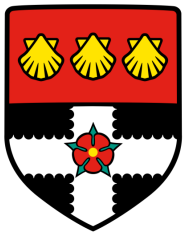

University of Reading

Department of Meteorology

Convective-scale hybrid data assimilation over the Maritime Continent

Joshua Chun Kwang Lee

Supervisors: Ross Noel Bannister, Javier Amezcua

A thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy

PhD Atmosphere, Oceans and Climate

August 3, 2024

Declaration

I, Joshua Chun Kwang Lee, of the Department of Meteorology, University of Reading, confirm that this is my own work and figures, tables, equations, code snippets, artworks, and illustrations in this report are original and have not been taken from any other person's work, except where the works of others have been explicitly acknowledged, quoted, and referenced.

Joshua Chun Kwang Lee
August 3, 2024

Abstract

Data assimilation is an important component of numerical weather prediction (NWP); it is used to estimate realistic initial conditions. In recent years, hybrid data assimilation, which combines the traditional climatological representation of background error covariances with an ensemble representation, has gained significant traction. This thesis develops and applies hybrid ensemble-variational data assimilation to a convective-scale tropical simplified model (ABC-DA) and an NWP system over the western Maritime Continent (SINGV-DA) to explore its benefits and how it can be better designed for the tropics.

Firstly, the hybrid ensemble-variational data assimilation approach (hybrid-En3DVar) was implemented in ABC-DA. Generally, hybrid-En3DVar outperformed the traditional 3DVar data assimilation approaches. The improvements were sensitive to ensemble size, and were less pronounced when the ensemble was very small. The results also highlighted how the sub-optimal background error covariance model in 3DVar, which uses geostrophic balance as a balance constraint, led to erroneous analysis increments in meridional wind and negatively impacted the forecasts.

Secondly, the hybrid ensemble-variational data assimilation approach was implemented in SINGV-DA. Without proper tuning, the initial hybrid-En3DVar setup had a relatively neutral impact compared to 3DVar. However, tuning the weightings for hybrid-En3DVar and time-shifting of ensemble perturbations were vital for improving precipitation forecasts and forecast fits to radiosonde humidity and wind compared to 3DVar. The results also highlighted how the autocovariance structures varied between variables, including the presence of robust cross-correlation structures between the moisture and temperature-related variables in the ensemble-derived background error matrix, which could have physical significance over the Maritime Continent.

Thirdly, the localisation aspect of En3DVar was modified to allow for variable-dependent localisation (prescribing different localisation length-scales for different vari-

ables) and selective multivariate localisation (knocking-out by localisation certain multivariate error covariances) in ABC-DA. This was to address two limitations of traditional ensemble-variational data assimilation approaches which were pertinent over the tropics. Using selective multivariate localisation was beneficial, particularly when covariances associated with hydrostatic balance were retained and when zonal wind errors were decoupled from the mass errors in the cross-covariances. The results also showed that variable-dependent localisation could be beneficial if the localisation length-scales were well-tuned for each variable. Both these enhancements within a pure EnVar framework reduced the forecast root-mean-square errors by about 3-4% for zonal wind and mass variables. These benefits may also apply to hybrid-En3DVar.

Acknowledgements

First and foremost, I would like to thank my PhD supervisors — Ross Bannister and Javier Amezcua — for their patient and dedicated support over the past three years. Our fruitful discussions during bi-weekly meetings on my Wednesday nights (cutting across three different time zones from -6 GMT to +8 GMT!) have been a driving force for me to continue this research remotely, and I would not have this thesis today otherwise. These three years really flew by and I have learnt so much from you both. A big thanks to my Monitoring Committee members Alison Fowler and Chris Holloway for their useful suggestions, identifying strengths and weaknesses of the research, and keeping me on track. Finally, I am grateful for the community in the Department of Meteorology in Reading (including DARC), for providing a nurturing environment during my BSc and MSc years, and now albeit remotely — my PhD years.

I would also like to thank the National Environment Agency (Singapore) for sponsoring my studies. This research topic was inspired by discussions with colleagues Hans Huang, Erland Källén, and Dale Barker, all of whom I credit for revealing to me the world of data assimilation which I have stumbled into. Thanks to Anurag Dipankar and Hugh Zhang who have allowed me flexibility to work on my PhD while juggling work commitments and responsibilities. Shoutout to colleagues from across the Unified Model Partnership (especially, but non-exhaustively) — Marco Milan, Douglas Boyd, Adam Maycock, Andrew Lorenc from the Met Office, Monika Krysta, Fiona Smith, Peter Steinle, Chun-Hsu Su from BoM, and Adam Clayton (previously Met Office) from KIAPS, who have through in-depth or brief conversations kept me thinking about how to best conduct data assimilation for the Maritime Continent.

Last but not least, I would like to thank God for guiding me through the ups and downs of this journey. To my wife — Grace — and my family, you all have been a beacon of light, constantly reminding me to never give up and press on through countless weekday nights and weekends. Thank you for your understanding and sacrifice whenever I have to prioritise PhD work. I could not have done this without you all.

Authorship of papers

This thesis is structured around three papers. They are unmodified from the published manuscripts, other than being re-formatted in accordance with the thesis chapters and with minor typographical adjustments to maintain consistency throughout the thesis. All appendices (if any) for each paper are included as Supporting Information in the corresponding chapter. The estimated percentage contribution of the candidate (JCKL) is provided below.

Lee, J. C. K., J. Amezcua, and R. N. Bannister, 2022: Hybrid ensemble-variational data assimilation in ABC-DA within a tropical framework. *Geoscientific Model Development*, **15**, 6197-6219, https://doi.org/10.5194/gmd-15-6197-2022.

Estimated contribution: 85%. JLCK developed the methodology, performed all analysis and wrote the manuscript with input and suggestions from JA and RNB. All authors commented on the manuscript and discussed the results at all stages. Two anonymous reviewers provided comments on one earlier version of the manuscript.

Lee, J. C. K., and D. M. Barker, 2023: Development of a hybrid ensemble–variational data assimilation system over the western Maritime Continent. *Weather and Forecasting*, **38(3)**, 425-444, https://doi.org/10.1175/WAF-D-22-0113.1.

Estimated contribution: 95%. JLCK developed the methodology, performed all analysis and wrote the manuscript. DMB commented on the manuscript and discussed the results at later stages of the manuscript. Three anonymous reviewers provided comments on two earlier versions of the manuscript.

Lee, J. C. K., J. Amezcua, and R. N. Bannister, 2024: Variable-dependent and selective multivariate localisation for ensemble-variational data assimilation in the tropics. *Monthly Weather Review*, **152(4)**, 1097-1118, https://doi.org/10.1175/MWR-D-23-0201.1.

Estimated contribution: 85%. JLCK developed the methodology, performed all analysis and wrote the manuscript with input and suggestions from JA and RNB. All authors commented on the manuscript and discussed the results at all stages. Three anonymous reviewers provided comments on two earlier versions of the manuscript.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis is about how variational and ensemble approaches to data assimilation can be used together for estimating the initial conditions for numerical weather prediction, focusing on the convective-scale application over the Maritime Continent[1]. This chapter first introduces the basic concepts of data assimilation to enable the work in this thesis to be understood, followed by the motivations behind this research scope and finally the aims of this thesis.

1.1 What is data assimilation?

Numerical weather prediction (NWP) systems can provide guidance on the weather conditions as early as ten days in advance (Zhang et al., 2019) for stakeholders such as businesses and governments. Accurate NWP forecasts inform decision making and risk assessments, which often translate to lives saved, impact mitigation or economic loss avoidance (Bauer et al., 2015). NWP is primarily an initial value problem; given the current weather conditions, a forecast model is used to predict the future weather conditions. Therefore to produce accurate NWP forecasts, an accurate short-range forecast model and a well-chosen starting point (initial conditions) are essential. The process to produce the initial conditions, e.g., current atmospheric state, is known as data assimilation.

The main idea of data assimilation is to combine observational information with prior information to produce a best estimate of the state (Lorenc, 1986). Observational information can be derived from a myraid of sensors, both active and passive from remote sensing instruments, as well as in-situ measurements. However, the state is

[1]The term Maritime Continent is a nickname for the Indo-Pacific archipelago, encompassing many islands, peninsulas, and the surrounding seas of Southeast Asia.

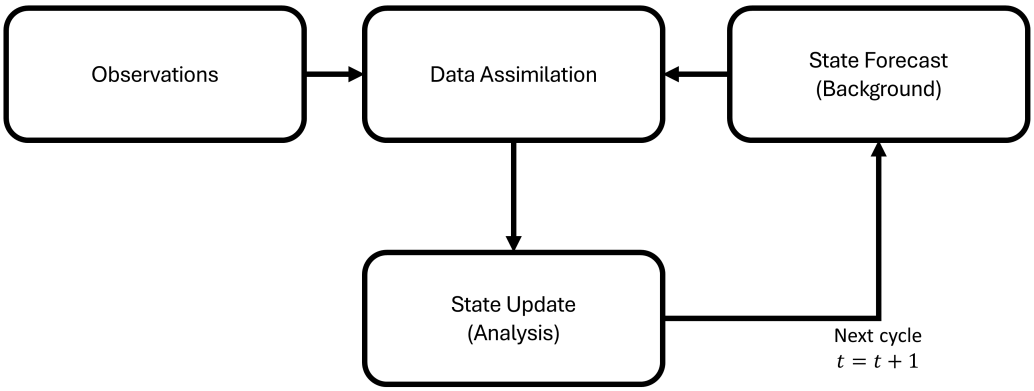

Figure 1.1: Schematic diagram of the data assimilation workflow at each cycle at time t.

often not fully observed, as observations of all variables are not always available at a given time. The prior information, known as the background, is derived from a previous short-range forecast, which contains information of the full state at the time of assimilation. The best estimate of the state is known as the analysis — this is the starting point for NWP forecasts and is retrieved by combining both observational and prior information in an optimal manner. Figure 1.1 shows the typical operational data assimilation framework, where observations are frequently incorporated into the NWP system via data assimilation. This procedure is repeated or cycled to ensure that the analysis is updated with the latest observations so that the most recent NWP forecast can be more accurate.

