilation test (SVT). The final quality-controlled dataset (QC-dataset) consisted of all observations that passed the SVT. We now provide a brief description of each QC test.

The CRT checked if an observation was within a specified tolerance of a location-

specific climatology. For this dataset, we used MIDAS daily temperature data (Met Office, 2006) to create monthly climatology datasets. These datasets were constructed by determining the maximum and minimum air temperature of each MIDAS station active during February to April 2018 from pre-2018 data. The CRT was implemented by comparing the observation to the nearest (in terms of great circle distance) MIDAS station monthly climatology dataset. If the observation was within a 2°C tolerance of the climatological range of the MIDAS station, then the observation was passed.

The SIT examined portions of vehicle-specific time-series to check whether the vehicle sensor was stuck on an air temperature value. This test required a vehicle identifier to determine observations that came from the same source. (This may be unavailable in other crowdsourced observation studies due to data privacy concerns). The SIT was implemented by comparing an observation with all other observations from the same vehicle that occurred within a 15-minute time-window. If there was at least one observation that had a different value of air temperature to the tested observation, then the tested observation was passed. This test is essentially a simplified version of a persistence test (see Zahumenskỳ (2004) for guidelines) that is able to account for any short journeys undertaken by participants during the trial and the low precision of the data.

The GPS test compared the location of an observation, denoted the test observation, relative to a prior observation from the same vehicle, denoted the reference observation, to evaluate the plausibility of the observation location metadata. The reference observation was at most 30 minutes before the test observation. As with the SIT, a vehicle identifier was required to determine if observations came from the same source. The GPS test was implemented by calculating the great-circle distance between the test and reference observations, d_{test}. Then d_{test} was compared with the maximum and minimum distances estimated using the speed and time metadata for the vehicle. The maximum distance was estimated by

$$d^e_{max} = max(v_{test}, v_{ref}) \times \Delta t, \tag{6.3.1}$$

where v_{test} and v_{ref} are the speeds of the test and reference observations respectively, and Δt is the time-gap between the two observations. Similarly, the minimum distance was estimated by

$$d^e_{min} = min(v_{test}, v_{ref}) \times \Delta t. \tag{6.3.2}$$

The test observation passed the GPS test provided $\Gamma_{min} d^e_{min} \leq d_{test} \leq \Gamma_{max} d^e_{max}$ where $\Gamma_{min} = 0.6$ and $\Gamma_{max} = 1.3$ are minimum and maximum multiplicative tolerance constants, respectively. For justification of the choice of Γ_{min} and Γ_{max}, we

QC test	Number of tested observations	Number of passed observations	Number of flagged observations	Number of untested observations
Climatological range test	32179	32129	50	0
Stuck sensor test	32179	30124	2008	47
GPS test	32179	20162	11181	836
Sensor ventilation test	19094	17425	1669	0

Table 6.1: Summary of the results from all QC tests. The observations untested by the SIT and GPS test are due to a lack of reference observations. The observations passed by the SVT form the QC-dataset.

refer the reader to Bell et al. (2021b). Test observations with $\Delta t < 1$ minute or $max(v_{test}, v_{ref}) < 25$km/h were passed if $d_{test} \leq \Gamma_{max} d^{e}_{max}$ as they were expected to be close to the reference observation. (The specific choice of 25km/h is related to the sensor ventilation test discussed in the next paragraph). If a test observation did not have an observation from the same vehicle that occurred at most 30 minutes prior, then it was left unclassified by the GPS test and became the reference observation for the next test observation in the vehicle time-series.

The SVT was the final QC test which was applied to the observations that passed all previous tests. This test involved checking that the speed metadata for each observation was above a predetermined sensor ventilation threshold, v_{sensor}. Examining the speed-temperature pairs of the filtered dataset (not shown) revealed that the largest air temperatures (above 26°C) occurred for speeds below 25km/h. We therefore set $v_{sensor} = 25$km/h. An observation passed the SVT if it had speed greater than v_{sensor}. Hence, any observations that were passed by the GPS test with low speeds were flagged by the SVT.

The QC-dataset contains 17425 observations (25.6% of original dataset). A summary of the results of each QC test is provided in table 6.1. We note that the SIT and GPS test could not test every observation in the filtered dataset due to unavailable or unsuitable reference observations. The most discriminating test was the GPS test. The majority of observations flagged by the GPS test were likely the result of the 500m update distance default setting on the app. We also note that the SVT was a fairly discriminating test.

The QC approach taken with this dataset relied upon range validity and time-series tests. For crowdsourced observations, time-series tests may be unsuitable as instrument identification metadata may be unavailable due to data privacy concerns. This may be overcome with appropriate encryption techniques (e.g., Verheul et al., 2019) or by performing the QC locally on the sensing device (e.g., Hintz et al., 2019b). Furthermore, the use of spatial consistency QC tests, which do not require instrument identification, would be a suitable replacement for time-series-based tests provided there is a sufficient density of observations in a given area (e.g., Nipen et al., 2020).

6.4 Examination of the quality-controlled dataset

In this section we compare the QC-dataset with UKV model data, RWIS data, and MIDAS hourly data. Illustrative examples of the effect of sunny and rainy weather conditions on vehicle-based observations are presented in section 6.4.1, analysis of observation-minus-background (OMB) and observation-minus-observation (OMO) statistics are discussed in section 6.4.2, and vehicle-specific OMB statistics are examined in section 6.4.3. We use UKV model data as the background in the OMB datasets and MIDAS hourly data in the OMO dataset.

The effects of different meteorological factors are quantified through statistical analysis of OMB departures grouped by sunny, cloudy and rainy weather conditions and season. The sunny dataset will consist of observations that occur between 09:00 and 17:00 on days with at least 6 sunshine hours and less than 2mm of rainfall. Therefore, the observations are likely to be influenced by solar radiation incident on UK roads. The rainy dataset will consist of observations that occur between 09:00 and 17:00 on days with at least 5mm of rainfall and less than 2 hours of sunshine. The cloudy dataset will consist of observations that occur between 09:00 and 17:00 on days with less 2 hours sunshine and 2mm rainfall. We note that it is possible the model may not be cloudy at these observation locations. To obtain the weather-specific sub-datasets, we used the Met Office daily weather summaries (Met Office, 2018). The seasons we consider are Winter, defined as all data occurring between February 20th and March 20th 2018, and Spring, defined as all data occurring between March 21st and April 30th 2018. We note that these seasons do not conform to the usual definitions of meteorological winter and spring, but have been chosen due to the period of the Met Office trial and so that the Winter and Spring datasets each contain a similar number of observations.

6.4.1 Case studies on the effect of sunny and rainy weather conditions on vehicle-based observations

We now show three time-series of vehicle-based observations of air temperature, 10-minute UKV 1.5m-air-temperature and surface-air-temperature model data, and RWIS observations of air temperature. The routes traversed in each time-series began and ended in suburban areas and were predominantly on major roads and major routes in rural areas which occasionally crossed urban areas. The location of each time-series is shown in figure 6.2. We denote the time-series shown in figure 6.3 as S1, figure 6.4 as S2, and figure 6.5 as R1. S1 and S2 are illustrative examples of the effect of sunny weather and R1 is an illustrative example of the effect of rainy weather on vehicle-based observations of air temperature. We note that the same vehicle produced the observations in S1 and R1, but a different vehicle produced the observations in S2. We also note that the large data gaps in the three time series

are due to breaks in the journeys, and the two small data gaps in S2 are due to observations removed by the QC procedure.

For clarity, we will refer to the OMB data using 1.5m-air temperature for the background as aOMB and using surface-air temperature data for the background as sOMB. Furthermore, we denote the bias (mean) and standard deviation of an aOMB dataset as μ_a and σ_a respectively and the bias (mean) and standard deviation of a sOMB dataset as μ_s and σ_s respectively.

Figure 6.3 shows data from sunny weather conditions on March 25th 2018, including the time-series S1, UKV and RWIS data. The OMB summary statistics for S1 are shown in table 6.2. The sun rose at 06:52 and set at 18:22 on this day. The RWIS stations included in this time-series recorded a dry road-state when the vehicle passed the station. As a result of these conditions, we expect the sensible heat flux emitted by the road to be large and the road surface to have a noticeable heating effect on the air temperature above (see section 6.2.3). Additionally, we expect the surface-air temperature to be larger than the 1.5m-air temperature as the sensible heat flux emitted by the UKV surface will also be large. The mean difference between the interpolated RWIS observations and the nearest-in-time UKV model data reveals that RWIS observations are in most agreement with 1.5m-air temperature and in least agreement with surface-air temperature. There is a clear separation between UKV 1.5m-air temperature and surface-air temperature at the start of the time-series that gradually decreases as the net radiation absorbed by the UKV surface decreases. The vehicle-based observations generally lie between the model fields as seen by the difference in sign between the biases, μ_a and μ_s. The vehicle-based observations on average agree most with surface-air temperature as $|\mu_s| < |\mu_a|$. This is consistent with the height of the vehicle sensor which is likely to be between the model field heights of 0m and 1.5m but closer to 0m than 1.5m. Calculating the standard deviation of the sOMB and aOMB departures shows that the sOMB departures are more variable as $\sigma_a < \sigma_s$. We hypothesise that the variability of the UKV sensible heat flux induced by the sunny weather conditions is the mechanism responsible for the larger sOMB variability.

	Time-series		
Summary Statistics	S1	S2	R1
Number of observations	212	193	259
μ_a °C	1.44	0.02	0.65
σ_a °C	0.71	0.90	0.71
μ_s °C	−0.46	−0.27	0.37
σ_s °C	1.21	1.71	0.64

Table 6.2: Summary of the OMB statistics for the three time-series shown in figure 6.2 using UKV 1.5m-air temperature and surface-air temperature as the background. The uncertainty in the mean for each time-series is less than 0.1°C.

Figure 6.2: Map of the United Kingdom showing the location of the three time-series discussed in section 6.4.1. The red squares show the location of cities passed by or near to the routes travelled in the three time-series. The black diamonds show the location of the RWIS stations passed on each journey. The two orange lines correspond to the sunny weather time-series and the blue line corresponds to the rainy weather time-series. The time-series S1 began near Exeter and travelled north towards Manchester. The time-series R1 travelled the same initial route as S1, but headed east from Birmingham towards Cambridge. The time-series S2 began in Edinburgh and travelled along the coast to Newcastle-upon-Tyne and then the vehicle travelled further inland and south towards Nottingham.

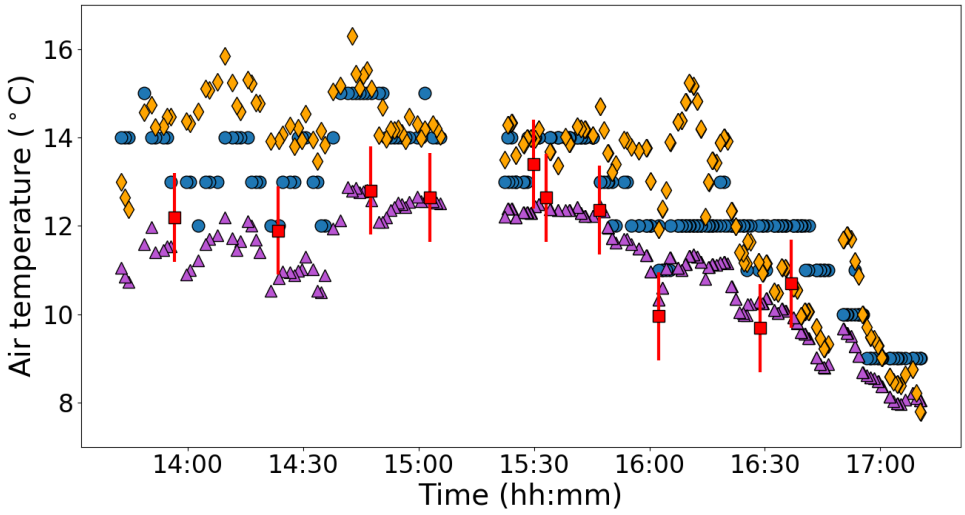


Figure 6.3: Time-series S1 of 212 vehicle-based observations of air temperature (blue circles) from a single vehicle driving along the M5 motorway on 25th March 2018 during sunny weather. Also shown are UKV 1.5m-air temperature (purple triangles) and UKV surface-air temperature (orange diamonds) linearly interpolated to the time and horizontal location of the vehicle observations, and RWIS observations of air temperature (red squares) linearly interpolated to the time the vehicle passed a station. The 1°C RWIS error bar represents the uncertainty used to assimilate RWIS observations into the UKV.

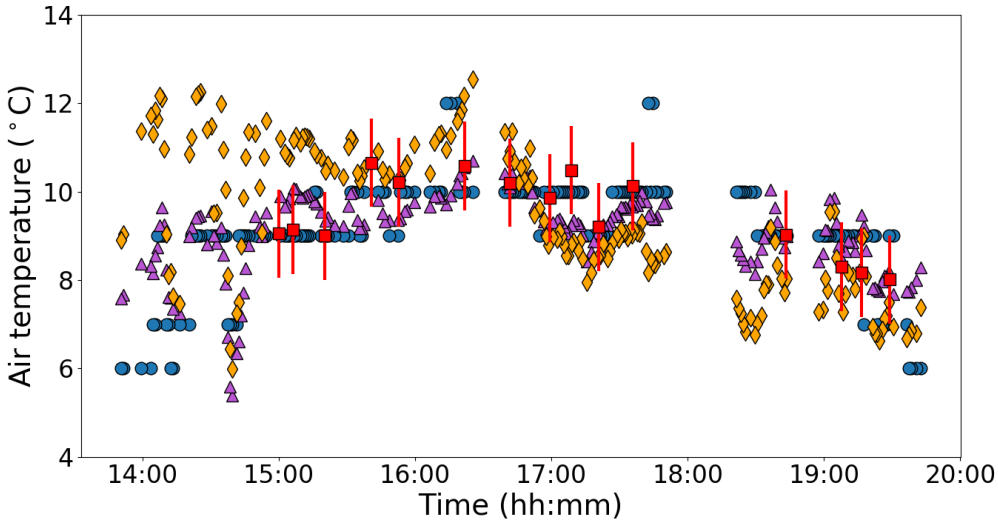


Figure 6.4: Time-series S2 of 193 vehicle-based observations of air temperature (blue circles) from a single vehicle driving along the A1 and the M1 motorway on April 5th 2018 during sunny weather. Also shown are UKV 1.5m-air temperature (purple triangles) and UKV surface-air temperature (orange diamonds) interpolated to the time and horizontal location of the vehicle observations, and RWIS observations of air temperature (red squares) interpolated to the time the vehicle passed a station. The 1°C RWIS error bar represents the uncertainty used to assimilate RWIS observations into the UKV.

Figure 6.4 shows data from sunny weather conditions on April 5th 2018, including the time-series S2, UKV and RWIS data. The OMB summary statistics for S2 are shown in table 6.2. The sun rose at 05:27 and set at 18:40 on this day. The RWISs included in this time-series recorded a dry road-state when the vehicle passed the station. Similarly to the data in figure 6.3, the RWIS observations are in most agreement with the UKV 1.5m-air temperature and in least agreement with surface-air temperature. The surface-air temperature is larger than the 1.5m-air temperature for the first half of this time-series. From approximately 17:00 we see that 1.5m-air temperature is greater than surface-air temperature. We hypothesise that this is due to the stabilisation of the boundary layer (e.g. Stull, 1988, pp 499–542). In contrast to S1, the vehicle-based observations are closest to UKV 1.5m-air temperature at the beginning of the time-series even though the sensible heat flux emitted from the road surface is expected to be greatest during this period. Possible reasons for this include cool breezes from the North Sea influencing the vehicle-based observations during the beginning of the time-series (see route map in figure 6.2) or because the air temperature is measured by a different vehicle's instrument. Furthermore, the difference between the biases μ_a and μ_s is large for S1 and small for S2. This, however, is likely due to the large number of observations that occurred during the evening for S2 when the temperature gradient between the surface and 1.5m is expected to be small. Considering the observations from the first 3 hours of S2 only, when the net solar radiation absorbed by the road and UKV surface is expected to be large, we find that the difference between the biases μ_a and μ_s is more profound. We note that the standard deviations σ_a and σ_s are larger for S2 than S1. This is likely due to the following two reasons. The first reason is the relatively long temporal length of the S2 time-series. The second reason is the possible transition to the nocturnal boundary layer as the dynamics induced by solar heating and the generation of convective plumes begins to cease and surface layer starts to become stably stratified. However, it is also plausible that the placement of the external air-temperature sensors on the two vehicles is contributing to this behaviour.

Figure 6.5 shows data from rainy weather conditions on March 30th 2018, including the time-series R1, UKV and RWIS data. The OMB summary statistics for R1 are shown in table 6.2. The RWIS stations included in this time-series recorded either a wet road-state or trace amounts of water at the road surface when the vehicle passed the station. As a result of these conditions, we expect the sensible heat emitted by the road to be small and the road surface to have a reduced effect on the air temperature above (see section 6.2.3). We note that the drop in air temperature between 12:30 and 13:30 is caused by an increase in altitude and an occluded front. The mean difference between the interpolated RWIS observations and the nearest-in-time UKV data reveals that RWIS observations are now in greater agreement with surface-air temperature than 1.5m-air temperature. The two UKV model fields are similar

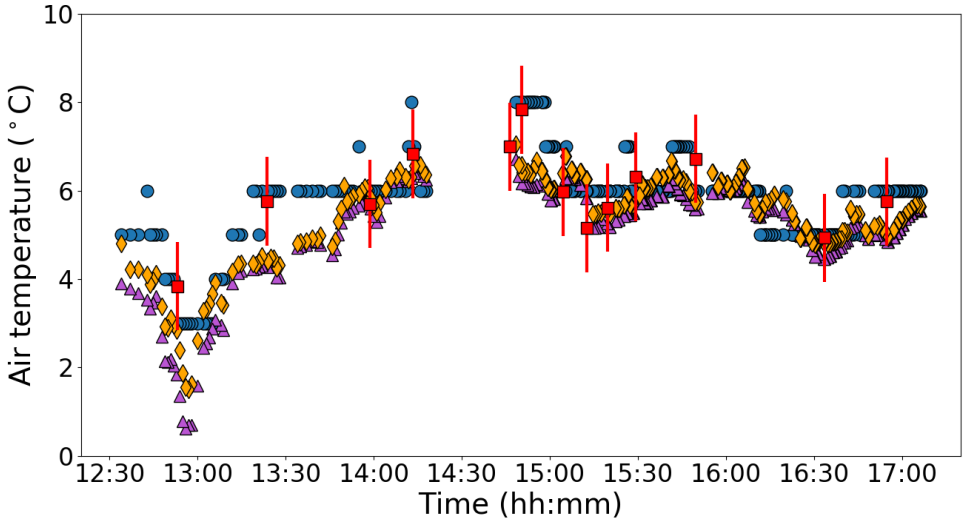


Figure 6.5: Time-series R1 of 259 vehicle-based observations of air temperature (blue circles) from a single vehicle driving along the M5 and M42 motorways and the A5 on 30th March 2018 during rainy weather conditions. Also shown are UKV 1.5m air temperature (purple triangles) and UKV surface air temperature (orange diamonds) interpolated to the time and horizontal location of the vehicle observations and RWIS observations of air temperature (red squares) interpolated to the time the vehicle passed a station. The 1°C RWIS error bar represents the uncertainty used to assimilate RWIS observations into the UKV.

throughout the time-series with multiple segments where the vehicle-based observations are greater than both fields. The vehicle-based observations are on average greater than the UKV model data as the biases $\mu_a, \mu_s > 0$°C, but agree more with surface-air temperature as $\mu_a > \mu_s$. This indicates that there are additional factors affecting the vehicle-based observations. Potential explanations for this behaviour are given in section 6.4.2.4. We note that while the aOMB departures are more variable than the sOMB departures (i.e. $\sigma_a > \sigma_s$), they are similar in size.

The effect of the sensible heat emitted by the road and UKV surfaces can be observed through comparison of the S1 and R1 time-series shown in figures 6.3 and 6.5, respectively. In sunny weather, the sensible heat emitted by the road and UKV surface will be large, resulting in a stronger vertical air-temperature gradient between the surface and the 1.5m height. In rainy weather, the sensible heat emitted by the road and UKV surface will be small, leading to a vertical air-temperature profile that is more uniform. Hence, the difference between the biases μ_a and μ_s will be larger in sunny weather conditions than rainy weather conditions. The OMB standard deviations calculated for each time-series show a negligible difference for σ_a and a noticeable difference for σ_s between the two time-series. For the sOMB standard deviation σ_s, we see that it is smaller for rainy weather and larger for sunny weather. This is likely because the variability of the sensible heat emitted by the UKV surface will be greater in sunny weather than rainy weather. However, there

may be other contributing factors such as the difference in observation operator error between the two time-series.

6.4.2 Statistical analysis of observation-minus-background and observation-minus-observation departures

In this section we investigate the uncertainty present in the QC-dataset through statistical analysis of OMB and OMO departures. The OMB datasets will be partitioned into weather-specific and seasonal sub-datasets so that we may examine how the OMB uncertainty changes with weather conditions and season. As there are only 347 observations within 1.5km of a MIDAS station, we will not split the OMO dataset into weather-specific and seasonal sub-datasets. We now discuss the characteristics of each dataset.

6.4.2.1 QC-dataset OMB and OMO statistics

The OMB and OMO statistics corresponding to the QC-dataset are given in table 6.3. Examining the OMB statistics shows that the vehicle-based observations are in poorer agreement with 1.5m-air temperature than surface-air temperature as the biases satisfy $|\mu_a| > |\mu_s|$. This is expected as external air-temperature sensors likely measure air temperature nearer to the surface than to a height of 1.5m. As $\mu_a > 0°\text{C}$ and $\mu_s > 0°\text{C}$, vehicle-based observations are on average warmer than both UKV model fields despite measuring the air temperature between them. Possible reasons for this behaviour are discussed in section 6.4.2.4. For the standard deviations, we have that $\sigma_a < \sigma_s$, showing that the sOMB dataset is more variable than the aOMB dataset. This is also visible in the aOMB and sOMB distributions shown by the histograms in figure 6.6a. While the distributions overlap substantially, a higher peak is seen for the aOMB distribution, whereas the sOMB distribution has a larger left tail. It is likely that the background uncertainty of surface-air temperature is greater than 1.5m-air temperature due to the simplifying assumptions made by the UKV in modelling the SEB of a grid box (see section 6.3.2.2).

Examining the OMO statistics reveals that vehicle-based observations are on average warmer than MIDAS observations. Comparing the OMO departure bias, μ_m, with the biases μ_a and μ_s obtained from the QC-dataset OMB departures, we find that $\mu_s < \mu_m < \mu_a$. Hence, the vehicle-based observations on average agree more with MIDAS data than 1.5m-air temperature but still agree most with surface-air temperature. While it is plausible that vehicle-based observations will generally agree more with MIDAS data than the UKV 1.5m-air temperature model data, we note the following two issues with these calculations. Firstly, the MIDAS data were not interpolated to the location of the vehicle-based observations and there will likely be differences in elevation between the two. As air temperature is expected to change

with elevation in the surface layer (e.g. Stull, 1988, pp 9–19), the omission of any vertical interpolation may cause a bias. Secondly, while the variability of the OMO dataset is less than the variability of any OMB dataset shown in table 6.3, it is calculated with far fewer observations making the statistic less reliable.

6.4.2.2 Weather-specific OMB datasets

The OMB statistics for each weather-specific dataset are given in table 6.3. For all weather types, the bias μ_a is positive showing that vehicle-based observations are on average warmer than UKV 1.5m-air temperature regardless of weather conditions. For the sunny and cloudy datasets, the bias μ_s is negative, whereas for the rainy dataset μ_s is positive. This agrees with the results of the three time-series discussed in section 6.4.1. This also suggests that the vehicle-based observations studied in this paper may be colder on average than UKV surface-air temperature in dry conditions. The smallest differences between the biases μ_a and μ_s occurs for the rainy dataset while the largest difference occurs for the sunny dataset which is also seen in the S1 and R1 time-series discussed in section 6.4.1. For the rainy dataset biases, we have that $\mu_a > 0°\text{C}$ and $\mu_s > 0°\text{C}$ which indicates the vehicle-based observations are on average warmer than the two UKV model fields. This suggests that there are other influencing factors on the vehicle-based observations that are not represented in the UKV as vehicles measure the air temperature between these two heights. Potential explanations for this behaviour are discussed in section 6.4.2.4.

Inspection of the weather-specific standard deviations reveals that, for sunny and cloudy weather conditions, σ_s is noticeably larger than σ_a. This difference in variability is shown in the histograms for the sunny dataset (figure 6.6b) and the cloudy dataset (figure 6.6c). The sunny aOMB distribution is unimodal and the sunny sOMB distribution is bimodal. The bimodal structure may be due to intermittent cloud cover on the days with less sunshine hours or the relatively small size of the sunny dataset. Similarly to the QC-dataset, the cloudy aOMB and sOMB distributions overlap substantially, but a higher peak is seen for the aOMB distribution and a larger left tail is seen for the sOMB distribution. For rainy weather conditions, the standard deviations σ_a and σ_s are similar leading to similar aOMB and sOMB distributions as shown in figure 6.6d.

Overall, the variability for the weather-specific datasets agrees with the variability calculated for the time-series discussed in section 6.4.1. We also find that the standard deviation σ_s is larger for the sunny sOMB dataset than for the cloudy sOMB dataset. This behaviour is likely the result of the increased variability of the sensible heat emitted by roads and the UKV surface during sunny weather conditions due to the larger amount of solar radiation absorbed by the two surfaces. This also suggests that the uncertainty of vehicle-based observations may be greatest in sunny

weather conditions due to the combination of the radiative effects on the vehicle sensor and the representation uncertainty. Conversely, the aOMB standard deviation σ_a is largest for the cloudy dataset and not the sunny dataset. This, however, may be due to changes in the sky-view factor due to variable cloud cover or because the cloudy dataset has more observations than the sunny dataset. We hypothesise that the rainy dataset standard deviations σ_s and σ_a are similar for the following three reasons. Rain increases the availability of water at the UKV surface which reduces the emitted sensible heat flux and hence the vertical air-temperature profile will be more uniform. There is also little to no sun at the times and locations of the vehicle-based observations in the rainy dataset resulting in negligible radiation reflected by the road surface incident on the vehicle temperature sensor. Finally, the rainy dataset is the smallest of our three weather-specific datasets and so the OMB statistics are the least robust.

6.4.2.3 Seasonal OMB datasets

The OMB statistics for the seasonal datasets are given in table 6.3 and the histograms are plotted in figures 6.6e (Spring) and 6.6f (Winter). We include information on the seasonal datasets to provide a baseline for the vehicle-specific analysis in section 6.4.3. For the seasonal datasets, the vehicle-based observations are on average greater than the UKV model fields except for surface-air temperature in Spring where they are approximately the same. Comparing the biases of the seasonal datasets we see that μ_s is smaller and μ_a is greater in Spring than in Winter. Inspection of the standard deviations of the seasonal datasets reveals that σ_s is larger than σ_a in both Winter and Spring. Comparing the seasonal OMB statistics, we find that the sOMB standard deviation σ_s is larger for the Spring dataset than for the Winter dataset. This agrees with the results of the weather-specific datasets as the Spring dataset contains more sunny days than the Winter dataset.

6.4.2.4 Discussion of the uncertainty exhibited by the OMB datasets

In this section we discuss several possible contributions to the uncertainty exhibited in the OMB statistics shown in table 6.3.

- The vehicle-based observations of air temperature are precise to 1°C. The details of the observation processing by the OBD system and app are unknown. However, an indication of the expected size of the processing errors can be obtained by considering the quantization error from a typical procedure that rounds to the nearest integer. In this case, the root-mean-squared quantization error is $\sqrt{1/12}$ °C and may be a positive or negative error (Widrow et al., 1996).
- As discussed in section 6.2.2, the external air-temperature sensors may exhibit

OMB and OMO statistics				
Dataset	Number of observations	Departure	Mean (°C)	Standard deviation (°C)
QC-dataset	17425	aOMB	0.67	1.24
		sOMB	0.12	1.59
	347	OMO	0.29	0.95
Sunny	1878	aOMB	1.05	1.15
		sOMB	−0.77	1.86
Cloudy	2366	aOMB	0.84	1.41
		sOMB	−0.14	1.74
Rainy	840	aOMB	0.56	0.97
		sOMB	0.30	1.00
Winter	7798	aOMB	0.48	1.29
		sOMB	0.26	1.55
Spring	9627	aOMB	0.82	1.17
		sOMB	0.01	1.61

Table 6.3: Summary of the OMB and OMO departure statistics for each dataset. The uncertainty in the mean for each dataset is less than 0.1°C.

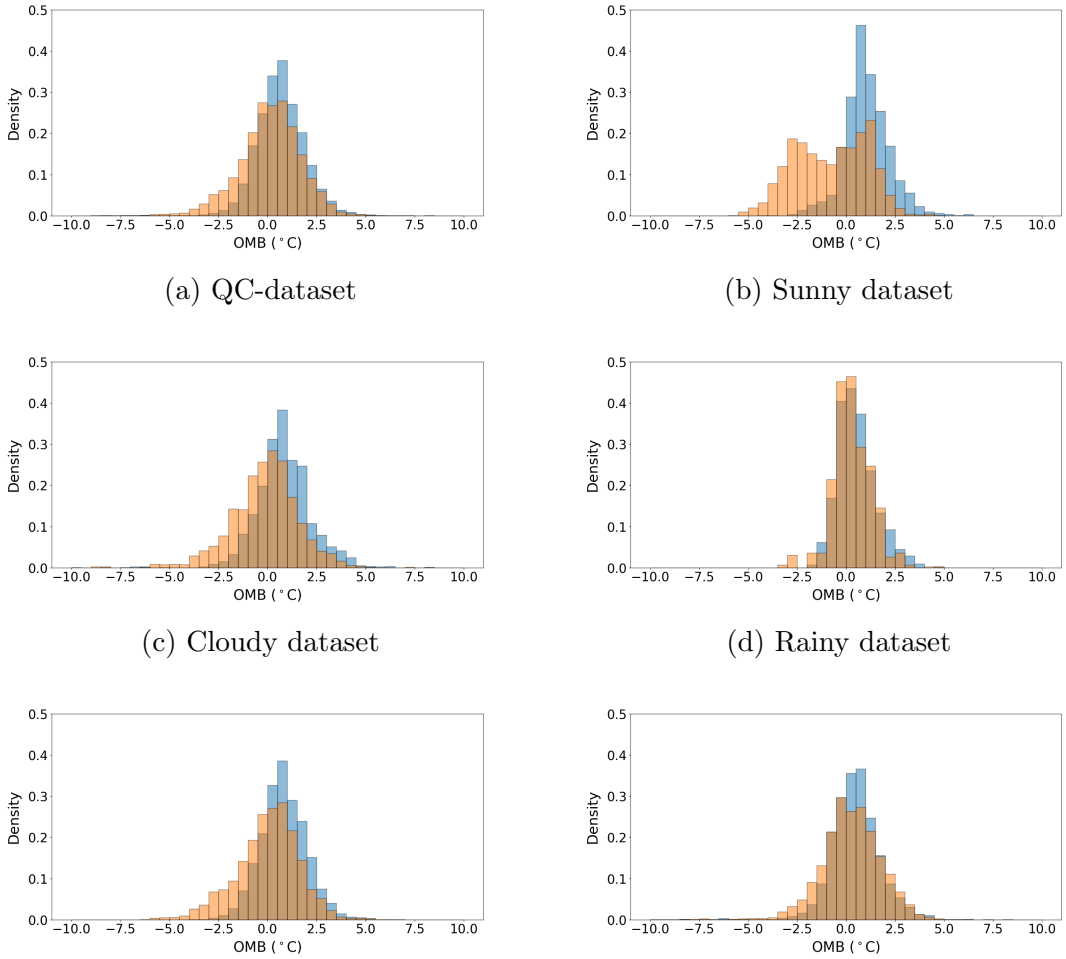
a warm bias from extraneous sources. For instance, the sensor may be in close proximity to the vehicle engine or the location of the sensor may be inadequate for sensor ventilation. A rough estimate of the bias due to the vicinity of the engine is 5°C - 25°C (Mercelis et al., 2021). However there are several other factors that may be influencing this estimate (e.g. differing vehicles).

- In unstable atmospheric conditions, the vertical air-temperature gradient between the surface and the 1.5m height will be negative. When the sensible heat emitted by the road and UKV surfaces is large, the vertical air-temperature gradient will be large. Therefore, surface-air temperature will be warmer on average than vehicle-based observations which likely measure air temperature between 20cm and 100cm above the road surface. Similarly, 1.5m-air temperature will be cooler on average than vehicle-based observations.

- The vehicle-based observations have not been corrected to the elevation of the model grid box. In NWP, it is common to correct surface observations using a standard adiabatic lapse rate of 0.0065°C m^{-1} (Dutra et al., 2020; Cosgrove et al., 2003).

- The road surface temperature is highly variable due to sky-view factors and traffic (Chapman and Thornes, 2011). As noted in section 6.2.3, these factors can change the local temperature by as much as 2 or 3°C.

- The difference in sensible heat emitted by the road surface and the UKV surface may contribute to the OMB biases. As discussed in section 6.3.2.2, these model errors will vary depending on the land-surface (e.g. urban/rural) but are difficult to quantify.



(a) QC-dataset

(b) Sunny dataset

(c) Cloudy dataset

(d) Rainy dataset

(e) Spring dataset

(f) Winter dataset

Figure 6.6: OMB histograms datasets corresponding to the datasets in table 6.3. Bins of width 0.5°C have been used for each histogram. The blue bars correspond to the aOMB bins and the orange bars correspond to the sOMB bins.

6.4.3 Vehicle specific observation-minus-background departure distributions

Throughout the Met Office trial, 31 vehicles were used to produce vehicle-based observations. To investigate whether error statistics differ with vehicle, we plot OMB histograms for each of the 12 vehicles with the most observations between 09:00 and 17:00 in the QC-dataset in figure 6.7. We use only observations between 09:00 and 17:00 so that the boundary layer is likely to be unstable (i.e. UKV surface-air temperature is greater than the 1.5m-air temperature). This is to avoid the complications of interpreting the OMB statistics experienced with the S2 time-series discussed in section 6.4.1 and so that more observations can be classified into the weather-specific data types discussed in section 6.4.2. The OMB statistics are summarised in table 6.4. Also included in table 6.4 is the percentage of observations for each vehicle occurring in each weather-specific and seasonal dataset discussed in section 6.4.2. We note that it is difficult to draw definitive conclusions in this examination for two reasons. Firstly, many of the vehicles experience different weather conditions and there are many observations which we are unable to classify into a weather type. Secondly, there are only 12 vehicles with an acceptable number of observations that we can examine.

The majority of histograms shown in figure 6.7 resemble normal distributions. The aOMB and sOMB distributions of vehicles (ii), (iii), (iv), (x), and (xii) are qualitatively similar with the visual distinction between them a result of the difference in means. The remaining vehicles have noticeably different aOMB and sOMB distributions.

Examining the biases of the vehicle-specific OMB distributions shows that there are some vehicles which agree more with UKV 1.5m air temperature than surface air temperature as $|\mu_a| < |\mu_s|$. For these vehicles it is possible that the external air temperature sensor is located closer to 1.5m height than the road surface or there are additional unknown factors affecting the vehicle-based observations as in S2 discussed in section 6.4.1. The values of the biases μ_a and μ_s also vary substantially between vehicles. The sign of μ_s also varies with vehicle whereas only vehicle (x) has negative μ_a which suggests that some element of the vehicle's external air-temperature sensor or processing procedure may have a cold bias.

Except for vehicle (xii), we find that the aOMB dataset is less variable than the sOMB dataset (i.e. $\sigma_a < \sigma_s$) for all vehicles which agrees with the results obtained in sections 6.4.1 and 6.4.2. The values of the standard deviations σ_a and σ_s also vary substantially between vehicles. When stratifying the vehicles by seasonal contribution, we are able to find some agreement between the OMB dataset variability. This can be seen between the groups of vehicles (ii) and (vii), vehicles (iv), (viii) and (x), and vehicles (i) and (xi). Each vehicle in these groups contains similar ratios of

Winter to Spring observations. For vehicles (iii), (vi), (ix) and (xii), which contain predominantly Spring observations, we see that the aOMB standard deviation σ_a is similar except for vehicle (xii) and the sOMB standard deviation σ_s is similar except for vehicle (iii). This shows that vehicles may have similar OMB uncertainty if they have similar ratios of seasonal observations.

6.5 Conclusion

In this work we have examined a novel low-precision vehicle-based observation dataset obtained from a Met Office proof-of-concept trial. An overview of the quality-control (QC) applied to the vehicle-based observations was given. The data that passed QC were examined and compared with UKV 1.5m-air-temperature and surface-air-temperature model data, roadside weather information station (RWIS) data and hourly Met Office integrated data archive system (MIDAS) data. Using these datasets, we explored the characteristics of the vehicle-based observations that passed QC.

The QC procedure consisted of four tests which assessed different aspects of the vehicle-based observations. Reports that did not have an observation of air temperature or the necessary metadata to be tested were removed prior to the QC procedure. The climatological range test (CRT), stuck instrument test (SIT) and global positioning system (GPS) test were applied in parallel. Both the SIT and GPS test required vehicle identification, which may be unavailable in other crowd-sourced and opportunistic datasets due to privacy concerns. While the majority of observations passed the CRT and SIT, a substantial number of observations were flagged by the GPS test due to unsuitable GPS update settings on the smartphone application and poor GPS signal. The observations that passed these three QC tests were put through a final sensor ventilation test (SVT) which flagged observations from vehicles driving below a predetermined sensor ventilation threshold speed. The SVT flagged a sizable amount of data relative to the amount tested. The final quality-controlled dataset (QC-dataset) consisted of 25.6% of the observations obtained from the Met Office trial.

Using the QC-dataset, we investigated the uncertainty present in vehicle-based observations by analysing two observation-minus-background datasets. One dataset used UKV 1.5m-air temperature as the background (aOMB) and the other dataset that used UKV surface-air temperature as the background (sOMB). Examining the OMB statistics of the QC-dataset we found that the vehicle-based observations of air temperature were on average greater than the UKV model data and agreed more with surface-air temperature than 1.5m-air temperature. This is expected as

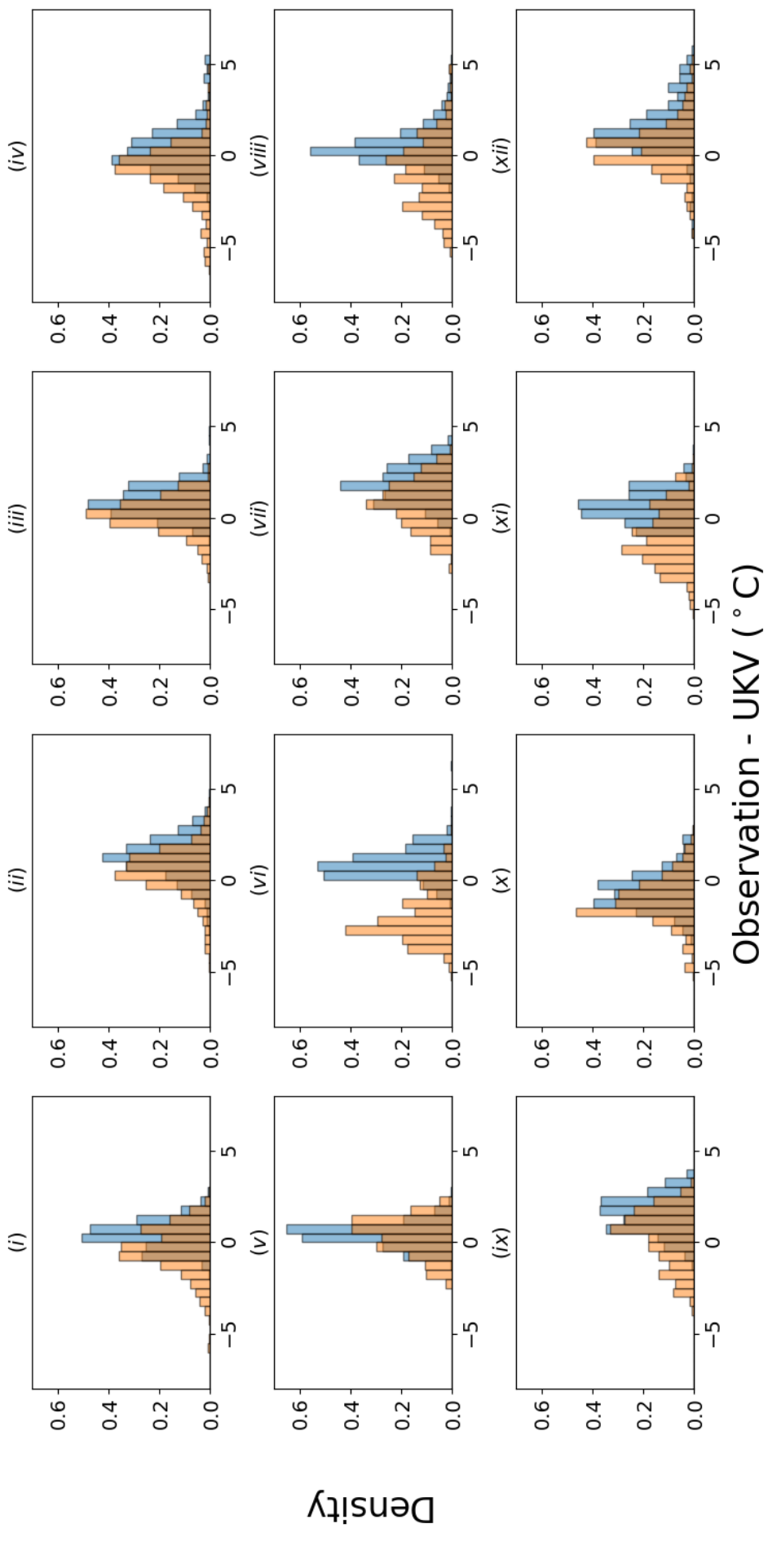


Figure 6.7: OMB histograms for the 12 vehicles with the most observations between 9am and 5pm in the QC-dataset. The aOMB distributions are shown by the blue bins, the sOMB distributions are shown by the orange bins, and the overlap in distributions is shown by the composite. The same bins of width 0.5°C have been used for each histogram. Information on the statistics and observation meteorological conditions for each vehicle is given in table 6.4.