To optimally combine observational and background information, knowledge about their error statistics (e.g., mean squared error over a population) or uncertainties must be available. This is critical because the background error (also known as forecast error) statistics control how the observational information is spread spatially and between variables (Bannister, 2008*a*), and how the observational and background information are weighted. If the observation errors are much larger than the background errors, the analysis is weighted more towards the background. Conversely, if the background errors are much larger than the observation errors, the analysis is fit closer to the observations (i.e., 'pulled' towards observations at their locations).

Typically, observation errors are a combination of instrument and representativity errors. Instrument errors are associated with the precision of the instrument, which

include both random and systematic errors (biases). Representativity errors are errors arising when observations resolve spatial scales that the forecast model cannot (Daley, 1993), or when interpolating the background to the observed locations or transforming to the observed quantity (Cohn, 1997) so that the model's equivalent can be compared with the observed value.

For forecast errors, it is not trivial how they may be estimated, since the 'true' state cannot be known at every time instance and therefore the forecast errors cannot be explicitly computed by taking the difference between the state forecast and the 'true' state. The forecast errors must therefore be estimated and modelled (Bannister, 2008*a*). A number of approaches have been proposed to find proxies for the forecast errors, including using differences between NWP forecasts (forecast differences; Parrish and Derber, 1992), differences between observations and forecasts (Hollingsworth and Lönnberg, 1986), or differences between ensemble (running multiple iterations) members (Evensen, 1994), although each with its own limitations.

To further simplify the data assimilation problem, the observation and forecast errors are also often assumed to be Gaussian and mutually uncorrelated, although this may not always hold true and require more complex approaches (Bocquet et al., 2010, Vetra-Carvalho et al., 2018, Poterjoy, 2022). The Gaussian assumption implies that their probability density functions can be represented by two moments (mean and standard deviation). The uncorrelated assumption avoids the need to account for error relationships between the observation errors and forecast errors, which is justified because their sources of errors are supposed to be entirely independent (Bouttier and Courtier, 1999).

As part of having a well-chosen starting point, an appropriate dynamically balanced and smooth analysis during initialisation is important for an accurate NWP forecast, as this prevents shocks to the cycling NWP system due to discontinuities or quantities which are severely imbalanced when the forecast is started. For example, introducing a large pressure spike at an observed location through data assimilation without balancing with other fields will trigger spurious waves, very much like the ripples from suddenly dropping a stone into a pond. For balancing the atmospheric analysis, knowledge of the behaviour of the atmosphere such as adherence to governing equations (i.e., conservation of mass, momentum and energy) can be helpful (Bannister, 2008*b*). For example in the mid-latitudes, geostrophic theory suggests that the pressure and wind fields should be corrected in a manner that preserves geostrophic balance on the large

scales. By careful design of the data assimilation approach, such as how the forecast error correlates between variables (i.e., multivariate forecast error relationships), one can constrain the corrections to the background (known as the analysis increments) to preserve balance (e.g., ensuring a geostrophically balanced wind correction together with a pressure correction). This naturally leads to a more balanced analysis, eliminating or reducing unrealistic shocks in the forecast.

The basic ideas of data assimilation and these concepts of Gaussian, mutually uncorrelated errors, and dynamical balance are further expounded on in the next section. They are also referenced during the explanation of the motivations behind the thesis topic and how the topic relates to previous literature (Sections 1.3 and 1.4).

1.2 Data assimilation approaches

Data assimilation can be introduced as a Bayesian estimation problem. Most, if not all data assimilation approaches are underpinned by Bayes' theorem (Lewis et al., 2006), given by:

$$P(\mathbf{x}|\mathbf{y}) = \frac{P(\mathbf{y}|\mathbf{x})P(\mathbf{x})}{P(\mathbf{y})}, \tag{1.1}$$

where P is the probability density function dependent on $\mathbf{x}$, $\mathbf{y}$ or their conditionals. $\mathbf{x}$ denotes the state to be determined (e.g., atmospheric state vector) and $\mathbf{y}$ denotes the observations of the system (e.g., radiosonde observations at each pressure level). Bayes' theorem then states that the posterior probability density of the state conditioned on the observations $P(\mathbf{x}|\mathbf{y})$ can be computed using the prior probability density of the model state $P(\mathbf{x})$ and the likelihood of the observations conditioned on the model state $P(\mathbf{y}|\mathbf{x})$, normalised by the marginal probability density of the observations $P(\mathbf{y})$. This marginal probability density is independent of the $\mathbf{x}$, so it can be viewed as a normalising constant. The prior (also known as the background; Section 1.1) represents information about the possible model states (e.g., a short-range forecast, a first guess of the state). The likelihood represents information about possible observations of the model state. Finally, the posterior (also known as the analysis; Section 1.1) can be interpreted as reinforcing a common ground between what is observed and what the first guess is, so it summarises the two sources of information to get an updated version of the probability density of the model state.

Traditionally, data assimilation approaches can be broadly classified into three cat-

egories (Lorenc, 1986): variational methods (Section 1.2.1), ensemble Kalman-based methods (Section 1.2.2; for atmospheric/oceanic data assimilation), and non-linear methods. More recently, a new category of data assimilation approaches leveraging neural networks to directly map sparse observations into an analysis have emerged (Chen et al., 2023, Vaughan et al., 2024). There also exist hybrid approaches where two methods are combined to leverage the strengths of individual methods while addressing their limitations, but all can still be classified into one of the four categories. For example, the use of an ensemble to estimate the forecast error statistics for variational methods is called the hybrid ensemble-variational method (Section 1.2.3); this method is classified under variational methods since the underlying variational principle (see Section 1.2.1) remains the same. Conversely, samples of forecast differences (Parrish and Derber, 1992) which are proxies to compute forecast error statistics (as mentioned in Section 1.1) for variational methods can also be added as ensemble perturbations to a forecast ensemble for Kalman-based methods (e.g., in Kretschmer et al., 2015, Kotsuki and Bishop, 2022); this method is classified under ensemble Kalman-based methods (see Section 1.2.2). As an introduction, the basic concepts of variational and ensemble Kalman-based methods are discussed along with the concept of a specific hybrid approach which is used in this work — hybrid ensemble-variational data assimilation.

1.2.1 Variational methods

In variational methods, the aim is to find the state that maximises $P(\mathbf{x}|\mathbf{y})$. This is also known as the maximum *a posteriori* probability. $P(\mathbf{x}|\mathbf{y})$ is proportional to the product of the $P(\mathbf{y}|\mathbf{x})$ and $P(\mathbf{x})$, and the normalising constant $P(\mathbf{y})$ is ignored.

Under Gaussian error assumptions, the data assimilation problem can be simplified (as highlighted in Section 1.1). From the definition of a Gaussian probability density function, the probability density function of the prior satisfies:

$$P(\mathbf{x}) \propto \exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{x}^b)^\top \mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}^b)\right\}, \tag{1.2}$$

where $\mathbf{x}^b$ is the background state and $\mathbf{B}$ is the background error covariance matrix. The background is assumed to be the mean of this distribution. The background error covariance matrix contains the statistics of the background errors; the cross-covariances between variables and autocovariances for each variable. It represents how the forecast errors are correlated in space (i.e., on the atmospheric state grid) and between variables. As such, it is a very large matrix; for an operational NWP system, the state size is the product of the number of gridpoints and variables, $\mathcal{O}(10^9)$, and the number of

elements of $\mathbf{B}$ is the square of the state size. As seen later in the thesis, $\mathbf{B}$ is sometimes static (constant) but can also be made to vary in time, as would the background error statistics in reality.

Similarly under Gaussian error assumptions, the probability density function of the likelihood conditioned on the $\mathbf{x}$ satisfies:

$$P(\mathbf{y}|\mathbf{x}) \propto \exp\left\{-\frac{1}{2}[\mathbf{y}-\mathcal{H}(\mathbf{x})]^\top \mathbf{R}^{-1}[\mathbf{y}-\mathcal{H}(\mathbf{x})]\right\}, \tag{1.3}$$

where $\mathbf{R}$ is the observation error covariance matrix and $\mathcal{H}$ is the observation operator which maps from state space to observation space so that a comparison can be made between the observation and the model's equivalent. For now, $P(\mathbf{y}|\mathbf{x})$ is regarded as a function of $\mathbf{y}$ for a fixed $\mathbf{x}$. It describes the probability density of possible observations and the mean of the distribution is assumed as $\mathcal{H}(\mathbf{x})$. Note that $\mathcal{H}$ can be non-linear (e.g., involving roots, powers, trigonometric functions), but in variational methods the linearisation of $\mathcal{H}$ about a reference state is often performed (see Eq. (1.6), and description within Chapters 2 and 3). The observation error covariance matrix represents how the observation errors are correlated and their error variances along the diagonal. In many NWP systems, the observation errors are assumed to be uncorrelated, so $\mathbf{R}$ is diagonal (i.e., only elements of the matrix on the diagonal are non-zero).