	Vehicle											
Summary statistics	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)	(ix)	(x)	(xi)	(xii)
Number of observations	1533	1522	910	904	780	593	562	502	497	421	342	212
μ_a °C	0.43	1.24	0.89	0.40	0.42	0.91	1.80	0.59	1.51	−0.56	0.61	1.56
σ_a °C	0.90	1.06	0.92	1.14	0.79	0.91	1.00	1.03	1.00	1.04	0.91	1.23
μ_s °C	−0.42	0.44	0.14	−0.84	0.36	−1.87	0.78	−0.94	0.29	−1.17	−1.03	0.40
σ_s °C	1.17	1.19	0.98	1.25	1.02	1.21	1.15	1.36	1.22	1.14	1.26	1.20
Sunny percentage	0%	23%	22%	3%	34%	64%	13%	9%	36%	16%	22%	43%
Cloudy percentage	25%	23%	44%	18%	7%	10%	31%	26%	12%	47%	35%	38%
Rainy percentage	3%	10%	27%	20%	0%	0%	6%	11%	10%	6%	9%	0%
Unclassified percentage	72%	44%	7%	59%	59%	26%	54%	54%	42%	30%	34%	19%
Winter percentage	63%	41%	1%	47%	98%	10%	28%	47%	19%	47%	68%	0%
Spring percentage	37%	59%	99%	53%	2%	90%	72%	53%	81%	53%	32%	100%

Table 6.4: Summary of the OMB statistics for each vehicle in figure 6.7. The uncertainty in the mean is at most 0.1°C for each vehicle OMB dataset. Also shown is the percentage of observations from a specific vehicle that appear in the weather-specific and seasonal datasets discussed in section 6.4.2. Observations which do not appear in any of the weather-specific datasets are accounted for in the unclassified field on this table.

the vehicle-based observations likely measure air temperature nearer to the surface than to a height of 1.5m. However, there are several possible contributing factors to this result such as the quantization error, the difference in height between the vehicle-based observations and the UKV grid box, or heat from the vehicle engine contaminating the vehicle-based observations. We also found that the sOMB uncertainty is greater than the aOMB uncertainty for the QC-dataset. This is likely because the UKV surface-energy balance has a much stronger influence on surface-air temperature than 1.5m-air temperature.

To examine how the vehicle-based observation uncertainty changes with weather conditions, we grouped the OMB datasets by sunny, cloudy and rainy weather. For the sunny and cloudy OMB datasets, UKV surface-air temperature was on average greater than the vehicle-based observations with the magnitude of the sOMB bias greatest for the sunny dataset when the vertical air-temperature gradient near the surface is large. However, as shown in an illustrative time-series example, vehicle-based observations can be in greater agreement with 1.5m-air temperature than surface-air temperature in sunny weather conditions. Possible explanations for this include the vehicle location (e.g., sea breeze effects) or the placement of the sensor used by the vehicle. For the rainy OMB dataset, when the vertical air-temperature gradient near the surface is small, we found that the vehicle-based observations were on average warmer than both UKV model fields. Inspecting the variability of the OMB datasets, we found that the sOMB variability is greatest in sunny weather conditions and smallest in rainy weather conditions. The aOMB variability was larger for the cloudy dataset than for the sunny dataset but still smallest for the rainy dataset. The large aOMB variability for the cloudy dataset may be due to changes in the sky-view factor caused by variable cloud cover or because it was the largest of the weather-specific datasets. These results strongly suggest that the uncertainty of vehicle-based observations of air temperature is weather-dependent. In particular, the uncertainty of vehicle-based observations will be largest in sunny weather conditions and smallest in rainy weather conditions.

To determine the effect of different vehicles on vehicle-based observations, the OMB datasets were grouped by vehicle. Due to the large number of observations with unclassified weather conditions and the different proportions of weather conditions experienced by vehicles, we were unable to distinguish between meteorological and vehicle effects on the vehicle-specific OMB distributions effectively. However, we note that vehicles with similar proportions of seasonal data may exhibit similar OMB variability. In order to determine the influence of the vehicle on the OMB statistics, further investigations into vehicle-based observations using a larger dataset must be conducted.

Vehicle-based observations are a potentially abundant source of low-cost, high-

resolution meteorological information. However, there are several issues regarding data collection and privacy in addition to the uncertainty characteristics which must be addressed before they may be utilised. We first note that an increase in precision of the vehicle-based observations will allow improved QC, understanding of the characteristics of the data, and value for NWP. For QC, an alternative to vehicle identification must be used such that privacy concerns are mitigated. For assimilation, the uncertainty and bias associated with each observation must be sufficiently evaluated. This will require further trials to assess the effect of local meteorological conditions and the vehicle sensing instrument. Despite these issues, vehicle-based observations are a promising opportunistic dataset for convection-permitting data assimilation.

6.6 Chapter summary

In this chapter, we addressed the second research question given in chapter 1: *what are the characteristics of vehicle-based observations of air temperature?* To investigate the characteristics of vehicle-based observations, we compared them with convection-permitting NWP model data, roadside weather information station observations and Met Office integrated data archive system observations. We have shown that the vehicle-based observation–model comparisons are most variable in sunny conditions and least variable in rainy conditions. Hence, the uncertainty of vehicle-based observation–model comparisons is likely weather-dependent. In addition, the statistics of vehicle-based observation–model comparisons changes when grouped by vehicle. As the vehicle-specific datasets are small in size, we conclude that the uncertainty of vehicle-based observation–model comparisons is possibly vehicle-dependent.

To correctly assimilate vehicle-based observations, the uncertainty due to unresolved scales must be accounted for in the assimilation system. Having established and gained some experience of the error due to unresolved scales with a real dataset, in the next chapter we focus on mathematical methods to account for uncertainty due to unresolved scales in data assimilation.

Chapter 7

Accounting for observation uncertainty and bias due to unresolved scales with the Schmidt-Kalman filter

The second aim of this thesis is to examine how observation uncertainty due to unresolved scales is accounted for in data assimilation. In this chapter we answer the third research question given in chapter 1: *can the Schmidt-Kalman filter effectively treat observation error and bias due to unresolved scales?* In particular, we wish to determine:

- Which regimes of observation uncertainty is the Schmidt-Kalman filter most suitable for?
- How can biases due to unresolved scales be treated with the Schmidt-Kalman filter?

The remainder of this chapter, except for the chapter summary in section 7.11, is strongly based on the paper Bell et al. (2020).

Abstract:

Data assimilation combines observations with numerical model data, to provide a best estimate of a real system. Errors due to unresolved scales arise when there is a spatio-temporal scale mismatch between the processes resolved by the observations and model. We present theory on error, uncertainty and bias due to unresolved scales for situations where observations contain information on smaller scales than can be represented by the numerical model. The Schmidt-Kalman filter, which accounts for the uncertainties in the unrepresented processes, is investigated and

compared with an optimal Kalman filter that treats all scales, and a suboptimal Kalman filter that accounts for the large-scales only. The equation governing true analysis uncertainty is reformulated to include representation uncertainty for each filter. We apply the filters to a random walk model with one variable for large-scale processes and one variable for small-scale processes. Our new results show that the Schmidt-Kalman filter has the largest benefit over a suboptimal filter in regimes of high representation uncertainty and low instrument uncertainty but performs worse than the optimal filter. Furthermore, we review existing theory showing that errors due to unresolved scales often result in representation error bias. We derive a novel bias-correcting form of the Schmidt-Kalman filter and apply it to the random walk model with biased observations. We show that the bias-correcting Schmidt-Kalman filter successfully compensates for representation error biases. Indeed, it is more important to treat an observation bias than an unbiased error due to unresolved scales.

7.1 Introduction

In atmospheric data assimilation, observations are combined with numerical model data, weighted by their respective error statistics, to provide a best estimate of the current atmospheric state, known as the *analysis*. This is achieved through comparison of observations with the numerical model equivalent of those observations. The errors associated with the observation-model comparison are the instrument error and representation error (Janjić et al., 2018). The representation error consists of the pre-processing error, the observation operator error and the *error due to unresolved scales* that occurs when there is a mismatch between the numerical model resolution and the scales resolved by the observation. The error due to unresolved scales depends on the observation footprint, which could be smaller or larger than the model grid, depending on the observation type and choice of model. For models which contain information on scales smaller than those observed, the standard approach to account for scale-mismatch would be to average the model state over the observation area (Janjić et al., 2018). However, for the purposes of this paper, we focus only on situations where the observation information content includes smaller scales than can be resolved by the model. In order to obtain the best analysis from these observations the representation error must be treated correctly by the data assimilation system.

Methods of accounting for uncertainty due to unresolved scales include, for example, prediction through ensemble statistics (Karspeck, 2016; Satterfield et al., 2017) and the use of a stochastic superparameterization (Grooms et al., 2014). In this manuscript we will consider two approaches: the standard approach where the uncertainty due to unresolved scales is included in the observation error covariance

matrix (e.g., Hodyss and Satterfield, 2016; Fielding and Stiller, 2019) and an alternative approach where unresolved processes are considered in state space and hence accounted for through the state error covariance (Janjić and Cohn, 2006).

Compensating for representation error through the standard approach involves using an observation error covariance matrix that takes into account both the instrument and representation uncertainty. This can then be used within a standard variational or sequential data assimilation scheme. Estimates of the observation uncertainty may be obtained using a statistical method, to estimate the entire observation error covariance matrix (e.g., Desroziers et al., 2005; Stewart et al., 2014; Waller et al., 2016c,a; Cordoba et al., 2017). Alternatively each component of the representation error statistics can be estimated separately and then combined with the instrument error covariance. For example the error due to unresolved scales may be approximated using high resolution observations (Oke and Sakov, 2008) or high resolution model data (Daley, 1993; Liu and Rabier, 2002; Waller et al., 2014b; Schutgens et al., 2016).

The Schmidt-Kalman filter (SKF) (Schmidt, 1966) is an example of a filter which uses the statistics of the unresolved processes in state space, without ever evaluating the unresolved state itself, to compensate for the error due to unresolved scales. This approach allows for consideration of flow-dependent correlations between the resolved errors and the unresolved processes at the cost of additional assumptions, approximations and increased computational expense. Janjić and Cohn (2006) have shown that the SKF can produce positive results despite the approximations and assumptions required for implementation in a geophysical context. In this paper we provide new results that determine in which observation and model uncertainty regimes the SKF performs best. In addition we compare the SKF to two other Kalman filtering approaches.

The SKF is deemed a suboptimal filter as it does not minimise the mean-square-error of its estimated states (Janjić and Cohn, 2006). In contrast, the Kalman filter that treats all scales is deemed optimal (for linear models and Gaussian statistics) (Nichols, 2010). In practice, suboptimal filters that do not treat all scales are often used. The analysis error covariances propagated by suboptimal filters are not representative of the true error statistics due to omitted or incorrectly specified filter components. As such, the true analysis error equations have been derived to evaluate the performance of suboptimal filters (e.g., Brown and Sage, 1971; Asher and Reeves, 1975; Asher et al., 1976). In this article we reformulate previous theory on true analysis error equations to include representation error (section 7.4) and evaluate the performance of the SKF.

A further issue noted by Janjić and Cohn (2006) is the potential for the representation error to be biased. This is because the error due to unresolved scales is

sequentially correlated in time and correlated with the state resolved by the model. Other authors have circumvented this bias by careful construction of their numerical model (Janjić and Cohn, 2006). However, in operational data assimilation, most observations are biased and the innovations need to be corrected or the bias accounted for within the assimilation (Dee, 2005). Bias correction can be incorporated into the data assimilation algorithm by augmenting the state vector with a bias term (Friedland, 1969; Jazwinski, 1970; Ignagni, 1981) which can be estimated along with the state variables. This method of bias correction is commonly used with variational data assimilation systems (e.g., Derber and Wu, 1998; Dee, 2004; Zhu et al., 2014; Eyre, 2016) but has also been applied with ensemble data assimilation systems (e.g., Fertig et al., 2009; Miyoshi et al., 2010; Aravéquia et al., 2011). To the best of our knowledge a bias correction scheme has yet to be implemented in conjunction with the SKF; in section 7.7 we introduce a bias-correcting SKF as a new method to compensate for biases due to unresolved scales.

In summary, the objective of this paper is to investigate under which model and observation error regimes the SKF is most effective. The theoretical aspects of representation error will be reviewed in section 7.2 with particular emphasis on the error due to unresolved scales. Section 7.3 details how the SKF can be used to account for error due to unresolved scales and introduces the optimal Kalman filter (OKF) and a reduced-state Kalman filter (RKF). In section 7.4 we state the standard true analysis error equation and reformulate it to include representation error for each filter.

To evaluate the performance of the SKF in a numerical example we use a Gaussian random walk model. The numerical experiment methodology and model formulation are described in section 7.5 and results are presented in section 7.6. Our results show that the SKF provides the largest improvement in performance compared with the RKF when there is large error variance due to unresolved scales and small instrument error variance. In section 7.7 we discuss observation bias correction schemes in sequential data assimilation and introduce a novel SKF with bias correction scheme. The methodology and model formulation for the numerical experiments with biased observations is discussed in section 7.8 and results are presented in section 7.9. Our results show the SKF with bias correction can simultaneously treat observation biases and compensate for the error due to unresolved scales. We summarise and draw conclusions from our results in section 7.10.

7.2 Theoretical framework

In this section we introduce a theoretical framework and the notation used in this paper. We begin by describing a numerical model (section 7.2.1) and observations (section 7.2.2). In data assimilation, the error statistics used in filters may not reflect

the true uncertainties they are intended to model. To help distinguish between these two sets of statistics throughout this manuscript we will define any true error statistics with a tilde ($\sim$). Error statistics used in or obtained from filter calculations will be referred to as *perceived* error statistics and have no tilde.

The mathematical framework used to examine the error due to unresolved scales in this manuscript is to estimate the projection of some state from a high, but finite dimensional real vector space, onto a lower dimensional subspace using observations and knowledge of the system dynamics, following a similar philosophy to Liu and Rabier (2002) and Waller et al. (2014b). Our approach differs from that of Janjić and Cohn (2006) which begins from the standpoint of infinite dimensional function spaces.

7.2.1 Model configuration

In this section we introduce the perfect and forecast models. We assume that the phase-space for the large-scale dynamics is a subspace of the phase-space for the full high dimensional system. The complement of the subspace for the large-scales will correspond to the phase-space for the small-scale dynamics. The notation for the models will be in a partitioned form that separates the large and small scales. In particular, we denote the true, complete state at time t_k as $\left((\boldsymbol{x}^{l,t})^T \;\; (\boldsymbol{x}^{s,t})^T\right)^T \in \mathbb{R}^{N_t}$ such that $\boldsymbol{x}^{l,t} \in \mathbb{R}^{N_l}$, $\boldsymbol{x}^{s,t} \in \mathbb{R}^{N_s}$ and $N_t = N_l + N_s$. Here, and throughout this paper, any component with a t-superscript indicates that it is a true variable. The l- and s-superscripts correspond to the large- and small-scale processes within the complete system dynamics. (We have deviated from the resolved/unresolved nomenclature of Janjic and Cohn 2006 for clarity, since the different filters used in our experiments resolve different scales).

An ideal linear model for the true state of a finite dimensional process can be expressed through the dynamical system

$$\begin{pmatrix} \boldsymbol{x}^{l,t} \\ \boldsymbol{x}^{s,t} \end{pmatrix}_k = \begin{pmatrix} \mathbf{M}^{l,t} & \mathbf{M}^{ls,t} \\ \mathbf{M}^{sl,t} & \mathbf{M}^{s,t} \end{pmatrix} \begin{pmatrix} \boldsymbol{x}^{l,t} \\ \boldsymbol{x}^{s,t} \end{pmatrix}_{k-1}, \tag{7.2.1}$$

such that the matrix blocks $\mathbf{M}^{l,t} \in \mathbb{R}^{N_l \times N_l}$, $\mathbf{M}^{ls,t} \in \mathbb{R}^{N_l \times N_s}$, $\mathbf{M}^{s,t} \in \mathbb{R}^{N_s \times N_s}$ and $\mathbf{M}^{sl,t} \in \mathbb{R}^{N_s \times N_l}$. From a numerical modelling perspective, this partitioned description of the dynamics would be suited to a pseudospectral discretization (e.g. Fourier modes). An illustrative example is given in figure 7.1.

In numerical weather prediction (NWP), the true models that govern the evolution of the atmosphere are unknown and have to be approximated. For our approximation of the true dynamical system (7.2.1), we assume that any subgrid-scale parameterizations used to approximate the contribution from the small-scale processes to the

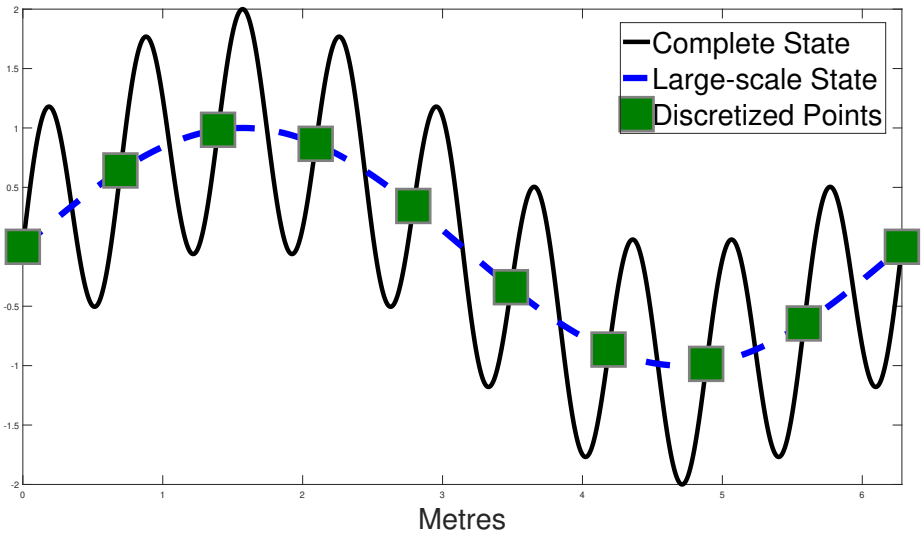


Figure 7.1: Illustrative example of the partitioned formulation of the dynamics. The complete state (solid black line) is the sum of the large-scale state (blue dashed line) and a higher frequency wave corresponding to the small-scale state. At the points of discretization (green squares), the complete state and the large-scale state are equal.

large-scale state are contained within the large-scale model (Janjić and Cohn, 2006; Janjić et al., 2018). Hence, the model block $\mathbf{M}^{ls} = \mathbf{0}_{N_l \times N_s}$ and our approximate dynamical model describing the complete system satisfies

$$\begin{pmatrix} \boldsymbol{x}^{l,t} \\ \boldsymbol{x}^{s,t} \end{pmatrix}_k = \begin{pmatrix} \mathbf{M}^{l} & \mathbf{0}_{N_l \times N_s} \\ \mathbf{M}^{sl} & \mathbf{M}^{s} \end{pmatrix} \begin{pmatrix} \boldsymbol{x}^{l,t} \\ \boldsymbol{x}^{s,t} \end{pmatrix}_{k-1} - \begin{pmatrix} \boldsymbol{\eta}^{l} \\ \boldsymbol{\eta}^{s} \end{pmatrix}_k . \tag{7.2.2}$$

In (7.2.2) each model block has the same dimensions as its true model counterpart. The large- and small-scale model errors are given by $\boldsymbol{\eta}^l \in \mathbb{R}^{N_l}$ and $\boldsymbol{\eta}^s \in \mathbb{R}^{N_s}$ respectively. Model errors are assumed to be random and unbiased with covariance given by

$$\widetilde{\mathbf{Q}} = \begin{pmatrix} \widetilde{\mathbf{Q}}^{ll} & \widetilde{\mathbf{Q}}^{ls} \\ \widetilde{\mathbf{Q}}^{sl} & \widetilde{\mathbf{Q}}^{ss} \end{pmatrix} . \tag{7.2.3}$$

Here, using $\langle \cdot \rangle$ to indicate the mathematical expectation over the corresponding error distribution, the matrices $\widetilde{\mathbf{Q}}^{ll} \equiv \langle \boldsymbol{\eta}^l (\boldsymbol{\eta}^l)^T \rangle \in \mathbb{R}^{N_l \times N_l}$, $\widetilde{\mathbf{Q}}^{ss} \equiv \langle \boldsymbol{\eta}^s (\boldsymbol{\eta}^s)^T \rangle \in \mathbb{R}^{N_s \times N_s}$ and $\widetilde{\mathbf{Q}}^{ls} \equiv \langle \boldsymbol{\eta}^l (\boldsymbol{\eta}^s)^T \rangle \in \mathbb{R}^{N_l \times N_s}$ (with $\widetilde{\mathbf{Q}}^{ls} = (\widetilde{\mathbf{Q}}^{sl})^T$) are the true model error covariances of the large-scale, the small-scale and cross-covariances between the large- and small-scale, respectively. We note that for the purposes of this work, the model error distribution is assumed to be stationary, so that $\widetilde{\mathbf{Q}}$ is not a function of time.

Analogously, the complete forecast state $\left((\boldsymbol{x}^{l,f})^T \quad (\boldsymbol{x}^{s,f})^T\right)^T \in \mathbb{R}^{N_t}$ satisfies

$$\begin{pmatrix} \boldsymbol{x}^{l,f} \\ \boldsymbol{x}^{s,f} \end{pmatrix}_k = \begin{pmatrix} \mathbf{M}^l & \mathbf{0}_{N_l \times N_s} \\ \mathbf{M}^{sl} & \mathbf{M}^s \end{pmatrix} \begin{pmatrix} \boldsymbol{x}^{l,f} \\ \boldsymbol{x}^{s,f} \end{pmatrix}_{k-1}. \tag{7.2.4}$$

The forecast errors can then be defined as

$$\begin{pmatrix} \mathbf{e}^{l,f} \\ \mathbf{e}^{s,f} \end{pmatrix}_k \equiv \begin{pmatrix} \boldsymbol{x}^{l,f} \\ \boldsymbol{x}^{s,f} \end{pmatrix}_k - \begin{pmatrix} \boldsymbol{x}^{l,t} \\ \boldsymbol{x}^{s,t} \end{pmatrix}_k = \begin{pmatrix} \mathbf{M}^l & \mathbf{0}_{N_l \times N_s} \\ \mathbf{M}^{sl} & \mathbf{M}^s \end{pmatrix} \begin{pmatrix} \mathbf{e}^{l,f} \\ \mathbf{e}^{s,f} \end{pmatrix}_{k-1} + \begin{pmatrix} \boldsymbol{\eta}^l \\ \boldsymbol{\eta}^s \end{pmatrix}_k, \tag{7.2.5}$$

where $\mathbf{e}^{l,f} \in \mathbb{R}^{N_l}$ and $\mathbf{e}^{s,f} \in \mathbb{R}^{N_s}$ are the large- and small-scale forecast errors respectively. The true forecast error covariance is denoted

$$\widetilde{\mathbf{P}}^f_k = \begin{pmatrix} \widetilde{\mathbf{P}}^{ll,f} & \widetilde{\mathbf{P}}^{ls,f} \\ \widetilde{\mathbf{P}}^{sl,f} & \widetilde{\mathbf{P}}^{ss,f} \end{pmatrix}_k. \tag{7.2.6}$$

Here, $\widetilde{\mathbf{P}}^{ll,f}_k \equiv \langle \mathbf{e}^{l,f}_k (\mathbf{e}^{l,f}_k)^T \rangle \in \mathbb{R}^{N_l \times N_l}$, $\widetilde{\mathbf{P}}^{ss,f}_k \equiv \langle \mathbf{e}^{s,f}_k (\mathbf{e}^{s,f}_k)^T \rangle \in \mathbb{R}^{N_s \times N_s}$ and $\widetilde{\mathbf{P}}^{ls,f}_k \equiv \langle \mathbf{e}^{l,f}_k (\mathbf{e}^{s,f}_k)^T \rangle \in \mathbb{R}^{N_l \times N_s}$ (with $\widetilde{\mathbf{P}}^{ls,f}_k = (\widetilde{\mathbf{P}}^{sl,f}_k)^T$) are the true forecast error covariances of the large-scale, the small-scales and cross-covariances between the large- and small-scale processes, respectively.

This formulation of the complete finite-dimensional dynamics allows us to consider several filters with different approaches to the treatment of large- and small-scales. Moreover, we can consider the interactions between scales and the effect they have on the modelling of observations.

7.2.2 Observations and their uncertainties

In this section we express the equations relating the observations, $\boldsymbol{y}_k \in \mathbb{R}^p$, to the model state in a partitioned form and describe their uncertainties. For the rest of this section, we assume that the model state and observations are valid at the same time, and drop the time subscript, k. At time t_k, the observations are related to the true model state as

$$\boldsymbol{y} = \begin{pmatrix} \mathbf{H}^{l,t} & \mathbf{H}^{s,t} \end{pmatrix} \begin{pmatrix} \boldsymbol{x}^{l,t} \\ \boldsymbol{x}^{s,t} \end{pmatrix} + \boldsymbol{\epsilon}, \tag{7.2.7}$$

where $\boldsymbol{\epsilon} \in \mathbb{R}^p$ is the instrument error, assumed to be random and unbiased with covariance $\widetilde{\mathbf{R}}^I = \langle \boldsymbol{\epsilon}\boldsymbol{\epsilon}^T \rangle \in \mathbb{R}^{p \times p}$ and $\mathbf{H}^{l,t} \in \mathbb{R}^{p \times N_l}$ and $\mathbf{H}^{s,t} \in \mathbb{R}^{p \times N_s}$ are the true linear observation operators which map the large- and small-scale states into observation space respectively. The observation operator $\begin{pmatrix} \mathbf{H}^{l,t} & \mathbf{H}^{s,t} \end{pmatrix}$ is the (linear) finite-dimensional counterpart to the continuum observation operator of Janjić and Cohn (2006). We will not consider nonlinear observation operators in the remainder of this paper.

Throughout this paper, we assume that there is no pre-processing error. Hence, we will be concerned with the two cases described in sections 7.2.2.1 (all scales analysed) and 7.2.2.2 (large scales analysed) below. Case 1 shows the form of the representation error for filters that resolve all scales and is pertinent to the theoretical optimal Kalman filter discussed in section 7.3.2. Case 2 shows the form of the representation error for filters typically used in operational practice and is pertinent to the reduced-state Kalman filter and the Schmidt-Kalman filter discussed in sections 7.3.3 and 7.3.4, respectively.

7.2.2.1 Case 1: All scales analysed

In this case we assume that both the large- and small-scale states are estimated. The total observation error (observation departure from the true state), $\mathbf{e}^o$, can be expressed as

$$\mathbf{e}^o = \boldsymbol{y} - \begin{pmatrix} \mathbf{H}^l & \mathbf{H}^s \end{pmatrix} \begin{pmatrix} \boldsymbol{x}^{l,t} \\ \boldsymbol{x}^{s,t} \end{pmatrix}, \tag{7.2.8}$$

where $\mathbf{H}^l \in \mathbb{R}^{p \times N_l}$ and $\mathbf{H}^s \in \mathbb{R}^{p \times N_s}$ are the blocks of the observation operator used by the filter, acting on the large- and small-scale state components, respectively. Using (7.2.7), we rewrite $\mathbf{e}^o$ as

$$\begin{aligned} \mathbf{e}^o &= \begin{pmatrix} \mathbf{H}^{l,t} & \mathbf{H}^{s,t} \end{pmatrix} \begin{pmatrix} \boldsymbol{x}^{l,t} \\ \boldsymbol{x}^{s,t} \end{pmatrix} + \boldsymbol{\epsilon} - \begin{pmatrix} \mathbf{H}^l & \mathbf{H}^s \end{pmatrix} \begin{pmatrix} \boldsymbol{x}^{l,t} \\ \boldsymbol{x}^{s,t} \end{pmatrix}, \\ &= \left(\mathbf{H}^{l,t} - \mathbf{H}^l\right) \boldsymbol{x}^{l,t} + \left(\mathbf{H}^{s,t} - \mathbf{H}^s\right) \boldsymbol{x}^{s,t} + \boldsymbol{\epsilon}, \\ &= \boldsymbol{\gamma}^l + \boldsymbol{\gamma}^s + \boldsymbol{\epsilon}, \end{aligned} \tag{7.2.9}$$

where $\boldsymbol{\gamma}^l \equiv \left(\mathbf{H}^{l,t} - \mathbf{H}^l\right) \boldsymbol{x}^{l,t}$ is the large-scale observation operator error and $\boldsymbol{\gamma}^s \equiv \left(\mathbf{H}^{s,t} - \mathbf{H}^s\right) \boldsymbol{x}^{s,t}$ is the small-scale observation operator error. Thus, the representation error for this case consists solely of observation operator error, $\boldsymbol{\gamma}^l + \boldsymbol{\gamma}^s$. The observation operator errors, $\boldsymbol{\gamma}^l$ and $\boldsymbol{\gamma}^s$, will each be assumed to be unbiased, so that in this case, the representation error is also unbiased. The representation error covariance for this case will be denoted by $\widetilde{\mathbf{R}}^G = \langle (\boldsymbol{\gamma}^l + \boldsymbol{\gamma}^s)(\boldsymbol{\gamma}^l + \boldsymbol{\gamma}^s)^T \rangle \in \mathbb{R}^{p \times p}$. The total observation error covariance is given by $\widetilde{\mathbf{R}} = \widetilde{\mathbf{R}}^I + \widetilde{\mathbf{R}}^G$, where we have assumed that the representation error and instrument error are mutually uncorrelated.

7.2.2.2 Case 2: Large scales analysed

In this case, we assume that only the large-scale state is estimated such that

$$\mathbf{e}^o = \boldsymbol{y} - \mathbf{H}^l \boldsymbol{x}^{l,t}, \tag{7.2.10}$$

where the observation operator used consists only of the block acting on the large-scales. The decomposition of $\mathbf{e}^o$ can be obtained by setting $\mathbf{H}^s = \mathbf{0}_{p\times N_s}$ in (7.2.9):

$$\mathbf{e}^o = \boldsymbol{\gamma}^l + \mathbf{H}^{s,t}\boldsymbol{x}^{s,t} + \boldsymbol{\epsilon}. \tag{7.2.11}$$

The filter observation operator does not act on the small scales, so the term $\boldsymbol{\gamma}^s$ is replaced by $\mathbf{H}^{s,t}\boldsymbol{x}^{s,t}$, the error due to unresolved scales. The representation error for Case 2 is thus $\boldsymbol{\gamma}^l + \mathbf{H}^{s,t}\boldsymbol{x}^{s,t}$ with covariance $\widetilde{\mathbf{R}}^H = \langle(\boldsymbol{\gamma}^l + \mathbf{H}^{s,t}\boldsymbol{x}^{s,t})(\boldsymbol{\gamma}^l + \mathbf{H}^{s,t}\boldsymbol{x}^{s,t})^T\rangle \in \mathbb{R}^{p\times p}$. Equations (7.2.10) and (7.2.11) are analogous to equation (1) in Janjić et al. (2018) with the pre-processing error omitted. The complete observation error covariance for Case 2 is given by $\widetilde{\mathbf{R}} = \widetilde{\mathbf{R}}^I + \widetilde{\mathbf{R}}^H$, where we have assumed that the representation error and instrument error are mutually uncorrelated. As in Case 1, $\boldsymbol{\gamma}^l$ is assumed to be unbiased. However, we will see in section 7.2.3 that the expected value of $\mathbf{H}^{s,t}\boldsymbol{x}^{s,t}$ is likely to be non-zero.

7.2.3 Bias due to unresolved scales

Analysing only the large-scales will result in an error due to unresolved scales (section 7.2.2.2) that is sequentially correlated in time and correlated with the resolved state, leading to a potential bias (Janjić and Cohn, 2006). Assuming that the large-scale observation operator is unbiased, taking the expectation of the error due to unresolved scales, (7.2.11), (and reintroducing the time subscript k) results in

$$\langle\mathbf{e}^o\rangle = \langle\mathbf{H}^{s,t}\boldsymbol{x}^{s,t}_k\rangle, \tag{7.2.12}$$

where $\langle\cdot\rangle$ denotes the mathematical expectation over the distribution of representation errors at time k. Using dynamical system (7.2.1), repeated substitution for the equation governing $\boldsymbol{x}^{s,t}$ into the expected error due to unresolved scales yields

$$\langle\mathbf{H}^{s,t}\boldsymbol{x}^{s,t}_k\rangle = \langle\mathbf{H}^{s,t}(\underline{\mathbf{M}^{sl,t}\boldsymbol{x}^{l,t}_{k-1}} + \mathbf{M}^{s,t}(\underline{\mathbf{M}^{sl,t}\boldsymbol{x}^{l,t}_{k-2}} + \mathbf{M}^{s,t}(\ldots(\underline{\mathbf{M}^{sl,t}\boldsymbol{x}^{l,t}_{0}} + \mathbf{M}^{s,t}(\boldsymbol{x}^{s,t}_0))\ldots)))\rangle, \tag{7.2.13}$$

Here the underlined terms represent the contribution from the large scales. For many non-trivial models, these terms will not be identically zero, and potentially introduce a bias even if the initial value for the small-scale state is zero, $\boldsymbol{x}^{s,t}_0 = \mathbf{0}$. For example, Janjić and Cohn (2006) solved a model of non-divergent linear advection on a sphere using a truncated expansion in spherical harmonics. Introducing a shear flow results in a dynamical system where the unresolved small-scales do not directly influence the resolved large-scales, but the large-scales influence the small-scales. This yields an error and bias due to unresolved scales. Janjić and Cohn (2006) were able to mitigate the bias using specific initial conditions. However, this experimental freedom would not be available in less-idealized situations. Therefore, when accounting for the

unresolved scales in data assimilation we must also determine and treat any bias arising. In the new results in sections 7.5-7.6 below, we carefully construct our model to avoid bias due to unresolved scales. However, we revisit this problem in sections 7.7-7.9 where we use filters with bias-correction schemes.

7.3 Sequential linear filters and representation uncertainty

In this section we describe the general linear filtering framework that we use for data assimilation in our theoretical investigations and numerical experiments. We consider three filters in more detail: an optimal Kalman filter (OKF) that takes account of all scales; a reduced-state Kalman filter (RKF) that disregards the small-scales; and the Schmidt-Kalman filter (SKF) that provides analyses of the large-scale state through consideration of both the large- and small-scale uncertainties.

7.3.1 A linear filter

A linear filter algorithm can be divided into analysis update and model prediction steps. The general form of the analysis update at time t_k, is given by

$$\boldsymbol{x}_k^a = \boldsymbol{x}_k^f + \mathbf{K}_k \mathbf{d}_k^{o,f}, \tag{7.3.1}$$

where $\boldsymbol{x}_k^a$ is the analysis (state estimate), $\boldsymbol{x}_k^f$ is the forecast state, $\mathbf{K}_k$ is the gain matrix and $\mathbf{d}_k^{o,f} = \boldsymbol{y}_k - \mathbf{H}_k \boldsymbol{x}_k^f$ is the innovation, defined as the observation-minus-forecast departure. In this general setting we have not defined the dimensions of the vectors and matrices in (7.3.1), as this will depend on the specific choice of filter. For example, the state $\boldsymbol{x}$ in equation (7.3.1) could be either the complete state $\left((\boldsymbol{x}^l)^T \;\; (\boldsymbol{x}^s)^T\right)^T$ or just the large-scale state $\boldsymbol{x}^l$. There are various approaches to determine the gain matrix which will be discussed in sections 7.3.2, 7.3.3 and 7.3.4.

The perceived analysis error covariance update calcluated by the filter at time t_k is given by

$$\mathbf{P}_k^a = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)\, \mathbf{P}_k^f, \tag{7.3.2}$$

where $\mathbf{I}$ is the identity matrix, $\mathbf{H}_k$ is the observation operator and $\mathbf{P}_k^f$ is the perceived forecast error covariance. Equation (7.3.2) is known as the short-form of the analysis error covariance update. This equation only provides the correct estimate of the analysis error covariance if the background and observation error statistics used in the filter reflect the true error statistics. The use of a suboptimal gain and the short form update equation (7.3.2) will result in the filter producing incorrect error

statistics. The true error statistics will be derived in section 7.4.

For the model prediction step, the forecast state at time t_{k+1} is evolved from the analysis at the previous time-step and is given by

$$\boldsymbol{x}^f_{k+1} = \mathbf{M}\boldsymbol{x}^a_k, \tag{7.3.3}$$

where $\mathbf{M}$ is a linear model. A model error term is not included in the forecast state update as linear filters estimate the mean state and we have assumed that the model error is unbiased. However, the error in the model $\mathbf{M}$ is accounted for in the forecast error covariance update given by

$$\mathbf{P}^f_{k+1} = \mathbf{M}\mathbf{P}^a_k\mathbf{M}^T + \mathbf{Q}, \tag{7.3.4}$$

which will be discussed further in section 7.4. We note that (7.3.4) will only produce correct error statistics when $\mathbf{P}^a_k$ and $\mathbf{Q}$ are equal to their true statistics counterparts.

Equations (7.3.1) - (7.3.4) form the core components of the linear filter algorithm. In the following sections we discuss three linear filters, each based on the Kalman filter (Kalman, 1960), that we will use in this paper. Table 7.1 summarizes the key vectors and matrices used in these three Kalman filters.

7.3.2 The Optimal Kalman filter (OKF)

For the optimal Kalman filter (OKF), we assume that we are able to model the processes for all scales and know the correct error statistics for the initial state, observations and model. Therefore, the perceived error statistics for the OKF will be equivalent to the true error statistics. The OKF simultaneously updates the large- and small-scale states, $\boldsymbol{x}^l \in \mathbb{R}^{N_l}$ and $\boldsymbol{x}^s \in \mathbb{R}^{N_s}$, so that the analysis update takes the form

$$\begin{pmatrix}\boldsymbol{x}^{l,a}\\ \boldsymbol{x}^{s,a}\end{pmatrix} = \begin{pmatrix}\boldsymbol{x}^{l,f}\\ \boldsymbol{x}^{s,f}\end{pmatrix} + \begin{pmatrix}\mathbf{K}^l\\ \mathbf{K}^s\end{pmatrix}\left[\boldsymbol{y} - \begin{pmatrix}\mathbf{H}^l & \mathbf{H}^s\end{pmatrix}\begin{pmatrix}\boldsymbol{x}^{l,f}\\ \boldsymbol{x}^{s,f}\end{pmatrix}\right], \tag{7.3.5}$$

cf. (7.3.1). The gain matrix for the OKF is partitioned into large- and small-scale components $\mathbf{K}^l \in \mathbb{R}^{N_l\times p}$ and $\mathbf{K}^s \in \mathbb{R}^{N_s\times p}$ respectively, and is given in Table 7.1. This is the optimal Kalman gain which minimises the trace of the analysis error covariance (e.g., Nichols, 2010). The analysis error covariance update is calculated using (7.3.2) with state error covariances with the same block structure as the forecast error covariance (7.2.6).

As the OKF filters all scales, the total observation error is described by Case 1 (all scales analysed, section 7.2.2.1). Hence, the observation error covariance for the

Table 7.1: The filter matrices and vectors for the Optimal Kalman filter (OKF), reduced-state Kalman filter (RKF) and Schmidt Kalman filter (SKF). The equations for the three filters are obtained through substituting these terms into (7.3.1) - (7.3.4). As discussed in section 7.3.4, we note that while the SKF uses $\mathbf{H}^l$ in the innovation only, the complete observation operator is used in the calculation of $\mathbf{K}^l$ and $\mathbf{D}$.