Under an additional assumption that the observation and background errors are mutually uncorrelated (as highlighted in Section 1.1), the posterior probability density function (Eq. (1.1)) then satisfies:

$$P(\mathbf{x}|\mathbf{y}) \propto \exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{x}^b)^\top \mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}^b) - \frac{1}{2}[\mathbf{y}-\mathcal{H}(\mathbf{x})]^\top \mathbf{R}^{-1}[\mathbf{y}-\mathcal{H}(\mathbf{x})]\right\}. \tag{1.4}$$

For the variational procedure, $\mathbf{x}$ now becomes the variable and $\mathbf{y}$ becomes fixed, containing the specific observations. In variational methods, this approach is further framed as a minimisation problem, where to maximise $P(\mathbf{x}|\mathbf{y})$, the minus exponent is minimised as a cost function:

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\mathbf{x}^b)^\top \mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}^b) + \frac{1}{2}[\mathbf{y}-\mathcal{H}(\mathbf{x})]^\top \mathbf{R}^{-1}[\mathbf{y}-\mathcal{H}(\mathbf{x})], \tag{1.5}$$

where J is the cost function (also referred as the penalty or objective function). The analysis is therefore the state that minimises Eq. (1.5) and maximises Eq. (1.4). This is the 3DVar (three-dimensional variational; all observations valid at analysis time) cost

function.

To minimise the cost function, most variational systems employ an incremental approach (Courtier et al., 1994), which involves iteratively linearising $\mathcal{H}$ around a reference state ($\mathbf{x}^r$) and framing the problem in terms of increments to $\mathbf{x}^r$ in a series of outer loops. For 3DVar, a single outer loop is typically used (and so $\mathbf{x}^r = \mathbf{x}^b$). To illustrate, the incremental form of the 3DVar cost function is:

$$J(\delta\mathbf{x}) = \frac{1}{2}\delta\mathbf{x}^\top\mathbf{B}^{-1}\delta\mathbf{x} + \frac{1}{2}(\mathbf{H}\delta\mathbf{x} - \mathbf{d})^\top\mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x} - \mathbf{d}), \tag{1.6}$$

where $\mathbf{x} = \mathbf{x}^r + \delta\mathbf{x}$, $\mathbf{x}^r$ is a reference state, $\delta\mathbf{x}$ is the state increment (equal to the analysis increment since there is only one outer loop), and $\mathbf{H}$ is the observation operator linearised around $\mathbf{x}^r$. We define the innovation:

$$\mathbf{d} = \mathbf{y} - \mathcal{H}(\mathbf{x}^b). \tag{1.7}$$

After finding the optimal $\delta\mathbf{x}$, the analysis $\mathbf{x}^a$ can be computed using $\mathbf{x}^a = \mathbf{x}^b + \delta\mathbf{x}$. Thus far for simplicity, the time notation has been omitted (all observations are assumed valid at analysis time). However, Chapter 2 presents a derivation including the time notation for First-Guess-at-Appropriate-Time (3DVar-FGAT) when observations are not all valid at analysis time, but at a later time within some assimilation window (e.g., 1 hour).

For the rest of this section, the incremental formulation is used to discuss additional concepts. As mentioned in Section 1.1, $\mathbf{B}$ is vital for conducting accurate and balanced data assimilation, as it controls the weighting between the observational and background information. However, it is impossible to explicitly specify $\mathbf{B}$, let alone compute $\mathbf{B}^{-1}$ due to its large size. Therefore, it must be modelled or estimated. There are two main outcomes of modelling $\mathbf{B}$: (i) it controls the autocorrelations (spatial) and cross-correlations (multivariate), and (ii) it simplifies the minimisation of the cost function for some approaches where $\mathbf{B}^{-1}$ need not be computed. Bannister (2008*a*,*b*) provide an overview of the complexities of estimating and modelling $\mathbf{B}$.

To illustrate (i), the gradient of the cost function ∇J can be computed by differentiating Eq. (1.6) with respect to $\delta\mathbf{x}$, giving:

$$\nabla J(\delta\mathbf{x}) = \mathbf{B}^{-1}\delta\mathbf{x} + \mathbf{H}^\top\mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x} - \mathbf{d}). \tag{1.8}$$

At the minimum, $\nabla J = 0$, so $\delta\mathbf{x}$ can be factorised and the equation can be re-shuffled, giving:

$$\delta\mathbf{x} = (\mathbf{B}^{-1} + \mathbf{H}^\top\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^\top\mathbf{R}^{-1}\mathbf{d}, \tag{1.9}$$

and using the Sherman-Morrison-Woodbury formula:

$$\delta\mathbf{x} = \mathbf{B}\mathbf{H}^\top(\mathbf{H}\mathbf{B}\mathbf{H}^\top + \mathbf{R})^{-1}\mathbf{d}. \tag{1.10}$$

The impact of $\mathbf{B}$ can be seen when considering only one observation at location index i. The innovation $\mathbf{d}$ becomes a scalar d (recall Eq. (1.7) compares the observations with model's equivalent observations), $\mathbf{H}$ has only one row, and $\mathbf{R}$ has only one element associated with the observation error variance σ_o^2. This yields:

$$\delta\mathbf{x}_i = \mathbf{B}_i \left(\frac{d}{\sigma_b^2 + \sigma_o^2} \right), \tag{1.11}$$

where σ_b^2 is the background error variance at the location index i and $\mathbf{B}_i$ is the i^{th} column of $\mathbf{B}$ which contain all the error cross-covariances with respect to location index i. Note how the analysis increment is then proportional to the error cross-covariances from $\mathbf{B}$. Also, note how the weighting towards the observation or background (Section 1.1; i.e., how much it is 'pulled' to the observation) is dependent on the background error variance and observation error variance. Since $\lim_{\sigma_o \to \infty} \frac{d}{\sigma_b^2 + \sigma_o^2} = 0$, the analysis increment is very small if σ_o is very large. Equations (1.9) to (1.11) are developed in order to illustrate this property of $\mathbf{B}$. In practice, the cost function uses a gradient descent algorithm to find $\delta\mathbf{x}$.

To illustrate (ii), a control variable transform $\mathbf{U}$ is first introduced. Essentially, the aim is to solve the minimisation in a different variable space (referred to as control variable space, represented by a control vector, $\delta\boldsymbol{\chi}$), before transforming back to model state space using $\mathbf{U}$. In $\delta\boldsymbol{\chi}$ space after applying the control variable transform, the control variables are assumed to have uncorrelated errors with unit variance. From Eq. (1.6), substituting $\delta\mathbf{x} = \mathbf{U}\delta\boldsymbol{\chi}$ yields:

$$J(\delta\boldsymbol{\chi}) = \frac{1}{2}\delta\boldsymbol{\chi}^\top\delta\boldsymbol{\chi} + \frac{1}{2}(\mathbf{H}\mathbf{U}\delta\boldsymbol{\chi} - \mathbf{d})^\top\mathbf{R}^{-1}(\mathbf{H}\mathbf{U}\delta\boldsymbol{\chi} - \mathbf{d}). \tag{1.12}$$

Note how $\mathbf{B}^{-1}$ is no longer explicitly required in the cost function by employing the control variable transform, and therefore $\mathbf{B} = \mathbf{U}\mathbf{U}^\top$ is implied by the variational algorithm when $\mathbf{U} = \mathbf{B}^{\frac{1}{2}}$. As stated above, the minimisation to get to Eq. (1.11) is effectively done in $\delta\boldsymbol{\chi}$ space.

The next question is: How then should the transformation to another variable space

be determined? There are an infinite number of square roots that satisfy $\mathbf{B} = \mathbf{U}\mathbf{U}^{\top}$, and there is freedom to make approximations. One key principle of modelling $\mathbf{B}$ is to preserve balance when choosing appropriate variables to be analysed, as highlighted in Section 1.1. Some studies apply statistical relationships (e.g., regression), and others use known balance operators (see Section 1.3.1 for examples). Regardless, the eventual choice of variables (often called parameters) are assumed to have uncorrelated errors.

In many data assimilation systems, $\mathbf{U}$ incorporates a parameter transform $\mathbf{U}_p$ (controls the multivariate aspects), a vertical transform $\mathbf{U}_v$ and a horizontal transform $\mathbf{U}_h$ (e.g., $\mathbf{U} = \mathbf{U}_p\mathbf{U}_v\mathbf{U}_h$). Each component can be modelled independently. Key things to note are:

- The analysis increment $\delta\mathbf{x}$ depends on $\mathbf{U}$ and a control vector $\delta\boldsymbol{\chi}$; $\delta\mathbf{x} = \mathbf{U}\delta\boldsymbol{\chi}$,
- $\delta\boldsymbol{\chi}$ depends on the choice of control variables, e.g., with hydrometeor variables, or with zonal/meridional wind (see Section 1.4.1 for examples),
- $\mathbf{U}$ depends on the control variables and whether homogeneity (using same spatial correlations throughout the domain) and isotropy (using spatial correlations only as a function of distance) is assumed, depending on the transform order of $\mathbf{U}_v$ or $\mathbf{U}_h$. If $\mathbf{U}_h$ is applied first on $\delta\boldsymbol{\chi}$, the horizontal correlations are homogeneous, but allows $\mathbf{U}_v$ (controlling vertical correlations) to vary in gridpoint space. The limitations of homogeneous and isotropic correlations are discussed in Section 1.3.1.

1.2.2 Ensemble Kalman-based methods

The Kalman filter (Kalman, 1960) is an optimal state estimation algorithm which is a specific implementation of Bayes' theorem with Gaussian error assumptions. It is a minimum variance estimator which aims to minimise the analysis error variances. At each time point, the Kalman filter performs (i) a forecast step and (ii) an update step. The update step propagates and updates the first two moments (mean and variances) of the Gaussian distribution, unlike variational methods which only update the first moment (mean). In traditional Kalman filters, the forecast step and update step are performed when the forecast model and observation operator are linear, or linearised about an estimate of the current mean (then referred to as the extended Kalman filter). This linearisation is similar to the linearisation about a reference state which is used in variational methods.