	OKF	RKF	SKF
State: $\boldsymbol{x}$	$\begin{pmatrix}\boldsymbol{x}^l\\ \boldsymbol{x}^s\end{pmatrix} \in \mathbb{R}^{N_t}$	$\boldsymbol{x}^l \in \mathbb{R}^{N_l}$	$\boldsymbol{x}^l \in \mathbb{R}^{N_l}$
State error covariance: $\mathbf{P}$	$\begin{pmatrix}\mathbf{P}^{ll} & \mathbf{P}^{ls}\\ \mathbf{P}^{sl} & \mathbf{P}^{ss}\end{pmatrix} \in \mathbb{R}^{N_t\times N_t}$	$\mathbf{P}^{ll} \in \mathbb{R}^{N_l\times N_l}$	$\begin{pmatrix}\mathbf{P}^{ll} & \mathbf{P}^{ls}\\ \mathbf{P}^{sl} & \mathbf{C}^{s}\end{pmatrix} \in \mathbb{R}^{N_t\times N_t}$
Observation operator: $\mathbf{H}$	$\begin{pmatrix}\mathbf{H}^l & \mathbf{H}^s\end{pmatrix} \in \mathbb{R}^{p\times N_t}$	$\mathbf{H}^l \in \mathbb{R}^{p\times N_l}$	$\begin{pmatrix}\mathbf{H}^l & \mathbf{H}^s\end{pmatrix} \in \mathbb{R}^{p\times N_t}$
Observation Error Covariance: $\mathbf{R}$	$\mathbf{R}^I + \mathbf{R}^G$	$\mathbf{R}^I + \mathbf{R}^H$	$\mathbf{R}^I + \mathbf{R}^G$
Innovation: $\mathbf{d}^{o,f}$	$\boldsymbol{y} - \begin{pmatrix}\mathbf{H}^l & \mathbf{H}^s\end{pmatrix}\begin{pmatrix}\boldsymbol{x}^{l,f}\\ \boldsymbol{x}^{s,f}\end{pmatrix}$	$\boldsymbol{y} - \mathbf{H}^l\boldsymbol{x}^{l,f}$	$\boldsymbol{y} - \mathbf{H}^l\boldsymbol{x}^{l,f}$
Innovation covariance: $\mathbf{D}$	$\begin{pmatrix}\mathbf{H}^l & \mathbf{H}^s\end{pmatrix}\begin{pmatrix}\mathbf{P}^{ll,f} & \mathbf{P}^{ls,f}\\ \mathbf{P}^{sl,f} & \mathbf{P}^{ss,f}\end{pmatrix}\begin{pmatrix}(\mathbf{H}^l)^T\\ (\mathbf{H}^s)^T\end{pmatrix} + \mathbf{R}$	$\mathbf{H}^l\mathbf{P}^{ll,f}(\mathbf{H}^l)^T + \mathbf{R}$	$\begin{pmatrix}\mathbf{H}^l & \mathbf{H}^s\end{pmatrix}\begin{pmatrix}\mathbf{P}^{ll,f} & \mathbf{P}^{ls,f}\\ \mathbf{P}^{sl,f} & \mathbf{C}^{s}\end{pmatrix}\begin{pmatrix}(\mathbf{H}^l)^T\\ (\mathbf{H}^s)^T\end{pmatrix} + \mathbf{R}$
Kalman gain: $\mathbf{K}$	$\begin{pmatrix}\mathbf{K}^l\\ \mathbf{K}^s\end{pmatrix} = \begin{pmatrix}\mathbf{P}^{ll,f} & \mathbf{P}^{ls,f}\\ \mathbf{P}^{sl,f} & \mathbf{P}^{ss,f}\end{pmatrix}\begin{pmatrix}(\mathbf{H}^l)^T\\ (\mathbf{H}^s)^T\end{pmatrix}\mathbf{D}^{-1}$	$\mathbf{K}^l = \left(\mathbf{P}^{ll,f}(\mathbf{H}^l)^T\right)\mathbf{D}^{-1}$	State update $\mathbf{K}^l = \left(\mathbf{P}^{ll,f}(\mathbf{H}^l)^T + \mathbf{P}^{ls,f}(\mathbf{H}^s)^T\right)\mathbf{D}^{-1}$ Covariance update $\begin{pmatrix}\mathbf{K}^l\\ \mathbf{K}^s\end{pmatrix} = \begin{pmatrix}\left(\mathbf{P}^{ll,f}(\mathbf{H}^l)^T + \mathbf{P}^{ls,f}(\mathbf{H}^s)^T\right)\mathbf{D}^{-1}\\ \mathbf{0}_{N_s\times p}\end{pmatrix}$
Forecast model: $\mathbf{M}$	$\begin{pmatrix}\mathbf{M}^l & \mathbf{0}_{N_l\times N_s}\\ \mathbf{M}^{sl} & \mathbf{M}^s\end{pmatrix} \in \mathbb{R}^{N_t\times N_t}$	$\mathbf{M}^l \in \mathbb{R}^{N_l\times N_l}$	State update $\mathbf{M}^l \in \mathbb{R}^{N_l\times N_l}$ Covariance update $\begin{pmatrix}\mathbf{M}^l & \mathbf{0}_{N_l\times N_s}\\ \mathbf{M}^{sl} & \mathbf{M}^s\end{pmatrix} \in \mathbb{R}^{N_t\times N_t}$
Model Error Covariance: $\mathbf{Q}$	$\begin{pmatrix}\mathbf{Q}^{ll} & \mathbf{Q}^{ls}\\ \mathbf{Q}^{sl} & \mathbf{Q}^{ss}\end{pmatrix} \in \mathbb{R}^{N_t\times N_t}$	$\mathbf{Q}^{ll} \in \mathbb{R}^{N_l\times N_l}$	$\begin{pmatrix}\mathbf{Q}^{ll} & \mathbf{Q}^{ls}\\ \mathbf{Q}^{sl} & \mathbf{Q}^{ss}\end{pmatrix} \in \mathbb{R}^{N_t\times N_t}$

OKF is $\mathbf{R} = \mathbf{R}^I + \mathbf{R}^G$.

For the OKF forecast step we use the matrix

$$\begin{pmatrix} \mathbf{M}^l & \mathbf{0}_{N_l \times N_s} \\ \mathbf{M}^{sl} & \mathbf{M}^s \end{pmatrix} \tag{7.3.6}$$

as our forecast model in (7.3.3) and the partitioned model error covariance given in Table 7.1 in the forecast error covariance prediction (7.3.4).

In summary, the analysis and forecast updates for the OKF state and covariance are a partitioned form of (7.3.1) - (7.3.4). By treating all scales in the assimilation the OKF has no error due to unresolved scales in the associated observation equation. However, due to computational constraints and inadequate knowledge of small-scale processes it is not possible to apply this technique in practice. Hence, methods that approximate the influence of small-scale processes must be employed instead.

7.3.3 The Reduced-state Kalman filter (RKF)

The suboptimal Kalman filter which estimates only the large-scale state and completely neglects the modelling of small-scale processes will be referred to as the reduced-state Kalman filter (RKF).

The analysis and forecast update equations for the RKF are simply the linear filter equations (7.3.1)-(7.3.4) where, as described in Table 7.1, we use the large-scale state, error covariances and observation operator. Thus, the forecast innovation is

$$\mathbf{d}^{o,f} = \boldsymbol{y} - \mathbf{H}^l \boldsymbol{x}^{l,f} = \boldsymbol{\epsilon} + \boldsymbol{\gamma}^l + \mathbf{H}^{s,t} \boldsymbol{x}^{s,t} - \mathbf{H}^l \mathbf{e}^{l,f}, \tag{7.3.7}$$

where the second equality can be established by adding and subtracting the term $\mathbf{H}^l \boldsymbol{x}^{l,t}$. Assuming each error has zero-mean, taking the expectation of the outer product of (7.3.7) yields the true innovation covariance (i.e. all contributing error covariances are true error statistics). However, the innovation covariance used by the RKF is given by

$$\mathbf{D} = \mathbf{H}^l \mathbf{P}^{ll,f} (\mathbf{H}^l)^T + \mathbf{R}^I + \mathbf{R}^H, \tag{7.3.8}$$

where the large-scale forecast error covariance, instrument error covariance and representation error covariance are perceived error statistics. The influence of any small-scale processes is now accounted for through the representation error covariance $\mathbf{R}^H$ which needs to be approximated.

Reduced-state methods form an attractive approach in situations where computational expense is an important consideration. However, it is necessary to approx-

imate the representation error covariance, $\mathbf{R}^H$. Hence, the Kalman gain for the RKF will not minimise the analysis error covariance and the filter will be suboptimal.

7.3.4 The Schmidt-Kalman filter (SKF)

The Schmidt-Kalman Filter (SKF) estimates only the large-scale state, but the statistics of any unmodelled processes are used to determine the Kalman gain for the filtered state (Schmidt, 1966; Janjić and Cohn, 2006). A summary of the relevant equations is included in Table 7.1.

As only the large-scale state is estimated the forecast innovation is computed using the large-scale state, $\boldsymbol{x}^{l,f}$, and observation operator, $\mathbf{H}^l$. To determine the innovation covariance we start with the innovation (7.3.7) and add and subtract the term $\begin{pmatrix}\mathbf{H}^l & \mathbf{H}^s\end{pmatrix}\begin{pmatrix}(\boldsymbol{x}^{l,t})^T & (\boldsymbol{x}^{s,t})^T\end{pmatrix}^T$. This allows us to write the innovation in the form

$$\mathbf{d}^{o,f} = \boldsymbol{\epsilon} + \boldsymbol{\gamma}^l + \boldsymbol{\gamma}^s + \begin{pmatrix}\mathbf{H}^l & \mathbf{H}^s\end{pmatrix}\begin{pmatrix}-\mathbf{e}^{l,f} \\ \boldsymbol{x}^{s,t}\end{pmatrix}. \tag{7.3.9}$$

The innovation is now written in terms of the observation errors corresponding to case 1 (where all scales are analysed, see section 7.2.2.1), the large-scale forecast error mapped into observation space, $\mathbf{H}^l\mathbf{e}^{l,f}$, and the term $\mathbf{H}^s\boldsymbol{x}^{s,t}$, the true small-scale state mapped into observation space. Assuming each error and $\boldsymbol{x}^{s,t}$ has zero-mean, taking the expectation of the outer product of (7.3.9) gives the true innovation covariance. The innovation covariance used by the SKF is given by

$$\mathbf{D} = \begin{pmatrix}\mathbf{H}^l & \mathbf{H}^s\end{pmatrix}\begin{pmatrix}\mathbf{P}^{ll,f} & \mathbf{P}^{ls,f} \\ \mathbf{P}^{sl,f} & \mathbf{C}^s\end{pmatrix}\begin{pmatrix}(\mathbf{H}^l)^T \\ (\mathbf{H}^s)^T\end{pmatrix} + \mathbf{R}^I + \mathbf{R}^G. \tag{7.3.10}$$

Here, we have abused our notation, to write $\mathbf{P}^{ls,f}$ as the perceived approximation of $\langle -\mathbf{e}^{l,f}\left(\boldsymbol{x}^{s,t}\right)^T\rangle$, such that $\mathbf{P}^{sl,f} = \left(\mathbf{P}^{ls,f}\right)^T$. Using this notation for the cross-covariances of the SKF is common amongst other literature on the filter (e.g., Janjić and Cohn, 2006; Janjić et al., 2018). Following Janjić and Cohn (2006), we employ a prescribed error covariance $\mathbf{C}^s$ as a time-independent approximation of $\langle \boldsymbol{x}^{s,t}\left(\boldsymbol{x}^{s,t}\right)^T\rangle$. We note that as the small-scale error covariance is prescribed, the innovation covariance is an inexact approximation. The innovation covariance for the SKF is theoretically the same as the innovation covariance for the RKF (7.3.8) but expressed in a different form that includes contributions from the small scale processes.

The analysis state update for the SKF is given by

$$\boldsymbol{x}_k^{l,a} = \boldsymbol{x}_k^{l,f} + \mathbf{K}_k^l\left(\boldsymbol{y}_k - \mathbf{H}_k^l\boldsymbol{x}_k^{l,f}\right), \tag{7.3.11}$$

where $\mathbf{K}_k^l = \left(\mathbf{P}_k^{ll,f}\left(\mathbf{H}^l\right)^T + \mathbf{P}_k^{ls,f}\left(\mathbf{H}^s\right)^T\right)\mathbf{D}_k^{-1}$ is the Schmidt-Kalman large-scale gain. To obtain an analysis error covariance update equation we augment $\mathbf{K}^l$ with $\mathbf{K}^s = \mathbf{0}_{N_s \times p}$ and substitute into equation (7.3.2). This is justified as the unfiltered state is assumed to have a small magnitude. Large instrument uncertainty associated with observations of the small-scale state or a small-magnitude observation operator for this state would also justify this assumption (Simon, 2006, pp 309 – 312). As the short-form analysis error covariance update for the SKF is not symmetric, we calculate $\mathbf{P}^{ll,a}$ and $\mathbf{P}^{ls,a}$ through the short-form update only and set $\mathbf{P}_k^{sl,a} = \left(\mathbf{P}_k^{ls,a}\right)^T$. Thus, the SKF analysis error covariance update equations are

$$\mathbf{P}_k^{ll,a} = \left(\mathbf{I}_{N_l} - \mathbf{K}_k^l\mathbf{H}_k^l\right)\mathbf{P}_k^{ll,f} - \mathbf{K}_k^l\mathbf{H}_k^s\mathbf{P}_k^{sl,f}, \tag{7.3.12}$$

$$\mathbf{P}_k^{ls,a} = \left(\mathbf{I}_{N_l} - \mathbf{K}_k^l\mathbf{H}_k^l\right)\mathbf{P}_k^{ls,f} - \mathbf{K}_k^l\mathbf{H}_k^s\mathbf{C}^s, \tag{7.3.13}$$

$$\mathbf{P}_k^{sl,a} = \left(\mathbf{P}_k^{ls,a}\right)^T. \tag{7.3.14}$$

We note that the term $-\mathbf{K}_k^l\mathbf{H}_k^s$ will usually be non-zero for the SKF. This term couples the large-scale uncertainty to the small-scale variability. If this term were zero, the large-scale state and uncertainty estimates produced by the RKF and SKF may still differ because of the differing innovation covariances between the filters.

The SKF treatment of the forecast step has a similar philosophy to the analysis step. The state prediction (7.3.3) is obtained through evolving the large-scale state $\boldsymbol{x}^l$ with the large-scale forecast model $\mathbf{M}^l$:

$$\boldsymbol{x}_{k+1}^{l,f} = \mathbf{M}^l\boldsymbol{x}_k^{l,t}. \tag{7.3.15}$$

The large-scale and cross-covariance blocks of the forecast error covariance are calculated using the complete forecast model and model error covariance in (7.3.4). The SKF forecast error covariance update equations are

$$\mathbf{P}_{k+1}^{ll,f} = \mathbf{M}^l\mathbf{P}_k^{ll,a}\left(\mathbf{M}^l\right)^T + \mathbf{Q}^l, \tag{7.3.16}$$

$$\mathbf{P}_{k+1}^{ls,f} = \mathbf{M}^l\left(\mathbf{P}_k^{ll,a}\left(\mathbf{M}^{sl}\right)^T + \mathbf{P}_k^{ls,a}\left(\mathbf{M}^s\right)^T\right), \tag{7.3.17}$$

$$\mathbf{P}_{k+1}^{sl,f} = \left(\mathbf{P}_{k+1}^{ls,f}\right)^T. \tag{7.3.18}$$

The prescribed small-scale error covariance $\mathbf{C}^s$ is assumed constant in time and is not updated.

The appeal of the SKF is in its ability to compensate for small-scales without estimation of the small-scale state. Practical implementation of the SKF would require the filter to be adapted to nonlinear models. However, even for linear systems, the models evolving the small-scale processes would be unknown and their influence on

the error covariances would need to be quantified. Additionally, the propagation of the state cross-covariances poses a considerable computational cost.

7.3.5 Discussion

The OKF, SKF and RKF represent three different approaches for dealing with observation uncertainty due to unresolved scales (see Table 7.1). The OKF analyses all scales, thus avoiding the error due to unresolved scales altogether, while the RKF completely disregards the small-scale processes and accounts for the error due to unresolved scales through the representation error covariance matrix. The SKF, however, takes a compromise approach where only the large-scale state is estimated, but the uncertainty in all-scales is accounted for in the estimation. Additionally, the SKF accounts for the flow-dependence of the correlations between the large-scale errors and small-scale processes (albeit approximately) through the cross-covariances in the analysis and forecast error covariances given in equations (7.3.13), (7.3.14), (7.3.17) and (7.3.18). Applications where it is a poor approximation to neglect these cross-covariances will benefit the most from using the SKF (as opposed to the RKF where these cross-covariances are neglected).

7.4 True Analysis Error Equations

A standard metric for assessing the quality of a data assimilation scheme is through examination of the magnitude of its analysis errors (e.g., Liu and Rabier, 2002). Under an unrealistic and restrictive set of conditions the Kalman filter is known to be optimal in a minimum mean-square-error sense and to produce the true error statistics describing its analysis and forecast (e.g., Todling and Cohn, 1994; Nichols, 2010). In contrast, both the SKF and RKF described in section 7.3 will incorrectly estimate the true analysis and forecast error covariances due to their treatment of the small-scales in the filter calculations. In this section we extend the existing literature on the true analysis error equations to include representation error so that we may evaluate the analysis obtained through the SKF and RKF.

To obtain the true analysis error equation for a linear filter we assume that we have exact knowledge of the truth and that both the true and filter models and observation operators are linear. Under this regime, the true analysis error at time t_k has been derived by Moodey (2013) and is given by

$$\mathbf{e}_k^a \equiv \boldsymbol{x}_k^a - \boldsymbol{x}_k^t = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\,\mathbf{M}\mathbf{e}_{k-1}^a + (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\,\boldsymbol{\eta}_k + \mathbf{K}_k\mathbf{e}_k^o, \tag{7.4.1}$$

where $\mathbf{e}_k^a$ is the analysis error, $\boldsymbol{\eta}_k$ is the model error (see section 7.2.1), $\mathbf{e}_k^o$ is the total observation error which will be specified for different cases in subsections 7.4.1 - 7.4.2 and $\mathbf{K}_k$ and $\mathbf{H}_k$ are the Kalman gains and observation operators for the

analysis state updates respectively. Therefore, the Kalman gain is calculated from the error statistics perceived by a filter. We assume that $\mathbf{e}^a_{k-1}$, $\boldsymbol{\eta}_k$ and $\mathbf{e}^o_k$ each have zero-mean and are mutually uncorrelated. (We note that this assumption excludes a consideration of bias due to unresolved scales. However, this is considered further in section 7.7). Under these assumptions the true analysis error covariance is obtained through taking the expectation of the outer product of equation (7.4.1) with itself to give

$$\widetilde{\mathbf{P}}^a_k \equiv \langle \mathbf{e}^a_k \left(\mathbf{e}^a_k\right)^T \rangle = \left(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k\right) \widetilde{\mathbf{P}}^f_k \left(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k\right)^T + \mathbf{K}_k\widetilde{\mathbf{R}}_k\mathbf{K}^T_k, \tag{7.4.2}$$

where

$$\widetilde{\mathbf{P}}^f_k \equiv \langle \mathbf{e}^f_k \left(\mathbf{e}^f_k\right)^T \rangle = \mathbf{M}\widetilde{\mathbf{P}}^a_{k-1}\mathbf{M}^T + \widetilde{\mathbf{Q}}. \tag{7.4.3}$$

Here we remind the reader that we have used tildes to indicate true error covariances, to help distinguish these from the covariances perceived by the filters, which may be suboptimal. Equation (7.4.2) is known as the Joseph-formula (e.g. Gelb, 1974, pp 305–306). The true analysis error covariance (7.4.2) is valid for any gain matrix. The Joseph-formula is equivalent to the short form analysis error covariance (7.3.2) for the optimal case (OKF) in exact arithmetic.

The true analysis error covariance can be calculated separately from the assimilation. In subsections 7.4.1 - 7.4.2 we use equations (7.4.1) and (7.4.2) to determine the true analysis error equations and error covariances for Cases 1 and 2 described in sections 7.2.2.1 and 7.2.2.2. Case 1 corresponds to analysing all scales and includes the OKF. Case 2 corresponds to analysing the large-scale state only and includes the RKF and SKF. Table 7.2 summarizes the matrices and vectors used in the true error calculations.

7.4.1 Case 1: true analysis error covariance when all scales are filtered

To obtain the true analysis error equation we assume that we have complete knowledge of all scales as with the OKF. As in section 7.2.2.1 the observation error will consist of instrument error, $\boldsymbol{\epsilon}$, and the observation operator error for large- and small-scales, $\boldsymbol{\gamma}^l + \boldsymbol{\gamma}^s$. Under these assumptions the true analysis error equation will be a partitioned form of equation (7.4.1) and the true analysis error covariance will be a partitioned form of (7.4.2) with the components given in column 1 of Table 7.2.

Table 7.2: Matrices and vectors used in the true error calculations for Case 1 and 2 described in sections 7.4.1 and 7.4.2. The tildes indicate true error covariances. Case 1 corresponds to analysing all scales and includes the OKF. Case 2 corresponds to analysing the large scales only and includes the RKF and SKF. The true analysis error equation, analysis error covariance and forecast error covariance for each case are obtained by substituting the corresponding components into equations (7.4.1), (7.4.2) and (7.4.3) respectively.

	Case 1 (OKF)	Case 2 (SKF and RKF)
Analysis Errors: $\mathbf{e}^a$	$\begin{pmatrix}\mathbf{e}^{l,a}\\ \mathbf{e}^{s,a}\end{pmatrix}$	$\begin{pmatrix}\mathbf{e}^{l,a}\\ \mathbf{e}^{s,a}\end{pmatrix}$
Model Errors: $\boldsymbol{\eta}$	$\begin{pmatrix}\boldsymbol{\eta}^l\\ \boldsymbol{\eta}^s\end{pmatrix}$	$\begin{pmatrix}\boldsymbol{\eta}^l\\ \boldsymbol{\eta}^s\end{pmatrix}$
Observation Errors: $\mathbf{e}^o$	$\boldsymbol{\epsilon}+\boldsymbol{\gamma}^l+\boldsymbol{\gamma}^s$	$\boldsymbol{\epsilon}+\boldsymbol{\gamma}^l+\mathbf{H}^{s,t}\boldsymbol{x}^{s,t}$
Kalman Gain: $\mathbf{K}$	$\begin{pmatrix}\mathbf{K}^l\\ \mathbf{K}^s\end{pmatrix}$	$\begin{pmatrix}\mathbf{K}^l\\ \mathbf{0}_{N_s\times p}\end{pmatrix}$
Observation Operator: $\mathbf{H}$	$\begin{pmatrix}\mathbf{H}^l & \mathbf{H}^s\end{pmatrix}$	$\begin{pmatrix}\mathbf{H}^l & \mathbf{0}_{p\times N_s}\end{pmatrix}$
Model: $\mathbf{M}$	$\begin{pmatrix}\mathbf{M}^l & \mathbf{0}_{N_l\times N_s}\\ \mathbf{M}^{sl} & \mathbf{M}^s\end{pmatrix}$	$\begin{pmatrix}\mathbf{M}^l & \mathbf{0}_{N_l\times N_s}\\ \mathbf{M}^{sl} & \mathbf{M}^s\end{pmatrix}$
State Error Covariance: $\widetilde{\mathbf{P}}$	$\begin{pmatrix}\widetilde{\mathbf{P}}^{ll} & \widetilde{\mathbf{P}}^{ls}\\ \widetilde{\mathbf{P}}^{sl} & \widetilde{\mathbf{P}}^{ss}\end{pmatrix}$	$\begin{pmatrix}\widetilde{\mathbf{P}}^{ll} & \widetilde{\mathbf{P}}^{ls}\\ \widetilde{\mathbf{P}}^{sl} & \widetilde{\mathbf{P}}^{ss}\end{pmatrix}$
Observation Error Covariance: $\widetilde{\mathbf{R}}$	$\widetilde{\mathbf{R}}^I+\widetilde{\mathbf{R}}^G$	$\widetilde{\mathbf{R}}^I+\widetilde{\mathbf{R}}^H$
Model Error Covariance: $\widetilde{\mathbf{Q}}$	$\begin{pmatrix}\widetilde{\mathbf{Q}}^{ll} & \widetilde{\mathbf{Q}}^{ls}\\ \widetilde{\mathbf{Q}}^{sl} & \widetilde{\mathbf{Q}}^{ss}\end{pmatrix}$	$\begin{pmatrix}\widetilde{\mathbf{Q}}^{ll} & \widetilde{\mathbf{Q}}^{ls}\\ \widetilde{\mathbf{Q}}^{sl} & \widetilde{\mathbf{Q}}^{ss}\end{pmatrix}$

7.4.2 Case 2: true analysis error covariance when only large-scales are filtered

The true analysis error equation for case 2 applies to filters that estimate the large-scale state only like the RKF and SKF. Using equation (7.4.1) and the state gain matrices and observation operators, the large-scale analysis error for the RKF and SKF is given by

$$\mathbf{e}_k^{l,a} \equiv \boldsymbol{x}_k^{l,t} - \boldsymbol{x}_k^{l,t} = \left(\mathbf{I}_{N_l} - \mathbf{K}_k^l \mathbf{H}_k^l\right) \mathbf{M}^l \mathbf{e}_{k-1}^{l,a} + \left(\mathbf{I}_{N_l} - \mathbf{K}_k^l \mathbf{H}_k^l\right) \boldsymbol{\eta}_k^l + \mathbf{K}_k^l \left(\boldsymbol{\epsilon}_k + \boldsymbol{\gamma}_k^l + \mathbf{H}_k^{s,t} \boldsymbol{x}_k^{s,t}\right). \tag{7.4.4}$$

We note that the observation errors correspond to case 2 described in 7.2.2.2 as both the RKF and SKF filter the large-scale state only. Hence, the effect of the small-scale processes on $\mathbf{e}_k^{l,a}$ in equation (7.4.4) is determined through the error due to unresolved scales $\mathbf{H}^{s,t}\boldsymbol{x}_k^{s,t}$. We observe that the true large-scale error covariance may thus be written in terms of the representation error covariance as

$$\widetilde{\mathbf{P}}_k^{ll,a} \equiv \langle \mathbf{e}_k^{l,a} (\mathbf{e}_k^{l,a})^T \rangle = \left(\mathbf{I}_{N_l} - \mathbf{K}_k^l \mathbf{H}_k^l\right) \widetilde{\mathbf{P}}_k^{ll,f} \left(\mathbf{I}_{N_l} - \mathbf{K}_k^l \mathbf{H}_k^l\right)^T + \mathbf{K}_k^l \left(\widetilde{\mathbf{R}}_k^I + \widetilde{\mathbf{R}}_k^H\right) \left(\mathbf{K}_k^l\right)^T. \tag{7.4.5}$$

However, the true error statistics contributing to the true analysis error covariance are unknown in practice making the use of (7.4.5) to evaluate filter performance unfeasible. For theoretical experiments where most error statistics can be prescribed, determining $\widetilde{\mathbf{R}}_k^H$ requires a Monte Carlo approach due to its dependence on $\boldsymbol{x}^{s,t}$. Alternatively, a different form of the analysis error equation may be more practical.

Assuming we know the true behaviour for the small-scales we can express the true analysis error equation for the RKF and SKF as

$$\begin{pmatrix} \mathbf{e}^{l,a} \\ \mathbf{e}^{s,a} \end{pmatrix}_k = \begin{pmatrix} \mathbf{I}_{N_l} - \mathbf{K}_k^l \mathbf{H}_k^l & \mathbf{0}_{N_l \times N_s} \\ \mathbf{0}_{N_s \times N_l} & \mathbf{I}_{N_s} \end{pmatrix} \begin{pmatrix} \mathbf{e}^{l,f} \\ \mathbf{e}^{s,f} \end{pmatrix}_k + \begin{pmatrix} \mathbf{K}_k^l \\ \mathbf{0}_{N_s \times p} \end{pmatrix} \left(\boldsymbol{\epsilon}_k + \boldsymbol{\gamma}_k^l + \mathbf{H}^{s,t} \boldsymbol{x}_k^{s,t}\right), \tag{7.4.6}$$

where $\mathbf{e}_k^{l,f} = \mathbf{M}\mathbf{e}_{k-1}^{l,t} + \boldsymbol{\eta}_k^l$ and $\mathbf{e}_k^{s,f} = \mathbf{M}^{sl}\mathbf{e}_{k-1}^{l,t} + \mathbf{M}^s \mathbf{e}_{k-1}^{s,a} + \boldsymbol{\eta}_k^s$. We note that, as the small-scale state isn't estimated, the small-scale gain is a zero matrix of dimension $N_s \times p$. We also note that, while the large- and small-scale errors ostensibly appear uncoupled in equation (7.4.6), they are in fact coupled as $\mathbf{e}_k^{s,a}$ and $\mathbf{e}_k^{s,f}$ each depend on $\boldsymbol{x}_k^{s,t}$. Adding and subtracting the term $\mathbf{K}_k^l \mathbf{H}_k^s \mathbf{e}_k^{s,f}$ from the large-scale component

of the analysis error, (7.4.6) may be written as

$$\begin{pmatrix}\mathbf{e}^{l,a}\\ \mathbf{e}^{s,a}\end{pmatrix}_k = \left(\mathbf{I}_{N_t} - \begin{pmatrix}\mathbf{K}_k^l\\ \mathbf{0}_{N_s\times p}\end{pmatrix}\begin{pmatrix}\mathbf{H}^l & \mathbf{H}^s\end{pmatrix}\right)\begin{pmatrix}\mathbf{e}^{l,f}\\ \mathbf{e}^{s,f}\end{pmatrix}_k + \begin{pmatrix}\mathbf{K}_k^l\\ \mathbf{0}_{N_s\times p}\end{pmatrix}\left(\boldsymbol{\epsilon}_k + \boldsymbol{\gamma}_k^l + \mathbf{H}^{s,t}\boldsymbol{x}_k^{s,t} + \mathbf{H}^s\mathbf{e}_k^{s,f}\right). \tag{7.4.7}$$

Using the definitions of the small-scale observation operator error (7.2.9), the error due to unresolved scales (7.2.11) and the small-scale forecast error (7.3.9) we find that

$$\mathbf{H}^{s,t}\boldsymbol{x}_k^{s,t} + \mathbf{H}^s\mathbf{e}_k^{s,f} = \boldsymbol{\gamma}_k^s + \mathbf{H}^s\boldsymbol{x}_k^{s,f}. \tag{7.4.8}$$

Thus the right-hand-side of (7.4.7) can be evaluated without knowledge of the error due to unresolved scales specifically. Instead, this can be written in terms of the observation operator error and a small-scale forecast:

$$\begin{pmatrix}\mathbf{e}^{l,a}\\ \mathbf{e}^{s,a}\end{pmatrix}_k = \left(\mathbf{I}_{N_t} - \begin{pmatrix}\mathbf{K}_k^l\\ \mathbf{0}_{N_s\times p}\end{pmatrix}\begin{pmatrix}\mathbf{H}^l & \mathbf{H}^s\end{pmatrix}\right)\begin{pmatrix}\mathbf{e}^{l,f}\\ \mathbf{e}^{s,f}\end{pmatrix}_k + \begin{pmatrix}\mathbf{K}_k^l\\ \mathbf{0}_{N_s\times p}\end{pmatrix}\left(\boldsymbol{\epsilon}_k + \boldsymbol{\gamma}_k^l + \boldsymbol{\gamma}_k^s + \mathbf{H}^s\boldsymbol{x}_k^{s,f}\right). \tag{7.4.9}$$

The partitioned case 2 error equation (7.4.9) can be used to obtain the true analysis error covariance for the SKF and RKF without knowing the full representation error covariance $\widetilde{\mathbf{R}}^H$. However, when using this form of the analysis error equation to obtain the true error statistics the correlations between $\boldsymbol{x}^{s,f}$ and $\mathbf{e}^{s,f}$ may be non-negligible. We note that, while $\boldsymbol{x}^{s,f}$ and $\boldsymbol{x}^{s,t}$ will also be unknown in practice, they could be approximated offline with high-resolution models.

7.5 Experiment Methodology

7.5.1 Gaussian random walk model

We now consider the methodology for numerical experiments where we apply the three filters to the simple model system

$$\begin{aligned}\begin{pmatrix}x^l\\ x^s\end{pmatrix}_{k+1} &= \begin{pmatrix}1 & 0\\ \mathrm{M}^{sl} & \exp(-1/2)\end{pmatrix}\begin{pmatrix}x^l\\ x^s\end{pmatrix}_k - \begin{pmatrix}\eta^l\\ \eta^s\end{pmatrix}_k,\\ y_k &= \begin{pmatrix}1 & 1\end{pmatrix}\begin{pmatrix}x^l\\ x^s\end{pmatrix}_k + \epsilon_k,\end{aligned} \tag{7.5.1}$$

such that $\eta_k^l \sim (0, \mathrm{Q}^l)$, $\eta_k^s \sim (0, \mathrm{Q}^s)$, and $\epsilon_k \sim (0, \mathrm{R}^I)$ (Brown and Hwang, 2012, pp 192). This system uses one variable for the large-scale state, x^l, and one variable for the small-scale state, x^s. The large-scale state x^l and small-scale state x^s are random

walk variables driven by the errors η^l and η^s whose structures are determined by the variances Q^l and Q^s respectively. There are no cross-covariances in the model error statistics. The model component M^{sl} is the contribution from the large-scale processes to the small-scale state. The observations will be taken to be the sum of the large- and small-scale states plus instrument error.

The random walk model (7.5.1) will first be used for a "nature run" from which observations can be created. The filters described in section 7.3 will then be used to assimilate these observations and the true large-scale analysis error variance calculated at the end of the assimilation window. As the RKF and SKF are suboptimal, they propagate inexact error variances. Therefore, the true error variances for the RKF and SKF are calculated using (7.4.9) to provide a comparison between their performances.

Through our experimental design we are able to easily control the magnitude of the observation error due to unresolved scales by adjusting Q^s. The relationship between Q^s and the error due to unresolved scales is described in section 7.5.3. This framework also allows for the determination of the optimal C^s as well as the sensitivity of the SKF to this modelled variance.

7.5.2 Initial conditions and filter parameters

For our experiments, we choose the initial conditions for the true state (nature run) to be $x_0^{l,t} = 10$ and $x_0^{s,t} = 0$ so that the true resolved state is noticeably larger than the true unresolved state. Setting the small-scale true state to zero also ensures that the representation errors are initially unbiased.

We set the initial conditions for the forecast state and forecast error covariance to be

$$\boldsymbol{x}_0^f = \begin{pmatrix} x_0^{l,f} \\ x_0^{s,f} \end{pmatrix} = \begin{pmatrix} x_0^l + \alpha^l \\ x_0^s + \alpha^s \end{pmatrix} \text{ and } \mathbf{P}_0^f = \begin{pmatrix} \mathrm{P}_0^{ll,f} & \mathrm{P}_0^{ls,f} \\ \mathrm{P}_0^{sl,f} & \mathrm{P}_0^{ss,f} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0.1 \end{pmatrix} \tag{7.5.2}$$

where $\alpha^l \sim \mathcal{N}(0, \mathrm{P}_0^{ll,f})$ and $\alpha^s \sim \mathcal{N}(0, \mathrm{P}_0^{ss,f})$ are perturbations from the true states. We have assumed that the initial large- and small-scale forecast errors are uncorrelated.

For our first set of experiments we set the model component $\mathrm{M}^{sl} = 0$ so that the representation errors remain unbiased throughout the assimilation. The large-scale model error variance, $\mathrm{Q}^l = 1$, is used throughout our experiments while Q^s will vary for different experiments.

Observations are assimilated every time-step. The true observation operator is $\mathbf{H} = (1 \;\; 1)$ which is used by all three filters; this ensures that there is no observation operator error. Unless otherwise specified, the instrument error variance is set to

$\mathrm{R}^I = 0.1$, and $\widetilde{\mathrm{R}}^I_k = \mathrm{R}^I$ so that each filter correctly accounts for the instrument error. For the RKF, we set $\mathrm{R}^H = 0$ so that the filter completely ignores the small-scale processes. The prescribed small-scale error variance C^s is varied throughout our experiments.

To calculate the true analysis error variance with (7.4.9), we neglect the variance of $\boldsymbol{x}^{s,f}$ and the correlations between $\boldsymbol{x}^{s,f}$ and $\mathbf{e}^{s,f}$ as the solution for $\boldsymbol{x}^{s,f}$ is exponentially decaying with time. This method of calculating the true analysis error covariance has been validated against a Monte Carlo approach.

7.5.3 Determining the small-scale variability over the assimilation window

For the SKF and RKF, which do not update the small-scale state, the true small-scale analysis and forecast error statistics are equal at the same time-step. Setting $\mathrm{M}^{sl} = 0$, the true small-scale error covariance, denoted $\widetilde{\mathbf{P}}^{ss}$, is evolved through the difference equation

$$\widetilde{\mathrm{P}}^{ss}_k = \mathbf{M}^s \widetilde{\mathrm{P}}^{ss}_{k-1} (\mathbf{M}^s)^T + \widetilde{\mathbf{Q}}^s. \tag{7.5.3}$$

For the Gaussian random walk model, we may use a scalar version of this equation, given by

$$\widetilde{\mathrm{P}}^{ss}_k = (\mathrm{M}^s)^2 \widetilde{\mathrm{P}}^{ss}_{k-1} + \widetilde{\mathrm{Q}}^s = (\mathrm{M}^s)^{2k} \widetilde{\mathrm{P}}^t_0 + \sum_{n=0}^{k-1} (\mathrm{M}^s)^{2n} \widetilde{\mathrm{Q}}^s, \tag{7.5.4}$$

where $\widetilde{\mathrm{P}}^{ss}_0 = \widetilde{\mathrm{P}}^{ss,f}_0$ and $\mathrm{M}^s = \exp(-1/2)$. Noting that the summation term in (7.5.4) is a geometric series we can express $\widetilde{\mathrm{P}}^{ss}_k$ as

$$\widetilde{\mathrm{P}}^{ss}_k = e^{-k} \widetilde{\mathrm{P}}^{ss}_0 + \frac{1 - e^{-k}}{1 - e^{-1}} \widetilde{\mathrm{Q}}^s. \tag{7.5.5}$$

For large k, the first term in equation (7.5.5) decays to zero while the second term tends to the limit $\widetilde{\mathrm{Q}}^s/(1 - e^{-1})$. Hence, after a burn-in period, the error due to unresolved scales for the SKF and RKF is primarily determined by the size of $\widetilde{\mathrm{Q}}^s$ and increases each time-step.

7.6 Numerical Experiments

In this section we apply the OKF, RKF and SKF to the random walk model defined in section 7.5.1 with filter parameters and error statistics assumptions detailed in section 7.5.2.

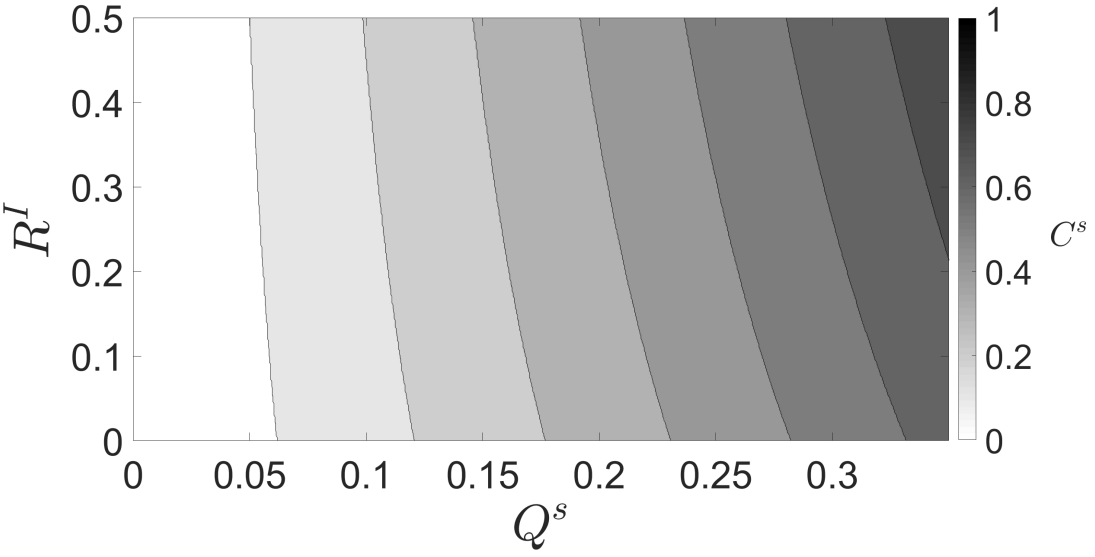


Figure 7.2: Contour plot of the values of C^s that give the minimum true large-scale analysis error variance at the final assimilated observation for the SKF for different ratios of Q^s and R^I.

7.6.1 Determining the optimal C^s

Before using the SKF, we first need to approximate the matrix C^s (see Table 7.1). To find the optimal value of C^s over the whole assimilation window we carry out numerical experiments for a range of values of R^I and Q^s. Both of these parameters will affect the magnitude of the true large-scale analysis error variance. For each (R^I, Q^s) parameter pair, we test a number of values of C^s to determine the value of C^s which gives the smallest true large-scale analysis error variance at the final assimilated observation. As we are calculating the variances only, the calculation is deterministic and the choice of noise realisation is irrelevant. For this experiment, we assimilate 15 observations. We start with $C^s = 0$ and increase C^s in steps of $\Delta C^s = 0.001$ until $C^s = 1$.

The optimal values of C^s that produce the minimum large-scale analysis error variance for the SKF at the final time-step are shown in figure 7.2. The optimal value of C^s increases as both R^I and Q^s increase. In particular, the optimal value of C^s is most sensitive to any increase in Q^s as the error due to unresolved scales is primarily determined by this error variance in our model. While not as sensitive, we find that large R^I also affects the optimal value of C^s. This is because the optimal value of C^s is a function of R^I and Q^l after the initial time. We also find that for small R^I and Q^s the optimal value of C^s over the whole assimilation window is similar to $\widetilde{P}^{ss}$ given by equation (7.5.5) for the final time-step. For large R^I and Q^s, the optimal value of C^s is approximately 1.4 times larger than $\widetilde{P}^{ss}$ evaluated at the final time-step.

In operational settings we would not be able to optimise C^s in this way. However,

it may be possible to approximate part of the representation error covariance using high resolution observations (Oke and Sakov, 2008) or model data (Waller et al., 2014b) and use these approximate representation error values to guide the choice of C^s. To mimic this situation in our experiments, we create an ensemble of 50,000 realizations of x^s for the length of the assimilation windows using the random walk model (7.5.1) and calculate the variance, averaged over the whole ensemble and time. The variance of this ensemble will be denoted S. The variance, S, represents an approximation to the total small-scale variability over the assimilation window. We now compare the values of C^s computed in Figure 7.2 with the values of S.

Figure 7.3(a) shows the optimal C^s values when $\mathrm{R}^I = 0.1$ (dashed line) and $\mathrm{R}^I = 0.5$ (dotted line) for different values of Q^s. The grey region shows all points between S and 2S. As Q^s is increased, the variance S also increases. Both optimal C^s lines lie within the shaded region for nearly all Q^s. We note that when there is little small-scale variability (i.e. $\mathrm{Q}^s \approx 0$) the optimal C^s values are less than S but both are close to zero. Figure 7.3(b) shows the effect of changing C^s on the SKF true large-scale analysis error variance (solid line) when $\mathrm{R}^I = 0.1$ and $\mathrm{Q}^s = 0.35$ in comparison to the true large scale analysis error variances for the RKF and OKF. Thus, for these experiments, a reasonable rule of thumb to avoid areas where the SKF under- or overcompensates for the error due to unresolved scales, is to choose $\mathrm{S} < \mathrm{C}^s < 2\mathrm{S}$.

7.6.2 Comparison of the SKF with the RKF and OKF

Using the optimal values of C^s calculated in figure 7.2, we now carry out experiments comparing the SKF and RKF for a range of values of R^I and Q^s relative to the OKF. The results are illustrated in terms of relative error percentage for the RKF in figure 7.4(a) and the SKF in figure 7.4(b). To generate these plots, $\mathrm{R}^I \in [0, 0.5]$ and $\mathrm{Q}^s \in [0, 0.35]$ have been discretised into 50 equally-spaced points. The relative error percentage is defined as

$$\text{relative error percentage} = \frac{\left|\widetilde{\mathrm{P}}^{ll,a}_{RKF/SKF} - \widetilde{\mathrm{P}}^{ll,a}_{OKF}\right|}{\widetilde{\mathrm{P}}^{ll,a}_{OKF}} \times 100\%, \tag{7.6.1}$$

where $|\cdot|$ indicates the absolute value and each term is evaluated at the end of the assimilation window. In these experiments, the SKF always has a true analysis error variance smaller than or equal to the RKF. When there is no error due to unresolved scales (i.e. $\mathrm{Q}^s = 0$) we have that $\mathrm{C}^s = 0$ is the optimal value for C^s and the SKF would reduce to the RKF. The largest relative error percentages for both the RKF and SKF occur when there is large uncertainty due to unresolved scales (large Q^s) and small R^I and the smallest differences are when Q^s is small. We also find that larger values of R^I limit the difference in performance between the RKF

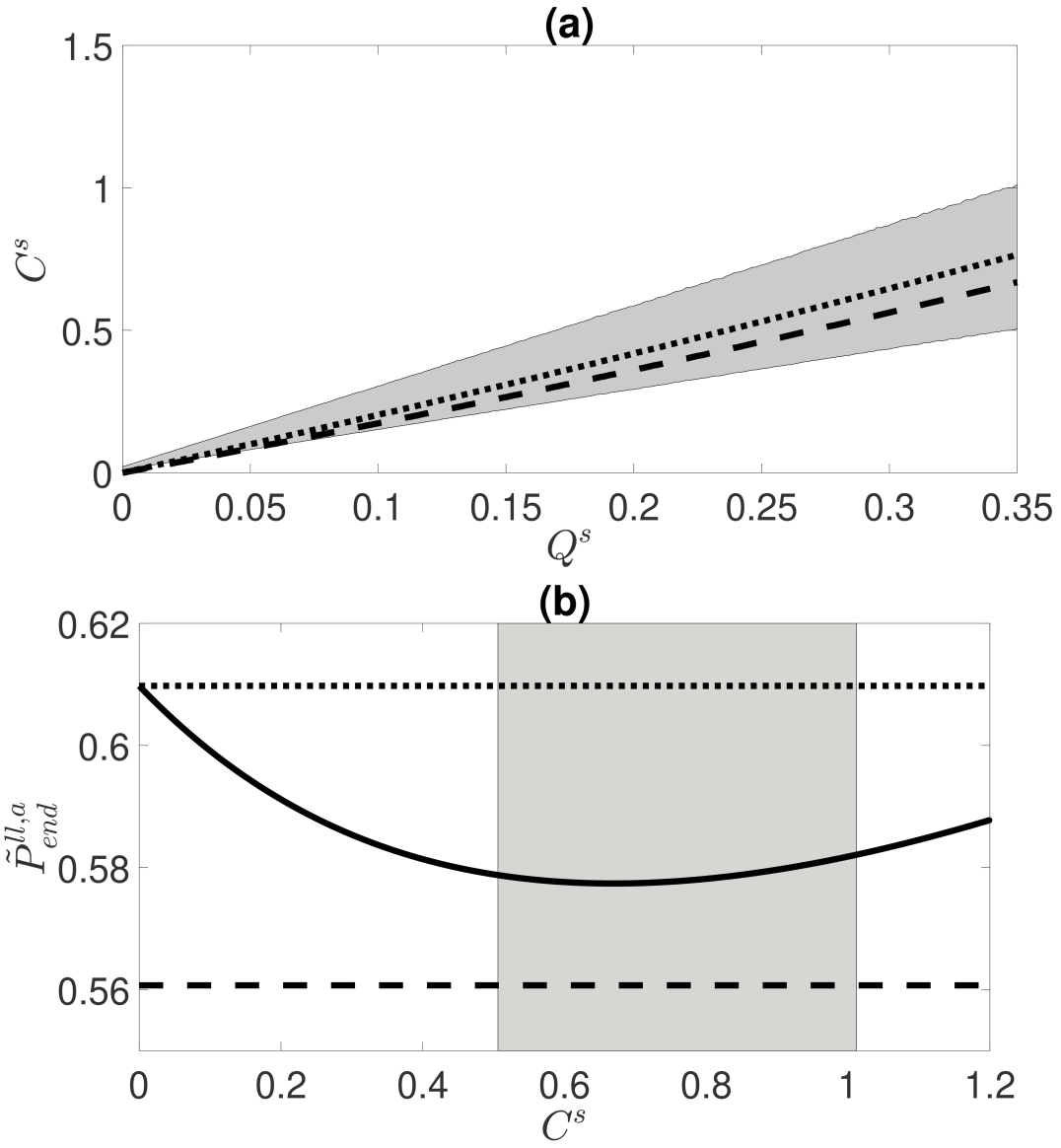


Figure 7.3: (a): The optimal C^s values when $R^I = 0.1$ (dashed line) and $R^I = 0.5$ (dotted line) as functions of Q^s. The grey region shows all points between S (lower edge) and 2S (upper edge) for different values of Q^s. (b): The effect of changing C^s on the final true large-scale analysis error covariance for the SKF (solid line) when $Q^s = 0.35$ and $R^I = 0.1$. Also shown are the OKF and RKF true large-scale analysis error variance (lower dashed line and upper dotted line respectively). The grey region shows all points between $C^s = S$ (left edge) and $C^s = 2S$ (right edge). The optimal value of C^s (i.e. the minimum of the solid line) lies in this region.

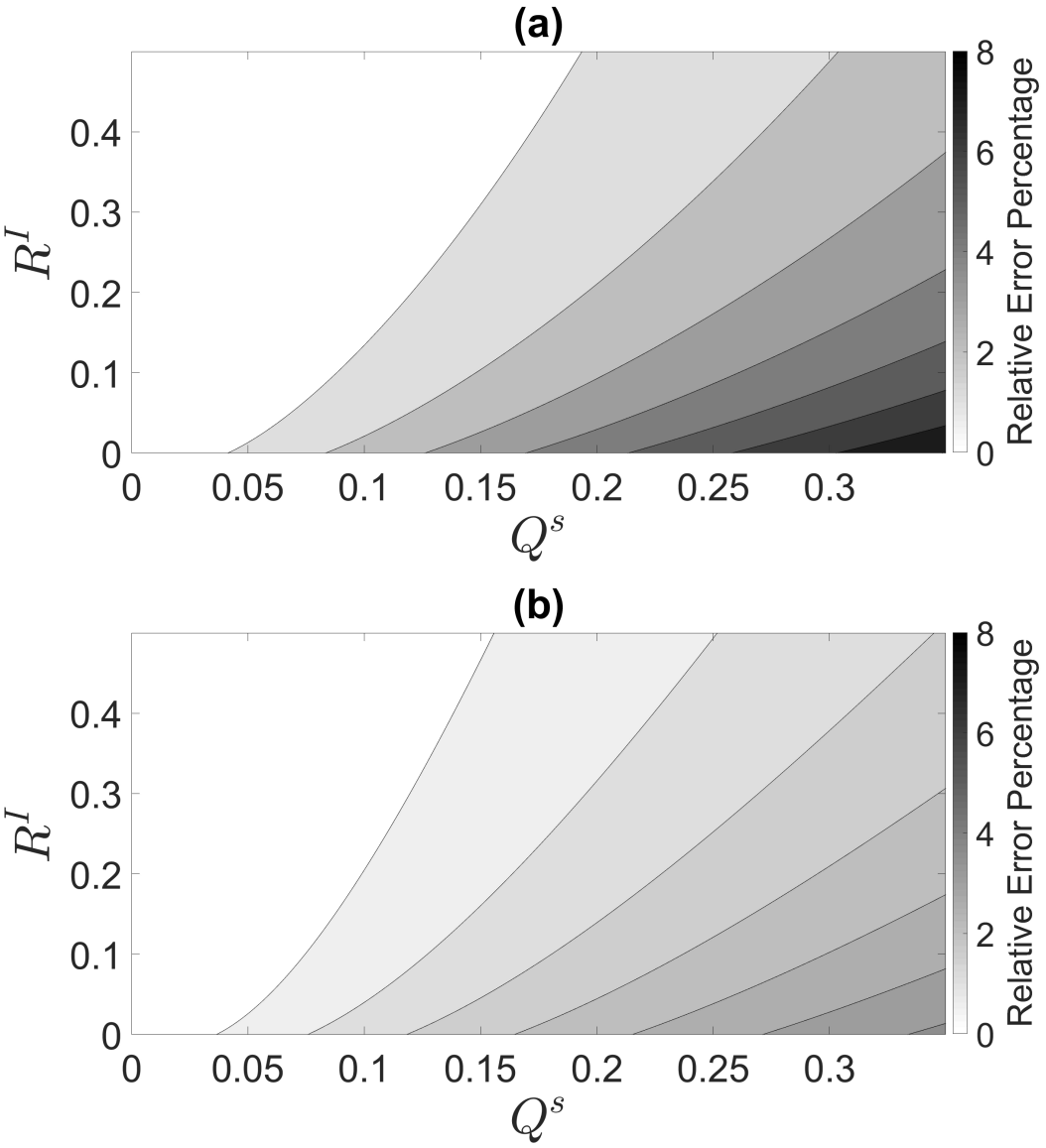


Figure 7.4: (a): Comparison of the RKF to the OKF at the final time-step in terms of relative error percentage given by equation (7.6.1). (b): Comparison of the SKF with optimal C^s to the OKF at the final time-step in terms of relative error percentage.

and SKF with the OKF. Therefore, the benefits of using the SKF are most apparent when there is considerable error due to unresolved scales and small instrument error. Comparing figure 7.4(a) to figure 7.4(b) we see that, for any fixed value of R^I, as the uncertainty due to unresolved scales is increased, the improvement of the SKF over the RKF will also increase.

To examine the performance perceived by the filter we compare it to the true performance of the filter at the final time-step. Before discussing the results, we note that the SKF perceived analysis error variance will not be a smooth field for the (R^I, Q^s) parameter pairs considered. This is because in section 7.6.1 the optimal value of C^s was calculated to a limited precision of 0.001.

In figure 7.5 we plot the difference between the perceived and true analysis error variance at the final time-step. We note that the magnitude of the difference between the perceived and true error variances is smallest for large R^I and Q^s. Here, the SKF (RKF) perceived error variance is approximately 1.25 (0.5) times the size of the true error variance. The SKF perceived-minus-truth difference shown in panel (a) is always positive for non-negligible representation uncertainty. This shows the SKF is a conservative filtering strategy when compensating for observations exhibiting error due to unresolved scales. As both R^I and Q^s are increased the SKF perceived-minus-truth difference increases. This is due to two reasons. The first reason is because the perceived analysis error variance, $\mathrm{P}^{ll,a}$, increases with larger C^s as it is calculated using the short form update (7.3.2) and the optimal C^s will be larger for higher values of R^I and Q^s. The second reason is because, for non-negligible representation uncertainty, the true analysis error variance, $\widetilde{\mathrm{P}}^{ll,a}$, will decrease as C^s approaches its optimal value. An illustrative case is provided by figure 7.3(b) for high representation uncertainty and low instrument uncertainty. Figure 7.5(b) shows the RKF perceived-minus-truth difference. This is always negative for non-negligible representation uncertainty. This shows the RKF is an overconfident filtering strategy for observations exhibiting error due to unresolved scales. The RKF is most overconfident in regimes of low instrument uncertainty and high representation uncertainty.

7.7 Representation error bias correction through state augmentation

Up to this point we have not considered observation bias in our numerical experiments. However, in operational data assimilation, most observations or their respective observation operators exhibit systematic errors which are referred to as biases. A common approach for correcting observation biases online in a Kalman filter algorithm is to augment the state vector with a bias term (Dee, 2005; Fertig et al., 2009). The bias state will then be estimated and evolved along with the state variables through the data assimilation algorithm (Friedland, 1969).

In section 7.2.3 we showed that observation errors may exhibit a representation error bias when there is a contribution from the large-scale processes to the small-scale state (i.e., when $\mathbf{M}^{sl} \neq \mathbf{0}_{N_s \times N_l}$). Throughout the remainder of this manuscript we only consider bias due to unresolved scales which is linked to the state-space representation of the small-scale processes. Since we know the exact form and origin of the observation bias in this study we may treat it as a model bias. Therefore, we

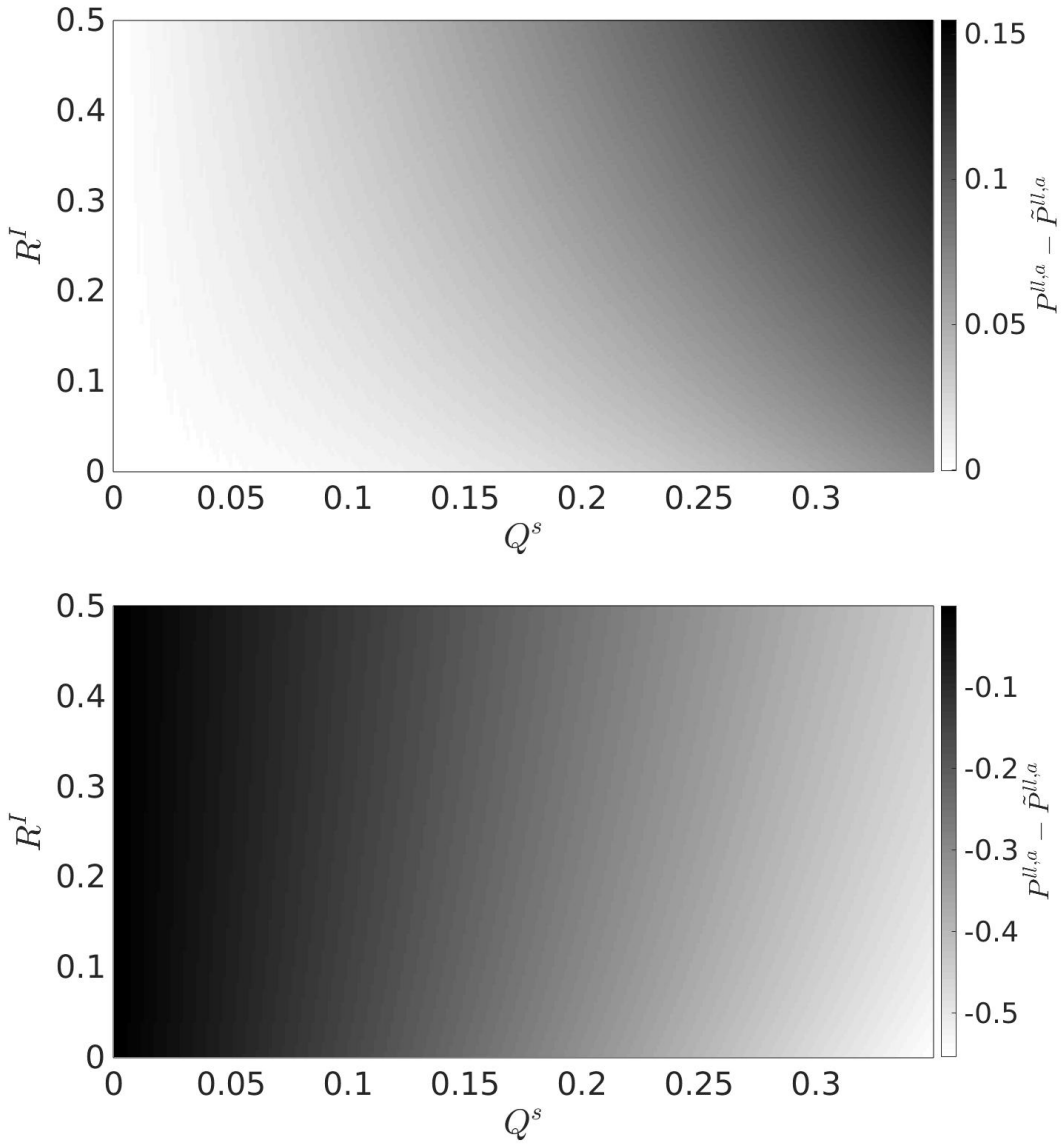


Figure 7.5: (a): Difference between the perceived and true analysis error variances at the final time-step for the SKF with optimal C^s. (b): Difference between the perceived and true analysis error variances at the final time-step for the RKF

consider the augmented state vector $\boldsymbol{x}$ with form

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}^l \\ \boldsymbol{x}^\beta \end{pmatrix}, \tag{7.7.1}$$

where $\boldsymbol{x}^\beta \in \mathbb{R}^{N_s}$ is the bias state. Thus, $\boldsymbol{x}^\beta$ is intended to represent $\langle \boldsymbol{x}^{s,t} \rangle$ (cf. (7.2.13)). For bias correction through state augmentation, we require a prior estimate of the bias and a model to forecast it. Using (7.2.2), the forecast model for the bias state is given by

$$\boldsymbol{x}_k^\beta = \mathbf{M}^{sl} \boldsymbol{x}_{k-1}^l + \mathbf{M}^s \boldsymbol{x}_{k-1}^\beta, \tag{7.7.2}$$

where we have assumed the model for the bias to be perfect. Random noise can be added to (7.7.2) to indicate that the bias evolution model is not perfect (Ménard, 2010) but is not explored here. In operational centres, the model for individual sources contributing to the bias will be unknown and models describing the total bias will be used instead. These models for the bias will be obtained from assumptions imposed on the bias such as assuming it evolves slowly or is constant in time (e.g., Lea et al., 2008). In cases such as these, the bias estimate will likely be poor as the variation of the bias with the evolution of the large-scale processes will be completely unaccounted for.

We now examine how a bias correction scheme can be implemented in conjunction with the SKF (section 7.7.1) and the RKF (section 7.7.2) to correct a bias due to unresolved scales. Table 7.3 summarizes the components for these two filters which are then substituted into the filter equations detailed in section 7.3.1.

7.7.1 The Schmidt-Kalman filter with observational bias correction

Bias correction through state augmentation is a common method used in operational centres but use of a bias correction scheme with the SKF, which will be denoted SKFbc, is novel.

We assume that we have knowledge of the processes for all scales. We further assume that we have a model and prior estimate for the bias. The filtered state vector for the SKFbc is given by equation (7.7.1) and only includes the large-scale state and the bias term. The small-scale state is split into a biased and unbiased component, i.e.

$$\boldsymbol{x}^s = \boldsymbol{x}^\beta + \boldsymbol{x}^\delta, \tag{7.7.3}$$

such that $\langle \boldsymbol{x}^s \rangle = \boldsymbol{x}^\beta$. The unbiased small-scale processes, $\boldsymbol{x}^\delta$, will be accounted for through their statistics. The full observation operator for the SKFbc is given

Table 7.3: The filter matrices and vectors for the SKFbc and RKFbc. The equations for the the two filters are obtained through substituting these terms into (7.3.1) - (7.3.4).

	SKFbc	RKFbc
State: $\boldsymbol{x}$	$\begin{pmatrix}(\boldsymbol{x}^l)^T & (\boldsymbol{x}^\beta)^T\end{pmatrix}^T \in \mathbb{R}^{N_t}$	$\begin{pmatrix}(\boldsymbol{x}^l)^T & (\boldsymbol{x}^\beta)^T\end{pmatrix}^T \in \mathbb{R}^{N_t}$
State error covariance: $\mathbf{P}$	$\begin{pmatrix}\mathbf{P}^{ll} & \mathbf{P}^{l\beta} & \mathbf{P}^{l\delta} \\ \mathbf{P}^{\beta l} & \mathbf{P}^{\beta\beta} & \mathbf{P}^{\beta\delta} \\ \mathbf{P}^{\delta l} & \mathbf{P}^{\delta\beta} & \mathbf{C}^{\delta}\end{pmatrix} \in \mathbb{R}^{N_a \times N_a}$	$\begin{pmatrix}\mathbf{P}^{ll} & \mathbf{P}^{l\beta} \\ \mathbf{P}^{\beta l} & \mathbf{P}^{\beta\beta}\end{pmatrix} \in \mathbb{R}^{N_t \times N_t}$
Observation operator: $\mathbf{H}$	$\begin{pmatrix}\mathbf{H}^l & \mathbf{H}^\beta & \mathbf{H}^\delta\end{pmatrix} \in \mathbb{R}^{p \times N_a}$	$\begin{pmatrix}\mathbf{H}^l & \mathbf{H}^\beta\end{pmatrix} \in \mathbb{R}^{p \times N_t}$
Observation error covariance: $\mathbf{R}$	$\mathbf{R}^I + \mathbf{R}^G \in \mathbb{R}^{p \times p}$	$\mathbf{R}^I + \mathbf{R}^H \in \mathbb{R}^{p \times p}$
Innovation: $\mathbf{d}^{o,f}$	$\boldsymbol{y} - \begin{pmatrix}\mathbf{H}^l & \mathbf{H}^\beta\end{pmatrix}\begin{pmatrix}\boldsymbol{x}^{l,f} \\ \boldsymbol{x}^{\beta,f}\end{pmatrix}$	$\boldsymbol{y} - \begin{pmatrix}\mathbf{H}^l & \mathbf{H}^\beta\end{pmatrix}\begin{pmatrix}\boldsymbol{x}^{l,f} \\ \boldsymbol{x}^{\beta,f}\end{pmatrix}$
Innovation covariance: $\mathbf{D}$	$\begin{pmatrix}\mathbf{H}^l & \mathbf{H}^\beta & \mathbf{H}^\delta\end{pmatrix}\begin{pmatrix}\mathbf{P}^{ll,f} & \mathbf{P}^{l\beta,f} & \mathbf{P}^{l\delta,f} \\ \mathbf{P}^{\beta l,f} & \mathbf{P}^{\beta\beta,f} & \mathbf{P}^{\beta\delta,f} \\ \mathbf{P}^{\delta l,f} & \mathbf{P}^{\delta\beta,f} & \mathbf{C}^{\delta}\end{pmatrix}\begin{pmatrix}(\mathbf{H}^l)^T \\ (\mathbf{H}^\beta)^T \\ (\mathbf{H}^\delta)^T\end{pmatrix} + \mathbf{R}$	$\begin{pmatrix}\mathbf{H}^l & \mathbf{H}^\beta\end{pmatrix}\begin{pmatrix}\mathbf{P}^{ll,f} & \mathbf{P}^{l\beta,f} \\ \mathbf{P}^{\beta l,f} & \mathbf{P}^{\beta\beta,f}\end{pmatrix}\begin{pmatrix}(\mathbf{H}^l)^T \\ (\mathbf{H}^\beta)^T\end{pmatrix} + \mathbf{R}$
Kalman gain: $\mathbf{K}$	State update $\begin{pmatrix}\mathbf{K}^l \\ \mathbf{K}^\beta\end{pmatrix} = \begin{pmatrix}\mathbf{P}^{ll,f}(\mathbf{H}^l)^T + \mathbf{P}^{l\beta,f}(\mathbf{H}^\beta)^T + \mathbf{P}^{l\delta,f}(\mathbf{H}^\delta)^T \\ \mathbf{P}^{\beta l,f}(\mathbf{H}^l)^T + \mathbf{P}^{\beta\beta,f}(\mathbf{H}^\beta)^T + \mathbf{P}^{\beta\delta,f}(\mathbf{H}^\delta)^T\end{pmatrix}\mathbf{D}^{-1}$ Covariance update $\begin{pmatrix}\mathbf{K}^l \\ \mathbf{K}^\beta \\ \mathbf{K}^\delta\end{pmatrix} = \begin{pmatrix}\mathbf{P}^{ll,f}(\mathbf{H}^l)^T + \mathbf{P}^{l\beta,f}(\mathbf{H}^\beta)^T + \mathbf{P}^{l\delta,f}(\mathbf{H}^\delta)^T \\ \mathbf{P}^{\beta l,f}(\mathbf{H}^l)^T + \mathbf{P}^{\beta\beta,f}(\mathbf{H}^\beta)^T + \mathbf{P}^{\beta\delta,f}(\mathbf{H}^\delta)^T \\ \mathbf{0}_{N_s \times p}\end{pmatrix}\mathbf{D}^{-1}$	$\begin{pmatrix}\mathbf{K}^l \\ \mathbf{K}^\beta\end{pmatrix} = \begin{pmatrix}\mathbf{P}^{ll,f}(\mathbf{H}^l)^T + \mathbf{P}^{l\beta,f}(\mathbf{H}^\beta)^T \\ \mathbf{P}^{\beta l,f}(\mathbf{H}^l)^T + \mathbf{P}^{\beta\beta,f}(\mathbf{H}^\beta)^T\end{pmatrix}\mathbf{D}^{-1}$
Forecast model: $\mathbf{M}$	State update $\begin{pmatrix}\mathbf{M}^l & \mathbf{0}_{N_l \times N_s} \\ \mathbf{M}^{sl} & \mathbf{M}^s\end{pmatrix} \in \mathbb{R}^{N_l \times N_l}$ Covariance update $\begin{pmatrix}\mathbf{M}^l & \mathbf{0}_{N_l \times N_s} & \mathbf{0}_{N_s \times N_s} \\ \mathbf{M}^{sl} & \mathbf{M}^s & \mathbf{0}_{N_s \times N_s} \\ \mathbf{0}_{N_s \times N_l} & \mathbf{0}_{N_s \times N_s} & \mathbf{M}^s\end{pmatrix} \in \mathbb{R}^{N_a \times N_a}$	$\begin{pmatrix}\mathbf{M}^l & \mathbf{0}_{N_l \times N_s} \\ \mathbf{M}^{sl} & \mathbf{M}^s\end{pmatrix} \in \mathbb{R}^{N_t \times N_t}$
Model error covariance: $\mathbf{Q}$	$\begin{pmatrix}\mathbf{Q}^{ll} & \mathbf{0}_{N_s \times N_s} & \mathbf{Q}^{ls} \\ \mathbf{0}_{N_s \times N_l} & \mathbf{0}_{N_s \times N_s} & \mathbf{0}_{N_s \times N_s} \\ \mathbf{Q}^{sl} & \mathbf{0}_{N_s \times N_s} & \mathbf{Q}^{ss}\end{pmatrix} \in \mathbb{R}^{N_a \times N_a}$	$\begin{pmatrix}\mathbf{Q}^{ll} & \mathbf{0}_{N_l \times N_s} \\ \mathbf{0}_{N_s \times N_l} & \mathbf{0}_{N_s \times N_s}\end{pmatrix} \in \mathbb{R}^{N_t \times N_t}$

by

$$\begin{pmatrix} \mathbf{H}^l & \mathbf{H}^\beta & \mathbf{H}^\delta \end{pmatrix} \in \mathbb{R}^{p \times N_a}, \tag{7.7.4}$$

where $\mathbf{H}^\beta \in \mathbb{R}^{p \times N_s}$ and $\mathbf{H}^\delta \in \mathbb{R}^{p \times N_s}$ are the linear observation operators which map the bias and unbiased small-scale states into observation space respectively. However, as with the SKF, the analysis update equation (7.3.1) uses a forecast innovation that takes no account of the unbiased small-scales,

$$\mathbf{d}^{o,f} = \boldsymbol{y} - \begin{pmatrix} \mathbf{H}^l & \mathbf{H}^\beta \end{pmatrix} \begin{pmatrix} \boldsymbol{x}^{l,f} \\ \boldsymbol{x}^{\beta,f} \end{pmatrix}. \tag{7.7.5}$$

This innovation is unbiased and hence the large-scale analysis errors are also unbiased. The Kalman gain for the SKFbc consists of a large-scale gain $\mathbf{K}^l \in \mathbb{R}^{N_l \times p}$ and a bias estimate gain $\mathbf{K}^\beta \in \mathbb{R}^{N_s \times p}$ given by

$$\begin{pmatrix} \mathbf{K}^l \\ \mathbf{K}^\beta \end{pmatrix} = \begin{pmatrix} \mathbf{P}^{ll,f}(\mathbf{H}^l)^T + \mathbf{P}^{l\beta,f}(\mathbf{H}^\beta)^T + \mathbf{P}^{l\delta,f}(\mathbf{H}^\delta)^T \\ \mathbf{P}^{\beta l,f}(\mathbf{H}^l)^T + \mathbf{P}^{\beta\beta,f}(\mathbf{H}^\beta)^T + \mathbf{P}^{\beta\delta,f}(\mathbf{H}^\delta)^T \end{pmatrix} \mathbf{D}^{-1}, \tag{7.7.6}$$

where $\mathbf{P}^{l\beta} \in \mathbb{R}^{N_l \times N_s}$ is the perceived cross-covariance between the large-scale errors and bias estimate errors, $\mathbf{P}^{l\delta} \in \mathbb{R}^{N_l \times N_s}$ is the perceived cross-covariance between the large-scale errors and unbiased small-scale errors, $\mathbf{P}^{\beta\beta} \in \mathbb{R}^{N_s \times N_s}$ is the perceived covariance of the bias estimate errors and $\mathbf{P}^{\beta\delta} \in \mathbb{R}^{N_s \times N_s}$ is the perceived cross-covariance between the bias estimate errors and unbiased small-scale errors. The perceived augmented innovation covariance $\mathbf{D}$ is given in Table 7.3. This increases the uncertainty the filter attributes to the forecast innovation compared with the standard SKF. The additional uncertainty is a result of the errors accrued in the estimation of the bias. The term $\mathbf{H}^s\mathbf{C}^\delta(\mathbf{H}^s)^T$ in the SKFbc innovation error covariance corresponds to the variability of the unbiased small-scale processes.

The SKFbc equations are obtained from augmenting the large-scale terms with bias terms and defining the cross-covariance terms appropriately. The analysis state update for the SKFbc is then

$$\begin{pmatrix} \boldsymbol{x}^{l,t} \\ \boldsymbol{x}^{\beta,a} \end{pmatrix}_k = \begin{pmatrix} \boldsymbol{x}^{l,f} \\ \boldsymbol{x}^{\beta,f} \end{pmatrix}_k + \begin{pmatrix} \mathbf{K}^l \\ \mathbf{K}^\beta \end{pmatrix}_k \left(\boldsymbol{y}_k - \begin{pmatrix} \mathbf{H}^l & \mathbf{H}^\beta \end{pmatrix}_k \begin{pmatrix} \boldsymbol{x}^{l,f} \\ \boldsymbol{x}^{\beta,f} \end{pmatrix}_k \right). \tag{7.7.7}$$

To obtain the analysis error covariance update we augment the gain (7.7.6) with $\mathbf{K}^\delta = \mathbf{0}_{N_s \times p}$ and substitute into the short-form analysis error covariance update (7.3.2). To mirror the SKF analysis error covariance update equations (7.3.12)-

(7.3.14), we express the SKFbc analysis error covariance update equations as

$$\begin{pmatrix} \mathbf{P}^{ll,a} & \mathbf{P}^{l\beta,a} \\ \mathbf{P}^{\beta l,a} & \mathbf{P}^{\beta\beta,a} \end{pmatrix}_k = \begin{pmatrix} \mathbf{I}_{N_l} - \mathbf{K}^l\mathbf{H}^l & -\mathbf{K}^l\mathbf{H}^\beta \\ -\mathbf{K}^\beta\mathbf{H}^l & \mathbf{I}_{N_s} - \mathbf{K}^\beta\mathbf{H}^\beta \end{pmatrix}_k \begin{pmatrix} \mathbf{P}^{ll,f} & \mathbf{P}^{l\beta,f} \\ \mathbf{P}^{\beta l,f} & \mathbf{P}^{\beta\beta,f} \end{pmatrix}_k - \begin{pmatrix} \mathbf{K}^l\mathbf{H}^\delta\mathbf{P}^{\delta l,f} & \mathbf{K}^l\mathbf{H}^\delta\mathbf{P}^{\delta\beta,f} \\ \mathbf{K}^\beta\mathbf{H}^\delta\mathbf{P}^{\delta l,f} & \mathbf{K}^\beta\mathbf{H}^\delta\mathbf{P}^{\delta\beta,f} \end{pmatrix}_k, \tag{7.7.8}$$

$$\begin{pmatrix} \mathbf{P}^{l\delta,a} \\ \mathbf{P}^{\beta\delta,a} \end{pmatrix}_k = \begin{pmatrix} \mathbf{I}_{N_l} - \mathbf{K}^l\mathbf{H}^l & -\mathbf{K}^l\mathbf{H}^\beta \\ -\mathbf{K}^\beta\mathbf{H}^l & \mathbf{I}_{N_s} - \mathbf{K}^\beta\mathbf{H}^\beta \end{pmatrix}_k \begin{pmatrix} \mathbf{P}^{l\delta,f} \\ \mathbf{P}^{\beta\delta,f} \end{pmatrix}_k - \begin{pmatrix} \mathbf{K}^l\mathbf{H}^\delta\mathbf{C}^\delta \\ \mathbf{K}^\beta\mathbf{H}^\delta\mathbf{C}^\delta \end{pmatrix}_k, \tag{7.7.9}$$

$$\begin{pmatrix} \mathbf{P}^{\delta l,a} & \mathbf{P}^{\delta\beta,a} \end{pmatrix}_k = \begin{pmatrix} \left(\mathbf{P}^{l\delta,f}\right)^T & \left(\mathbf{P}^{\beta\delta,f}\right)^T \end{pmatrix}_k. \tag{7.7.10}$$

Since in the context of the SKFbc the complete model evolving all scales is assumed to be known, it is appropriate to update the bias term using this model (7.7.2). Thus, the forecast state update is given by

$$\begin{pmatrix} \boldsymbol{x}^{l,f} \\ \boldsymbol{x}^{\beta,f} \end{pmatrix}_{k+1} = \begin{pmatrix} \mathbf{M}^l & \mathbf{0}_{N_l\times N_s} \\ \mathbf{M}^{sl} & \mathbf{M}^s \end{pmatrix} \begin{pmatrix} \boldsymbol{x}^{l,t} \\ \boldsymbol{x}^{\beta,a} \end{pmatrix}_k. \tag{7.7.11}$$

For the forecast error covariance we need the model for the unbiased small-scale processes. To determine this model we use the definition (7.7.3) together with the bias evolution equation (7.7.2), to give

$$\boldsymbol{x}_k^\delta = \mathbf{M}^s\boldsymbol{x}_{k-1}^\delta + \boldsymbol{\eta}_k^s. \tag{7.7.12}$$

Note that the small-scale model error $\boldsymbol{\eta}^s$ is assumed to be unbiased. To mirror the SKF forecast error covariance update equations (7.3.16)-(7.3.18), we express the SKFbc forecast error covariance updates as

$$\begin{pmatrix} \mathbf{P}^{ll,f} & \mathbf{P}^{l\beta,f} \\ \mathbf{P}^{\beta l,f} & \mathbf{P}^{\beta\beta,f} \end{pmatrix}_{k+1} = \begin{pmatrix} \mathbf{M}^l & \mathbf{0}_{N_l\times N_s} \\ \mathbf{M}^{sl} & \mathbf{M}^s \end{pmatrix} \begin{pmatrix} \mathbf{P}^{ll,a} & \mathbf{P}^{l\beta,a} \\ \mathbf{P}^{\beta l,a} & \mathbf{P}^{\beta\beta,a} \end{pmatrix}_k \begin{pmatrix} \mathbf{M}^l & \mathbf{0}_{N_l\times N_s} \\ \mathbf{M}^{sl} & \mathbf{M}^s \end{pmatrix}^T, \tag{7.7.13}$$

$$\begin{pmatrix} \mathbf{P}^{l\delta,f} \\ \mathbf{P}^{\beta\delta,f} \end{pmatrix}_{k+1} = \begin{pmatrix} \mathbf{M}^l & \mathbf{0}_{N_l\times N_s} \\ \mathbf{M}^{sl} & \mathbf{M}^s \end{pmatrix} \begin{pmatrix} \mathbf{P}^{l\delta,a} \\ \mathbf{P}^{\beta\delta,a} \end{pmatrix}_k (\mathbf{M}^s)^T, \tag{7.7.14}$$

$$\begin{pmatrix} \mathbf{P}^{\delta l,f} & \mathbf{P}^{\delta\beta,f} \end{pmatrix}_{k+1} = \begin{pmatrix} \mathbf{P}^{l\delta,f} \\ \mathbf{P}^{\beta\delta,f} \end{pmatrix}_{k+1}^T. \tag{7.7.15}$$

The prescribed unbiased small-scale error covariance $\mathbf{C}^\delta$ is assumed constant in time and is not updated.

The SKFbc allows us to correct biases due to unresolved scales and consider the effects of the unbiased small-scale processes on the large-scale state. A key ad-

vantage in this method is that the cross-correlations between the large-scale errors and small-scale errors are retained. However, the SKF is a computationally expensive procedure. This issue is exacerbated by the use of state augmentation for bias correction.

7.7.2 The reduced-state Kalman filter with observation bias correction

To save on the computational expense incurred by the SKFbc we can disregard the unbiased small-scale processes to obtain the reduced-state Kalman filter with bias correction (RKFbc). As before, we augment the large scale state vector with a bias term, so that the estimated state is given by (7.7.1). The observation operator is also augmented and takes the form,

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}^l & \mathbf{H}^\beta \end{pmatrix}. \tag{7.7.16}$$

As with the SKFbc, the forecast innovation (7.7.5) used in the analysis update (7.3.1) takes no account of unbiased small scale error. Therefore, a properly specified observation error covariance for the RKFbc contains both instrument error and representation error (i.e. $\mathbf{R} = \mathbf{R}^I + \mathbf{R}^H$).

Similarly to the SKFbc, the Kalman gain for the RKFbc consists of a large-scale gain $\mathbf{K}^l \in \mathbb{R}^{N_l \times p}$ and a bias estimate gain $\mathbf{K}^\beta \in \mathbb{R}^{N_s \times p}$ given by

$$\begin{pmatrix} \mathbf{K}^l \\ \mathbf{K}^\beta \end{pmatrix} = \begin{pmatrix} \mathbf{P}^{ll,f}(\mathbf{H}^l)^T + \mathbf{P}^{l\beta,f}(\mathbf{H}^\beta)^T \\ \mathbf{P}^{\beta l,f}(\mathbf{H}^l)^T + \mathbf{P}^{\beta\beta,f}(\mathbf{H}^\beta)^T \end{pmatrix} \mathbf{D}^{-1}, \tag{7.7.17}$$

where the perceived innovation covariance $\mathbf{D} \in \mathbb{R}^{p \times p}$ is given in Table 7.3. Thus, the analysis state update for the RKFbc is obtained through substitution of the gain matrix (7.7.17) and forecast innovation (7.7.5) into the linear filter analysis state update equation (7.3.1). Likewise, the analysis error covariance update equation is obtained through substitution of the gain matrix (7.7.17) into the short form analysis error covariance update (7.3.2) along with the observation operator (7.7.16).

For the RKF we assumed knowledge of the large-scale processes only. Hence, the model for the bias due to unresolved scales would be unknown and further assumptions required for the observation bias correction scheme. Nevertheless, to provide a direct comparison we will use the same model as the SKFbc (7.7.11) for the forecast state update. This model and a consistent model error covariance matrix (see Table 7.3) are used for the augmented analysis error covariance update (7.3.2).

Comparison of the OKF column in table 7.1 and the RKFbc column in table 7.3 shows the two filters have similar components as a result of the bias correction

through state augmentation approach. The key difference between the two filters is the model error covariance expressions. The OKF accounts for the uncertainty in all scales and so uses the full model error covariance. The RKFbc accounts for the uncertainty in the large-scales and the estimate of the bias. Since the model for the bias (7.7.2) has been assumed perfect the RKFbc will only account for large-scale model error. We note that, as no knowledge of the small-scale processes is assumed for the RKFbc, the forecast model will differ in practice from that of the OKF as the RKFbc bias forecast model would come from additional assumptions placed on the bias.

The RKFbc is a computationally cheaper alternative to the SKFbc for online bias correction that takes no account of unbiased small-scale errors, except through the choice of observation error covariance.

7.7.3 True analysis error equations for bias correcting filters

The true analysis error equation for the SKFbc and RKFbc will differ from the case 2 true analysis error equation (7.4.4) due to the innovation (7.7.5). The change will only be in the large-scale part of the true analysis error equations as the small-scale state is not analysed by either filter. The large-scale true analysis error equation for the bias correction filters is obtained from subtracting $\mathbf{K}_k^l \mathbf{H}^\beta \boldsymbol{x}_k^{\beta,f}$ from the case 2 true analysis error equation (7.4.4) which produces

$$\mathbf{e}_k^{l,a} = \left(\mathbf{I}_{N_l} - \mathbf{K}_k^l \mathbf{H}_k^l\right) \mathbf{M}^l \mathbf{e}_{k-1}^{l,a} + \left(\mathbf{I}_{N_l} - \mathbf{K}_k^l \mathbf{H}_k^l\right) \boldsymbol{\eta}_k^l + \mathbf{K}_k^l \left(\boldsymbol{\epsilon}_k + \boldsymbol{\gamma}_k^l + \mathbf{H}_k^{s,t} \boldsymbol{x}_k^{s,t} - \mathbf{H}_k^\beta \boldsymbol{x}_k^{\beta,f}\right). \tag{7.7.18}$$

We note that $\mathbf{H}_k^{s,t} \boldsymbol{x}_k^{s,t} - \mathbf{H}_k^\beta \boldsymbol{x}_k^{\beta,f}$ has zero-mean. Taking the expectation of the outer product of $\mathbf{e}_k^{l,a}$ in equation (7.7.18), the true large-scale analysis error covariance for the bias correcting filters is given by

$$\begin{aligned}\widetilde{\mathbf{P}}_k^{ll,a} = &\left(\mathbf{I}_{N_l} - \mathbf{K}_k^l \mathbf{H}_k^l\right) \widetilde{\mathbf{P}}_k^{ll,f} \left(\mathbf{I}_{N_l} - \mathbf{K}_k^l \mathbf{H}_k^l\right)^T \\ &+ \mathbf{K}_k^l \left(\widetilde{\mathbf{R}}_k^I + \left(\mathbf{H}_k^{s,t} \boldsymbol{x}_k^{s,t} - \mathbf{H}_k^\beta \boldsymbol{x}_k^{\beta,f}\right) \left(\mathbf{H}_k^{s,t} \boldsymbol{x}_k^{s,t} - \mathbf{H}_k^\beta \boldsymbol{x}_k^{\beta,f}\right)^T\right) \left(\mathbf{K}_k^l\right)^T.\end{aligned} \tag{7.7.19}$$

The difference between the true analysis error covariance for the non-bias correcting filters and (7.7.19) is that $\widetilde{\mathbf{R}}_k^H$ has been replaced with $\left(\mathbf{H}_k^{s,t} \boldsymbol{x}_k^{s,t} - \mathbf{H}_k^\beta \boldsymbol{x}_k^{\beta,f}\right) \left(\mathbf{H}_k^{s,t} \boldsymbol{x}_k^{s,t} - \mathbf{H}_k^\beta \boldsymbol{x}_k^{\beta,f}\right)^T$, which corresponds to the uncertainty due to unresolved scales and the uncertainty in the estimate of the bias. Similarly to (7.4.5), equation (7.7.19) is still dependent on $\boldsymbol{x}^{s,t}$ and so a different form may be more suitable.