For atmospheric data assimilation, the state space includes many variables and gridpoints, so traditional Kalman filters which require an explicit computation of large error covariance matrices are impractical. Therefore, the Kalman filter is often employed together with an efficient ensemble implementation; these are termed as ensemble Kalman-based methods (Evensen, 1994). While traditional Kalman filters are tailored for linear dynamical systems, ensemble Kalman-based methods offer *a priori* error statistics from an ensemble which does not require the use of linearised forecast models. While the update step involves linearisation, the forecast step in ensemble Kalman-based methods is performed using the full non-linear dynamical model; each ensemble member is propagated forward in time to offer time-dependent error statistics for estimating $\mathbf{B}$.

Evensen (2006) provides a thorough introduction to the ensemble Kalman-based methods. Essentially, the *a priori* error statistics from an ensemble allows for the representation of $P(\mathbf{x})$ using the ensemble mean (or control member) and covariance. This comes from the background ensemble (formed by having each ensemble member propagated forward in time). To form the analysis ensemble (using the update step) representing $P(\mathbf{x}|\mathbf{y})$, it can be done either stochastically (Houtekamer and Mitchell, 1998, Burgers et al., 1998), or deterministically (Bishop et al., 2001, Anderson, 2001, Whitaker and Hamill, 2002). The stochastic approach uses 'perturbed observations' to treat the observations as random variables (i.e., generate an ensemble sample of observations), but this introduces more sampling noise. The deterministic approach does not have 'perturbed observations' — it does not require the generation of random numbers in the update and forecast step after the initialisation of the ensemble, and deals with the update step using a transform matrix (non-unique, thus there exist different flavours). The deterministic approach is therefore an implementation of a Kalman square-root filter (Bierman, 1977). Tippett et al. (2003) provides a summary of the different possible transform matrices compared using the Kalman square-root filter framework.

In this thesis, ensemble Kalman-based methods are not used so the full ensemble Kalman equations are not presented. The focus instead is on leveraging the *a priori* error statistics offered by an ensemble — which could be propagated using ensemble Kalman-based methods (not used) or other ensemble methods — in variational methods. The time-dependent error statistics (consistent with the weather conditions at a particular time) from the ensemble forecast step hybridised into the variational method provides the main advantage over the variational methods where the same $\mathbf{B}$

is used at each data assimilation cycle without being updated.

The tradeoff of using of an ensemble to estimate $\mathbf{B}$ (see mathematical representation below), however, is that there are usually far fewer ensemble members, N, than the degrees of freedom of $\mathbf{B}$ (i.e., the number of parameters that can vary independently, which in this case is the square of the state size). As such, $\mathbf{B}$ calculated in this manner from the ensemble is rank-deficient. Therefore, while using an ensemble to estimate $\mathbf{B}$ can generate inhomogeneous and anisotropic covariance structures (vis-à-vis variational methods which may not), it is affected by sampling noise. One way to address it is to introduce localisation (knocking-out) of spurious long-range correlations. This may be achieved using a Schur multiplication of a localisation matrix $\mathbf{L}$ (a correlation/cross-correlation matrix), as proposed by Houtekamer and Mitchell (2001). Localisation is an important aspect of ensemble methods, and more details of how it is relevant in this thesis are covered in Section 2.3.2. The key points are:

- The *a priori* error statistics populate a rectangular matrix $\mathbf{X}_t^f$, whose columns contain the scaled (by a factor of $\frac{1}{\sqrt{N-1}}$) differences between ensemble forecasts and the ensemble mean (also referred as ensemble perturbations),
- The ensemble-derived background error covariance matrix is $\mathbf{P}_e^f[t] = \mathbf{X}_t^f \mathbf{X}_t^{f\top}$,
- $\mathbf{B}$ is retrieved from $\mathbf{L}$ and $\mathbf{P}_e^f[t]$; $\mathbf{B} = \mathbf{L} \circ \mathbf{P}_e^f[t]$. If no localisation is applied, $\mathbf{B} = \mathbf{P}_e^f[t]$. Since $\mathbf{P}_e^f[t]$ is time-dependent, the ensemble-derived $\mathbf{B}$ is referred to as $\mathbf{B}_e$ (with the time-dependence subsumed under the ensemble subscript e; see Section 1.2.3).

In Chapters 2, 3 and 4, ensemble Kalman-based methods are not used to propagate or generate the ensemble. For Chapters 2 and 4, which are based on a simplified model of the atmosphere (Section 1.6.1), ensemble bred vectors (Balci et al., 2012) are used to update the ensemble-derived error covariances (see Section 2.3.3) in a cycling framework. Ensemble bred vectors 'breed' perturbations by considering the fastest growing error directions (or modes). It does not take into account the observation network when forming the analysis ensemble, but the method is simple to implement with a low computational cost while ensuring that the ensemble perturbations sample the space of analysis errors. For Chapter 3, which is based on a realistic model of the atmosphere (Section 1.6.2), a downscaler ensemble is used to update the ensemble-derived error covariances. A regional downscaler (also known as limited area model; LAM) ensemble uses the initial conditions provided (and updated) by a global ensemble, and runs the ensemble forecasts at a higher resolution to represent the

forecast errors (and the error statistics) at that scale (see Section 1.4 and Fig. 1.4 below for an LAM illustration and description). Error growth information is therefore propagated across cycles only via the global ensemble and may be inconsistent at smaller scales compared to ensemble Kalman-based methods, but the ensemble is cheaper to implement for real-time applications.

In respective chapters, the details on the design of localisation are described as localisation is key to exploring some of the research questions posed in the chapters. The alpha control variable approach of Lorenc (2003) is a very efficient way to build localisation into $\mathbf{B}_e$ without explicitly constructing $\mathbf{P}_e^f[t]$ and $\mathbf{L}$, through modification of the cost function (see Section 1.2.3). All chapters use the alpha control variable approach. Another approach by Buehner (2005) applies the localisation directly on the ensemble perturbations, but this is not used in this thesis. Wang et al. (2007) demonstrates the mathematical equivalence of the Lorenc (2003) and Buehner (2005) approaches. Chapters 2 and 3 use the traditional formulation of Lorenc (2003), but Chapter 4 further modifies the design of localisation and assesses its impact.

1.2.3 Hybrid ensemble-variational methods

Hybrid ensemble-variational methods encompass a wide spectrum of combinations of ensemble or variational methods due to subtleties in their derivation and usage. Bannister (2017) provides a comprehensive review of operational ensemble-variational methods. These include methods which use the ensemble without a linearised model (four-dimensional ensemble-variational, known as 4DEnVar; Liu et al., 2008), and with a linearised model (ensemble four-dimensional variational, known as En4DVar; Clayton et al., 2013) in the variational algorithm, a hybrid gain approach (Penny, 2014) as opposed to the typical hybrid covariance approach, and for each hybrid variant, the use of different ensemble setups. Bannister (2017) also discusses various ways of combining a climatological $\mathbf{B}$ (one that is calibrated using climatological perturbations), which is hereafter referred to as $\mathbf{B}_c$, with an ensemble-derived $\mathbf{B}$ (one that is prescribed using an ensemble, with or without localisation as discussed in Section 1.2.2), which is hereafter referred to as $\mathbf{B}_e$.

In this section, the hybrid covariance approach (specifically for ensemble three-dimensional variational, known as En3DVar) is described since this is used in Chapters 2 and 3. This entails a linear combination of $\mathbf{B}_c$ and $\mathbf{B}_e$, in the form following Hamill and Snyder (2000):

$$\mathbf{B}_h = \beta_c^2 \mathbf{B}_c + \beta_e^2 \mathbf{B}_e, \tag{1.13}$$

where β_c^2 and β_e^2 are (positive) scalar weights often determined empirically. These weights are often chosen to add to unity, but this is not a necessary condition. As discussed in Section 1.2.2, $\mathbf{B}_e$ offers time-dependent error statistics which may be consistent with the weather conditions at a particular time, while $\mathbf{B}_c$ offers robust climatological error statistics with smaller sampling noise. Therefore, $\mathbf{B}_h$ is expected to benefit from both advantages.

While this approach (Eq. (1.13)) computes $\mathbf{B}_h$ explicitly, it is often computationally unfeasible for many data assimilation systems where the degrees of freedom of the state is very large (too many variables, gridpoints and their cross-components). Nonetheless, it is possible to adopt an alternative approach, the alpha control variable approach (Lorenc, 2003) so as to enable hybrid ensemble-variational data assimilation. As mentioned above, this approach is used throughout this thesis. Starting from the pre-conditioned 3DVar cost function of variational methods (reproduced Eq. (1.12)) from Section 1.2.1:

$$J(\delta\boldsymbol{\chi}) = \frac{1}{2}\delta\boldsymbol{\chi}^\top\delta\boldsymbol{\chi} + \frac{1}{2}(\mathbf{HU}\delta\boldsymbol{\chi} - \mathbf{d})^\top\mathbf{R}^{-1}(\mathbf{HU}\delta\boldsymbol{\chi} - \mathbf{d}), \tag{1.14}$$

it can be extended to include additional ensemble-related components. The extension enables the use of ensemble-derived error statistics (Eq. (1.15b)), given by:

$$\begin{aligned} J(\delta\boldsymbol{\chi}, \boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2, \ldots, \boldsymbol{\alpha}^N) = \overbrace{\frac{1}{2}\delta\boldsymbol{\chi}^\top\delta\boldsymbol{\chi}}^{J_b} + \overbrace{\frac{1}{2}(\mathbf{H}\delta\mathbf{x} - \mathbf{d})^\top\mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x} - \mathbf{d})}^{J_o} \\ + \overbrace{\frac{1}{2}\sum_{k=1}^{N}\boldsymbol{\alpha}^{k\top}\mathbf{L}^{-1}\boldsymbol{\alpha}^k}^{J_e} \end{aligned} \tag{1.15a}$$