Using the definitions of the small-scale observation operator error (7.2.9), the er-

ror due to unresolved scales (7.2.11), the small-scale forecast error (7.3.9) and the identity (7.4.8) we can rewrite (7.7.18) as

$$\begin{aligned}\begin{pmatrix}\mathbf{e}^{l,a}\\ \mathbf{e}^{s,a}\end{pmatrix}_k &= \left(\mathbf{I}_{N_t} - \begin{pmatrix}\mathbf{K}_k^l\\ \mathbf{0}_{N_s\times p}\end{pmatrix}\begin{pmatrix}\mathbf{H}^l & \mathbf{H}^s\end{pmatrix}\right)\begin{pmatrix}\mathbf{e}^{l,f}\\ \mathbf{e}^{s,f}\end{pmatrix}_k \\ &\quad + \begin{pmatrix}\mathbf{K}_k^l\\ \mathbf{0}_{N_s\times p}\end{pmatrix}\left(\boldsymbol{\epsilon}_k + \boldsymbol{\gamma}_k^l + \boldsymbol{\gamma}_k^s + \mathbf{H}_k^s\boldsymbol{x}_k^{s,f} - \mathbf{H}_k^\beta\boldsymbol{x}_k^{\beta,f}\right), \end{aligned} \tag{7.7.20}$$

which is analogous to (7.4.9). In order to use (7.7.20) to obtain the true error statistics the correlations between $\mathbf{e}_k^{s,f}$ and $\mathbf{H}_k^s\boldsymbol{x}_k^{s,f} - \mathbf{H}_k^\beta\boldsymbol{x}_k^{\beta,f}$ must be considered.

7.8 Experimental methodology for bias correcting filters

7.8.1 Gaussian random walk model

To investigate the performance of the SKFbc and the RKFbc we will apply them to the random walk model detailed in section 7.5.1. To introduce a bias into the observations we will set the contribution from the large-scale processes to the small-scale state M^{sl} to be nonzero in (7.5.1). As in (7.5.1), the true observation operator for the large- and small-scale states is $\mathbf{H} = (1 \;\; 1)$ and consequently the observation operator for the bias state $\mathrm{H}^\beta = 1$ and the unbiased small-scale state $\mathrm{H}^\delta = 1$.

To calculate the true large-scale analysis error variance of the RKFbc and SKFbc we proceed as discussed in section 7.7.3. Noting that $x_0^{s,f}$ and $x_0^{\beta,f}$ are forecast by the same equation and tend to the same bias for large times we may neglect the variance of $x_k^{s,f} - x_k^{\beta,f}$ and the correlations between $x_k^{s,f} - x_k^{\beta,f}$ and $e_k^{s,f}$ as they will be small at the end of the assimilation window.

7.8.2 Initial conditions and filter parameters

The random walk model with $\mathrm{M}^{sl} = 0.05$ is used to create a reference or truth trajectory for the large- and small-scale states. For our experiments we set $x_0^{l,t} = 10$ and $x_0^{s,t} = \mathrm{M}^{sl}x_0^{l,t}/(1-\exp(-1/2))$. This choice for the small-scale truth is the limit of $x^{s,t}$ for the deterministic version of the random walk model (i.e. (7.5.1) with no model noise). Using these initial conditions, the model equivalent of the observations will be biased at each time-step. The initial prior large- and small-scale estimates are set as

$$\begin{pmatrix}x_0^{l,f}\\ x_0^{s,f}\end{pmatrix} = \begin{pmatrix}x_0^{l,t} + \alpha^l\\ x_0^{s,t} + \alpha^s\end{pmatrix}, \tag{7.8.1}$$

where $\alpha^l \sim \mathcal{N}(0, \mathrm{P}_0^{ll,f})$ and $\alpha^s \sim \mathcal{N}(0, \mathrm{P}_0^{ss,f})$ where we take $\mathrm{P}_0^{ll,f} = 1$ and $\mathrm{P}_0^{ss,f} = 0.1$. Similarly, we set the initial prior bias estimate as $x_0^{\beta,f} = x_0^{s,t} + \alpha^\beta$ where $\alpha^\beta \sim \mathcal{N}(0, \mathrm{P}_0^{ss,f})$. We take the initial cross-covariances between the forecast errors for the large-scale and bias state errors to be zero. The modelled unbiased small-scale error variance C^δ for the SKFbc will be varied for our experiments. We also take the unbiased small-scale errors to be initially uncorrelated with large-scale and bias estimate forecast errors. We set the large-scale model error variance as $\mathrm{Q}^l = 1$ throughout our experiments while Q^s will be varied. Unless otherwise specified, the instrument error variance will be set to $\mathrm{R}^I = 0.1$. For the RKFbc, we set $\mathrm{R}^H = 0$ so that the filter completely ignores the unbiased small-scale processes.

7.9 Numerical experiments with bias correcting filters

7.9.1 Comparison between bias correcting filters and non-bias correcting filters

We now consider the case of assimilating biased observations with standard and bias correcting filters. Figure 7.6 shows the analyses created by the SKF and SKFbc when assimilating biased observations for a single realization of the background, observation and model errors where $\mathrm{Q}^s = 0.3$. As optimal modelled small-scale error variances have not been calculated for the random walk model with $\mathbf{M}^{sl} \neq 0$, we set $\mathrm{C}^\delta = \mathrm{C}^s = 0.1$. These are suboptimal choices for both filters which results in a small difference in the true analysis error variances between the SKFbc (SKF) and RKFbc (RKF). Panel (a) shows an almost constant offset between the solutions of the bias-correcting and non-bias correcting schemes. Furthermore, calculating the time-average of the squared analysis errors we find the SKF error is over four times larger than the SKFbc error.

For this realization, the time-average of the squared analysis errors for the RKFbc and the SKFbc are the same to two decimal places. However, the SKFbc does have a smaller true large-scale analysis error variance than the RKFbc and the difference increases more as C^δ is more optimally chosen. The same is true for the RKF and SKF with modelled variance C^s.

In figure 7.6(b) we see the bias value estimated by the SKFbc and the small-scale true model solution for a particular realization, which is dominated by noise. The bias state $x_k^{\beta,a}$ is intended to estimate the expected value of the small-scale state evolved with the filter forecast model such that it is unaffected by small-scale noise (see (7.7.3)). Using (7.2.2) we see that $\langle x_k^{s,t} \rangle = \langle \mathrm{M}^{sl} x_{k-1}^{l,t} + \mathrm{M}^s x_{k-1}^{s,t} + \eta_k^s \rangle$ where the angular brackets indicate the mathematical expectation over the distribution of

the small-scale model errors. Here, we have plotted $x_k^{s,t}$ which is dependent on the large-scale noise and small-scale noise (cf. (7.2.2)). From this panel we see that the bias estimate is consistent with the small-scale true model solution.

Additional experiments using persistence as the forecast model for the bias state with the SKFbc have been carried out. We find that the time-average of the SKFbc squared analysis errors is approximately three times smaller than the time-average of the SKF squared analysis errors without bias correction. Nevertheless, the mean-square analysis errors for the SKFbc with the persistence bias model are more than 50% larger than when using (7.7.2). Additionally, when using persistence as the forecast model for the bias state in the RKFbc we find the time-average of the squared analysis errors is also approximately three times less than the SKF error. Hence, for this system it is more important to treat the bias due to unresolved scales than compensate for the unbiased error due to unresolved scales.

7.9.2 Determining the optimal C^δ over the assimilation window

In this section we determine the optimal C^δ over the whole assimilation window.

For the SKF, we found that the choice of C^s was key to the performance of the filter. We follow a similar procedure to section 7.6.1 to find the optimal values of the unbiased small-scale error covariance C^δ. Our experiments have an assimilation window of 15 time-steps with an observation assimilated at each time-step. To find the optimal C^δ for given parameter values for R^I and Q^s, we calculate the true large-scale analysis error variance of the SKFbc for C^δ ranging from 0 to 1 in steps of $\Delta\mathrm{C}^\delta = 0.001$ and save the value that produces the smallest variance at the final time-step.

Figure 7.7 shows the optimal C^δ for different values of Q^s and R^I. The behaviour is qualitatively similar to C^s with the SKF shown in figure 7.2 but numerical comparison is not meaningful as a different model is used. In particular, the size of C^δ is primarily determined by the magnitude of Q^s. However, we find that an increase in R^I can also result in a larger C^δ being optimal. If M^{sl} is increased, the optimal C^δ decreases as the uncertainty caused by the contribution from the large-scale processes to the small-scale state becomes more important.

7.9.3 Comparison of the bias correction filters

In this section we evaluate the performance of the SKFbc and RKFbc relative to the OKF and examine their perceived error variances.

We now compare the SKFbc and RKFbc with the OKF in terms of relative error percentage (7.6.1), plotted in figure 7.8. The SKFbc provides most improvement over

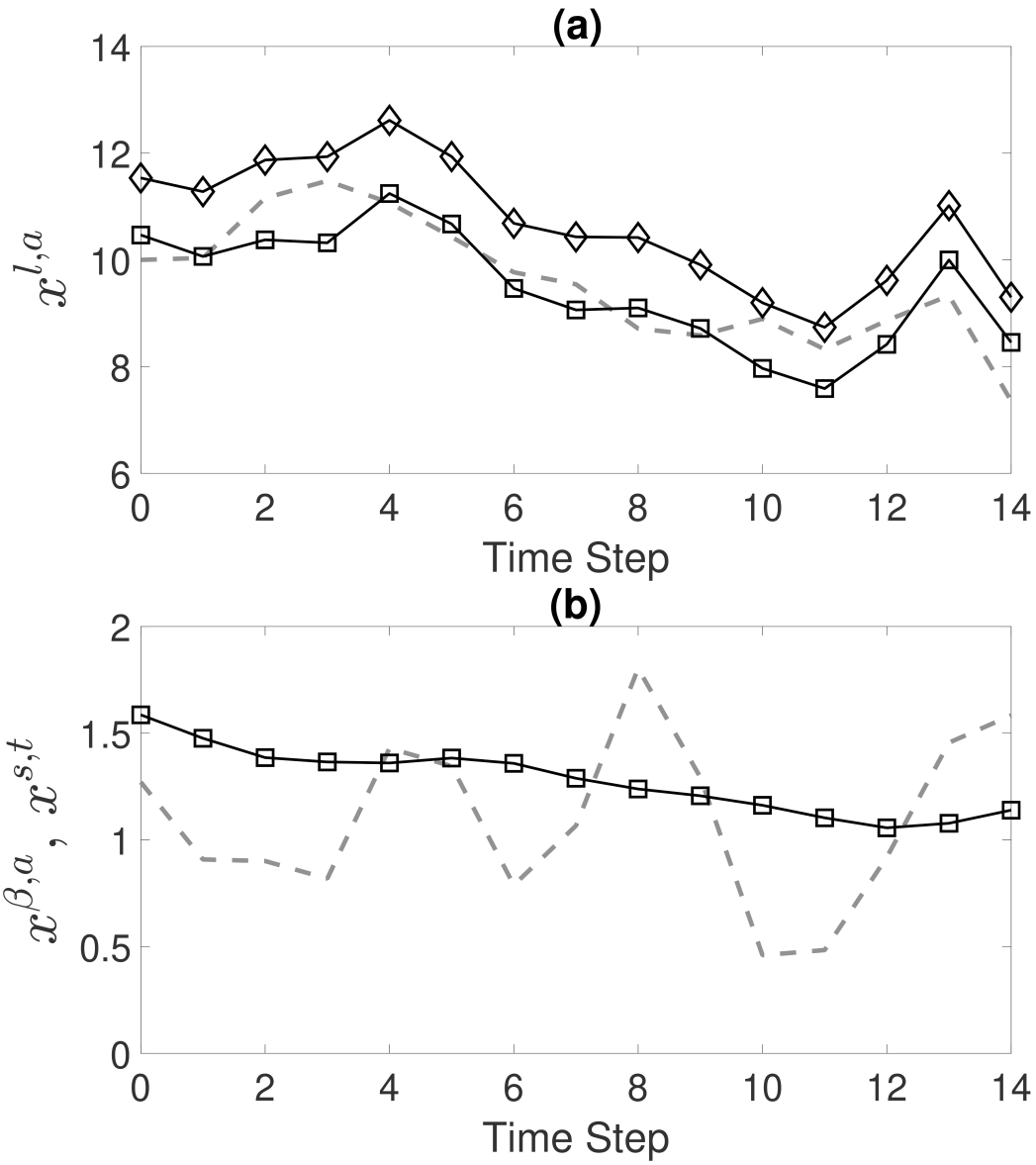


Figure 7.6: **(a)**: The large-scale analysis for the SKFbc (square markers) and SKF (diamond markers) obtained through assimilation of biased observations to recreate the true large-scale state (grey dashed line). For this realization the large-scale analysis mean-square-error for the SKFbc is 0.29 and for the SKF is 1.53. **(b)**: The SKFbc bias analysis estimate (square markers) and the true small-scale state (grey dashed line) for the same realization as the top part of this figure.

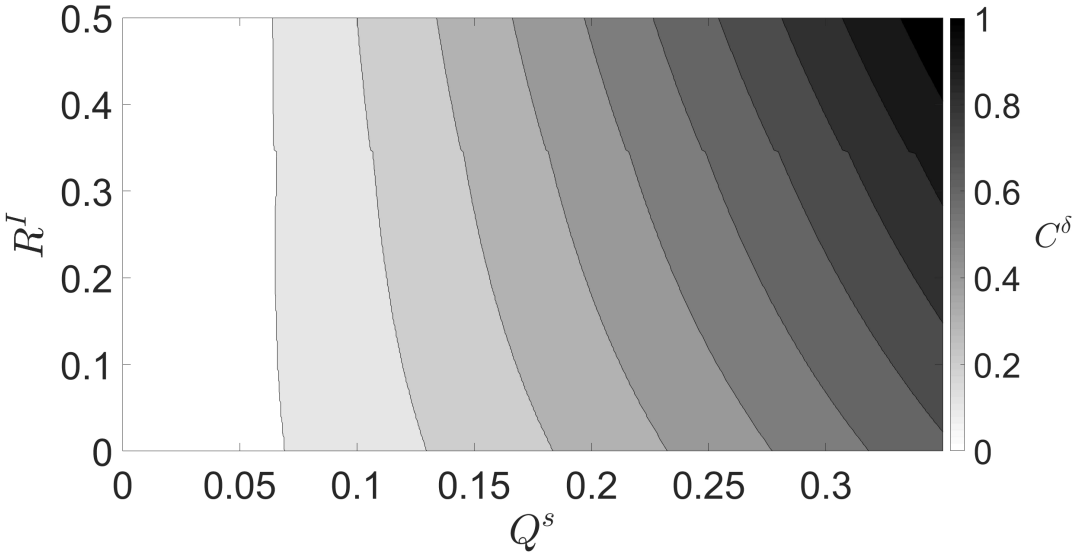


Figure 7.7: The values of C^δ which give the minimum true large-scale analysis error variance at the end of the assimilation window for the SKFbc.

the RKFbc for large Q^s and small R^I. This behaviour is qualitatively similar to the comparison between the RKF and SKF with the OKF shown in figure 7.4. We have also examined the perceived and true analysis error variances for the RKFbc and SKFbc (not plotted). The results are qualitatively similar to those given in section 7.6.2 for the RKF and SKF. Indeed, for non-negligible representation uncertainty the SKFbc (RKFbc) is a conservative (overconfident) filtering strategy as the perceived-minus-truth difference is positive (negative).

7.10 Summary and conclusion

Observations of the atmospheric state may contain information on spatio-temporal scales unable to be represented by a numerical model. The resulting error caused by this scale mismatch between the observations and numerical model is known as the error due to unresolved scales. To obtain accurate analyses from assimilation of these observations requires that the data assimilation algorithm correctly account for this error.

In this work we have considered the ability of linear filters to compensate for the error due to unresolved scales. We considered a finite-dimensional true state which could be partitioned into a large-scale state resolved by a numerical model and a small-scale state unresolved by a numerical model. The representation error was defined in this framework and a bias due to unresolved scales was shown to occur when there is a contribution from the large-scale processes to the small-scale state.

For our experiments we considered three filters: the Schmidt-Kalman filter (Janjić and Cohn, 2006) that analyses the large-scales but models the uncertainty on all

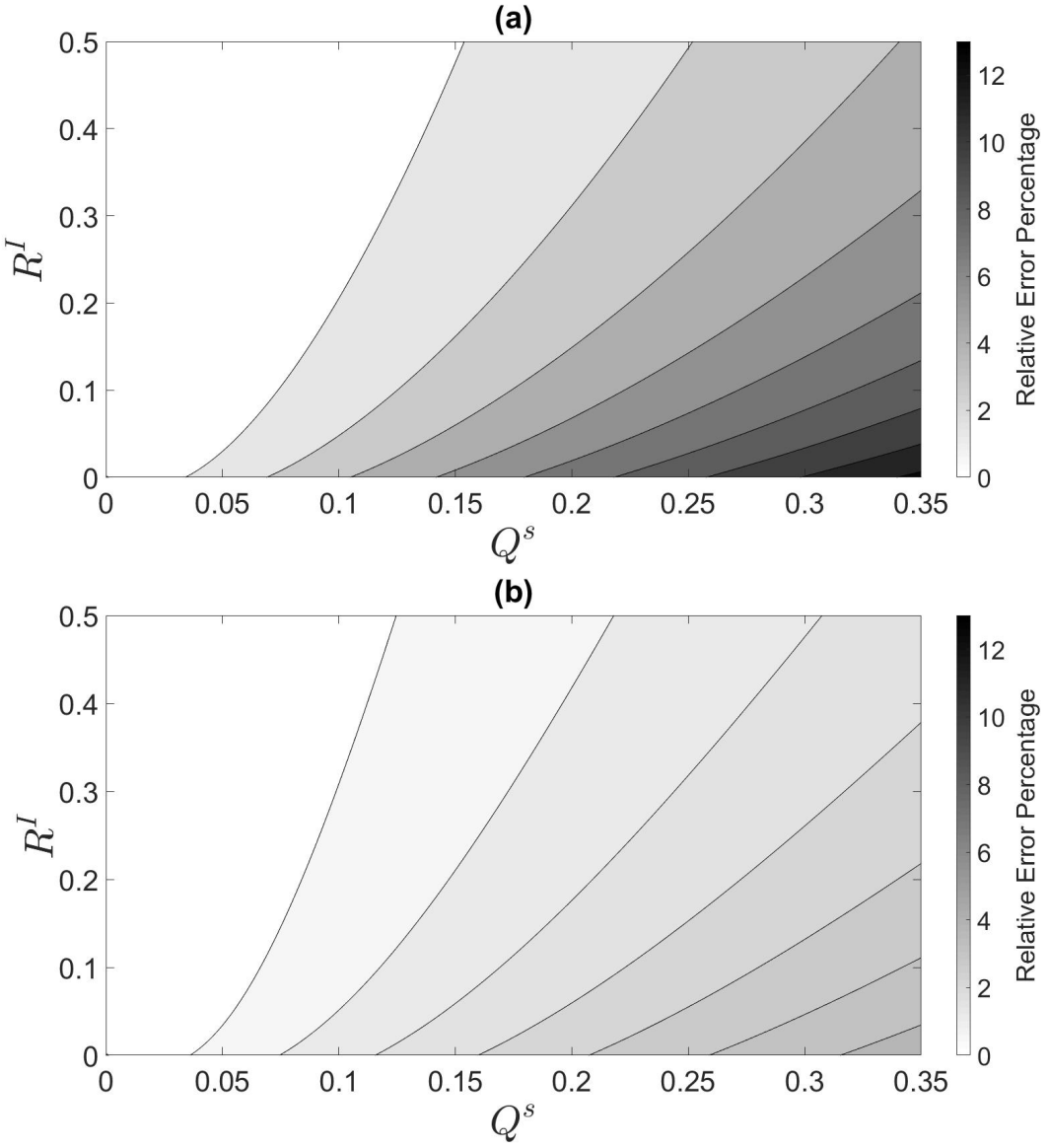


Figure 7.8: (a): Comparison of the RKFbc to the OKF in terms of relative error percentage given by equation (7.6.1). (b): Comparison of the SKFbc with optimal C^{δ} to the OKF at the final time-step in terms of relative error percentage.

scales; the optimal Kalman filter, which analyses all scales, and a reduced-state Kalman filter, which completely disregards the small-scale processes.

The three filters were tested numerically on a random walk model with one variable for the large-scale processes and one variable for the small-scale processes. The observations were taken to be the sum of the large- and small-scale states with added noise to simulate instrument error. To obtain the best performance from the Schmidt-Kalman filter we had to tune the modelled small-scale error covariance to compensate for the variability of the small-scale processes which grew over the first half of the assimilation window. The Schmidt-Kalman filter works best in regimes of high error due to unresolved scales and low instrument error provided a suitable approximate small-scale error covariance is used. Examination of the perceived error variances revealed the analysis uncertainty calculated by the Schmidt-Kalman filter is greater than the true analysis uncertainty when accounting for error due to unresolved scales.

The novel use of the Schmidt-Kalman filter with an observation bias-correction scheme was introduced as a means to correct the bias due to unresolved scales. The Schmidt-Kalman filter with a bias correction scheme proved to be a suitable method to treat observation biases and compensate for due to unresolved scales. In our experiments we found it was more important to treat an observation bias than to compensate for an unbiased error due to unresolved scales.

An important note to make regarding these experiments was that we had complete knowledge of the small-scale processes. This allowed for minimal approximations to be made to implement the Schmidt-Kalman filter and to tune the modelled error variances. In an operational setting, where all the small-scale processes are likely to be unknown, further approximations would be required. Assuming that the true small-scale state is forecast by a simple model such as persistence may be a suitable approximation for implementation. Additionally, the Schmidt-Kalman filter is a computationally expensive method due to the augmentation and propagation of the state error covariances. This must also be addressed before the filter could be considered for large problems.

7.11 Chapter summary

In this chapter, we addressed the third research question given in chapter 1: *can the Schmidt-Kalman filter effectively treat observation error and bias due to unresolved scales?* The Schmidt-Kalman filter can effectively treat observation error due to unresolved scales and was shown to be most effective in regimes of high uncertainty due to unresolved scales and low instrument uncertainty. Additionally, by incorporating a bias correction scheme into the Schmidt-Kalman filter, we have created a

novel filter that can correct observation biases and compensate for an unbiased error due to unresolved scales simultaneously. We also found that it is more important to treat an observation bias than account for an unbiased error due to unresolved scales.

The Schmidt-Kalman filter provides an alternative method to compensate for observation uncertainty due to unresolved scales in data assimilation for linear systems. In the next chapter we develop a new ensemble formulation of the Schmidt-Kalman filter suitable for nonlinear dynamical systems such as those encountered in numerical weather prediction.

Chapter 8

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

In this chapter we build upon the investigation of the Schmidt-Kalman filter in chapter 7 to answer the fourth research question given in chapter 1: *how can the Schmidt-Kalman filter be adapted for nonlinear models?* In particular, we wish to investigate:

- How can the small-scale variability utilised in the Schmidt-Kalman filter be represented in an ensemble formulation?
- How does the ensemble formulation of the Schmidt-Kalman filter perform in comparison to standard ensemble Kalman filters?

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 modelled. 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). The ETSKF compensates for the observation uncertainty due to unresolved scales in nonlinear dynamical systems by sampling a small-scale ensemble from the

representation error covariance. To test the ETSKF, we use a simple nonlinear system of ordinary differential equations with two timescales known as the swinging spring (or elastic pendulum). The performance of the ETSKF is evaluated through comparison with two versions of the ensemble transform Kalman filter (ETKF). The first version of the ETKF completely disregards the uncertainty due to unresolved scales. The second version of the ETKF includes the uncertainty due to unresolved scales in the observation error covariance matrix. Our results show that the ETSKF is most beneficial when there is more uncertainty due to unresolved scales than instrument uncertainty. In such situations, the ETSKF performs similarly to the ETKF that includes the uncertainty due to unresolved scales in the observation error covariance. Indeed, use of small-scale ensemble statistics has potential as a new approach to compensate for observation uncertainty due to unresolved scales in nonlinear dynamical systems.

8.1 Introduction

In ensemble data assimilation, the true state of a dynamical system is estimated using observational data and an ensemble of forecasts obtained from numerical models (Houtekamer and Zhang, 2016; Bannister, 2017; Vetra-Carvalho et al., 2018). This is achieved through consideration of the uncertainty associated with the observations and the uncertainty characterised by the forecast ensemble. The observation uncertainty can be attributed to two sources: the instrument uncertainty and the representation uncertainty. The instrument uncertainty arises from any errors in the observation process caused by the measuring instrument. The representation uncertainty is caused by the misrepresentation of an observation by the modelling system (Janjić et al., 2018). The representation uncertainty can be further divided into three components: pre-processing uncertainty, observation operator uncertainty, and the uncertainty due to unresolved scales. The pre-processing uncertainty is caused by the incorrect preparation of an observation for assimilation. The observation operator uncertainty is the result of an incorrect or approximate mapping of the state variables to observation space in the analysis update. The uncertainty due to unresolved scales arises when there is a mis-match in scales observed and modelled.

The error due to unresolved scales is state-dependent and correlated (Janjić and Cohn, 2006). For example, for meso-scale numerical weather prediction (NWP) (12km grid spacing), Waller et al. (2014b) demonstrated that the error due to unresolved scales for specific humidity and temperature is state-dependent and spatially correlated. In a similar study for convection-permitting NWP (1.5km grid spacing), Waller et al. (2021) showed that the bias and uncertainty due to unresolved scales for temperature, specific humidity, zonal and meridional wind were greatest in lower model levels. Bell et al. (2021a) explored the characteristics of crowdsourced vehicle-

based observations of air temperature and showed that the uncertainty of the corresponding observation-model comparisons is likely to be weather-dependent. The representation uncertainty for Doppler radial winds at low heights consists primarily of uncertainty due to unresolved scales, but becomes dominated by the observation operator uncertainty as the height is increased (Waller et al., 2016c, 2019).

Methods to account for the error 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 error due to unresolved scales may be compensated for by averaging the model state over the observed area. When the reverse is true, the uncertainty caused by the small-scale processes must be accounted for in the data assimilation algorithm. This case is important for convection-permitting NWP, where high-resolution observations are used to provide information on the relevant scales for data assimilation (e.g., Sun et al., 2014; Ballard et al., 2016; Clark et al., 2016; Gustafsson et al., 2018; Dance et al., 2019; Waller et al., 2019). In this manuscript, we consider the uncertainty due to unresolved scales that occurs when an observation contains information on scales smaller than those resolved by the numerical model.

In order to produce an optimal analysis through assimilating high-resolution observations, the uncertainty due to unresolved scales must be properly accounted for by the data assimilation algorithm (Stewart et al., 2008, 2013; Stewart, 2010; Rainwater et al., 2015; Fowler et al., 2018; Simonin et al., 2019; Bell et al., 2020). There are several methods that may be used to compensate for uncertainty due to unresolved scales in data assimilation. The standard approach is to include the representation uncertainty in the observation error covariance (e.g., Hodyss and Satterfield, 2016; Fielding and Stiller, 2019). This may be achieved through estimation of the entire observation error covariance through statistical diagnostic methods (e.g., Schraff et al., 2016; Waller et al., 2016c,a; Cordoba et al., 2017; Simonin et al., 2019; Waller et al., 2019), or by estimation of the representation error covariance using high-resolution data (e.g., Daley, 1993; Liu and Rabier, 2002; Oke and Sakov, 2008; 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 and Vannitsem, 2011; Mitchell and Carrassi, 2015).

In this paper, we consider a different approach where the large-scale state resolved by a model is estimated and the statistics of the unresolved small-scale processes are considered. Such multiscale approaches have been successful for both variational (Li et al., 2015) and ensemble Kalman filter (Grooms et al., 2014) methods of data assimilation. A multiscale method of data assimilation that has been relatively unexplored for atmospheric data assimilation is the Schmidt-Kalman filter (SKF) (Schmidt, 1966). The potential of the SKF to compensate for uncertainty due

to unresolved scales in linear systems was first demonstrated by Janjić and Cohn (2006). Further experiments conducted by Bell et al. (2020) showed the SKF is most beneficial in regimes of high uncertainty due to unresolved scales and low instrument uncertainty. In order to use the SKF for atmospheric data assimilation, it must be adapted to an ensemble formulation 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. The ensemble transform Kalman filter (ETKF) is referred to as a deterministic filter as the analysis ensemble is constructed deterministically. The primary benefit of using a deterministic filter instead of a stochastic filter is that the sampling error caused by perturbing observations is eliminated (Tippett et al., 2003). Another benefit of the ETKF in comparison to other deterministic ensemble filters is the small computational cost of implementation (Vetra-Carvalho et al., 2018). The aim of this paper is to develop an ensemble transform formulation of the SKF suitable for applications with limited ensemble sizes.

The structure of this paper is as follows. In section 8.2 we discuss the model and observation configurations and provide a description of the ensemble transform Kalman filter. In section 8.3, we introduce the Schmidt-Kalman filter and extend it to an ensemble transform formulation. The experimental design that will be used to test the ensemble transform Schmidt-Kalman filter is described in section 8.4 and the results are given in section 8.5. We find that the ensemble transform Schmidt-Kalman filter is most beneficial when the uncertainty due to unresolved scales is greater than the instrument uncertainty. Furthermore, in our experiments the performance of the ensemble transform Schmidt-Kalman filter is similar to the ensemble transform Kalman filter that includes the uncertainty due to unresolved scales in the observation error covariance. In section 8.6 our results are summarised and our conclusions are drawn.

8.2 Theoretical framework

Following Bell et al. (2020), the mathematical framework used to examine the error due to unresolved scales is to estimate the projection of some state from a high, but finite-dimensional real vector space, onto a lower-dimensional subspace using knowledge of the system dynamics. We begin by describing the numerical model in section 8.2.1, the observations in section 8.2.2, and the standard ensemble transform Kalman filter in section 8.2.3.

8.2.1 Model configuration

In this section, we 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 $\boldsymbol{x}^t \in \mathbb{R}^{N_t}$ in the partitioned form

$$\boldsymbol{x}^t = \begin{pmatrix} \boldsymbol{x}^{l,t} \\ \boldsymbol{x}^{s,t} \end{pmatrix}, \tag{8.2.1}$$

where $\boldsymbol{x}^{l,t} \in \mathbb{R}^{N_l}$ is the true large-scale state, $\boldsymbol{x}^{s,t} \in \mathbb{R}^{N_s}$ 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.

For nonlinear dynamical systems, the perfect model evolving the true state is given by the coupled system

$$\begin{pmatrix} \boldsymbol{x}^{l,t} \\ \boldsymbol{x}^{s,t} \end{pmatrix}_k = \begin{pmatrix} \mathcal{M}^{l,t}(\boldsymbol{x}^{l,t}, \boldsymbol{x}^{s,t}) \\ \mathcal{M}^{s,t}(\boldsymbol{x}^{l,t}, \boldsymbol{x}^{s,t}) \end{pmatrix}_{k-1}, \tag{8.2.2}$$

where $\mathcal{M}^{l,t} : [\mathbb{R}^{N_l}, \mathbb{R}^{N_s}] \to \mathbb{R}^{N_l}$ and $\mathcal{M}^{s,t} : [\mathbb{R}^{N_l}, \mathbb{R}^{N_s}] \to \mathbb{R}^{N_s}$ are the true nonlinear models that map the true large- and small-scale states forward in time respectively. From a numerical modelling perspective, this partitioned description of the dynamics could be suited to a pseudospectral discretization (e.g. Fourier modes).

In NWP, approximate models for the large-scale dynamics that use subgrid-scale parameterizations to represent the influence of the small-scale processes are employed (Janjić and Cohn, 2006; Janjić et al., 2018) while models for the small-scale dynamics are disregarded. We therefore assume the imperfect forecast model has the partitioned form

$$\begin{pmatrix} \boldsymbol{x}^{l,t} \\ \boldsymbol{x}^{s,t} \end{pmatrix}_k = \begin{pmatrix} \mathcal{M}^{l}(\boldsymbol{x}^{l,t}) \\ \boldsymbol{x}^{s,t} \end{pmatrix}_{k-1} - \begin{pmatrix} \boldsymbol{\eta}^{l} \\ \boldsymbol{\eta}^{s} \end{pmatrix}_k, \tag{8.2.3}$$

where $\mathcal{M}^{l} : \mathbb{R}^{N_l} \to \mathbb{R}^{N_l}$ 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

$$\begin{pmatrix} \mathbf{Q}^{ll} & \mathbf{Q}^{ls} \\ \mathbf{Q}^{sl} & \mathbf{Q}^{ss} \end{pmatrix}. \tag{8.2.4}$$

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 $\mathbf{Q}^{sl} = \left(\mathbf{Q}^{ls}\right)^T$) is the cross-covariance between the large- and small-scale model errors. The assumption that the small-scale forecast model is persistence is related to the implementation of our novel filter which we discuss in section 8.3.3.

Analogously, the complete forecast state $\left((\boldsymbol{x}^{l,f})^T \;\; (\boldsymbol{x}^{s,f})^T\right)^T \in \mathbb{R}^{N_t}$ satisfies

$$\begin{pmatrix} \boldsymbol{x}^{l,f} \\ \boldsymbol{x}^{s,f} \end{pmatrix}_k = \begin{pmatrix} \mathcal{M}^l(\boldsymbol{x}^{l,f}) \\ \boldsymbol{x}^{s,f} \end{pmatrix}_{k-1}, \tag{8.2.5}$$

where the f-superscript indicates the forecast. We note that the small-scale forecast $\boldsymbol{x}^{s,f}$ 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

$$\begin{pmatrix} \mathbf{P}^{ll,f} & \mathbf{P}^{ls,f} \\ \mathbf{P}^{sl,f} & \mathbf{P}^{ss,f} \end{pmatrix}_k, \tag{8.2.6}$$

where each block has the same dimensions as the corresponding model error block. Using this formulation of the complete system dynamics, we may apply filters that analyse the large-scale state only but account for the small-scale processes in different ways.

8.2.2 Observation configuration

In this section we describe the observations in terms of the true partitioned state and discuss the observation uncertainty. The observations, $\boldsymbol{y} \in \mathbb{R}^p$, are assumed to be the sum of the true large- and small-scale states mapped into observation space. Assuming each component to be valid at time k so that we may drop the time subscript, the observations can then be expressed as

$$\boldsymbol{y} = \mathbf{h}^l\left(\boldsymbol{x}^{l,t}\right) + \mathbf{h}^s\left(\boldsymbol{x}^{s,t}\right) + \mathbf{e}^o, \tag{8.2.7}$$

where $\mathbf{h}^l : \mathbb{R}^{N_l} \to \mathbb{R}^p$ and $\mathbf{h}^s : \mathbb{R}^{N_s} \to \mathbb{R}^p$ are the imperfect large- and small-scale observation operators, and $\mathbf{e}^o \in \mathbb{R}^p$ is the observation error that is assumed to be random and unbiased. In this formulation, the error due to unresolved scales is given by $\mathbf{h}^s(\boldsymbol{x}^{s,t})$ and is not part of the observation error term $\mathbf{e}^o$. Throughout this manuscript, we assume there is no observation operator error or pre-processing error such that $\mathbf{e}^o$ is comprised entirely of instrument error. Therefore, the instrument uncertainty is given by $\mathbf{R}^I \equiv < \mathbf{e}^o(\mathbf{e}^o)^T >$ and the representation uncertainty is given by

$$\widetilde{\mathbf{R}}^H \equiv < \mathbf{h}^s(\boldsymbol{x}^{s,t})(\mathbf{h}^s(\boldsymbol{x}^{s,t}))^T >, \tag{8.2.8}$$

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.

8.2.3 Data assimilation using the ensemble transform Kalman filter

The ensemble transform Kalman filter (ETKF) (Bishop et al., 2001; Wang et al., 2004) is a commonly used filter for environmental data assimilation. The ETKF uses an ensemble of m forecasts, $\boldsymbol{x}^{l,f,(i)} \in \mathbb{R}^{N_l}$ for $i = 1, \ldots, m$, as a statistical sample to approximate the true state of a system. The forecast ensemble is combined with observations through their respective uncertainties to provide a best estimate of the true state known as the analysis. The ETKF algorithm consists of a correction step, known as the analysis, and a prediction step, known as the forecast. We first describe the ETKF analysis update equations where it can be assumed that all components are valid at the same time so that the time subscript may be dropped. The estimate of the true large-scale state of the system is given by the forecast ensemble mean,

$$\overline{\boldsymbol{x}}^{l,f} = \frac{1}{m} \sum_{i=1}^{m} \boldsymbol{x}^{l,f,(i)}, \tag{8.2.9}$$

where the (i) superscript denotes the i-th ensemble member. The uncertainty associated with $\overline{\boldsymbol{x}}^{l,f}$ is given by the large-scale forecast error covariance

$$\mathbf{P}^{ll,f} = \mathbf{X}^{l,f} \left(\mathbf{X}^{l,f}\right)^T, \tag{8.2.10}$$

where $\mathbf{X}^{l,f} \in \mathbb{R}^{N_l \times m}$ is the forecast perturbation matrix given by

$$\mathbf{X}^{l,f} = \frac{1}{\sqrt{m-1}} \left(\boldsymbol{x}^{l,f,(1)} - \overline{\boldsymbol{x}}^{l,f} \quad \ldots \quad \boldsymbol{x}^{l,f,(m)} - \overline{\boldsymbol{x}}^{l,f} \right). \tag{8.2.11}$$

The ETKF mean analysis state update is given by

$$\overline{\boldsymbol{x}}^{l,a} = \overline{\boldsymbol{x}}^{l,f} + \mathbf{K}^l \left(\boldsymbol{y} - \mathbf{H}^l \overline{\boldsymbol{x}}^{l,f} \right). \tag{8.2.12}$$

Here, $\mathbf{K}^l \in \mathbb{R}^{N_l \times p}$ is the Kalman gain and $\mathbf{H}^l \in \mathbb{R}^{p \times N_l}$ is a linear, large-scale observation operator. The Kalman gain is given by

$$\mathbf{K}^l = \mathbf{X}^{l,f} \left(\mathbf{Y}^{l,f}\right)^T \mathbf{D}^{-1}, \tag{8.2.13}$$

where $\mathbf{Y}^{l,f} = \mathbf{H}^l \mathbf{X}^{l,f}$ and $\mathbf{D} \in \mathbb{R}^{p \times p}$ is the innovation covariance given by

$$\mathbf{D} = \mathbf{Y}^{l,f} \left(\mathbf{Y}^{l,f}\right)^T + \mathbf{R}, \tag{8.2.14}$$

where $\mathbf{R} \in \mathbb{R}^{p \times p}$ is the observation error covariance assumed by the filter which may or may not contain a representation uncertainty component. In equations (8.2.12)–(8.2.14), we have used a linear form of observation operator for notational convenience, and we will continue to use this linear form throughout the rest of this article. However, the filter equations are straightforwardly generalizable to nonlinear observation operators as

$$\mathbf{Y}^{l,f} = \frac{1}{\sqrt{m-1}} \left(\mathbf{h}^l(\boldsymbol{x}^{l,f,(1)}) - \overline{\mathbf{h}^l(\boldsymbol{x}^{l,f})} \quad \dots \quad \mathbf{h}^l(\boldsymbol{x}^{l,f,(m)}) - \overline{\mathbf{h}^l(\boldsymbol{x}^{l,f})} \right). \tag{8.2.15}$$

The analysis perturbation matrix is given by

$$\mathbf{X}^{l,a} = \mathbf{X}^{l,f} \mathbf{T}, \tag{8.2.16}$$

where

$$\mathbf{T}\mathbf{T}^T = \mathbf{I}_m - \left(\mathbf{Y}^{l,f}\right)^T \mathbf{D}^{-1} \mathbf{Y}^{l,f} \tag{8.2.17}$$

$$= \left(\mathbf{I}_m + \left(\mathbf{Y}^{l,f}\right)^T \mathbf{R}^{-1} \mathbf{Y}^{l,f} \right)^{-1} \tag{8.2.18}$$

with $\mathbf{I}_m \in \mathbb{R}^{m \times m}$ the identity matrix. To obtain equation (8.2.18), we apply the Sherman-Morrison-Woodbury identity (Tippett et al., 2003) to $\mathbf{D}^{-1}$ in equation (8.2.17).