$$\text{with } \delta\mathbf{x} = \beta_c\mathbf{U}\delta\boldsymbol{\chi} + \beta_e\sum_{k=1}^{N}\mathbf{x}_t'^k \circ \boldsymbol{\alpha}^k, \tag{1.15b}$$

where J_b, J_o and J_e are the (climatological) background, observation and ensemble penalties respectively. $\boldsymbol{\alpha}^k$ and $\mathbf{x}_t'^k$ are the alpha field and scaled (by a factor of $\frac{1}{\sqrt{N-1}}$) ensemble perturbation (or error mode) respectively, for the k^{th} ensemble member out of a total of N members. $\mathbf{L}$ is a localisation matrix typically required for ensemble methods (Houtekamer and Mitchell, 2001), as discussed in the previous section. The details of how each $\mathbf{x}_t'^k$ can be computed is covered in Section 2.3.2. Minimisation of this cost function yields the $\delta\mathbf{x}$, which is a linear combination of the 3DVar analysis increment and

the ensemble perturbations weighted by associated alpha fields. As before in Section 1.2.1, the optimal analysis increment is added to the background (control member of the ensemble) to find $\mathbf{x}^a$. Details on how one might compute the gradient and the minimisation algorithm are described in Chapter 2. The key points are:

- J is minimised simultaneously with respect to $\delta\boldsymbol{\chi}$ and all $\boldsymbol{\alpha}^k$,
- The implied $\mathbf{B}_h$ from this approach is exactly $\mathbf{B}_h$ from Eq. (1.13), as shown by Wang et al. (2007),
- Minimising Eq. (1.15a) would be equivalent to replacing $\mathbf{B}$ with $\mathbf{B}_h$ in Eq. (1.6) and minimising it.

1.3 Data assimilation over the Maritime Continent

Although the underlying dynamical and thermodynamical equations represented in NWP models (e.g., Ullrich et al., 2017) are the same over various parts of the world, the modelled weather conditions can often vary regionally because of local topography, bathymetry, and land-sea contrasts. The flow regimes (e.g., Rossby numbers — the ratio of inertial forces to Coriolis forces) also vary in different regions. This has implications for the estimation of NWP forecast errors and their statistics, and therefore data assimilation in different regions. The dominant mode of diurnal variability due to different forcing (e.g., solar) in the tropics (as shown by Yang and Slingo, 2001) is different from the annular modes (seasonal) in the polar region, leading to different sources and estimates of forecast errors at various timescales. For example, the forecast errors arising from missing the peak of the diurnal cycle of convection is likely to be more severe in the tropics than in the polar region. Additionally, the deep tropics (very near the Equator) may adhere to large-scale balances like the weak temperature gradient balance (Sobel et al., 2001, Yano and Bonazzola, 2009) that are different from those in the mid-latitudes, like geostrophic balance. Logically, in the absence of a unified framework, the design of data assimilation approaches should be contextualised for each region by considering their unique characteristics so as to produce an appropriate dynamically balanced and accurate analysis.

1.3.1 Why focus on the Maritime Continent?

Numerous studies have incorporated data assimilation in regional NWP systems over many regions (Fillion et al., 2010, Gustafsson et al., 2014, Ito et al., 2016, Zhang et al., 2019, Milan et al., 2020, Bouyssel et al., 2022), including regions with limited

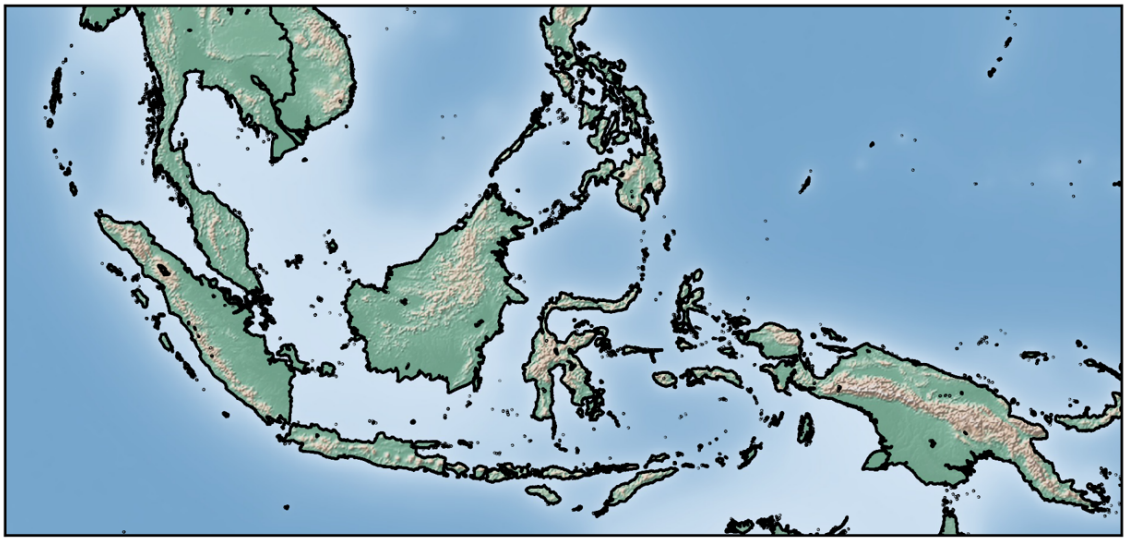

Figure 1.2: Maritime Continent domain, including the geographical variations of the terrain height as indicated by shaded relief.

observations such as the Antarctic (Barker, 2005), Western Africa (Montmerle et al., 2006), South America (Dillon et al., 2016), and the Middle East (Neyestani et al., 2021). However, very few institutes have set up data assimilation in a regional NWP system over the Maritime Continent (Fig. 1.2), so the performance of data assimilation approaches in this region remains relatively unexplored and may therefore not be optimised. In this light, there are opportunities for improvements by considering how the following three challenges which are pertinent to the region can be overcome.

Firstly, the Maritime Continent is a region that is considered data-sparse, especially over the adjacent seas surrounding the Malaysian Peninsula, Borneo, and islands of Indonesia. Figure 1.3 shows the in-situ observations assimilated in the European Centre for Medium-range Weather Forecasts (ECMWF) global NWP system for a single cycle (6-hour window). The observations (especially radiosonde and surface which sample the atmospheric boundary layer) have a sparser spatial distribution compared to over Europe and North America. Lee et al. (2021) further discussed how the wind observations have insufficient temporal resolution which under-samples the diurnal variability of weather in the region. From a data assimilation perspective, these limit the quality of the wind analysis because one would have missing information about the wind to frequently update the analysis. Information would then have to be somehow derived from observations of other variables (e.g., satellite observations of brightness temperature). However, even for other variables, there is still a lack of in-situ observations over the adjacent seas. Given these limitations, there is potential to assess different data assimilation approaches and find the one which best utilises the

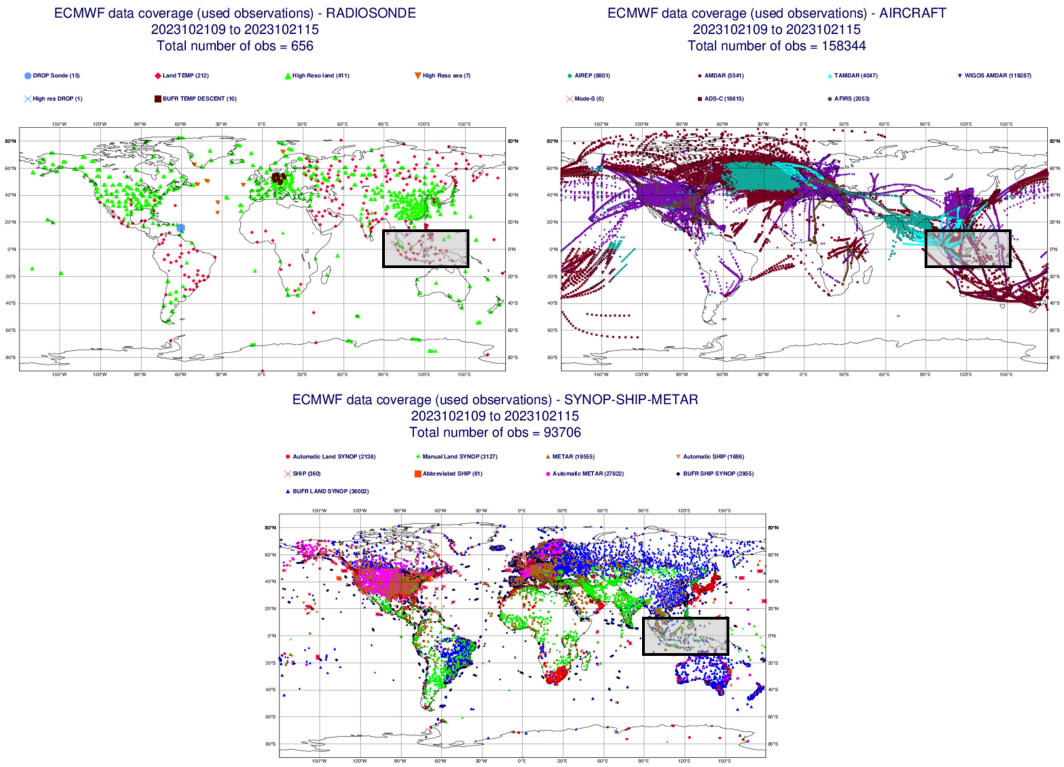

Figure 1.3: In-situ observations (radiosonde, aircraft and surface) assimilated in the ECMWF global NWP system for a single cycle (6-hour window), with the Maritime Continent region outlined in black. Image is courtesy of the ECMWF Observations Monitoring Portal.

given observation network for this region.

Secondly, it is unclear how balanced multivariate data assimilation for the Maritime Continent may be conducted, due to a lack of understanding of useful multivariate forecast error relationships for the region. Existing traditional variational data assimilation approaches (Section 1.2.1) using geostrophic balance in the control variable transform may not be useful because it does not hold in the tropics, yet ensemble Kalman-based approaches (Section 1.2.2) which estimate appropriate forecast error correlations explicitly between mass (e.g., pressure, temperature) and wind variables contain sampling noise which may still introduce imbalances during initialisation (Lorenc, 2003). Žagar et al. (2004) discussed how one might conduct balanced data assimilation for the tropics using tropical wave theory and how the different components (e.g., Kelvin waves, Rossby waves) were essential for determining the mass and wind forecast error relationships. Nonetheless, their results were only within the context of a simplified shallow water modelling framework. Chen et al. (2013) also explored the multivariate forecast error balance characteristics in the

tropics, but the analysis was based on statistical regression (computing the correlations explicitly from sample data) instead of dynamical balance constraints (using theory to determine the correlations). A better understanding of which multivariate relationships can be leveraged when performing data assimilation over the Maritime Continent may improve data assimilation for this region, but this requires further studies.

Thirdly, it is unclear how the forecast error characteristics (e.g., the spatial correlation distances and shape) may differ for each variable over the Maritime Continent, and how to account for these differences when conducting data assimilation. Existing traditional variational data assimilation approaches (Section 1.2.1) often make assumptions such as homogeneity and isotropy when modelling the forecast error statistics, which may be violated especially over the Maritime Continent with its inhomogeneous orography and surface type (e.g., inland, open sea or coastal). Lee and Huang (2022) demonstrated how in general over the Maritime Continent, homogeneity in the forecast error characteristics does not hold. Other existing ensemble Kalman-based approaches (Section 1.2.2) require localisation to reduce the sampling noise within the estimated forecast error statistics (as mentioned in Section 1.2.2), but traditional localisation does not allow localisation to depend on variable. Necker et al. (2020) found that in the mid-latitudes, different variables typically have their own unique forecast error characteristics and should be localised differently. Their results are likely also applicable over the Maritime Continent. Given these findings, there may be potential to re-design certain aspects of existing traditional approaches to improve data assimilation over the Maritime Continent. This is further explored in Chapter 4.

1.4 Convective-scale data assimilation

Convective-scale (often referred as convection-permitting) data assimilation refers to NWP data assimilation systems which are configured on a grid with a gridspacing of between 1 to 4km, where convection can be explicitly represented instead of parametrised (Hu et al., 2023). All leading existing operational global NWP data assimilation systems (e.g., Clayton et al., 2013, Bonavita et al., 2016) thus far are configured on a grid which has a coarser resolution due to computational constraints, so convective-scale data assimilation at the moment is relevant only to regional NWP data assimilation systems. Figure 1.4 shows how a regional NWP data assimilation system with a smaller gridspacing compared to global NWP systems might assimilate remotely sensed observations such as those from radar or satellites. The idea of a regional NWP model, or LAM, is to have a global model provide the starting initial

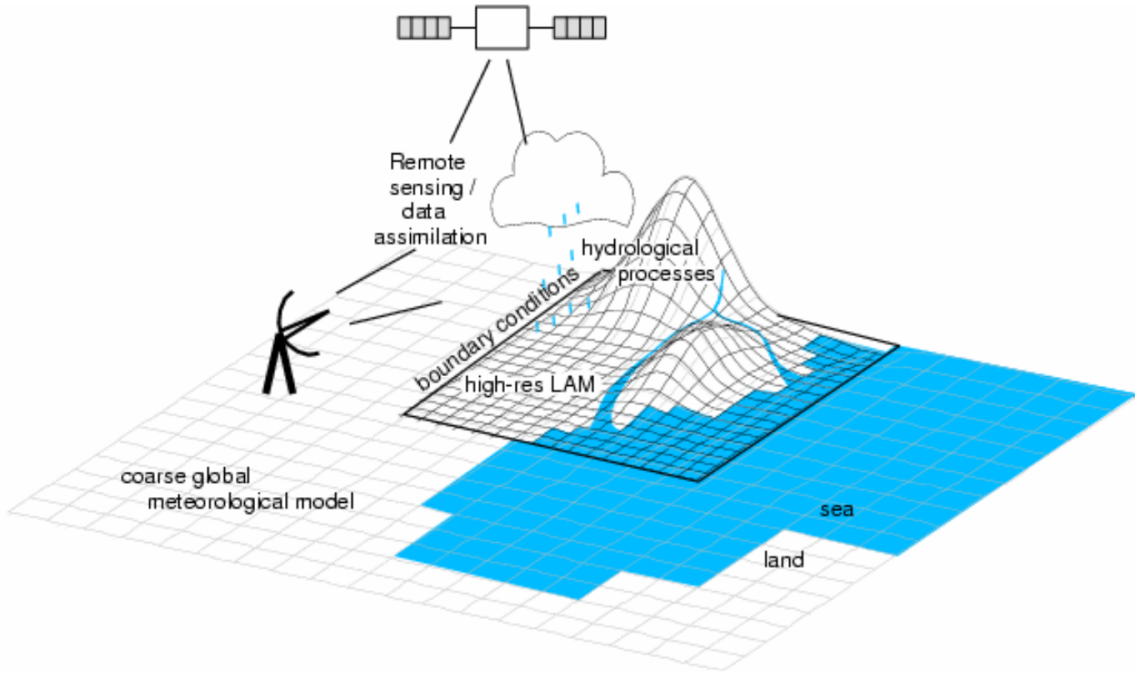

Figure 1.4: Illustration of convective-scale data assimilation of various remotely sensed observations in a high resolution regional NWP data assimilation system. Figure courtesy of Ross Bannister.

conditions that are reconfigured onto the regional NWP model grid. In a fully cycling LAM with data assimilation, only the lateral boundary conditions are updated from the global model continously, and the forecasts are produced at higher resolution (e.g., convective-scale) by the regional NWP model within its domain. Convective-scale data assimilation then involves using the convective-scale background from the regional NWP model and regional observations to produce a convective-scale analysis.

As the name also suggests, the main outcome of convective-scale data assimilation is to produce an analysis that accurately represents convective systems in the domain, without being heavily imbalanced. These convective systems are often transient (lasting an hour) and small-scale (10 to 20 km wide), so it is very challenging for regional NWP data assimilation systems to accurately represent their behaviour without additional observational information where the NWP model is lacking. Even with additional observational information, this assumes that the regional NWP models themselves are able to represent the processes, which to a certain extent holds true with a sufficiently small gridspacing, but the regional NWP models could then suffer from aliasing of the small-scale information onto the large-scales (Baxter et al., 2011) if the data assimilation method does not handle it. Therefore, to achieve the outcome desired by convective-scale data assimilation, there are two conditions to be met: (i) observations

of sufficient resolution (both temporally and spatially) must be available to capture the rapid fluctuations of the atmosphere that are related to convective systems, and (ii) the data assimilation algorithms must be able to assimilate the variety of these convective-scale observations without violating simplifications used in typical data assimilation approaches, such as uncorrelated observation errors, or aliasing issues.

1.4.1 Why focus on convective-scale data assimilation?

Heavy rainfall from convective systems can often cause severe flooding to occur, especially over the Maritime Continent (e.g., Tangang et al., 2008, Nuryanto et al., 2017). Regional NWP data assimilation systems therefore have an important role to play in providing specific guidance on the occurrence of such events, but the analysis must first be retrieved through convective-scale data assimilation.

Recent studies have attempted to better understand if the abovementioned two conditions (i) and (ii) are met to retrieve a high quality convective-scale analysis. For (i), the advances in satellite and radar technology over the past two decades have provided opportunities for the utilisation of convective-scale observations in regional NWP data assimilation systems (Seity et al., 2011, Simonin et al., 2014, Augros et al., 2016, Müller et al., 2017, Honda et al., 2018, Heng et al., 2020, Hawkness-Smith and Simonin, 2021, Ikuta et al., 2021, Wang et al., 2022). Additionally, novel observations such as unmanned aerial systems (Leuenberger et al., 2020) or crowdsourced observations (see Hintz et al., 2019 and references within) could potentially augment the convective-scale observation network. However, for (ii), there are still differing views on the fundamental principles to assimilate these observations in convective-scale data assimilation (Gustafsson et al., 2018). More exploratory work is required to inform the community on the research priorities for the following three areas, but this thesis focuses only on the third area below (see Section 1.5).