Lastly, using the singular value decomposition on $(\mathbf{Y}^{l,f})^T \mathbf{R}^{-1/2}$ we obtain

$$\mathbf{T} = \mathbf{U} \left(\mathbf{I}_m + \mathbf{\Sigma}\mathbf{\Sigma}^T \right)^{-1} \mathbf{U}^T \tag{8.2.19}$$

(Wang et al., 2004; Livings, 2005; Livings et al., 2008). Here, $\mathbf{U} \in \mathbb{R}^{m \times m}$ is an orthogonal matrix containing the left singular vectors of $(\mathbf{Y}^{l,f})^T \mathbf{R}^{-1/2}$ and $\mathbf{\Sigma} \in \mathbb{R}^{m \times p}$ is a rectangular diagonal matrix containing the corresponding non-zero singular values. The analysis ensemble is obtained by adding the analysis ensemble mean $\overline{\boldsymbol{x}}^{l,a}$ to each column of $(m-1)^{1/2} \mathbf{X}^{l,a}$.

In the forecast step, the analysis ensemble is evolved using the nonlinear forecast model. Reintroducing the time subscript, the forecast update is given by

$$\boldsymbol{x}_{k+1}^{l,f,(i)} = \mathcal{M}^l(\boldsymbol{x}_k^{l,a,(i)}) - \boldsymbol{\eta}_{k+1}^{l,(i)}, \tag{8.2.20}$$

where $\boldsymbol{\eta}^{l,(i)}$ is additive inflation or model error for the i-th ensemble member (Houtekamer and Zhang, 2016). We note that multiplicative inflation (e.g., Anderson and An-

derson, 1999; Zheng, 2009) could be used instead of additive inflation to ensure the magnitude of the forecast ensemble variance is sufficiently large. However, in this manuscript we use additive inflation as it is more in line with the work of Bell et al. (2020). In the following section, we extend the Schmidt-Kalman filter to a deterministic ensemble transform formulation.

8.3 The deterministic ensemble Schmidt-Kalman filter

In section 8.3.1 we describe the Schmidt-Kalman filter (Schmidt, 1966) and how it may be used to compensate for error due to unresolved scales. We then extend the Schmidt-Kalman filter to a novel ensemble transform formulation suitable for nonlinear models in section 8.3.2 and provide additional discussion on the filter implementation in section 8.3.3. Throughout this section, all components are assumed to be valid at the same time so we do not include any time subscripts.

8.3.1 The Schmidt-Kalman filter

The Schmidt-Kalman filter (SKF) is a linear filter that analyses the large-scale state only through consideration of the large-scale uncertainty and the variability of the small-scale processes (Janjić and Cohn, 2006; Bell et al., 2020). The analysis state update has a similar form to equation (8.2.12), where the large-scale gain matrix $\mathbf{K}^l$ is given by

$$\mathbf{K}^l = \begin{pmatrix} \mathbf{P}^{ll,f} & \mathbf{P}^{ls,f} \end{pmatrix} \begin{pmatrix} (\mathbf{H}^l)^T \\ (\mathbf{H}^s)^T \end{pmatrix} \left(\begin{pmatrix} \mathbf{H}^l & \mathbf{H}^s \end{pmatrix} \begin{pmatrix} \mathbf{P}^{ll,f} & \mathbf{P}^{ls,f} \\ \mathbf{P}^{sl,f} & \mathbf{C}^s \end{pmatrix} \begin{pmatrix} (\mathbf{H}^l)^T \\ (\mathbf{H}^s)^T \end{pmatrix} + \mathbf{R}^I \right)^{-1}. \tag{8.3.1}$$

Here, the forecast error covariance blocks $\mathbf{P}^{ll,f}$, $\mathbf{P}^{ls,f}$ and $\mathbf{P}^{sl,f}$ are defined in equation (8.2.6), $\mathbf{C}^s \in \mathbb{R}^{N_s \times N_s}$ is a climatological approximation to the true small-scale variability, $\langle \boldsymbol{x}^{s,t} (\boldsymbol{x}^{s,t})^T \rangle$, and the observation operator is in partitioned form where $\mathbf{H}^s \in \mathbb{R}^{p \times N_s}$ is the small-scale observation operator. We note that from equation (8.2.8), an approximation of the representation error covariance is given by

$$\mathbf{R}^H = \mathbf{H}^s \mathbf{C}^s (\mathbf{H}^s)^T. \tag{8.3.2}$$

The large-scale forecast error covariance $\mathbf{P}^{ll,f}$ and forecast cross-covariance $\mathbf{P}^{ls,f}$ are analysed using the gain matrix given by equation (8.3.1).

In the forecast step, the SKF evolves the large-scale analysis state $\boldsymbol{x}^{l,a}$, the large-scale analysis error covariance $\mathbf{P}^{ll,a}$, and the analysis cross-covariance $\mathbf{P}^{ls,a}$. Only knowledge of the large-scale model is required to evolve $\boldsymbol{x}^{l,a}$ whereas knowledge of

the complete system dynamics is required to evolve $\mathbf{P}^{ll,a}$ and $\mathbf{P}^{ls,a}$. For a detailed description of the SKF equations we refer the reader to Bell et al. (2020).

The difference between the SKF and the standard Kalman filter that analyses the large-scale state only can be seen through their treatment of the small-scale variability. The SKF accounts for the small-scale variability in state-space which allows for the consideration of the flow-dependent cross-covariances between the large-scale uncertainty and the small-scale variability. As a result of this, the observation error covariance for the SKF will consist of instrument uncertainty only. The standard Kalman filter may account for the small-scale variability in observation space (i.e. $\mathbf{R} = \mathbf{R}^I + \mathbf{R}^H$) or completely disregard it altogether (i.e. $\mathbf{R} = \mathbf{R}^I$) such that only the large-scale filter and model components are needed for its implementation. Therefore, the analysis and forecast error cross-covariances are completely disregarded by the standard Kalman filter. The disadvantage of using the SKF over the standard Kalman filter is the high computational cost associated with the augmentation of the state error covariances. Due to the small computational cost of implementation for the ETKF (Vetra-Carvalho et al., 2018), we believe an ensemble transform formulation of the SKF would be the most suitable ensemble adaptation. We now derive the ensemble transform SKF.

8.3.2 The ensemble transform Schmidt-Kalman filter (ETSKF)

To derive the ensemble transform SKF (ETSKF), the small-scale covariance $\mathbf{C}^s$ must be approximated through ensemble statistics in the analysis step. As $\boldsymbol{x}^{s,t}$ 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 varability, $\mathbf{C}^s$, and dividing by $\sqrt{m-1}$. Hence, the forecast cross-covariance is given by

$$\mathbf{P}^{ls,f} = \mathbf{X}^{l,f} \left(\mathbf{X}^s\right)^T. \tag{8.3.3}$$

Further discussion regarding the construction of the small-scale perturbation matrix is given in section 8.3.3.

We next describe the ETSKF analysis update equations. The gain matrix for the ensemble formulation of the Schmidt-Kalman filter is given by

$$\mathbf{K}^l = \left(\mathbf{X}^{l,f} \left(\mathbf{Y}^{l,f}\right)^T + \mathbf{X}^{l,f} \left(\mathbf{Y}^s\right)^T\right) \mathbf{D}^{-1}, \tag{8.3.4}$$

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

$$\mathbf{D} = \left(\mathbf{Y}^{l,f} + \mathbf{Y}^s\right)\left(\mathbf{Y}^{l,f} + \mathbf{Y}^s\right)^T + \mathbf{R}^I. \tag{8.3.5}$$

For the same reasons as discussed in section 8.3.1, 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 using the gain in equation (8.3.4) in the ETKF mean analysis state update given by equation (8.2.12).

To derive the analysis perturbation matrix, we start by writing the SKF large-scale analysis error covariance update,

$$\mathbf{P}^{ll,a} = \left(\mathbf{I}_{N_l} - \mathbf{K}^l\mathbf{H}^l\right)\mathbf{P}^{ll,f} - \mathbf{K}^l\mathbf{H}^s\mathbf{P}^{sl,f} \tag{8.3.6}$$

(equation (3.12), Bell et al., 2020), in terms of perturbation matrices to obtain

$$\begin{aligned}\mathbf{X}^{l,a}\left(\mathbf{X}^{l,a}\right)^T &= \left(\mathbf{I}_{N_l} - \left(\mathbf{X}^{l,f}\left(\mathbf{Y}^{l,f}\right)^T + \mathbf{X}^{l,f}\left(\mathbf{Y}^s\right)^T\right)\mathbf{D}^{-1}\mathbf{H}^l\right)\mathbf{X}^{l,f}\left(\mathbf{X}^{l,f}\right)^T \\ &\quad - \left(\mathbf{X}^{l,f}\left(\mathbf{Y}^{l,f}\right)^T + \mathbf{X}^{l,f}\left(\mathbf{Y}^s\right)^T\right)\mathbf{D}^{-1}\mathbf{H}^s\mathbf{X}^s\left(\mathbf{X}^{l,f}\right)^T \\ &= \mathbf{X}^{l,f}\left(\mathbf{I}_m - \left[\mathbf{Y}^{l,f} + \mathbf{Y}^s\right]^T\mathbf{D}^{-1}\left[\mathbf{Y}^{l,f} + \mathbf{Y}^s\right]\right)\left(\mathbf{X}^{l,f}\right)^T. \end{aligned} \tag{8.3.7}$$

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

$$\mathbf{T}\mathbf{T}^T = \mathbf{I}_m - \left[\mathbf{Y}^{l,f} + \mathbf{Y}^s\right]^T\mathbf{D}^{-1}\left[\mathbf{Y}^{l,f} + \mathbf{Y}^s\right]. \tag{8.3.8}$$

We now ask the reader to compare equation (8.3.8) with the definition of $\mathbf{T}$ for the ETKF given in equations (8.2.17)–(8.2.18). Similarly to equation (8.2.18), we apply the Sherman-Morrison-Woodbury identity on $\mathbf{D}^{-1}$ and then use the singular value decomposition on $\left[\mathbf{Y}^{l,f} + \mathbf{Y}^s\right]^T\left(\mathbf{R}^I\right)^{-1/2}$ to obtain

$$\mathbf{T} = \mathbf{U}\left(\mathbf{I}_m + \mathbf{\Sigma}\mathbf{\Sigma}^T\right)^{-1/2}\mathbf{U}^T, \tag{8.3.9}$$

where $\mathbf{U} \in \mathbb{R}^{m\times m}$ is now an orthogonal matrix containing the left singular vectors of $\left(\mathbf{Y}^{l,f} + \mathbf{Y}^s\right)^T\left(\mathbf{R}^I\right)^{-1/2}$ and $\mathbf{\Sigma} \in \mathbb{R}^{m\times p}$ is a rectangular diagonal matrix containing the corresponding non-zero singular values. Hence, the ETSKF analysis update equations are obtained by replacing $\mathbf{Y}^{l,f}$ with $\mathbf{Y}^{l,f} + \mathbf{Y}^s$ and $\mathbf{R}$ with $\mathbf{R}^I$ in the ETKF equations. The ETSKF forecast step is identical to the ETKF forecast step given by equation (8.2.20).

8.3.3 Discussion of the small-scale perturbation matrix

The ETSKF compensates for uncertainty due to unresolved scales through use of a small-scale perturbation matrix. Since the small-scale perturbation matrix is mapped into observation space in the ETSKF equations, 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 (8.3.2). For the additive inflation used in section 8.5, it is necessary to resample the small-scale ensemble every analysis step to ensure enough degrees of freedom to avoid filter divergence. However, if sufficiently large additive inflation is used, it may be possible to use the same small-scale ensemble in each analysis step. An approximation to the climatological uncertainty due to unresolved scales may be obtained in several ways (e.g., Daley, 1993; Liu and Rabier, 2002; Oke and Sakov, 2008; Schutgens et al., 2016; Waller et al., 2021). Sampling the small-scale ensemble in observation space is preferable as it circumvents the need for a small-scale observation operator $\mathbf{H}^s$.

To determine a strategy to sample the small-scale ensemble for the ETSKF, 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

$$\mathbf{P}^{ls,a} = \left(\mathbf{I} - \mathbf{K}^l\mathbf{H}^l\right)\mathbf{P}^{ls,f} - \mathbf{K}^l\mathbf{H}^s\mathbf{C}^s = \mathbf{X}^{l,f}\mathbf{T}\mathbf{T}^T(\mathbf{X}^s)^T \tag{8.3.10}$$

(equation (3.13), Bell et al., 2020), where the second equality is obtained by expressing the covariances in terms of perturbation matrices. We now define two sampling strategies for the ETSKF. The first sampling strategy is a computationally inexpensive method to sample the small-scale perturbation matrix, but neglects the evolution of the cross-covariances between the large-scale forecast uncertainty and the small-scale variability. The second sampling strategy provides an example of how the evolution of the cross-covariances may be accounted for, but is more computationally expensive than the first sampling strategy.

1. The small-scale perturbation matrix $\mathbf{Y}^s$ is sampled from $\mathcal{N}(\mathbf{0}, \mathbf{R}^H)$ such that the evolution of the cross-covariance between the large-scale uncertainty and small-scale variability is unaccounted for (i.e. random). The ETSKF that uses this sampling strategy will be referred to as the ETSKF-R.
2. The partitioned perturbation matrix $\begin{pmatrix}(\mathbf{Y}^{l,a})^T & (\mathbf{Y}^s)^T\end{pmatrix}^T$ is sampled from $\mathcal{N}(\mathbf{0}, \mathbf{\Psi})$

where

$$\boldsymbol{\Psi} = \begin{pmatrix} \mathbf{H}^l \mathbf{P}^{ll,a} \left(\mathbf{H}^l\right)^T & \mathbf{H}^l \mathbf{P}^{ls,a} \left(\mathbf{H}^s\right)^T \\ \mathbf{H}^s \mathbf{P}^{sl,a} \left(\mathbf{H}^l\right)^T & \mathbf{H}^s \mathbf{C}^s \left(\mathbf{H}^s\right)^T \end{pmatrix} = \begin{pmatrix} \mathbf{Y}^{l,f} \mathbf{T} \mathbf{T}^T \left(\mathbf{Y}^{l,f}\right)^T & \mathbf{Y}^{l,f} \mathbf{T} \mathbf{T}^T \left(\mathbf{Y}^s\right)^T \\ \mathbf{Y}^s \mathbf{T} \mathbf{T}^T \left(\mathbf{Y}^{l,f}\right)^T & \mathbf{R}^H \end{pmatrix}. \tag{8.3.11}$$

Here, the second equality is obtained using equations (8.3.2), (8.3.7) and (8.3.10). The $\mathbf{Y}^s$ obtained from this sampling will have consistent cross-covariances and will be used in the ETSKF algorithm. The $\mathbf{Y}^{l,a}$ obtained from this sampling will be discarded. We refer to the ETSKF using this sampling strategy as ETSKF-C.

In both cases, $\mathbf{Y}^s$ is sampled after $\mathbf{X}^{l,a}$ is calculated and is used in the following analysis step. We note that as the small-scale forecast model in equation (8.2.3) is persistence, the cross-covariances are correctly forecast by the ETSKF-C.

8.3.4 Discussion of the computational expense

Due to the use of a small-scale perturbation matrix $\mathbf{Y}^s$ in the analysis step, the ETSKF will have a greater computational cost than the ETKF. Provided the observation error covariance is diagonal, 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 using equation (8.2.20), which is identical to the ETKF forecast step. In the analysis step, the ETSKF equations are obtained by replacing $\mathbf{Y}^{l,f}$ with $\mathbf{Y}^{l,f} + \mathbf{Y}^s$ and $\mathbf{R}$ with $\mathbf{R}^I$ in the ETKF equations. Hence, the additional computational expense of the ETSKF in reference to the ETKF is caused by sampling the small-scale ensemble $\mathbf{Y}^s$ and adding it to $\mathbf{Y}^{l,f}$ each analysis step. Two methods for sampling $\mathbf{Y}^s$ were given in section 8.3.3. The first strategy involves sampling $\mathbf{Y}^s$ from $\mathbf{R}^H$. Using this strategy would result in the cheapest computational implementation of the ETSKF as $\mathbf{R}^H$, or a decomposition of $\mathbf{R}^H$ that may be used for random sampling such as the Cholesky decomposition, may be computed offline. The second strategy involves sampling $\left((\mathbf{Y}^{l,a})^T \ \ (\mathbf{Y}^s)^T\right)^T$ from $\boldsymbol{\Psi} \in \mathbb{R}^{2p \times 2p}$, given by equation (8.3.11). The computational expense of the second strategy is a substantially greater than the first sampling strategy as it requires the construction of $\boldsymbol{\Psi}$ every analysis step before the random sampling can occur. In operational weather prediction this would be computationally expensive.

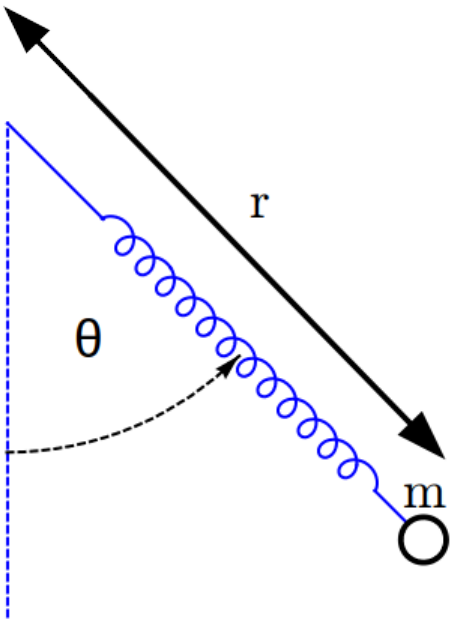


Figure 8.1: Schematic of the swinging spring. The angle θ is measured from the downward vertical and the length of the spring r is measured from the point of suspension to the mass.

8.4 Experimental design

8.4.1 The Swinging Spring model

To evaluate the performance of the ETSKF we will use the swinging spring (elastic pendulum) model (Lynch, 2002) shown in figure 8.1. The swinging spring consists of a mass m 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 along its length. Using polar coordinates, the nonlinear system of coupled ordinary differential equations describing the motion of the swinging spring is given by

$$\dot{\theta} = \frac{p_\theta}{mr^2}, \tag{8.4.1}$$

$$\dot{p_\theta} = -mgr\sin\theta, \tag{8.4.2}$$

$$\dot{r} = \frac{p_r}{m}, \tag{8.4.3}$$

$$\dot{p_r} = \frac{p_\theta^2}{mr^3} - k(r-\ell_0) + mg\cos\theta, \tag{8.4.4}$$

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

The swinging spring system has two equilibrium points: a stable equilibrium with

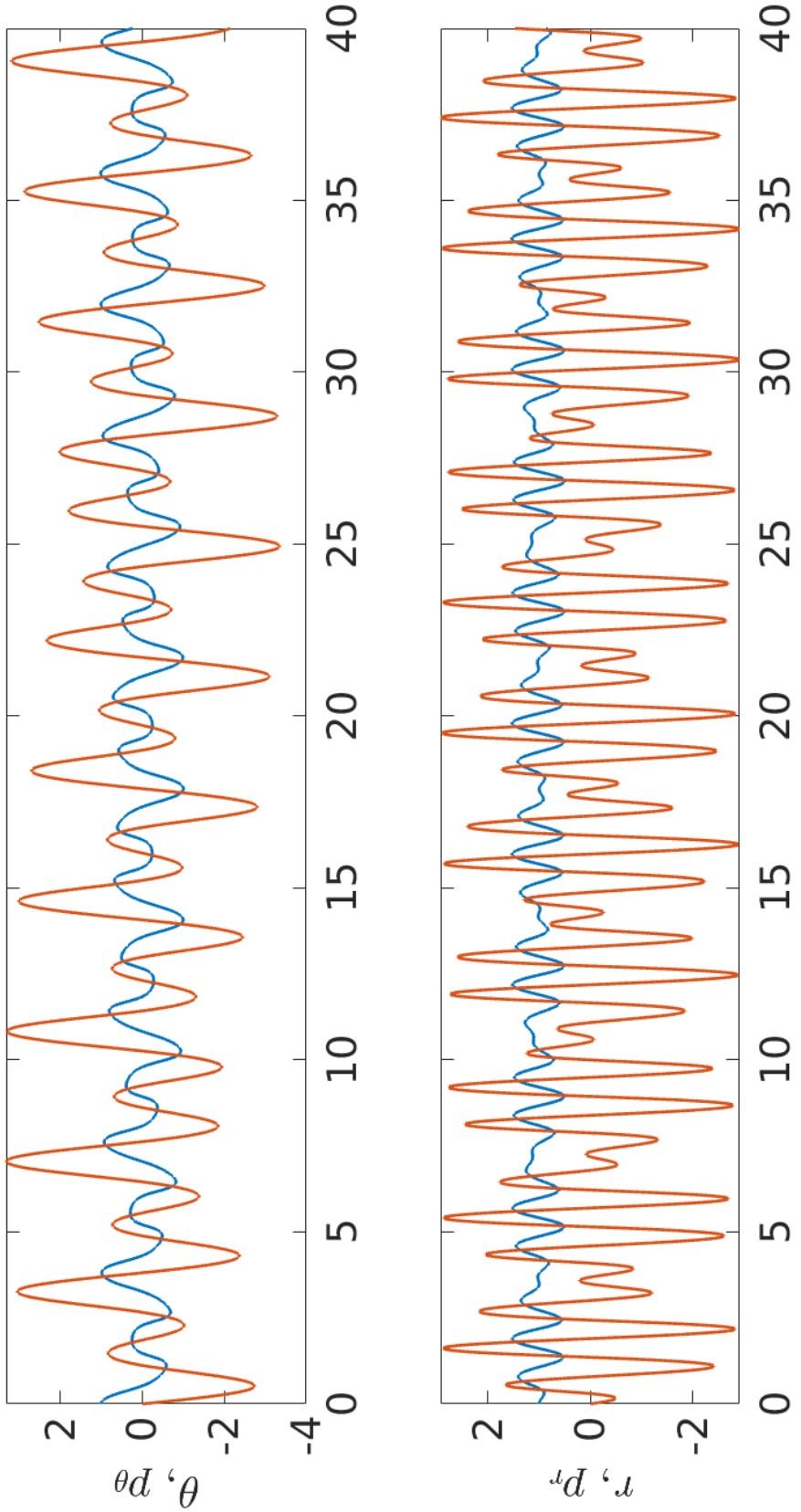


Figure 8.2: Typical behaviour of the swinging spring governed by equations (8.4.1)-(8.4.4) where $m = 1$, $\ell = 1$, $g = \pi^2$ and $k = 3\pi^2$. The top panel shows θ (blue) and p_θ (orange) and the bottom panel shows r (blue) and p_r (orange).

the spring stretched vertically below the point of suspension, and an unstable equilibrium where the spring is compressed vertically above the point of suspension. For our experiments, we consider only the stable equilibrium point. The equilibrium length of the spring is given by

$$\ell = \ell_0 + \frac{mg}{k}. \tag{8.4.5}$$

The frequency of the rotational motions is

$$\omega_\theta = \sqrt{\frac{g}{\ell}}, \tag{8.4.6}$$

and the frequency of the elastic motions is

$$\omega_r = \sqrt{\frac{k}{m}}. \tag{8.4.7}$$

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 experiments, we set $\ell = 1$, $m = 1$, $g = \pi^2$, and $k = 3\pi^2$ so that there is a separation in scales and substantial small-scale variability. Typical behaviour of the state variables is shown in figure 8.2. The sign of θ, p_θ and p_r change over time and all state variables are out of phase with each other. The frequency of θ and p_θ is lower than the frequency of r and p_r. Hence, θ and p_θ may be correctly represented with a coarser temporal resolution than r and p_r.

To obtain the true partitioned model in equation (8.2.2), we define $r = \ell + \rho$ where ρ is the displacement of the spring from the stable equilibrium with momentum p_ρ. We note that, since ℓ is fixed, the variability of r is completely associated with ρ. With this transformation, the swinging spring system of ODEs is given by

$$\dot{\theta} = \frac{p_\theta}{m(\ell+\rho)^2}, \tag{8.4.8}$$

$$\dot{p_\theta} = -mg(\ell+\rho)\sin\theta, \tag{8.4.9}$$

$$\dot{\ell} = 0, \tag{8.4.10}$$

$$\dot{\rho} = \frac{p_\rho}{m}, \tag{8.4.11}$$

$$\dot{p_\rho} = \frac{p_\theta^2}{m(\ell+\rho)^3} - k(\ell+\rho-l_0) + mg\cos\theta. \tag{8.4.12}$$

Using this form of the swinging spring, the large-scale state is given by $(\theta \;\; p_\theta \;\; \ell)^T$ and the small-scale state is given by $(\rho \;\; p_\rho)^T$.

To obtain the large-scale forecast model in equation (8.2.3), we neglect ρ from equations (8.4.8)-(8.4.10) such that the large- and small-scale processes are uncou-

pled:

$$\dot{\theta} = \frac{p_\theta}{m\ell^2}, \tag{8.4.13}$$

$$\dot{p_\theta} = -mg\ell \sin\theta, \tag{8.4.14}$$

$$\dot{\ell} = 0. \tag{8.4.15}$$

We note that the forecast model for θ and p_θ will have model error due to unresolved scales.

To integrate the true system in equations (8.4.8)-(8.4.12) and the forecast model in equations (8.4.13)-(8.4.15), we will use the MATLAB (version R2020b) *ode45* solver. The *ode45* function is a Runge-Kutta method with a variable timestep based on the Dormand-Prince $(4, 5)$ method (Dormand and Prince, 1980), where we use a relative error tolerance of 10^{-3} and an absolute error tolerance of 10^{-6}. The value of the systems will be recorded every 0.01 model seconds.

8.4.2 Observations and their uncertainty

For our experiments, we consider observations of θ^t and $r^t = \ell^t + \rho^t$. Expressing the observations $\boldsymbol{y} \in \mathbb{R}^2$ in the form of equation (8.2.7), we have

$$\boldsymbol{y} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \theta^t \\ p_\theta^t \\ \ell^t \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \rho^t \\ p_\rho^t \end{pmatrix} + \mathbf{e}^o, \tag{8.4.16}$$

where the first matrix-vector product corresponds to $\mathbf{H}^l \boldsymbol{x}^{l,t}$, the second matrix-vector product corresponds to the error due to unresolved scales $\mathbf{H}^s \boldsymbol{x}^{s,t}$, and $\mathbf{e}^o \in \mathbb{R}^2$ is the observation error. We remind the reader that in this instance the observation error consists only of instrument error.

To generate observations, we take the values of θ^t and r^t at specified time-steps and add observation error $\mathbf{e}^o$. The observation error will have distribution $\mathbf{e}^o \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ where σ^2 will be specified in each experiment.

In our experiments, we consider filters that analyse the large-scale state only. Hence, the observations of r^t will be treated as observations of ℓ^t with error due to unresolved scales ρ^t. We note that there is a bias due to unresolved scales as the time average of ρ^t is non-zero. The removal of the bias due to unresolved scales from the observations will be detailed in section 8.4.3.

8.4.3 Twin experiments

In our experiments we will use four filters: the ETKF that disregards uncertainty due to unresolved scales (i.e. $\mathbf{R} = \mathbf{R}^I$) which we denote the ETKF-LS, the ETKF that compensates for uncertainty due to unresolved scales (i.e. $\mathbf{R} = \mathbf{R}^I + \mathbf{R}^H$) which we will denote the ETKF-RH, the ETSKF-R and the ETSKF-C. The two versions of the ETSKF we use differ by their sampling strategies which are described in section 8.3.3.

To evaluate the performance of the ETSKF we will perform twin experiments where a trajectory obtained from the true model given by equations (8.4.1)-(8.4.4) will be recreated by assimilating imperfect observations of this true trajectory into the forecast. The bias due to unresolved scales will be removed from the observations before their assimilation so that we may focus on the ability of the ETSKF to compensate for uncertainty due to unresolved scales. The forecast will be evolved using the large-scale forecast model given by equations (8.4.13)-(8.4.15). Each experiment will consist of the following initialization steps.

1. Using the true model in equations (8.4.1)-(8.4.4), evolve the state $\begin{pmatrix} 1 & 0 & 1 & 0 \end{pmatrix}^T$ for 100 seconds and use the MATLAB (version R2020b) function *randi* to choose a starting point t_{start} for the assimilation. The truth trajectory is obtained by evolving the truth from time t_{start} for 10 model seconds.

2. Create observations with form given by equation (8.4.16) from the truth trajectory every 0.9 seconds and remove the bias due to unresolved scales. This observation frequency is chosen such that each filter produced good approximations of ℓ^t, but there was some difference in performance between the various filters. When the observations are more frequent, we found that ℓ^t was approximated extremely well by all filters and there was no discernible difference in performance. In the experiments with less frequent observations, we found that ℓ^t was approximated poorly by all 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 $\ell^t = 1$.

3. Using the large-scale forecast model in equations (8.4.13)-(8.4.15), evolve the state $\begin{pmatrix} 1 & 0 & 1+\zeta \end{pmatrix}^T$, 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 $\mathbf{e}_0^{l,f} \sim \mathcal{N}(\mathbf{0}, \mathbf{P}_0^{ll,f})$ to the mean forecast state at time t_{start} obtained in 3. Due to the low dimensionality of the forecast model, this ensemble size will be sufficient to circumvent the effects of sampling error on filter performance.

The initial forecast error covariance is set to

$$\mathbf{P}_0^{ll,f} = \begin{pmatrix} 0.2 & 0 & 0 \\ 0 & 0.6 & 0 \\ 0 & 0 & 0.2 \end{pmatrix}^2, \tag{8.4.17}$$

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 so that we may calculate statistics from the results of each experiment.

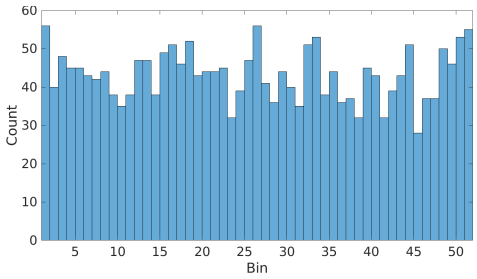
The ETKF-RH, ETSKF-R and ETSKF-C use a time-averaged approximation of the representation uncertainty. This approximation is obtained by taking the covariance of a single 100 model second truth run of $(r^t \quad p_r^t)$ used in the initialization procedure and mapping to observation space with $\mathbf{H}^s$. The climatological approximation of the representation uncertainty is given by

$$\mathbf{R}_{clim}^H = \begin{pmatrix} 0 & 0 \\ 0 & 0.28^2 \end{pmatrix}. \tag{8.4.18}$$

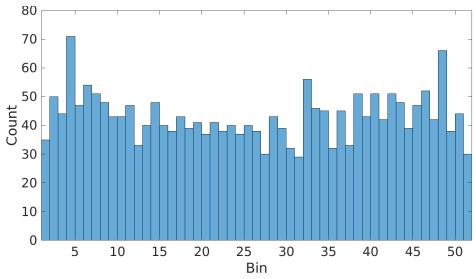
To account for large-scale model error and prevent ensemble collapse, model noise will be added to each ensemble member every 0.01 model seconds. The large-scale model-noise covariance $\mathbf{Q}^{ll}$ is chosen such that the rank histograms (Hamill, 2001) produced from the ETSKF-R and ETSKF-C forecast ensembles are uniform. To construct a rank histogram for the ETSKF, we first take the forecast ensemble

$$\mathcal{Y}^f = \begin{pmatrix} \mathbf{H}^l \boldsymbol{x}^{l,f,(1)} & \mathbf{H}^l \boldsymbol{x}^{l,f,(2)} & \dots & \mathbf{H}^l \boldsymbol{x}^{l,f,(m)} \end{pmatrix} + \sqrt{m-1}\mathbf{Y}^s, \tag{8.4.19}$$

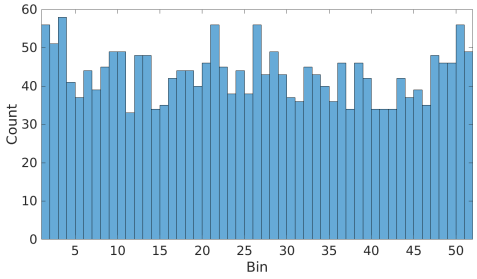
at the time of an observation. To account for the instrument uncertainty, we add random noise with distribution $\mathcal{N}(\mathbf{0}, \mathbf{R}^I)$ to each ensemble member. We note that we have already accounted for the representation uncertainty in the forecast ensemble through $\sqrt{m-1}\mathbf{Y}^s$. For each component of the observation vector, the corresponding forecast ensemble is sorted into ascending order and the bins $(-\infty, y^{f,(1)}], (y^{f,(1)}, y^{f,(2)}], \dots, (y^{f,(m)}, \infty)$ are defined where $y^{f,(i)}$ indicates an ensemble member in the ordered ensemble (i.e. $y^{f,(i)} < y^{f,(i+1)}$). The bin which the observation falls, known as the rank, is saved to a list and the process repeated for each observation assimilated over the 200 experiments by the filter. A histogram for each observed variable is then constructed from the ranks and information regarding the dispersion characteristics of the forecast ensemble is determined from the shape of the histogram.



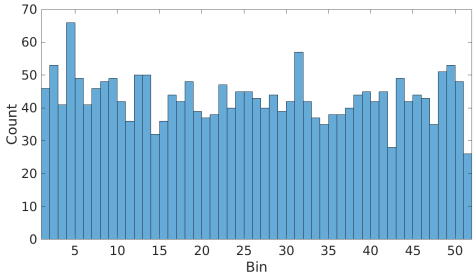
(a) Rank histogram for θ^f when $\mathbf{R}^I = 0.1^2\mathbf{I}$.



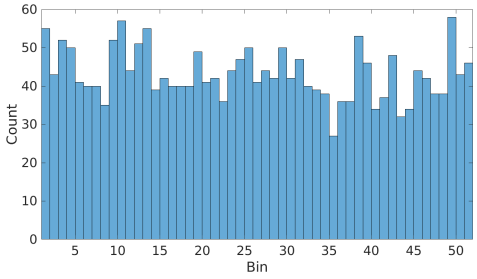
(b) Rank histogram for ℓ^f when $\mathbf{R}^I = 0.1^2\mathbf{I}$.



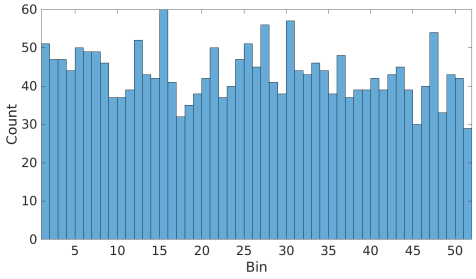
(c) Rank histogram for θ^f when $\mathbf{R}^I = 0.2^2\mathbf{I}$.



(d) Rank histogram for ℓ^f when $\mathbf{R}^I = 0.2^2\mathbf{I}$.



(e) Rank histogram for θ^f when $\mathbf{R}^I = 0.3^2\mathbf{I}$.



(f) Rank histogram for ℓ^f when $\mathbf{R}^I = 0.3^2\mathbf{I}$.

Figure 8.3: Rank histograms for the ETSKF-R constructed using the observations assimilated across 200 experiments using $\mathbf{Q}^{ll}$ given by equation (8.4.20).

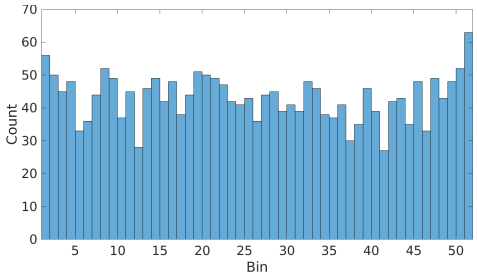
The $\mathbf{Q}^{ll}$ that will be used by each filter in our experiments is given by

$$\mathbf{Q}^{ll} = \begin{pmatrix} 0.05 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 0.001 \end{pmatrix}^2 . \tag{8.4.20}$$

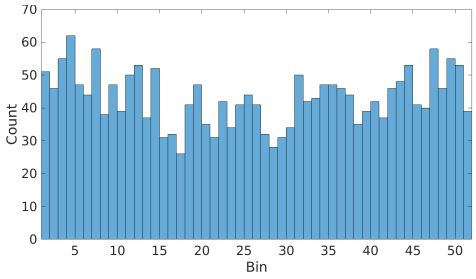
The rank histograms obtained by the ETSKF-R and ETSKF-C with 50 ensemble members when using equation (8.4.20) for the instrument uncertainty regimes we examine in section 8.5.1 are shown in figures 8.3 and 8.4 respectively. The rank histograms for both filters are fairly uniform for each instrument uncertainty regime indicating that this choice of $\mathbf{Q}^{ll}$ produces reliable ensembles for the ETSKF-R and ETSKF-C.

8.4.4 Performance metrics

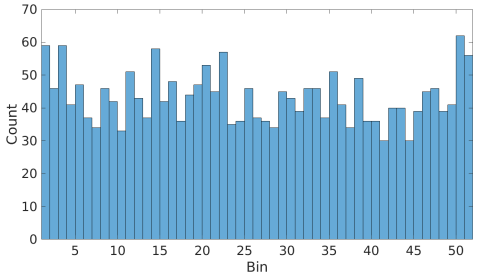
To assess the performance of the ETSKF and compare it with the ETKF we use two performance metrics. The first performance metric we use is the root-mean-square



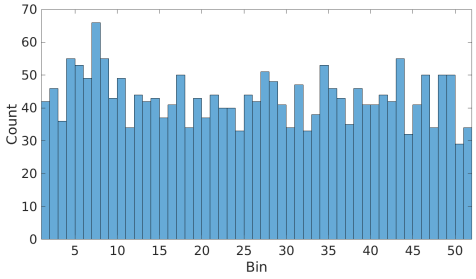
(a) Rank histogram for θ^f when $\mathbf{R}^I = 0.1^2\mathbf{I}$.



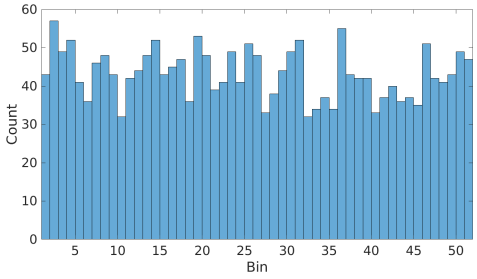
(b) Rank histogram for ℓ^f when $\mathbf{R}^I = 0.1^2\mathbf{I}$.



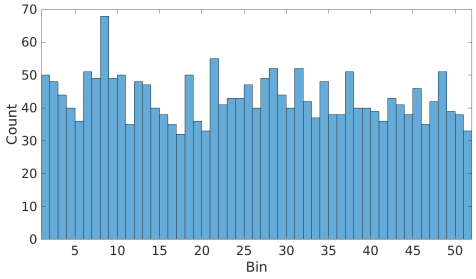
(c) Rank histogram for θ^f when $\mathbf{R}^I = 0.2^2\mathbf{I}$.



(d) Rank histogram for ℓ^f when $\mathbf{R}^I = 0.2^2\mathbf{I}$.



(e) Rank histogram for θ^f when $\mathbf{R}^I = 0.3^2\mathbf{I}$.



(f) Rank histogram for ℓ^f when $\mathbf{R}^I = 0.3^2\mathbf{I}$.

Figure 8.4: Rank histograms for the ETSKF-C constructed using the observations assimilated across 200 experiments using $\mathbf{Q}^{ll}$ given by equation (8.4.20).

error (RMSE) to determine the average forecast-mean error produced by a filter for a single component of the state vector. For a vector of length N_k, where N_k is the number of time-steps, the RMSE is given by

$$\text{RMSE} = \sqrt{\frac{\sum_{k=1}^{N_k}(\overline{x}_k^{l,f} - x_k^{l,t})^2}{N_k}}. \tag{8.4.21}$$

The RMSE will be calculated for each component of the state vector over the second half of the assimilation window and averaged over the 200 experiments.

The second performance metric we use is the continuous rank probability score (CRPS) which measures the difference between the true cumulative distribution function and the probabilistic forecast produced by each filter. To calculate the CRPS at time k, we first sort the ensemble members into ascending order and define the extra ensemble members $x^{l,f(0)} = -\infty$ and $x^{l,f(m+1)} = \infty$ (Hersbach, 2000). Assuming that each ensemble member is equally probable, the CRPS is given by

$$CRPS = \sum_{j=0}^{m} c_j, \tag{8.4.22}$$

where

$$c_j = \alpha_j \left(\frac{j}{m}\right)^2 + \beta_j \left(1 - \frac{j}{m}\right)^2. \tag{8.4.23}$$

Here, the j superscript corresponds to the ensemble member in the ordered ensemble (i.e. $x^{l,f,(j)} < x^{l,f,(j+1)}$). The values of α_j and β_j are given in table 8.1. The best possible CRPS value is 0 which corresponds to a perfect deterministic forecast. The CRPS will be calculated for each component of the state vector and averaged over the second half of the assimilation window to obtain the mean CRPS for that experiment. The mean CRPS will then be averaged over the 200 experiments.

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

$$\frac{A - B}{A} \times 100\% \tag{8.4.24}$$

where A corresponds to the value for the ETKF-LS and B corresponds to another filter.

$0 < j < m$	α_j	β_j
$x^{l,t} > x^{l,f,(j)}$	$x^{l,f,(j+1)} - x^{l,f,(j)}$	0
$x^{l,f,(j+1)} > x^{l,t} > x^{l,f,(j)}$	$x^{l,t} - x^{l,f,(j)}$	$x^{l,f,(j+1)} - x^{l,t}$
$x^{l,f,(j)} > x^{l,t}$	0	$x^{l,f,(j+1)} - x^{l,f,(j)}$

Table 8.1: Values for α_j and β_j in equation (8.4.23). The subscript j indicates the ensemble member in the ensemble arranged into ascending order (i.e. $x^{l,f,(j)} < x^{l,f,(j+1)}$).

8.5 Numerical experiments

8.5.1 Filter performance for different observation uncertainty regimes

In this section we consider the performance of the four filters for three observation uncertainty regimes. To assess the performance of each filter for different observation uncertainty regimes, we calculate the forecast RMSE and mean CRPS over the second half of the assimilation window for each experiment and then average over the 200 experiments. The average RMSE and mean CRPS of the four filters for each observation uncertainty regime is shown in table 8.2.