Firstly, the relative importance of which variables to observe for convective-scale data assimilation needs to be explored. One may work backwards from the weather phenomena of interest — convection and thunderstorms — to speculate the required variables. Logically, observations of environmental conditions favourable for convection (the precursors), or convection itself appear to be most relevant. However, there is still a limited amount of literature on the underlying fundamental processes and the predictability at these scales. WMO (2022) provides a guidance on the observation requirements for high-resolution NWP, but this needs to be further contextualised for the convective-scale application, particularly over the Maritime Continent (see

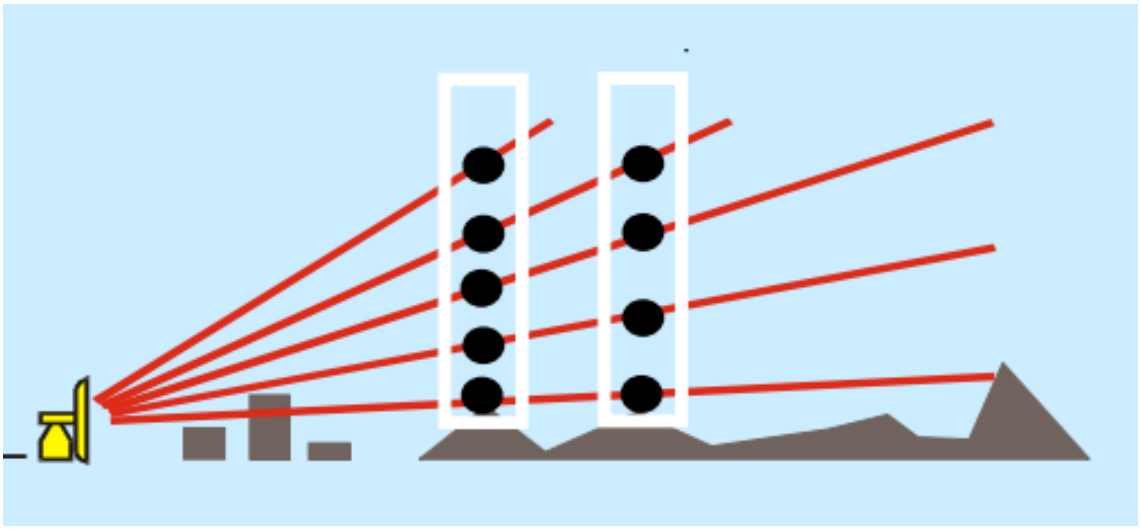

Figure 1.5: Illustration from Wattrelot et al. (2016) showing humidity pseudo-observation profiles derived from radar scans of reflectivity.

Section 1.3.1). Lee et al. (2021) previously showed that assimilating radiosonde wind information alone led to an improvement in the precipitation forecasts of a regional convective-scale NWP system over the Maritime Continent. They also found that assimilating additional moisture information in the lower troposphere was beneficial for representing the diurnal cycle of convection. Such studies are helpful for validating the initial speculations on the useful variables to observe. However, as highlighted by Gustafsson et al. (2018), more studies are still required to decide which initial condition variables are the most important, and on which scales.

Secondly, the treatment of observations and their errors in convective-scale data assimilation needs to be explored. In convective-scale data assimilation, there may be high density non-wind observations (e.g., radar reflectivity and crowdsourced observations) available within the NWP domain (although this is still not the case over the adjacent seas in the Maritime Continent). Due to the high density observation network, observation errors may no longer be assumed to be uncorrelated (e.g., due to correlations in representativity error stemming from the observation operator). Additionally, the observed variable (radar reflectivity or moisture-related variables) and their errors may not adhere to a Gaussian distribution. These add additional complexity in the data assimilation process. Fowler et al. (2018) explored the interactions of observation and forecast error correlations under various scenarios. They found that it was important to account for observation error correlations, especially for certain observation types which have long observation error correlations length-scales. Hawkness-Smith and Simonin (2021) discussed the use of a Huber norm to avoid

problems with convergence or over-penalising contributions from radar observations, particularly where there were large deviations between observations and the background. Wattrelot et al. (2016) discussed another sophisticated approach to assimilate high density radar observations, by involving humidity pseudo-observation profiles computed from radar reflectivity (see Fig. 1.5 for an illustration). For high density satellite observations, most studies have opted to thin the observations to avoid dealing with observation error correlations (e.g., Honda et al., 2018, Jones et al., 2020), but this might inadvertently omit vital convective-scale information. The presence of clouds might also affect the Gaussianity of the satellite observation statistics, which must be handled appropriately (e.g., Chan et al., 2023). Otherwise, these satellite observations might be rejected and convection-related information would be lost. All these complications warrant further studies to provide more evidence on the approaches to treat convective-scale observations.

Thirdly, the suitability and design of data assimilation approaches (e.g., variational, ensemble Kalman-based, non-linear; see Section 1.2) for the convective-scale problem needs to be explored. This area is complex and depends on the intricacies of each approach. Conceptually, the convective-scale atmospheric processes (e.g., convection, cloud microphysics) represented by convective-scale NWP are typically non-linear, so the forecasts may have non-Gaussian errors (Posselt and Bishop, 2018). Consequently, a suitable data assimilation method must be able to account for the dependence of the errors on the weather conditions (flow-dependence), and ideally also their non-linear growth in the forecast error statistics. For the variational approach (Section 1.2.1), some studies have included hydrometeor variables to be analysed together with the prognostic variables (Sun et al., 2021, Wang and Wang, 2021, Destouches et al., 2023). This enabled the direct assimilation of hydrometeor-related convective-scale observations, but this then requires knowledge of how the errors of thermodynamic and dynamic variables are correlated with those of the hydrometeor variables. Many studies using the variational approach conventionally represent the winds using the 'Helmholtz decomposition', namely representing streamfunction (or vorticity) and velocity potential (or divergence) in control variable space (see Section 1.2.1). Some studies have instead retained the zonal and meriodional wind as control variables to be analysed, which was believed to improve the analysis of convective-scale structures (Sun et al., 2016) because of their tighter background error spatial correlations and closer fit to high density observations. For the ensemble Kalman-based approach, the representation of the forecast error statistics using an ensemble explicitly allows for the time-dependence of the estimated statistics (as discussed in Section 1.2.2). This is favourable for convective-

scale data assimilation, but sampling noise must be reduced through localisation. Non-linear data assimilation approaches like particle filters are also promising (Poterjoy et al., 2017), but the application is not yet tractable for a full convective-scale NWP data assimilation system.

1.5 Aims of the thesis

Motivated by the lack of exploratory research in data assimilation over the tropics, especially the Maritime Continent (Section 1.3), and at convective scales (Section 1.4), this thesis seeks to address the following key research questions (RQs):

1. How does the performance of hybrid ensemble-variational data assimilation compare with traditional variational data assimilation over the Maritime Continent?

2. How can traditional ensemble-variational data assimilation approaches, in particular the localisation, be better designed to improve data assimilation and NWP over the tropics?

RQ 1 focuses on (i) the development of hybrid ensemble-variational data assimilation for a simplified model and full NWP system, and (ii) the comparison of hybrid ensemble-variational data assimilation with traditional data assimilation within those modelling frameworks. The justification of choice of these models and their details are covered in Section 1.6. During the development phase, a parallel-run ensemble also had to be implemented to support ensemble-variational data assimilation; the intricacies of the implementation are also discussed within the respective chapters. RQ 2 focuses on modifying the design of ensemble-variational methods, building on the results from RQ 1. The answers to these RQs may inform future development of convective-scale NWP systems over the Maritime Continent.

1.6 Modelling framework and data assimilation systems

The choice of modelling framework and data assimilation system for this thesis is highly dependent on the RQs to be answered. Simplified models allow for rapid testing and development of data assimilation approaches, but they are often not sufficiently complex to mirror certain aspects of data assimilation in full NWP models, or the real-world forecasting problem. For example, the Lorenz-63 model (Lorenz, 1963), the simplest non-linear model of convection (three coupled variables and how they change

in time), can permit multivariate data assimilation research, but it does not represent the spatial variability of the variables. The Lorenz-96 model (Lorenz, 1996) represents the spatial variability of one variable (e.g., waves) and can permit data assimilation research focusing on that aspect, but it is missing the multivariate aspects of full NWP models. Similarly, the one-dimensional shallow water model of Würsch and Craig (2014) can represent the spatial intermittency of convective-scale models, but its one-dimensionality in the horizontal prevents the investigation of important balances that occur in the vertical (e.g., hydrostatic balances).

Clearly, the tradeoff between model complexity and computational cost needs to be balanced to represent all necessary aspects for answering the RQs, yet not rendering the research prohibitively expensive that the RQs cannot be answered. To this end, a portion of this thesis will be explored using a simplified model (ABC-DA system; Section 1.6.1), and a portion with a full NWP model (SINGV-DA system; Section 1.6.2). Most of the development work in advancing data assimilation algorithms is undertaken with the ABC-DA system to enable rapid development, but the SINGV-DA system is also used to translate the lessons from the ABC-DA system to assess the applicability over the Maritime Continent.

1.6.1 Choice of simplified model — the ABC-DA system

To address the lack of simplified models for convective-scale data assimilation, Petrie et al. (2017) developed a simplified model from the compressible and non-hydrostatic three-dimensional Euler equations (see Holton, 1973), known as the ABC model (named after its three key parameters: the pure gravity wave frequency A, the controller of the acoustic wave speed B, and the constant of proportionality between pressure and density perturbations C). The prognostic equations of the ABC model are covered in Chapter 2. Bannister (2020) further implemented basic variational data assimilation approaches in the ABC model, termed as the ABC-DA system. The details are also described in Bannister (2020) and Chapter 2.