We first consider the case where $\mathbf{R}^I = 0.1^2\mathbf{I}$ such that the representation uncertainty is approximately three times that of the instrument uncertainty. To determine the statistical significance of the RMSE and mean CRPS results for each filter, we use the t-test to compare the results averaged across the 200 experiments. The difference in RMSE and mean CRPS for θ^f and p_θ^f between the ETKF-LS and each of the ETKF-RH, ETSKF-R and ETSKF-C were not found to be statistically significant. However, the differences in RMSE and mean CRPS for ℓ^f between the ETKF-LS and each other filter is statistically significant. From table 8.2 we see that the ETKF-RH, ETSKF-R and ETSKF-C have a large, positive percentage improvement in RMSE and mean CPRS for ℓ^f compared to the ETKF-LS. Hence, the ETKF-RH, ETSKF-R and ETSKF-C 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, ETSKF-R and ETSKF-C is not statistically significant. Therefore, the ETSKF-R and ETSKF-C perform similarly to the ETKF-RH in a high representation uncertainty and low instrument uncertainty regime.

We next examine the performance of the three filters relative to the ETKF-LS when the instrument uncertainty is increased to $\mathbf{R}^I = 0.2^2\mathbf{I}$ and $\mathbf{R}^I = 0.3^2\mathbf{I}$. Similarly to the experiments where $\mathbf{R}^I = 0.1^2\mathbf{I}$, the difference in RMSE and mean CRPS for each forecast averaged across 200 experiments between the ETKF-RH, ETSKF-R and ETSKF-C is not statistically significant when $\mathbf{R}^I = 0.2^2\mathbf{I}$ and $\mathbf{R}^I = 0.3^2\mathbf{I}$. When $\mathbf{R}^I = 0.2^2\mathbf{I}$, only the differences in mean CRPS for ℓ^f between the ETKF-

$\mathbf{R}^I$	Filter	θ^f		p_θ^f		ℓ^f	
		RMSE	mean CRPS	RMSE	mean CRPS	RMSE	mean CRPS
$0.1^2\mathbf{I}$	ETKF-LS	0.059	0.008	0.178	0.019	0.126	0.016
	ETKF-RH	4.51%	13.92%	−2.08%	11.86%	**17.01**%	**83.33**%
	ETSKF-R	4.35%	7.59%	−5.85%	6.19%	**28.22**%	**84.57**%
	ETSKF-C	6.86%	6.32%	−1.63%	4.64%	**19.0**%	**80.86**%
$0.2^2\mathbf{I}$	ETKF-LS	0.062	0.008	0.210	0.019	0.118	0.007
	ETKF-RH	2.57%	6.49%	−5.23%	4.64%	−0.17%	**58.88**%
	ETSKF-R	6.59%	5.19%	−0.19%	3.61%	6.59%	**48.53**%
	ETSKF-C	3.86%	−1.30%	−2.42%	−3.09%	7.85%	**54.41**%
$0.3^2\mathbf{I}$	ETKF-LS	0.063	0.008	0.238	0.022	0.122	0.005
	ETKF-RH	1.59%	8.43%	−8.42%	7.72%	−9.22%	26.09%
	ETSKF-R	1.59%	6.02%	−2.48%	8.63%	0.41%	19.57%
	ETSKF-C	−0.64%	1.20%	−3.16%	1.36%	0.74%	30.43%

Table 8.2: Summary of the results for the RMSE and mean CRPS obtained for each filter averaged across 200 experiments for three different instrument uncertainty regimes. We note that the same model error covariance was used for each filter and instrument uncertainty regime. Determination of the model error covariance is given in section 8.4.3. The results are given in relative percentage form (see equation (8.4.24)) for the ETKF-RH, ETSKF-R and ETSKF-C. Relative percentages in bold indicate when the improvement of the ETKF-RH, ETSKF-R or ETSKF-C over the ETKF-LS is statistically significant.

LS and each other filter is statistically significant. From table 8.2, we see that the ETKF-RH, ETSKF-R and ETSKF-C have a large, positive percentage improvement in mean CPRS for ℓ^f compared to the ETKF-LS. Hence, the ETKF-RH, ETSKF-R and ETSKF-C on average produce a more accurate probabilistic forecast than the ETKF-LS when the instrument uncertainty is approximately 2/3 the size of the representation uncertainty. When $\mathbf{R}^I = 0.3^2\mathbf{I}$, the difference in RMSE and mean CRPS between the four filters for any forecast variable is not statistically significant. Hence, the ETSKF-R and ETSKF-C are most beneficial when the representation uncertainty is greater than the instrument uncertainty. This agrees with the results of Bell et al. (2020) for a linear system who found that the SKF is most suited to regimes of high representation uncertainty and low instrument uncertainty.

8.5.2 Comparison of the ETSKF-R and ETSKF-C performance

To compare the sampling strategies employed by the ETSKF-R and ETSKF-C, we examine the gain matrices at observation times for the two filters using a single truth and observation realization. For the instrument uncertainty we choose $\mathbf{R}^I = 0.1^2\mathbf{I}$ such that we are in a high representation uncertainty and low instrument uncertainty regime. We will examine two realizations of the initial forecast ensemble and sampled small-scale ensembles. In the first realization, the ETSKF-R produces a smaller RMSE and mean CRPS for each large-scale state variable than the ETSKF-C. We refer to this case study as CS1. In the second realization, the ETSKF-C produces a smaller RMSE and mean CRPS for each large-scale state variable than the ETSKF-

R. We refer to this case study as CS2.

To examine the gain matrix for the ETSKF-R and the ETSKF-C, we express each component of the gain matrix in scalar form:

$$\mathbf{K}^l = \begin{pmatrix} \mathrm{K}^l_{\theta\theta} & \mathrm{K}^l_{\theta\ell} \\ \mathrm{K}^l_{p_\theta\theta} & \mathrm{K}^l_{p_\theta\ell} \\ \mathrm{K}^l_{\ell\theta} & \mathrm{K}^l_{\ell\ell} \end{pmatrix}. \tag{8.5.1}$$

Here, the subscripts indicate the weighted contribution from the innovation vector to the analysed state (i.e. $\mathrm{K}^l_{\theta\ell}$ weights the contribution of the ℓ-component of the innovation $\boldsymbol{y} - \mathbf{H}^l\overline{\boldsymbol{x}}^{l,f}$ to θ^a). The gains in the first column of equation (8.5.1) may be expressed as

$$\begin{aligned}\mathrm{K}^l_{\bullet\theta} = \frac{1}{det(\mathbf{D})}\Bigg\{ \left(\sum_{\lambda=1}^{m} \mathbf{X}^{l,f}_{\bullet,\lambda}\mathbf{X}^{l,f}_{\theta,\lambda}\right)\left[\left(\sum_{\phi=1}^{m}\left(\mathbf{X}^{l,f}_{\ell,\phi}+\mathbf{X}^{s}_{\rho,\phi}\right)\left(\mathbf{X}^{l,f}_{\ell,\phi}+\mathbf{X}^{s}_{\rho,\phi}\right)\right)+0.1^2\right] \\ -\left(\sum_{\psi=1}^{m}\mathbf{X}^{l,f}_{\bullet,\psi}\left(\mathbf{X}^{l,f}_{\ell,\psi}+\mathbf{X}^{s}_{\rho,\psi}\right)\left(\mathbf{X}^{l,f}_{\ell,\psi}+\mathbf{X}^{s}_{\rho,\psi}\right)\mathbf{X}^{l,f}_{\theta,\psi}\right)\Bigg\},\end{aligned} \tag{8.5.2}$$

where $\bullet$ is a placeholder for a chosen state variable and $det(\mathbf{D})$ is the determinant of the innovation covariance given by

$$\begin{aligned}det(\mathbf{D}) = \left[\left(\sum_{\lambda=1}^{m}\mathbf{X}^{l,f}_{\theta,\lambda}\mathbf{X}^{l,f}_{\theta,\lambda}\right)+0.1^2\right]\left[\left(\sum_{\phi=1}^{m}\left(\mathbf{X}^{l,f}_{\ell,\phi}+\mathbf{X}^{s}_{\rho,\phi}\right)\left(\mathbf{X}^{l,f}_{\ell,\phi}+\mathbf{X}^{s}_{\rho,\phi}\right)\right)+0.1^2\right] \\ -\left(\sum_{\psi=1}^{m}\mathbf{X}^{l,f}_{\theta,\psi}\left(\mathbf{X}^{l,f}_{\ell,\psi}+\mathbf{X}^{s}_{\rho,\psi}\right)\left(\mathbf{X}^{l,f}_{\ell,\psi}+\mathbf{X}^{s}_{\rho,\psi}\right)\mathbf{X}^{l,f}_{\theta,\psi}\right).\end{aligned} \tag{8.5.3}$$

We remind the reader that for our experiments $\mathbf{D}$ is a 2×2 matrix and we may therefore write the matrix inverse explicitly. Similarly, the second column of equation (8.5.1) may be expressed as

$$\begin{aligned}\mathrm{K}^l_{\bullet\ell} = \frac{1}{det(\mathbf{D})}\Bigg\{ \left(\sum_{\lambda=1}^{m}\mathbf{X}^{l,f}_{\bullet,\lambda}\left(\mathbf{X}^{l,f}_{\ell,\lambda}+\mathbf{X}^{s}_{\rho,\lambda}\right)\right)\left[\left(\sum_{\phi=1}^{m}\mathbf{X}^{l,f}_{\theta,\phi}\mathbf{X}^{l,f}_{\theta,\phi}\right)+0.1^2\right] \\ -\left(\sum_{\psi=1}^{m}\mathbf{X}^{l,f}_{\bullet,\psi}\mathbf{X}^{l,f}_{\theta,\psi}\mathbf{X}^{l,f}_{\theta,\psi}\left(\mathbf{X}^{l,f}_{\ell,\psi}+\mathbf{X}^{s}_{\rho,\psi}\right)\right)\Bigg\}.\end{aligned} \tag{8.5.4}$$

Each component of the gain matrix for the ETSKF-R and ETSKF-C is shown in figure 8.5 for CS1 and figure 8.6 for CS2. Regardless of filter or case study, θ^a is always close to the θ-observation. This is due to two reasons. The first reason is that the uncertainty of θ^f is much greater than the uncertainty of the θ-observations (i.e. $\sum_{\lambda=1}^{m}\mathbf{X}^{l,f}_{\theta,\lambda}\mathbf{X}^{l,f}_{\theta,\lambda} \gg 0.1^2$). Therefore, the numerator of equation (8.5.2) is ap-

proximately equal to $det(\mathbf{D})$ such that $\mathrm{K}^l_{\theta\theta} \approx 1$. The second reason is that the errors in θ^f and ℓ^f will be weakly correlated as ℓ^f is forecast by persistence. Hence, $\sum_{\lambda=1}^{m} \mathbf{X}^{l,f}_{\theta,\lambda}\mathbf{X}^{l,f}_{\ell,\lambda}$ will be small and therefore $\mathrm{K}^l_{\theta\ell}$ will also be small.

Both $\mathrm{K}^l_{p_\theta\theta}$ and $\mathrm{K}^l_{p_\theta\ell}$ vary in size throughout the assimilation window in both case studies. In most instances, p^a_θ has contributions from both components of the innovation vector as there is substantial additive inflation when the state is forecast and p^t_θ is not observed. It can be seen from figure 8.2 that p^t_θ changes sign over time and is out of phase with θ^t. We can expect the cross-covariances between θ^f and p^f_θ to reflect this behaviour. Examining $\mathrm{K}^l_{p_\theta\theta}$ for CS1 and CS2 in figures 8.5 and 8.6 respectively, we see that the qualitative behaviour of $\mathrm{K}^l_{p_\theta\theta}$ is consistent between filters and case studies.

Similarly to $\mathrm{K}^l_{\theta\ell}$, we see that $\mathrm{K}^l_{\ell\theta} \approx 0$ which is likely because the correlation between the errors in θ^f and ℓ^f is weak. A further implication of the weak correlation between θ^f and ℓ^f is that the sign of $\mathrm{K}^l_{\ell\ell}$ will usually be positive as the second term in the curly brackets in equation (8.5.4) will be small. The value of $\mathrm{K}^l_{\ell\ell}$ is largest near the start of the assimilation window and generally decreases as more observations are assimilated. This is expected as the ℓ^f forecast uncertainty is initially smaller than the total uncertainty of the ℓ-observations and there is little additive inflation. While there are differences in the $\mathrm{K}^l_{\ell\ell}$ values between the ETSKF-R and ETSKF-C, over the whole assimilation window they exhibit similar qualitative behaviour.

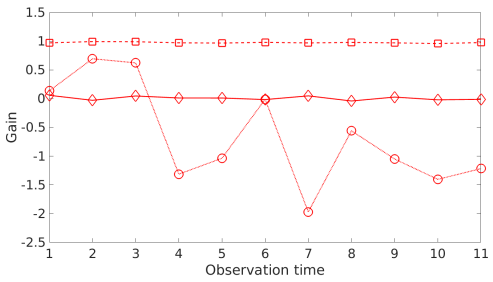
The key difference between the ETSKF-R and ETSKF-C is their treatment of the cross-covariances between the large-scale uncertainty and the small-scale variability. To determine the influence of these cross-covariances on the two filters, we examine the correlation between the large-scale forecast errors for each state variable and the small-scale variability given by

$$\frac{\sum_{\lambda=1}^{m} \mathbf{X}^{l,f}_{\bullet,\lambda}\mathbf{X}^{s}_{\rho,\lambda}}{\sqrt{\left(\sum_{\phi=1}^{m} \mathbf{X}^{l,f}_{\bullet,\phi}\mathbf{X}^{l,f}_{\bullet,\phi}\right)\left(\sum_{\psi=1}^{m} \mathbf{X}^{s}_{\rho,\psi}\mathbf{X}^{s}_{\rho,\psi}\right)}}. \tag{8.5.5}$$

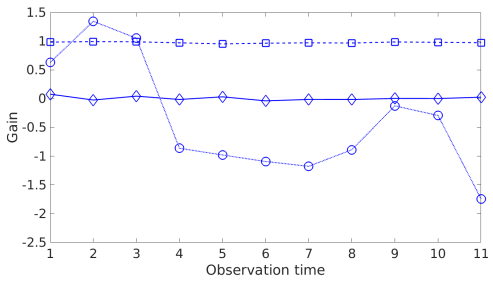
The correlations for the ETSKF-R and ETSKF-C in both case studies are shown in figure 8.7. In general, the correlations are weak for both filters which suggests that the cross-covariances will have a small effect on the performance of the filter. This explains why the difference in performance between the ETSKF-R and ETSKF-C was small for the experiments performed in section 8.5.1.

8.6 Conclusion

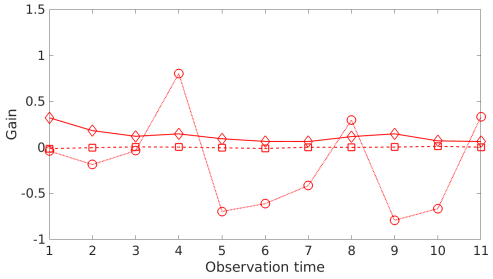
The error caused by a mismatch in scales observed and modelled is known as the error due to unresolved scales. In this work, we considered the situation where



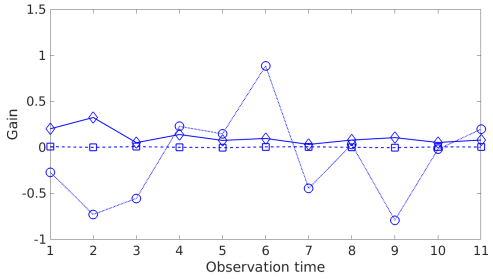
(a) $K^l_{\theta\theta}$, $K^l_{p_\theta\theta}$ and $K^l_{\ell\theta}$ for the ETSKF-R.



(b) $K^l_{\theta\theta}$, $K^l_{p_\theta\theta}$ and $K^l_{\ell\theta}$ for the ETSKF-C.

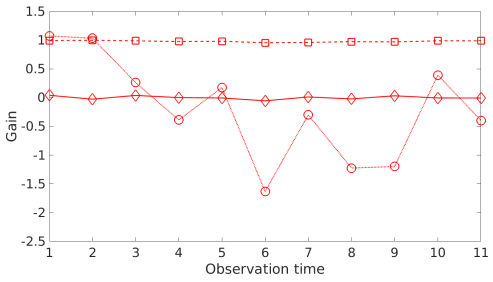


(c) $K^l_{\theta\ell}$, $K^l_{p_\theta\ell}$ and $K^l_{\ell\ell}$ for the ETSKF-R.

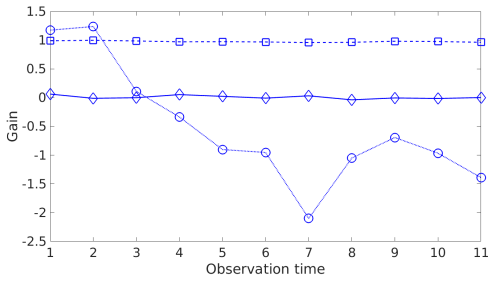


(d) $K^l_{\theta\ell}$, $K^l_{p_\theta\ell}$ and $K^l_{\ell\ell}$ for the ETSKF-C.

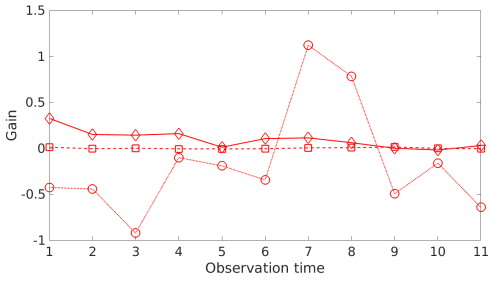
Figure 8.5: The components of the gain matrix at observation times for CS1 for the ETSKF-R and ETSKF-C. The weights $K^l_{\theta\bullet}$ are indicated by a dashed line with square markers, $K^l_{p_\theta\bullet}$ by a dotted line with circle markers, and $K^l_{\ell\bullet}$ by a solid line with diamond markers.



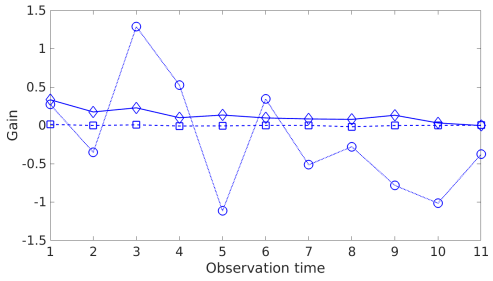
(a) $K^l_{\theta\theta}$, $K^l_{p_\theta\theta}$ and $K^l_{\ell\theta}$ for the ETSKF-R.



(b) $K^l_{\theta\theta}$, $K^l_{p_\theta\theta}$ and $K^l_{\ell\theta}$ for the ETSKF-C.

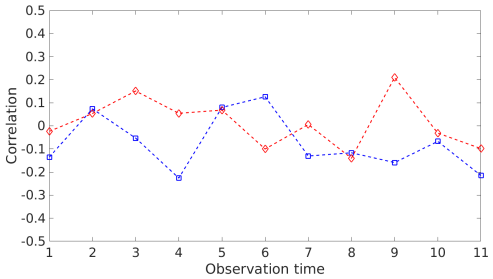


(c) $K^l_{\theta\ell}$, $K^l_{p_\theta\ell}$ and $K^l_{\ell\ell}$ for the ETSKF-R.

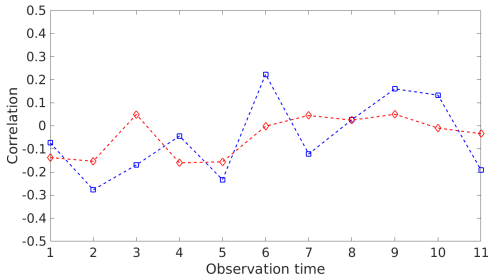


(d) $K^l_{\theta\ell}$, $K^l_{p_\theta\ell}$ and $K^l_{\ell\ell}$ for the ETSKF-C.

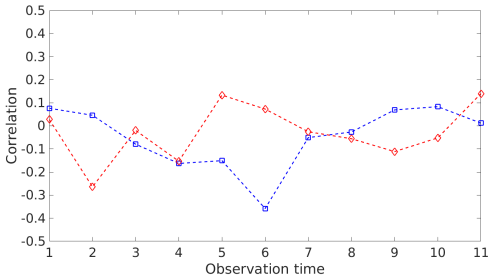
Figure 8.6: The components of the gain matrix at observation times for CS2 for the ETSKF-R and ETSKF-C. The weights $K^l_{\theta\bullet}$ are indicated by a dashed line with square markers, $K^l_{p_\theta\bullet}$ by a dotted line with circle markers, and $K^l_{\ell\bullet}$ by a solid line with diamond markers.



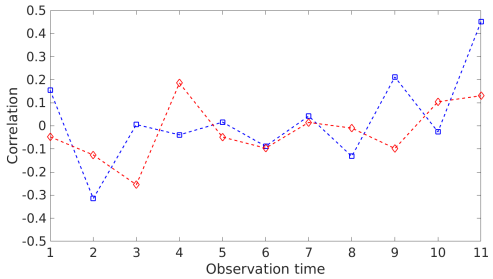
(a) Correlation between θ^f and ρ^f for CS1.



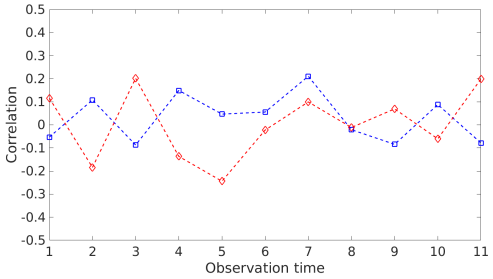
(b) Correlation between θ^f and ρ^f for CS2.



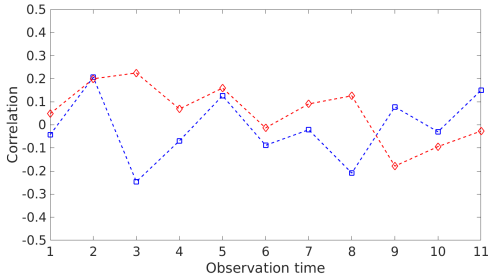
(c) Correlation between p_θ^f and ρ^f for CS1.



(d) Correlation between p_θ^f and ρ^f for CS2.



(e) Correlation between l^f and ρ^f for CS1.



(f) Correlation between l^f and ρ^f for CS2.

Figure 8.7: The correlation between the large-scale forecast errors and the small-scale variability. The ETSKF-R correlations are shown in red with diamond markers and the ETSKF-C correlations are shown in blue with square markers.

observations contain information on scales smaller than those represented by the numerical model. The standard approach to compensate for error due to unresolved scales in data assimilation is to include it in the observation error covariance matrix. An alternative approach is to estimate the large-scale state resolved by the model while considering the statistics of the small-scale processes unresolved by the model. An example of a filter capable of compensating for the error due to unresolved scales using the alternative approach is the Schmidt-Kalman filter.

In this work, we introduced a new ensemble transform formulation of the Schmidt-Kalman filter (ETSKF) where the small-scale variability is represented through an ensemble sampled from the representation error covariance. The sampled small-scale ensemble is added to the large-scale forecast ensemble to obtain an ensemble with uncertainty representative of all scales resolved by the observations.

To test the performance of the ETSKF, we performed experiments using the swinging spring (elastic pendulum) model. In the swinging spring model, the large-scale processes correspond to the rotational motions and the small-scale processes correspond to the elastic motions. Hence, observations of the length of the spring were treated as having uncertainty due to unresolved scales in our experiments. The filters we used in our experiments include an ensemble transform Kalman filter (ETKF) that disregarded the uncertainty due to unresolved scales (ETKF-LS), an ETKF that accounts for the uncertainty due to unresolved scales using the standard approach (ETKF-RH), and two versions of the ETSKF that differ in how they sample the small-scale ensemble. The first ETSKF samples the small-scale ensemble such that the cross-covariances between the large-scale errors and the small-scale variability are random (ETSKF-R). The second ETSKF samples the small-scale ensemble such that the cross-covariances are consistent (ETSKF-C). The four filters were implemented without localization as the dimensionality of the swinging spring system is small and we used large ensemble sizes in our experiments. In principle, localization for the ETSKF may be implemented in the same way as for the ETKF.

To provide a comparison between filters, we examined the improvement of the ETKF-RH, ETSKF-R and ETSKF-C relative to the ETKF-LS. The ETKF-RH, ETSKF-R and ETSKF-C showed statistically significant improvements over the ETKF-LS in a regime of high representation uncertainty and low instrument uncertainty. We also found that the difference in performance between the ETKF-RH, ETSKF-R and ETSKF-C was not statistically significant. Experiments using increased instrument uncertainty showed that the ETSKF-R and ETSKF-C provide most benefit over the ETKF-LS when the representation uncertainty is greater than the instrument uncertainty. Furthermore, when the representation uncertainty and instrument uncertainty were similar in magnitude, there was little to no benefit in

using the ETSKF-R or ETSKF-C over the ETKF-LS. This agrees with the results of Bell et al. (2020) who showed the Schmidt-Kalman filter is most beneficial in regimes of high uncertainty due to unresolved scales and low instrument uncertainty.

The effect of the sampling strategies employed by the ETSKF-R and ETSKF-C on the filter performance was examined for a single realization of the truth and observations in a high representation uncertainty and low instrument uncertainty regime. The correlation between the large-scale forecast errors and the small-scale variability was generally small for both filters and hence the sampling strategies would have little effect on their performance. This was reflected in the gain matrices of the two filters which shared several similar qualitative characteristics.

The ETSKF using either sampling method has been shown to be capable of compensating for uncertainty due to unresolved scales in nonlinear systems. However, further experiments using more complicated models are required to understand the robustness of the ETSKF, the importance of the cross-covariances, and its suitability for operational data assimilation.

8.7 Chapter summary

In this chapter, we addressed the fourth research question given in chapter 1: *how can the Schmidt-Kalman filter be adapted for nonlinear models?* A novel ensemble transform formulation of the Schmidt-Kalman filter (ETSKF) has been derived. The performance of the ETSKF using two different sampling strategies for the small-scale ensemble were evaluated against two forms of the ensemble transform Kalman filter (ETKF) using the swinging spring model. The first ETKF completely disregards the small-scale processes. The second ETKF includes the uncertainty due to unresolved scales as part of the observation error covariance matrix and is denoted the ETKF-RH. We found that the ETSKF using either sampling strategy is capable of compensating for uncertainty due to unresolved scales in nonlinear dynamical systems and performs similarly to the ETKF-RH. Furthermore, the ETSKF is most beneficial when the uncertainty due to unresolved scales is greater than the instrument uncertainty. We were unable to determine the superior sampling strategy for the ETSKF due to the weak correlation between the large- and small-scale processes in the swinging spring system. In the next chapter, we conclude this thesis by answering the research questions given in chapter 1 and discussing potential avenues of future work.

Chapter 9

Conclusions and future work

9.1 Conclusions

The objective of this thesis is to improve high-impact weather prediction via a better treatment of observation uncertainty in data assimilation. This is particularly important for low-cost, crowdsourced observations that are often obtained in urban areas. We investigated observation error due to unresolved scales from two perspectives. We first considered the uncertainty due to unresolved scales of crowdsourced vehicle-based observations of air temperature. This involved quality-controlling vehicle-based observations and exploring the characteristics of the data through comparison with other datasets. We next considered a largely unexplored approach to compensate for observation uncertainty due to unresolved scales in data assimilation. This involved investigating the ability of the Schmidt-Kalman filter to compensate for observation uncertainty due to unresolved scales and extending the filter to an ensemble transform formulation. We posed several research questions to be addressed by the thesis. The main conclusions for each of these questions are listed below.

1. **How can we quality-control the vehicle-based air-temperature dataset obtained from a Met Office proof-of-concept trial?**
   A quality-control procedure consisting of range-validity tests, a time-series check and a new GPS test was developed for the vehicle-based observations. Approximately 25% of the vehicle-based observations passed quality-control. Almost 54% of the original dataset was unable to be tested due to missing air-temperature measurements or invalid speed metadata and a further 12.4% failed quality-control due to inaccurate location metadata. Therefore, the quality of the available metadata is key for the quality-control of vehicle-based observations.

2. **What are the characteristics of vehicle-based observations of air tem-**

perature?

The characteristics of vehicle-based observations of air temperature were explored through comparison with UKV model data, roadside weather information station data and Met Office integrated data archive system data. For this dataset, we found that the standard deviation of the observation-model comparisons is approximately between 1.2°C and 1.6°C. Examining the statistics of the vehicle-based observation–model comparisons under different weather conditions showed that the variability is greatest in sunny weather conditions and smallest in rainy weather conditions. These results show that the uncertainty of vehicle-based observation–model comparisons is likely weather dependent. Examination of the observation–model comparisons for specific vehicles revealed that the uncertainty may also be vehicle-dependent.

3. **Can the Schmidt-Kalman filter effectively treat observation error and bias due to unresolved scales?**

   Using a random walk model with one variable for the large-scale processes and one variable for the small-scale processes, the Schmidt-Kalman filter was shown to produce better estimates of the large-scale state than a Kalman filter that completely disregards the small-scale processes. Furthermore, the Schmidt-Kalman filter is most suitable in regimes of high uncertainty due to unresolved scales and small instrument uncertainty. Observation biases due to unresolved scales and unbiased error due to unresolved scales may be accounted for by using the Schmidt-Kalman filter in conjunction with a bias correction scheme.

4. **How can the Schmidt-Kalman filter be adapted for nonlinear models?**

   A new ensemble transform formulation of the Schmidt-Kalman filter was derived which samples a small-scale ensemble from the representation error covariance to compensate for observation uncertainty due to unresolved scales. The performance of the ensemble transform Schmidt-Kalman filter using two different sampling methods was evaluated using the swinging spring model. The ensemble transform Schmidt-Kalman filter was shown to be most beneficial when the uncertainty due to unresolved scales is greater than the instrument uncertainty. This agrees with the results of the Schmidt-Kalman filter for a linear system as investigated for research question 3. In such situations, the ensemble transform Schmidt-Kalman filter was shown to perform similarly well to the ensemble transform Kalman filter that includes the uncertainty due to unresolved scales as part of the observation error covariance matrix. Due to the weak correlation between the large- and small-scale processes in the swinging spring system, we were unable to determine the superior sampling strategy.

9.2 Future work

The work produced for this thesis opens many interesting areas of future research. Potential avenues of future work for vehicle-based observations and the Schmidt-Kalman filter are described below.

9.2.1 Vehicle-based observations

The vehicle-based temperature dataset studied in this thesis was initially small in size and was further reduced through quality-control. An extension to our work would be to conduct a new trial with improved data collection (see section 5.6), more accurate metadata and more participants to increase the spatio-temporal density of the observations. Such a trial would allow for improved quality-control (e.g., Boyce et al., 2017; Siems-Anderson et al., 2019) and could reduce the percentage of observations immediately discarded due to poor metadata. The scope of trial may be increased through observing more meteorological variables through available built-in vehicle sensors or through use of externally mounted sensors. It would also be interesting to compare observations of the same variable obtained from built-in and externally-mounted sensors.

Using the data from the improved trial, a more extensive exploration of the characteristics of vehicle-based observations may be conducted. Further examination of the uncertainty of the observation-model comparisons for different weather conditions, seasons and vehicle make would help quantify the observation uncertainty for data assimilation. Provided additional metadata such as the height or sensor make was available, it may be feasible to quantify the contribution from each component of the observation error to the bias and uncertainty of the observation–model comparisons. The effects of elevation and atmospheric stability (i.e. day or night) on the uncertainty of the observation–model comparisons would provide other interesting lines of exploration.

Provided a sufficient amount of high-quality vehicle-based observations are obtained and their uncertainty has been adequately quantified, the next step in the study of vehicle-based observations would be to conduct assimilation experiments. While Siems-Anderson et al. (2020) have performed assimilation experiments using simulated vehicle-based observations, the spatio-temporal distribution of observations was homogeneous and the observation uncertainty consisted entirely of instrument uncertainty. It is likely that real vehicle-based observation networks will be heterogeneous and the observation uncertainty will vary with weather type. Such issues must be considered when assimilating vehicle-based observations and ultimately determine their benefit to convection-permitting forecasts.

9.2.2 The ensemble transform Schmidt-Kalman filter

There are a number of questions regarding the performance of the ensemble transform Schmidt-Kalman filter (ETSKF) that must be addressed before it may be considered in operational numerical weather prediction. The most important next step for the ETSKF is to examine the performance of the filter on an idealised model of the atmosphere. As the ETSKF is motivated by convection-permitting numerical weather prediction, a suitable candidate may be the modified rotating shallow water model of Kent et al. (2017). In addition, dynamical models of processes with strong correlations between the large- and small-scales, such as boundary layer turbulence (e.g., Kaimal and Finnigan, 1994, pp 32–65), would be suitable to determine the benefit of the cross-covariances in the ETSKF.

Covariance inflation and localisation are integral for the successful implementation of ensemble Kalman filter algorithms for systems of large dimensionality. For our implementation of the ETSKF, we used additive inflation with a climatological model error covariance. Further tuning experiments using additive and/or multiplicative inflation in the ETSKF could provide further insight into the optimal implementation of the filter. Localisation was not required for our experiments because of the low dimensionality of the forecast model and the large ensemble used in our experiments. In principle, localisation for the ETSKF may be implemented in the same manner as the ensemble transform Kalman filter (ETKF). However, the optimal localisation length-scales may differ between the ETSKF and the ETKF when applied to the same system due to the difference in treatment between the uncertainty due to unresolved scales. It would therefore be beneficial to examine how the optimal localisation length-scale for the ETSKF compares with that of the ETKF for different assimilation systems. It would also be interesting to examine how the treatment of the small-scale variability in the ETSKF affects adaptive inflation and localisation techniques.

Another interesting avenue of exploration for the ETSKF involves the approximation of the statistics of the small-scale processes. In this thesis, a sample was drawn from a climatological approximation of the representation error covariance to act as a small-scale ensemble. In operational settings, where ensemble sizes are limited, there will likely be substantial sampling error in this small-scale ensemble resulting in the uncertainty due to unresolved scales being poorly accounted for by the ETSKF. The implications of this sampling error and how to compensate for it are key to understanding the limitations of the ETSKF. It would also be interesting to consider alternative methods to obtain the statistics of the small-scale processes. For example, Grooms et al. (2014) use stochastic physics to provide estimates of the statistics of the small-scale processes. It has been suggested by Janjić et al. (2018) that this approach could be suitable for an ensemble formulation of the Schmidt-

Kalman filter.

Bibliography

Abdelhamid, S., Hassanein, H. S., and Takahara, G. (2014). Vehicle as a mobile sensor. *Procedia Computer Science*, 34:286–295.

Anandakumar, K. (1999). A study on the partition of net radiation into heat fluxes on a dry asphalt surface. *Atmospheric Environment*, 33(24-25):3911–3918.

Anderson, A. R., Chapman, M., Drobot, S. D., Tadesse, A., Lambi, B., Wiener, G., and Pisano, P. (2012). Quality of mobile air temperature and atmospheric pressure observations from the 2010 development test environment experiment. *Journal of Applied Meteorology and Climatology*, 51(4):691–701.

Anderson, J. (2009). Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus A: Dynamic meteorology and oceanography*, 61(1):72–83.

Anderson, J. L. (2001). An ensemble adjustment Kalman filter for data assimilation. *Monthly weather review*, 129(12):2884–2903.

Anderson, J. L. (2012). Localization and sampling error correction in ensemble Kalman filter data assimilation. *Monthly Weather Review*, 140(7):2359–2371.

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

Aravéquia, J. A., Szunyogh, I., Fertig, E. J., Kalnay, E., Kuhl, D., and Kostelich, E. J. (2011). Evaluation of a strategy for the assimilation of satellite radiance observations with the local ensemble transform Kalman filter. *Monthly Weather Review*, 139(6):1932–1951.

Asch, M., Bocquet, M., and Nodet, M. (2016). *Data assimilation: methods, algorithms, and applications*. SIAM.

Asher, R. B., Herring, K. D., and Ryles, J. C. (1976). Bias, variance, and estimation error in reduced order filters. *Automatica*, 12(6):589–600.

Asher, R. B. and Reeves, R. M. (1975). Performance evaluation of suboptimal filters. *IEEE Transactions on Aerospace and Electronic Systems*, AES-11(3):400–405.

Ballard, S. P., Li, Z., Simonin, D., and Caron, J.-F. (2016). Performance of 4D-Var NWP-based nowcasting of precipitation at the Met Office for summer 2012. *Quarterly Journal of the Royal Meteorological Society*, 142(694):472–487.

Bannister, R. (2017). A review of operational methods of variational and ensemble-variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 143(703):607–633.

Bathmann, K. and Collard, A. (2021). Surface-dependent correlated infrared observation errors and quality control in the FV3 framework. *Quarterly Journal of the Royal Meteorological Society*, 147(734):408–424.

Bauer, C. (2013). On the (in-) accuracy of GPS measures of smartphones: a study of running tracking applications. In *Proceedings of International Conference on Advances in Mobile Computing & Multimedia*, page 335. ACM.

Bell, S., Cornford, D., and Bastin, L. (2013). The state of automated amateur weather observations. *Weather*, 68(2):36–41.

Bell, S., Cornford, D., and Bastin, L. (2015). How good are citizen weather stations? Addressing a biased opinion. *Weather*, 70(3):75–84.

Bell, Z., Dance, S. L., and 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.

Bell, Z., Dance, S. L., and Waller, J. A. (2021a). Exploring the characteristics of a vehicle-based temperature dataset for convection-permitting numerical weather prediction. *Meteorological Applications*. In review. `https://arxiv.org/abs/2105.12526`.

Bell, Z., Dance, S. L., Waller, J. A., and O'Boyle, K. (2021b). Quality-control of vehicle-based temperature observations and future recommendations. *Forecasting Research Technical Report 644, Met Office, United Kingdom*.

Bennitt, G., Johnson, H., Weston, P., Jones, J., and Pottiaux, E. (2017). An assessment of ground-based GNSS Zenith Total Delay observation errors and their correlations using the Met Office UKV model. *Quarterly Journal of the Royal Meteorological Society*, 143(707):2436–2447.

BIPM, IEC, IFCC, ILAC, IUPAC, IUPAP, ISO, and OIML (2012). *International vocabulary of metrology – Basic and general concepts and associated terms (VIM)*, volume 3rd edn. JCGM 200:2012.

Bishop, C. and Hodyss, D. (2009a). Ensemble covariances adaptively localized with ECO-RAP. Part 1: tests on simple error models. *Tellus A: Dynamic Meteorology and Oceanography*, 61(1):84–96.

Bishop, C. and Hodyss, D. (2009b). Ensemble covariances adaptively localized with ECO-RAP. Part 2: a strategy for the atmosphere. *Tellus A: Dynamic Meteorology and Oceanography*, 61(1):97–111.

Bishop, C. H., Etherton, B. J., and Majumdar, S. J. (2001). Adaptive sampling with the ensemble transform Kalman filter. Part I: theoretical aspects. *Monthly weather review*, 129(3):420–436.

Blair, G. S., Bassett, R., Bastin, L., Beevers, L., Borrajo, M. I., Brown, M., Dance, S. L., Dionescu, A., Edwards, L., Ferrario, M. A., Fraser, R., Fraser, H., Gardner, S., Henrys, P., Hey, T., Homann, S., Huijbers, C., Hutchison, J., Jonathan, P., Lamb, R., Laurie, S., Leeson, A., Leslie, D., McMillan, M., Nundloll, V., Oyebamiji, O., Phillipson, J., Pope, V., Prudden, R., Reis, S., Salama, M., Samreen, F., Sejdinovic, D., Simm, W., Street, R., Thornton, L., Towe, R., Hey, J. V., Vieno, M., Waller, J., and Watkins, J. (2021). The role of digital technologies in responding to the grand challenges of the natural environment: The Windermere Accord. *Patterns*, 2(1):100156.

Bocquet, M. (2011). Ensemble Kalman filtering without the intrinsic need for inflation. *Nonlinear Processes in Geophysics*, 18(5):735–750.

Bonavita, M., Isaksen, L., and Hólm, E. (2012). On the use of EDA background error variances in the ECMWF 4D-Var. *Quarterly journal of the royal meteorological society*, 138(667):1540–1559.

Bormann, N. and Bauer, P. (2010). Estimates of spatial and interchannel observation-error characteristics for current sounder radiances for numerical weather prediction. Part I: methods and application to ATOVS data. *Quarterly Journal of the Royal Meteorological Society*, 136(649):1036–1050.

Bormann, N., Bonavita, M., Dragani, R., Eresmaa, R., Matricardi, M., and McNally, A. (2016). Enhancing the impact of IASI observations through an updated observation-error covariance matrix. *Quarterly Journal of the Royal Meteorological Society*, 142(697):1767–1780.

Bormann, N., Collard, A., and Bauer, P. (2010). Estimates of spatial and interchannel observation-error characteristics for current sounder radiances for numerical weather prediction. Part II: application to AIRS and IASI data. *Quarterly Journal of the Royal Meteorological Society*, 136:1051–1063.

Boyce, B., Weiner, G., Anderson, A., and Linden, S. (2017). Pikalert® system vehicle data translator (VDT) utilizing integrated mobile observations: Pikalert VDT enhancements, operations, & maintenance. Technical report, United States. Department of Transportation.

Brandt, S. (1999). *Data Analysis: Statistical and Computational Methods for Scientists and Engineers*. Springer.

Bremner, R. (2019). Ever-growing cars: why they keep on getting wider. *Autocar*. `https://www.autocar.co.uk/car-news/features/investigation-why-are-cars-becoming-so-wide`. Accessed: 14-01-2021.

Brown, R. and Sage, A. (1971). Analysis of modeling and bias errors in discrete-time state estimation. *IEEE Transactions on Aerospace and Electronic Systems*, AES-7(2):340–354.

Brown, R. G. and Hwang, P. Y. (2012). *Introduction to random signals and applied Kalman filtering: with MATLAB exercises*. John Wiley & Sons.

Burgers, G., Jan van Leeuwen, P., and Evensen, G. (1998). Analysis scheme in the ensemble Kalman filter. *Monthly weather review*, 126(6):1719–1724.

Buttell, R., Keer, L., Moles, M., Mansfield, A., Hanson, E., and Waring, P. (2020). *SWIS user guide v19*. Highways England.

Campbell, W. F., Satterfield, E. A., Ruston, B., and Baker, N. L. (2017). Accounting for correlated observation error in a dual-formulation 4D variational data assimilation system. *Monthly Weather Review*, 145(3):1019–1032.

Candille, G. and Talagrand, O. (2005). Evaluation of probabilistic prediction systems for a scalar variable. *Quarterly Journal of the Royal Meteorological Society*, 131(609):2131–2150.

Candy, B., Cotton, J., and Eyre, J. (2021). Recent results of observation data denial experiments. *Forecasting Research Technical Report 641, Met Office, United Kingdom*.

Carrassi, A. and 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.

Casati, B., Wilson, L., Stephenson, D., Nurmi, P., Ghelli, A., Pocernich, M., Damrath, U., Ebert, E., Brown, B., and Mason, S. (2008). Forecast verification: current status and future directions. *Meteorological Applications*, 15(1):3–18.

Chapman, L., Bell, C., and Bell, S. (2017). Can the crowdsourcing data paradigm take atmospheric science to a new level? A case study of the urban heat island of London quantified using Netatmo weather stations. *International Journal of Climatology*, 37(9):3597–3605.

Chapman, L. and Thornes, J. (2011). What spatial resolution do we need for a route-based road weather decision support system? *Theoretical and applied climatology*, 104(3-4):551–559.

Chapman, L. and Thornes, J. E. (2005). The influence of traffic on road surface temperatures: implications for thermal mapping studies. *Meteorological Applications*, 12(4):371–380.

Chapman, L., Thornes, J. E., and Bradley, A. V. (2001). Modelling of road surface temperature from a geographical parameter database. Part 1: statistical. *Meteorological Applications*, 8(4):409–419.

Chapman, M., Drobot, S., Jensen, T., Johansen, C., Mahoney III, W., Pisano, P., and McKeever, B. (2010). Diagnosing road weather conditions with vehicle probe data: results from Detroit IntelliDrive field study. *Transportation Research Record*, 2169(1):116–127.

Clark, M. R., Webb, J. D., and Kirk, P. J. (2018). Fine-scale analysis of a severe hailstorm using crowd-sourced and conventional observations. *Meteorological Applications*, 25(3):472–492.

Clark, P., Roberts, N., Lean, H., Ballard, S. P., and Charlton-Perez, C. (2016). Convection-permitting models: a step-change in rainfall forecasting. *Meteorological Applications*, 23(2):165–181.

Clark, P. A. (1998). The implementation and performance of a 1D model coupled to NWP forcing for low-cost site-specific forecasting. *AMS 2nd Urban Symposium, Albuquerque, New Mexico, Nov 1998*.

Clayton, A. M., Lorenc, A. C., and Barker, D. M. (2013). Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *Quarterly Journal of the Royal Meteorological Society*, 139(675):1445–1461.

Cordoba, M., Dance, S. L., Kelly, G. A., Nichols, N. K., and Waller, J. A. (2017). Diagnosing atmospheric motion vector observation errors for an operational high-resolution data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 143(702):333–341.

Cosgrove, B. A., Lohmann, D., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Marshall, C., Sheffield, J., Duan, Q., et al. (2003). Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project. *Journal of Geophysical Research: Atmospheres*, 108(D22).

Coulson, R., Evans, B., and Skea, A. (2012). Operational OpenRoad Verification. *16th International Road Weather Conference, Helsinki, Finland, May 2012*.

Courtier, P., Thépaut, J.-N., and Hollingsworth, A. (1994). A strategy for operational implementation of 4D-Var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, 120(519):1367–1387.

Daley, R. (1993). Estimating observation error statistics for atmospheric data assimilation. *Ann. Geophysicae*, 11:634–647.

Dance, S. L., Ballard, S. P., Bannister, R. N., Clark, P., Cloke, H. L., Darlington, T., Flack, D. L., Gray, S. L., Hawkness-Smith, L., Husnoo, N., 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.

De Rosnay, P., Drusch, M., Vasiljevic, D., Balsamo, G., Albergel, C., and Isaksen, L. (2013). A simplified extended Kalman filter for the global operational soil moisture analysis at ECMWF. *Quarterly Journal of the Royal Meteorological Society*, 139(674):1199–1213.

de Vos, L., Droste, A., Zander, M., Overeem, A., Leijnse, H., Heusinkveld, B., Steeneveld, G., and Uijlenhoet, R. (2019a). Hydrometeorological monitoring using opportunistic sensing networks in the Amsterdam metropolitan area. *Bulletin of the American Meteorological Society*, (2019).

de Vos, L., Leijnse, H., Overeem, A., and Uijlenhoet, R. (2017). The potential of urban rainfall monitoring with crowdsourced automatic weather stations in Amsterdam. *Hydrology and Earth System Sciences*, 21(2):765–777.

de Vos, L., Raupach, T., Leijnse, H., Overeem, A., Berne, A., and Uijlenhoet, R. (2018). High-resolution simulation study exploring the potential of radars, crowdsourced personal weather stations, and commercial microwave links to monitor small-scale urban rainfall. *Water Resources Research*, 54(12):10–293.

de Vos, L. W., Leijnse, H., Overeem, A., and Uijlenhoet, R. (2019b). Quality control for crowdsourced personal weather stations to enable operational rainfall monitoring. *Geophysical Research Letters*, 46(15):8820–8829.

Dee, D. P. (2004). Variational bias correction of radiance data in the ECMWF system. In *Proceedings of the ECMWF workshop on assimilation of high spectral resolution sounders in NWP, Reading, UK*, volume 28, pages 97–112.

Dee, D. P. (2005). Bias and data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3323–3343.

Derber, J. C. and Wu, W.-S. (1998). The use of TOVS cloud-cleared radiances in the NCEP SSI analysis system. *Monthly Weather Review*, 126(8):2287–2299.

Desroziers, G., Berre, L., Chapnik, B., and 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.

Devarakonda, S., Sevusu, P., Liu, H., Liu, R., Iftode, L., and Nath, B. (2013).

Real-time air quality monitoring through mobile sensing in metropolitan areas. In *Proceedings of the 2nd ACM SIGKDD international workshop on urban computing*, pages 1–8.

Donegan, B. (2017). Why your car thermometer is wrong. `https://weather.com/science/weather-explainers/news/car-thermometer-thermistor-temperature-wrong`. Accessed: 27-01-2020.

Dormand, J. R. and Prince, P. J. (1980). A family of embedded Runge-Kutta formulae. *Journal of computational and applied mathematics*, 6(1):19–26.

Drobot, S., Chapman, M., Lambi, B., Wiener, G., Anderson, A., et al. (2011). The vehicle data translator v3. 0 system description. Technical report, United States. Joint Program Office for Intelligent Transportation Systems.

Drobot, S., Chapman, M., Schuler, E., Wiener, G., William Mahoney, I., Pisano, P., and McKeever, B. (2010). Improving road weather hazard products with vehicle probe data: vehicle data translator quality-checking procedures. *Transportation Research Record*, 2169(1):128–140.

Droste, A., Pape, J.-J., Overeem, A., Leijnse, H., Steeneveld, G.-J., Van Delden, A., and Uijlenhoet, R. (2017). Crowdsourcing urban air temperatures through smartphone battery temperatures in São Paulo, Brazil. *Journal of Atmospheric and Oceanic Technology*, 34(9):1853–1866.

Droste, A. M., Heusinkveld, B. G., Fenner, D., and Steeneveld, G.-J. (2020). Assessing the potential and application of crowdsourced urban wind data. *Quarterly Journal of the Royal Meteorological Society*.

Dutra, E., Muñoz-Sabater, J., Boussetta, S., Komori, T., Hirahara, S., and Balsamo, G. (2020). Environmental lapse rate for high-resolution land surface downscaling: an application to ERA5. *Earth and Space Science*, 7(5):e2019EA000984.

El Gharamti, M. (2018). Enhanced adaptive inflation algorithm for ensemble filters. *Monthly Weather Review*, 146(2):623–640.

Essery, R., Best, M., Betts, R., Cox, P. M., and Taylor, C. M. (2003). Explicit representation of subgrid heterogeneity in a GCM land surface scheme. *Journal of Hydrometeorology*, 4(3):530–543.

Evensen, G. (1992). Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model. *Journal of Geophysical Research: Oceans*, 97(C11):17905–17924.

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5):10143–10162.

Eyre, J. (2016). Observation bias correction schemes in data assimilation systems: a theoretical study of some of their properties. *Quarterly Journal of the Royal Meteorological Society*, 142(699):2284–2291.

Eyre, J. R., English, S. J., and Forsythe, M. (2020). Assimilation of satellite data in numerical weather prediction. Part I: the early years. *Quarterly Journal of the Royal Meteorological Society*, 146(726):49–68.

Fertig, E., Baek, S.-J., Hunt, B., Ott, E., Szunyogh, I., Aravéquia, J., Kalnay, E., Li, H., and Liu, J. (2009). Observation bias correction with an ensemble Kalman filter. *Tellus A: Dynamic Meteorology and Oceanography*, 61(2):210–226.

Fiebrich, C. A., Morgan, C. R., McCombs, A. G., Hall Jr, P. K., and McPherson, R. A. (2010). Quality assurance procedures for mesoscale meteorological data. *Journal of Atmospheric and Oceanic Technology*, 27(10):1565–1582.

Fielding, M. and Stiller, O. (2019). Characterizing the representativity error of cloud profiling observations for data assimilation. *Journal of Geophysical Research: Atmospheres*, 124(7):4086–4103.

FierceElectronics (2014). Temperature sensors to generate $8 billion revenue by 2020. `https://www.fierceelectronics.com/components/temperature-sensors-to-generate-8-billion-revenue-by-2020`. Accessed: 08-01-2021.

Fowler, A., Dance, S., and Waller, J. (2018). On the interaction of observation and prior error correlations in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 144(710):48–62.

Friedland, B. (1969). Treatment of bias in recursive filtering. *IEEE Transactions on Automatic Control*, 14(4):359–367.

Frogner, I.-L., Andrae, U., Bojarova, J., Callado, A., Escribà, P., Feddersen, H., Hally, A., Kauhanen, J., Randriamampianina, R., Singleton, A., et al. (2019). HarmonEPS—the HARMONIE ensemble prediction system. *Weather and Forecasting*, 34(6):1909–1937.

Furrer, R. and Bengtsson, T. (2007). Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. *Journal of Multivariate Analysis*, 98(2):227–255.

Gandin, L. S. (1988). Complex quality control of meteorological observations. *Monthly Weather Review*, 116(5):1137–1156.

Garand, L., Heilliette, S., and Buehner, M. (2007). Interchannel error correlation associated with airs radiance observations: inference and impact in data assimilation. *Journal of applied meteorology and climatology*, 46(6):714–725.

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

Gauthier, P., Du, P., Heilliette, S., and Garand, L. (2018). Convergence issues in the estimation of interchannel correlated observation errors in infrared radiance data. *Monthly Weather Review*, 146(10):3227–3239.

Geer, A. J., Migliorini, S., and Matricardi, M. (2019). All-sky assimilation of infrared radiances sensitive to mid-and upper-tropospheric moisture and cloud. *Atmospheric Measurement Techniques*, 12(9):4903–4929.

Gelb, A. (1974). *Applied optimal estimation*. MIT press.

Greybush, S. J., Kalnay, E., Miyoshi, T., Ide, K., and Hunt, B. R. (2011). Balance and ensemble Kalman filter localization techniques. *Monthly Weather Review*, 139(2):511–522.

Grooms, I., Lee, Y., and Majda, A. J. (2014). Ensemble Kalman filters for dynamical systems with unresolved turbulence. *Journal of Computational Physics*, 273:435–452.

Gustafsson, N., Janjić, T., Schraff, C., Leuenberger, D., Weissmann, M., Reich, H., Brousseau, P., Montmerle, T., Wattrelot, E., Bučánek, A., 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.

Gustavsson, T., Bogren, J., and Green, C. (2001). Road climate in cities: a study of the Stockholm area, South-East Sweden. *Meteorological Applications*, 8(4):481–489.

Ha, S.-Y. and Snyder, C. (2014). Influence of surface observations in mesoscale data assimilation using an ensemble Kalman filter. *Monthly Weather Review*, 142(4):1489–1508.

Haberlandt, U. and Sester, M. (2010). Areal rainfall estimation using moving cars as rain gauges - a modelling study. *Hydrology and Earth System Sciences 14 (2010), Nr. 7*, 14(7):1139–1151.

Hamill, T. M. (2001). Interpretation of rank histograms for verifying ensemble forecasts. *Monthly Weather Review*, 129(3):550–560.

Hamill, T. M., Whitaker, J. S., and Snyder, C. (2001). Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Monthly Weather Review*, 129(11):2776–2790.

Hanson, G. S. (2016). Impact of assimilating surface pressure observations from smartphones on a regional, high-resolution ensemble forecast: observing system simulation experiments. *Pennsylvania State University Masters Thesis*.

Harrison, R. G. (2015). *Meteorological measurements and instrumentation.* , chapter 5, pages 90–95. John Wiley & Sons.

Harrison, R. G. and Burt, S. (2020). Shall I compare thee to a summer's day? Art thou more temperate?... Sometimes too hot the eye of heaven shines.... *Weather*, 75(6):172–174.

Healy, S. and White, A. (2005). Use of discrete Fourier transforms in the 1D-Var retrieval problem. *Quarterly Journal of the Royal Meteorological Society*, 131(605):63–72.

Hersbach, H. (2000). Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, 15(5):559–570.

Hertwig, D., Grimmond, S., Hendry, M. A., Saunders, B., Wang, Z., Jeoffrion, M., Vidale, P. L., McGuire, P. C., Bohnenstengel, S. I., Ward, H. C., et al. (2020). Urban signals in high-resolution weather and climate simulations: role of urban land-surface characterisation. *Theoretical and Applied Climatology*, pages 1–28.

Hess, B., Farahani, A. Z., Tschirschnitz, F., and von Reischach, F. (2012). Evaluation of fine-granular GPS tracking on smartphones. In *Proceedings of the First ACM SIGSPATIAL International Workshop on Mobile Geographic Information Systems*, pages 33–40. ACM.

Highways England (2018). Historical environmental sensor station data. Data obtained with permission from `https://www.swis.highwaysengland.co.uk/`. Accessed: 30-04-2020.

Highways England (2020). SWIS support. Personal communication on 11-09-2020.

Hintz, K., Vedel, H., Kaas, E., and Nielsen, N. (2020). Estimation of wind speed and soughness length using smartphones: method and quality assessment. *Journal of Atmospheric and Oceanic Technology*, 37(8):1319–1332.

Hintz, K. S., McNicholas, C., Randriamampianina, R., Williams, H. T., Macpherson, B., Mittermaier, M., Onvlee-Hooimeijer, J., and Szintai, B. (2021). Crowdsourced observations for short-range numerical weather prediction: report from EWLAM/SRNWP meeting 2019. *Atmospheric Science Letters*, page e1031.

Hintz, K. S., O'Boyle, K., Dance, S. L., Al-Ali, S., Ansper, I., Blaauboer, D., Clark, M., Cress, A., Dahoui, M., Darcy, R., et al. (2019a). Collecting and utilising crowdsourced data for numerical weather prediction: propositions from the meet-

ing held in Copenhagen, 4–5 December, 2018. *Atmospheric Science Letters*, page e921.

Hintz, K. S., Vedel, H., and Kaas, E. (2019b). Collecting and processing of barometric data from smartphones for potential use in numerical weather prediction data assimilation. *Meteorological Applications*, 26(4):733–746.

Hodyss, D. and Nichols, N. (2015). The error of representation: basic understanding. *Tellus A: Dynamic Meteorology and Oceanography*, 67(1):24822.

Hodyss, D. and Satterfield, E. (2016). Mathematical concepts of data assimilation. In Park, S. K. and Xu, L., editors, *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications*, volume 3, pages 177–194. Springer, Berlin. ISBN 978-3-319-43415-5-8.

Hollingsworth, A. and Lönnberg, P. (1986). The statistical structure of short-range forecast errors as determined from radiosonde data. Part i: the wind field. *Tellus A*, 38(2):111–136.

Houtekamer, P., Mitchell, H. L., and Deng, X. (2009). Model error representation in an operational ensemble Kalman filter. *Monthly Weather Review*, 137(7):2126–2143.

Houtekamer, P. and Zhang, F. (2016). Review of the ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 144(12):4489–4532.

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

Hu, Y., Almkvist, E., Gustavsson, T., and Bogren, J. (2019). Modeling road surface temperature from air temperature and geographical parameters — implication for the application of floating car data in a road weather forecast model. *Journal of Applied Meteorology and Climatology*, 58(5):1023–1038.

Hunt, B. R., Kostelich, E. J., and Szunyogh, I. (2007). Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena*, 230(1-2):112–126.

Huwald, H., Higgins, C. W., Boldi, M.-O., Bou-Zeid, E., Lehning, M., and Parlange, M. B. (2009). Albedo effect on radiative errors in air temperature measurements. *Water Resources Research*, 45(8).

Ignagni, M. (1981). An alternate derivation and extension of Friendland's two-stage Kalman estimator. *IEEE Transactions on Automatic Control*, 26(3):746–750.

Janjić, T., Bormann, N., Bocquet, M., Carton, J., Cohn, S., Dance, S., Losa, S., Nichols, N., Potthast, R., Waller, J., et al. (2018). On the representation er-

ror in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 144(713):1257–1278.

Janjić, T. and Cohn, S. E. (2006). Treatment of observation error due to unresolved scales in atmospheric data assimilation. *Monthly Weather Review*, 134(10):2900–2915.

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

Jolliffe, I. T. and Stephenson, D. B. (2012). *Forecast verification: a practitioner's guide in atmospheric science*. John Wiley & Sons.

Kaimal, J. C. and Finnigan, J. J. (1994). *Atmospheric boundary layer flows: their structure and measurement*. Oxford university press.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45.

Kalnay, E. (2003). *Atmospheric modeling, data assimilation and predictability*. Cambridge university press.

Karsisto, V. and Lovén, L. (2019). Verification of road surface temperature forecasts assimilating data from mobile sensors. *Weather and Forecasting*, 34(3):539–558.

Karsisto, V. and Nurmi, P. (2016). Using car observations in road weather forecasting. 18th International Road Weather Conference, Fort Collins, Colorado, Apr 2016.

Karspeck, A. R. (2016). An ensemble approach for the estimation of observational error illustrated for a nominal 1° global ocean model. *Monthly Weather Review*, 144(5):1713–1728.

Kent, T., Bokhove, O., and Tobias, S. (2017). Dynamics of an idealized fluid model for investigating convective-scale data assimilation. *Tellus A: Dynamic Meteorology and Oceanography*, 69(1):1369332.

Kirk, P. J., Clark, M. R., and Creed, E. (2021). Weather observations website. *Weather*, 76(2):47–49.

Knight, S., Smith, C., and Roberts, M. (2010). Mapping Manchester's urban heat island. *Weather*, 65(7):188–193.

Lange, H. and Janjić, T. (2016). Assimilation of Mode-S EHS aircraft observations in COSMO-KENDA. *Monthly Weather Review*, 144(5):1697–1711.

Lawless, A., Gratton, S., and Nichols, N. (2005). An investigation of incremental 4D-Var using non-tangent linear models. *Quarterly Journal of the Royal Meteorological Society*, 131(606):459–476.

Lea, D., Drecourt, J.-P., Haines, K., and Martin, M. (2008). Ocean altimeter assimilation with observational- and model-bias correction. *Quarterly Journal of the Royal Meteorological Society*, 134(636):1761–1774.

Lean, H. W., Clark, P. A., Dixon, M., Roberts, N. M., Fitch, A., Forbes, R., and Halliwell, C. (2008). Characteristics of high-resolution versions of the Met Office Unified Model for forecasting convection over the United Kingdom. *Monthly Weather Review*, 136(9):3408–3424.

Li, H., Kalnay, E., and Miyoshi, T. (2009). Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Quarterly Journal of the Royal Meteorological Society*, 135(639):523–533.

Li, Z., Feng, S., Liu, Y., Lin, W., Zhang, M., Toto, T., Vogelmann, A. M., and Endo, S. (2015). Development of fine-resolution analyses and expanded large-scale forcing properties: 1. Methodology and evaluation. *Journal of Geophysical Research: Atmospheres*, 120(2):654–666.

Limber, M., Drobot, S., Fowler, T., et al. (2010). Clarus quality checking algorithm documentation report. Technical report, United States. Intelligent Transportation Systems Standards Program.

Liu, Z.-Q. and 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.

Livings, D. (2005). Aspects of the ensemble Kalman filter. *Reading University Masters Thesis*.

Livings, D. M., Dance, S. L., and Nichols, N. K. (2008). Unbiased ensemble square root filters. *Physica D: Nonlinear Phenomena*, 237(8):1021–1028.

Lorenc, A. (1981). A global three-dimensional multivariate statistical interpolation scheme. *Monthly Weather Review*, 109(4):701–721.

Lorenc, A., Bell, R., and Macpherson, B. (1991). The meteorological office analysis correction data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, 117(497):59–89.

Lorenc, A. C. (2003). The potential of the ensemble Kalman filter for NWP — a comparison with 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, 129(595):3183–3203.

Lorenc, A. C. and Jardak, M. (2018). A comparison of hybrid variational data assimilation methods for global NWP. *Quarterly Journal of the Royal Meteorological Society*, 144(717):2748–2760.

Lou, T.-s., Chen, N.-h., Xiong, H., Li, Y.-x., and Wang, L. (2018). Ensemble consider Kalman filtering. In *2018 IEEE CSAA Guidance, Navigation and Control Conference (CGNCC)*, pages 1–5. IEEE.

Luo, X. and Hoteit, I. (2011). Robust ensemble filtering and its relation to covariance inflation in the ensemble Kalman filter. *Monthly Weather Review*, 139(12):3938–3953.

Lynch, P. (2002). The swinging spring: a simple model of atmospheric balance. In Norbury, J. and Roulstone, I., editors, *Large-scale Atmosphere-Ocean Dynamics II*, pages 64–108. Cambridge University Press.

Madaus, L. E., Hakim, G. J., and Mass, C. F. (2014). Utility of dense pressure observations for improving mesoscale analyses and forecasts. *Monthly Weather Review*, 142(7):2398–2413.

Madaus, L. E. and Mass, C. F. (2017). Evaluating smartphone pressure observations for mesoscale analyses and forecasts. *Weather and Forecasting*, 32(2):511–531.

Mahoney III, W. P. and O'Sullivan, J. M. (2013). Realizing the potential of vehicle-based observations. *Bulletin of the American Meteorological Society*, 94(7):1007–1018.

Mass, C. F. and Madaus, L. E. (2014). Surface pressure observations from smartphones: a potential revolution for high-resolution weather prediction? *Bulletin of the American Meteorological Society*, 95(9):1343–1349.

McNicholas, C. and Mass, C. F. (2018a). Impacts of assimilating smartphone pressure observations on forecast skill during two case studies in the Pacific Northwest. *Weather and Forecasting*, 33(5):1375–1396.

McNicholas, C. and Mass, C. F. (2018b). Smartphone pressure collection and bias correction using machine learning. *Journal of Atmospheric and Oceanic Technology*, 35(3):523–540.

Meier, F., Fenner, D., Grassmann, T., Otto, M., and Scherer, D. (2017). Crowdsourcing air temperature from citizen weather stations for urban climate research. *Urban Climate*, 19:170–191.

Ménard, R. (2010). Mathematical concepts of data assimilation. In Lahoz, W., Menard, R., and Khattatov, B., editors, *Data assimilation: making sense of observations*, pages 113–135. Springer.

Ménard, R. (2016). Error covariance estimation methods based on analysis residuals: theoretical foundation and convergence properties derived from simplified observation networks. *Quarterly Journal of the Royal Meteorological Society*, 142(694):257–273.

Menard, T., Miller, J., Nowak, M., and Norris, D. (2011). Comparing the GPS capabilities of the Samsung Galaxy S, Motorola Droid X, and the Apple iPhone for vehicle tracking using FreeSim_Mobile. In *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 985–990. IEEE.

Mercelis, S., Watelet, S., and Casteels, W. (2021). Personal communication on 03-03-2021.

Mercelis, S., Watelet, S., Casteels, W., Bogaerts, T., Van den Bergh, J., Reyniers, M., and Hellinckx, P. (2020). Towards detection of road weather conditions using large-scale vehicle fleets. In *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, pages 1–7. IEEE.

Merry, K. and Bettinger, P. (2019). Smartphone GPS accuracy study in an urban environment. *Public Library of Science*, 14(7).

Met Office (2006). MIDAS: UK daily temperature data. Data retieved from online CEDA archive https://catalogue.ceda.ac.uk/uuid/220a65615218d5c9cc9e4785a3234bd0, NCAS British Atmospheric Data Centre (Accessed on 16/06/2019).

Met Office (2011). Weather observations website. `https://wow.metoffice.gov.uk/`.

Met Office (2016). NWP-UKV: Met Office UK Atmospheric High Resolution Model data. Data retieved from online CEDA archive https://catalogue.ceda.ac.uk/uuid/f47bc62786394626b665e23b658d385f, NCAS British Atmospheric Data Centre. Accessed: 16-06-2019.

Met Office (2018). Daily weather summary. `https://www.metoffice.gov.uk/research/library-and-archive/publications/daily-weather-summary`. Accessed: 15-05-2020.

Met Office (2020). Past weather events. `https://www.metoffice.gov.uk/weather/learn-about/past-uk-weather-events`.

Michel, Y. (2018). Revisiting Fisher's approach to the handling of horizontal spatial correlations of observation errors in a variational framework. *Quarterly Journal of the Royal Meteorological Society*, 144(716):2011–2025.

Milan, M., Macpherson, B., Tubbs, R., Dow, G., Inverarity, G., Mittermaier, M., Halloran, G., Kelly, G., Li, D., Maycock, A., et al. (2020). Hourly 4D-Var in the Met Office UKV operational forecast model. *Quarterly Journal of the Royal Meteorological Society*, 146(728):1281–1301.

Mirza, A. K., Ballard, S. P., Dance, S. L., Maisey, P., Rooney, G. G., and Stone, E. K. (2016). Comparison of aircraft-derived observations with in situ research

aircraft measurements. *Quarterly Journal of the Royal Meteorological Society*, 142(701):2949–2967.

Mitchell, H. L. and Daley, R. (1997a). Discretization error and signal/error correlation in atmospheric data assimilation: (I). all scales resolved. *Tellus A*, 49(1):32–53.

Mitchell, H. L. and Daley, R. (1997b). Discretization error and signal/error correlation in atmospheric data assimilation: (II). the effect of unresolved scales. *Tellus A*, 49(1):54–73.

Mitchell, H. L. and Houtekamer, P. L. (2000). An adaptive ensemble Kalman filter. *Monthly Weather Review*, 128(2):416–433.

Mitchell, L. and 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.

Miyoshi, T., Kalnay, E., and Li, H. (2013). Estimating and including observation-error correlations in data assimilation. *Inverse Problems in Science and Engineering*, 21(3):387–398.

Miyoshi, T., Sato, Y., and Kadowaki, T. (2010). Ensemble Kalman filter and 4D-Var intercomparison with the Japanese operational global analysis and prediction system. *Monthly Weather Review*, 138(7):2846–2866.

Moodey, A. J. (2013). *Instability and regularization for data assimilation*. PhD thesis, University of Reading.

Muller, C., Chapman, L., Johnston, S., Kidd, C., Illingworth, S., Foody, G., Overeem, A., and Leigh, R. (2015). Crowdsourcing for climate and atmospheric sciences: current status and future potential. *International Journal of Climatology*, 35(11):3185–3203.

Nakamura, R. and Mahrt, L. (2005). Air temperature measurement errors in naturally ventilated radiation shields. *Journal of Atmospheric and Oceanic Technology*, 22(7):1046–1058.

Navon, I. M. (2009). Data assimilation for numerical weather prediction: a review. In *Data assimilation for atmospheric, oceanic and hydrologic applications*, pages 21–65. Springer.

Nerger, L., Danilov, S., Hiller, W., and Schröter, J. (2006). Using sea-level data to constrain a finite-element primitive-equation ocean model with a local SEIK filter. *Ocean Dynamics*, 56(5):634–649.

Nichols, N. K. (2010). Mathematical concepts of data assimilation. In Lahoz, W.,

Menard, R., and Khattatov, B., editors, *Data assimilation: making sense of observations*, pages 13–39. Springer.

Nipen, T. N., Seierstad, I. A., Lussana, C., Kristiansen, J., and Hov, Ø. (2020). Adopting citizen observations in operational weather prediction. *Bulletin of the American Meteorological Society*, 101(1):E43–E57.

Oke, P. R. and Sakov, P. (2008). Representation error of oceanic observations for data assimilation. *Journal of Atmospheric and Oceanic Technology*, 25(6):1004–1017.

Oke, T. R., Mills, G., Christen, A., and Voogt, J. A. (2017). *Urban climates.* , chapter 3, pages 44–49. Cambridge University Press.

Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J., Corazza, M., Kalnay, E., Patil, D., and Yorke, J. A. (2004). A local ensemble Kalman filter for atmospheric data assimilation. *Tellus A: Dynamic Meteorology and Oceanography*, 56(5):415–428.

Overeem, A., R. Robinson, J., Leijnse, H., Steeneveld, G.-J., P. Horn, B., and Uijlenhoet, R. (2013). Crowdsourcing urban air temperatures from smartphone battery temperatures. *Geophysical Research Letters*, 40(15):4081–4085.

Padarthy, M. and Heyns, E. (2019). Identification and classification of slippery winter road conditions using commonly available vehicle variables. *Transportation research record*, 2673(2):60–70.

Porson, A., Clark, P. A., Harman, I., Best, M., and Belcher, S. (2010). Implementation of a new urban energy budget scheme in the MetUM. Part I: description and idealized simulations. *Quarterly Journal of the Royal Meteorological Society*, 136(651):1514–1529.

Price, C., Maor, R., and Shachaf, H. (2018). Using smartphones for monitoring atmospheric tides. *Journal of Atmospheric and Solar-Terrestrial Physics*, 174:1–4.

Prusa, J. M., Segal, M., Temeyer, B. R., Gallus Jr, W. A., and Takle, E. S. (2002). Conceptual and scaling evaluation of vehicle traffic thermal effects on snow/ice-covered roads. *Journal of Applied Meteorology*, 41(12):1225–1240.

Pu, Z., Zhang, H., and Anderson, J. (2013). Ensemble Kalman filter assimilation of near-surface observations over complex terrain: comparison with 3DVAR for short-range forecasts. *Tellus A: Dynamic Meteorology and Oceanography*, 65(1):19620.

Public Health England (2021). Heatwave plan for England. `https://`

assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/888668/Heatwave_plan_for_England_2020.pdf.

Raanes, P. N., Carrassi, A., and Bertino, L. (2015). Extending the square root method to account for additive forecast noise in ensemble methods. *Monthly Weather Review*, 143(10):3857–3873.

Rabiei, E., Haberlandt, U., Sester, M., and Fitzner, D. (2013). Rainfall estimation using moving cars as rain gauges - laboratory experiments. *Hydrology and Earth System Sciences 17 (2013), Nr. 11*, 17(11):4701–4712.

Rada, E. C., Ragazzi, M., Brini, M., Marmo, L., Zambelli, P., Chelodi, M., and Ciolli, M. (2016). Perspectives of low-cost sensors adoption for air quality monitoring. In *Air Quality*, pages 29–40. Apple Academic Press.

Rainwater, S., Bishop, C. H., and Campbell, W. F. (2015). The benefits of correlated observation errors for small scales. *Quarterly Journal of the Royal Meteorological Society*, 141(693):3439–3445.

Richardson, S. J., Brock, F. V., Semmer, S. R., and Jirak, C. (1999). Minimizing errors associated with multiplate radiation shields. *Journal of Atmospheric and Oceanic Technology*, 16(11):1862–1872.

Saetra, Ø., Hersbach, H., Bidlot, J.-R., and Richardson, D. S. (2004). Effects of observation errors on the statistics for ensemble spread and reliability. *Monthly Weather Review*, 132(6):1487–1501.

Sasaki, Y. (01 Dec. 1970). Some basic formalisms in numerical variational analysis. *Monthly Weather Review*, 98(12):875 – 883.

Satterfield, E., Hodyss, D., Kuhl, D. D., and Bishop, C. H. (2017). Investigating the use of ensemble variance to predict observation error of representation. *Monthly Weather Review*, 145(2):653–667.

Schmidt, S. F. (1966). Application of state-space methods to navigation problems. In *Advances in control systems*, volume 3, pages 293–340. Elsevier.

Schraff, C., Reich, H., Rhodin, A., Schomburg, A., Stephan, K., Perianez, A., and Potthast, R. (2016). Kilometre-scale ensemble data assimilation for the COSMO model (KENDA). *Quarterly Journal of the Royal Meteorological Society*, 142(696):1453–1472.

Schutgens, N. A., Gryspeerdt, E., Weigum, N., Tsyro, S., Goto, D., Schulz, M., and Stier, P. (2016). Will a perfect model agree with perfect observations? The impact of spatial sampling. *Atmospheric Chemistry and Physics*, 16(10):6335–6353.

Siems-Anderson, A., Lee, J. A., Brown, B., Wiener, G., and Linden, S. (2020). Impacts of assimilating observations from connected vehicles into a numerical

weather prediction model. *Transportation Research Interdisciplinary Perspectives*, 8:100253.

Siems-Anderson, A. R., Walker, C. L., Wiener, G., Mahoney III, W. P., and Haupt, S. E. (2019). An adaptive big data weather system for surface transportation. *Transportation research interdisciplinary perspectives*, 3:100071.

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

Simonin, D., Waller, J. A., Ballard, S. P., Dance, S. L., and 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.

Snelder, M. and Calvert, S. (2016). Quantifying the impact of adverse weather conditions on road network performance. *European Journal of Transport and Infrastructure Research*, 16(1).

Steeneveld, G.-J., Koopmans, S., Heusinkveld, B., Van Hove, L., and Holtslag, A. (2011). Quantifying urban heat island effects and human comfort for cities of variable size and urban morphology in the Netherlands. *Journal of Geophysical Research: Atmospheres*, 116(D20).

Stewart, L. M. (2010). *Correlated observation errors in data assimilation*. PhD thesis, University of Reading.

Stewart, L. M., Dance, S., and Nichols, N. (2008). Correlated observation errors in data assimilation. *International journal for numerical methods in fluids*, 56(8):1521–1527.

Stewart, L. M., Dance, S. L., and 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.

Stewart, L. M., Dance, S. L., Nichols, N. K., Eyre, J. R., and Cameron, J. (2014). Estimating interchannel observation-error correlations for IASI radiance data in the Met Office system. *Quarterly Journal of the Royal Meteorological Society*, 140(681):1236–1244.

Stull, R. B. (1988). *An introduction to boundary layer meteorology*. Springer Science & Business Media.

Sun, J., Xue, M., Wilson, J. W., Zawadzki, I., Ballard, S. P., Onvlee-Hooimeyer, J., Joe, P., Barker, D. M., Li, P.-W., Golding, B., et al. (2014). Use of NWP for nowcasting convective precipitation: recent progress and challenges. *Bulletin of the American Meteorological Society*, 95(3):409–426.

Tabeart, J. M., Dance, S. L., Lawless, A. S., Migliorini, S., Nichols, N. K., Smith, F., and Waller, J. A. (2020a). The impact of using reconditioned correlated observation-error covariance matrices in the Met Office 1D-Var system. *Quarterly Journal of the Royal Meteorological Society*, 146(728):1372–1390.

Tabeart, J. M., Dance, S. L., Lawless, A. S., Nichols, N. K., and Waller, J. A. (2020b). Improving the condition number of estimated covariance matrices. *Tellus A: Dynamic Meteorology and Oceanography*, 72(1):1–19.

Talagrand, O. (1997). Assimilation of observations, an introduction. *Journal of the Meteorological Society of Japan. Ser. II*, 75(1B):191–209.

Talagrand, O. (2010). Variational assimilation. In Lahoz, W., Menard, R., and Khattatov, B., editors, *Data assimilation: making sense of observations*, pages 41–67. Springer.

Tchir, J. (2016). Why the car thermometer isn't affected by engine heat or wind. *The Globe and Mail*. `https://www.theglobeandmail.com/globe-drive/culture/commuting/why-the-car-thermometer-isnt-affected-by-engine-heat-or-wind/article28765614/`. Accessed: 16-06-2019.

Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M., and Whitaker, J. S. (2003). Ensemble square root filters. *Monthly Weather Review*, 131(7):1485–1490.

Todling, R. and Cohn, S. E. (1994). Suboptimal schemes for atmospheric data assimilation based on the Kalman filter. *Monthly Weather Review*, 122(11):2530–2557.

UK Government (2015). The highway code, road safety and vehicle rules. `https://www.gov.uk/speed-limits`.

Verheul, E., Hicks, C., and Garcia, F. D. (2019). IFAL: Issue First Activate Later certificates for V2X. In *2019 IEEE European Symposium on Security and Privacy (EuroS P)*, pages 279–293.

Vetra-Carvalho, S., Van Leeuwen, P. J., Nerger, L., Barth, A., Altaf, M. U., Brasseur, P., Kirchgessner, P., and 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.

Waller, J., Dance, S., and Nichols, N. (2017). On diagnosing observation-error statistics with local ensemble data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 143(708):2677–2686.

Waller, J. A. (2019). Personal communication.

Waller, J. A. (2020). Editorial: The use of unconventional observations in numerical weather prediction. *Meteorological Applications*, 27(5):e1948.

Waller, J. A., Ballard, S. P., Dance, S. L., Kelly, G., Nichols, N. K., and Simonin, D. (2016a). Diagnosing horizontal and inter-channel observation error correlations for SEVIRI observations using observation-minus-background and observation-minus-analysis statistics. *Remote Sensing*, 8(7):581.

Waller, J. A., Bauernschubert, E., Dance, S. L., Nichols, N. K., Potthast, R., and 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.

Waller, J. A., Dance, S. L., Lawless, A. S., and Nichols, N. K. (2014a). Estimating correlated observation error statistics using an ensemble transform Kalman filter. *Tellus A: Dynamic Meteorology and Oceanography*, 66(1):23294.

Waller, J. A., Dance, S. L., Lawless, A. S., Nichols, N. K., and Eyre, J. (2014b). Representativity error for temperature and humidity using the met office high-resolution model. *Quarterly Journal of the Royal Meteorological Society*, 140(681):1189–1197.

Waller, J. A., Dance, S. L., and Lean, H. W. (2021). Evaluating errors due to unresolved scales in convection perimitting numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*.

Waller, J. A., Dance, S. L., and Nichols, N. K. (2016b). Theoretical insight into diagnosing observation error correlations using observation-minus-background and observation-minus-analysis statistics. *Quarterly Journal of the Royal Meteorological Society*, 142(694):418–431.

Waller, J. A., Simonin, D., Dance, S. L., Nichols, N. K., and Ballard, S. P. (2016c). 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.

Wang, X. and Bishop, C. H. (2003). A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *Journal of Atmospheric Sciences*, 60(9):1140–1158.

Wang, X., Bishop, C. H., and Julier, S. J. (2004). Which is better, an ensemble of positive–negative pairs or a centered spherical simplex ensemble? *Monthly Weather Review*, 132(7):1590–1605.

Weston, P. (2011). Progress towards the implementation of correlated observation errors in 4D-Var. *Met Office Forecasting Research Technical Report*, 560.

Weston, P., Bell, W., and Eyre, J. (2014). Accounting for correlated error in the assimilation of high-resolution sounder data. *Quarterly Journal of the Royal Meteorological Society*, 140(685):2420–2429.

Whitaker, J. S. and Hamill, T. M. (2012). Evaluating methods to account for system errors in ensemble data assimilation. *Monthly Weather Review*, 140(9):3078–3089.

Widrow, B., Kollar, I., and Liu, M.-C. (1996). Statistical theory of quantization. *IEEE Transactions on instrumentation and measurement*, 45(2):353–361.

Wiggins, A. and Crowston, K. (2011). From conservation to crowdsourcing: a typology of citizen science. In *2011 44th Hawaii international conference on system sciences*, pages 1–10. IEEE.

Wolters, D. and Brandsma, T. (2012). Estimating the urban heat island in residential areas in the Netherlands using observations by weather amateurs. *Journal of Applied Meteorology and Climatology*, 51(4):711–721.

World Meteorological Organisation (2021). Urban cross-cutting focus. `https://public.wmo.int/en/our-mandate/focus-areas/urban-development-megacities`. Accessed: 12-04-2021.

World Meteorological Organization (2008). *Guide to meteorological instruments and methods of observation*. Secretariat of the World Meteorological Organization, Geneva, Switzerland, 7 edition.

World Meteorological Organization (2021). OSCAR - Observing Systems Capability Analysis and Review tool. `https://community.wmo.int/oscar`. Accessed: 01/06/2021.

Zahumenskỳ, I. (2004). Guidelines on quality control procedures for data from automatic weather stations. *World Meteorological Organization, Switzerland*.

Zeng, Y., Janjić, T., de Lozar, A., Welzbacher, C. A., Blahak, U., and Seifert, A. (2021). Assimilating radar radial wind and reflectivity data in an idealized setup of the COSMO-KENDA system. *Atmospheric Research*, 249:105282.

Zhang, F., Snyder, C., and Sun, J. (2004). Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Monthly Weather Review*, 132(5):1238–1253.

Zheng, X. (2009). An adaptive estimation of forecast error covariance parameters for Kalman filtering data assimilation. *Advances in Atmospheric Sciences*, 26(1):154–160.

Zhu, Y., Derber, J., Collard, A., Dee, D., Treadon, R., Gayno, G., and Jung, J. A. (2014). Enhanced radiance bias correction in the National Centers for Environ-

mental Prediction's gridpoint statistical interpolation data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 140(682):1479–1492.