Petrie et al. (2017) highlighted how the ABC model can represent large-scale geostrophically and hydrostatically balanced flow, but also permit intermittent convective-like behaviour. Figure 1.6 shows the ABC-DA domain that is used in Chapters 2 and 4, illustrating the spatial variability of zonal wind, one of its prognostic variables. Another feature of the ABC model is its ability to vary the Coriolis parameter since it is present in the Euler equations. This is critical for the data assimilation research over the Maritime Continent (in the deep tropics, very near the Equator with

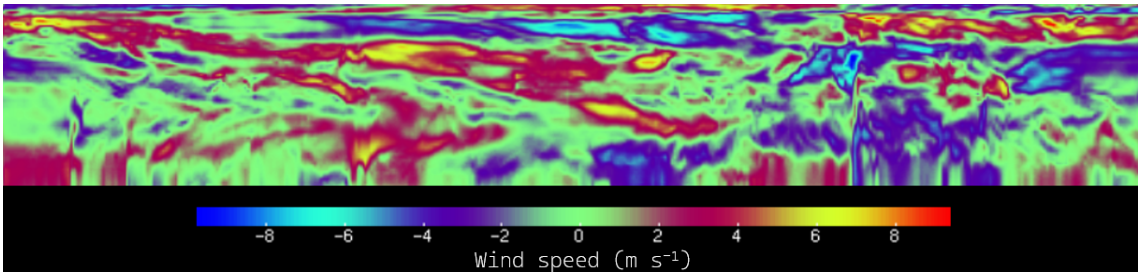

Figure 1.6: ABC-DA domain (longitude-height slice) with coloured contours illustrating spatial variability of zonal wind.

a small Coriolis parameter value), especially since data assimilation over the Maritime Continent is a unique problem (Section 1.3.1). A tropical framework (Chapter 2) can then be set up for the investigation of hybrid ensemble-variational data assimilation.

Evidently, the ABC model represents the intricacies unique to both convective-scale data assimilation and data assimilation over the Maritime Continent and is therefore suitable for answering the RQs in this thesis. However, one aspect that it does not capture is the moist dynamics of the tropical atmosphere, as it is inherently a dry model. Recent work has addressed this in a hydro-ABC model (Zhu and Bannister, 2023), but this version does not yet support data assimilation, unlike ABC-DA. Nonetheless, the tropical dry dynamics representing vertical wind (dry convection) and the mass-wind interactions are still relevant to explore within the ABC-DA system. This limitation is also discussed in Chapters 2 and 4.

1.6.2 Choice of NWP model — the SINGV-DA system

There are only a handful of research and operational centres that maintain a convective-scale NWP system over the western Maritime Continent. Very few centres have incorporated data assimilation, and these centres typical apply only traditional variational methods, partly due to the lack of a suitable high-resolution ensemble which is needed for the application of hybrid ensemble–variational methods.

The SINGV-DA system is operated by the Meteorological Service Singapore (Heng et al., 2020), and is a regional NWP data assimilation system which accounts for orography, land-sea contrasts and other intricacies over the western Maritime Continent (Fig. 1.7). It is sufficiently complex to answer the RQs, with some development work required (e.g., on the parallel-run ensemble) to facilitate hybrid ensemble-variational data assimilation.

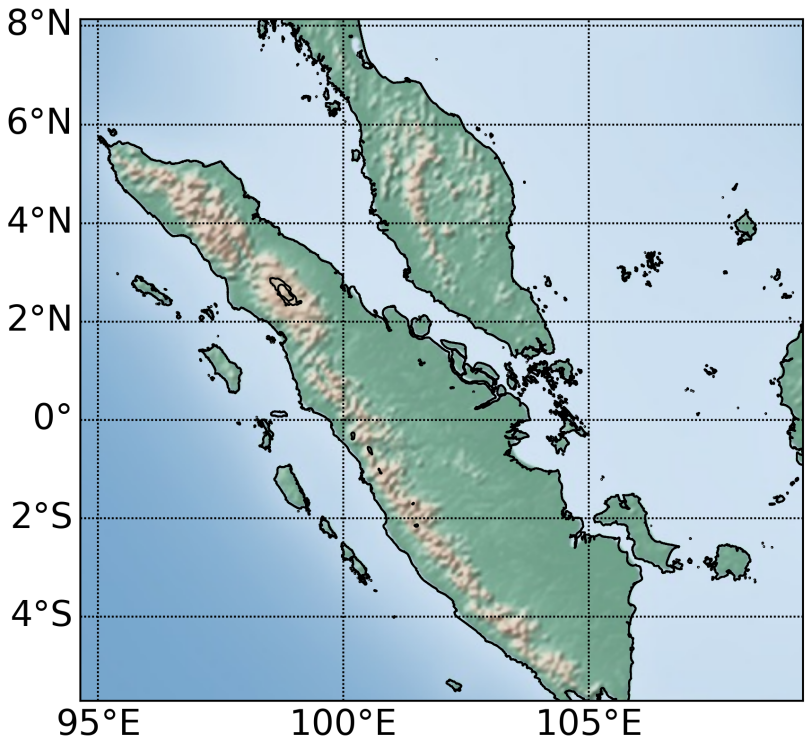

Figure 1.7: SINGV-DA domain, including geographical variations of the terrain height as indicated by shaded relief.

SINGV-DA is based on the United Kingdom Met Office (UKMO) Unified Model framework, which solves the non-hydrostatic deep atmosphere dynamical equations using a semi-implicit, semi-Lagrangian numerical scheme (see Tang et al., 2013 for details). SINGV-DA is a 3-hourly cycling 3DVar-FGAT data assimilation system (cycling illustration in Fig. 1.1), operating at a 1.5 km convective-scale resolution, and driven by ECMWF analysis and forecast data providing lateral boundary conditions (see Fig. 1.4). Assimilated observations are retrieved from the Global Telecommunication System, including conventional, satellite and satellite-derived observations. For brevity, the reader is referred to Heng et al. (2020) for the full observations availability and list. The observation error profiles for SINGV-DA are retrieved from the UKMO.

The setup of SINGV-DA over the western Maritime Continent makes it a natural choice for the investigation of convective-scale hybrid ensemble-variational data assimilation over the Maritime Continent. SINGV-DA is also used to provide the starting initialisation state for the ABC-DA system (see Chapter 2) so that the tropical framework of the ABC model starts with a state that adheres to tropical dynamics.

1.7 Thesis structure

The main body of the thesis consists of three individual papers, presented here as original manuscripts that have been re-formatted to maintain consistency throughout the thesis. Chapters 2 and 3 focus on RQ 1, and Chapter 4 focuses on RQ 2. Further specific literature is reviewed in respective chapters.

Chapter 2 contains the first paper (*Hybrid ensemble-variational data assimilation in ABC-DA within a tropical framework*; Lee et al., 2022), which was published in *Geoscientific Model Development* in August 2022. It describes the development of hybrid ensemble-variational data assimilation in the ABC-DA system (Petrie et al., 2017, Bannister, 2020) to explore RQ 1.

Chapter 3 contains the second paper (*Development of a hybrid ensemble-variational data assimilation system over the western Maritime Continent*; Lee and Barker, 2023), which was published in *Weather and Forecasting* in March 2023. It describes the development of hybrid ensemble-variational data assimilation in a full tropical convective-scale NWP system, SINGV-DA, to further explore RQ 1.

Chapter 4 contains the third paper (*Variable-dependent and selective multivariate localisation for ensemble-variational data assimilation in the tropics*; Lee et al., 2024), was published in *Monthly Weather Review* in April 2024. It describes how the localisation design can be modified to address limitations of traditional localisation approaches. The modified localisation designs are further tested using the ABC model to answer RQ 2.

Each chapter also contains individual questions to explore other aspects of data assimilation, particularly those related to the parallel-run ensemble (e.g., ensemble size/design) which feeds information to hybrid ensemble-variational data assimilation. These questions supplement the main research in this thesis.

Chapter 2

Hybrid ensemble-variational data assimilation in the ABC-DA within a tropical framework

This chapter concerns RQ1 posed in Section 1.5 and has been published in *Geoscientific Model Development* with the following reference:

Lee, J.C.K., Amezcua, J. and Bannister, R.N., 2022. Hybrid ensemble-variational data assimilation in ABC-DA within a tropical framework. *Geoscientific Model Development*, **15(15)**, pp. 6197-6219, https://doi.org/10.5194/gmd-15-6197-2022.

It is unmodified from the published manuscript, other than being re-formatted in accordance with the thesis chapters and with minor typographical adjustments to maintain consistency throughout the thesis.

Abstract

Hybrid ensemble-variational data assimilation (DA) methods have gained significant traction in recent years. These methods aim to alleviate the limitations and maximise the advantages offered by ensemble or variational methods. Most existing hybrid applications focus on the mid-latitudinal context; almost none have explored its benefits in the tropical context. In this article, hybrid ensemble-variational DA is introduced to a tropical configuration of a simplified non-hydrostatic convective-scale fluid dynamics model and its existing variational framework, the ABC-DA system.

The hybrid ensemble-variational DA algorithm is developed based on the alpha

control variable approach, often used in numerical weather prediction. Aspects of the algorithm such as localisation (used to mitigate sampling error caused by finite ensemble sizes) and weighting parameters (used to weight the ensemble and climatological contributions to the background error covariance matrix) are implemented. To produce the flow-dependent error modes (ensemble perturbations) for the ensemble-variational DA algorithm, an ensemble system is also designed for the ABC model, which is run alongside the hybrid DA system. A random field perturbations method is used to generate an initial ensemble, which is then propagated using the ensemble bred vectors method. This setup allows the ensemble to be centred on the hybrid control analysis. Visualisation software has been developed to focus on the diagnosis of the ensemble system.

To demonstrate the hybrid ensemble-variational DA in the ABC-DA system, sensitivity tests using observing system simulation experiments are conducted within a tropical framework. A 30-member ensemble was used to generate the error modes for the experiments. In general, the best performing configuration (with respect to the 'truth') for the hybrid ensemble-variational DA system used an 80%/20% weighting on the ensemble-derived/climatological background error covariance matrix contributions. For the horizontal wind variables though, full weight on the ensemble-derived background error covariance matrix (100%/0%) resulted in the smallest cycle-averaged analysis root-mean-square errors, mainly due to large errors in the meridional wind field when contributions from the climatological background error covariance matrix were involved, possibly related to a sub-optimal background error covariance model.

The ensemble bred vectors method propagated a healthy-looking DA-centred ensemble without bimodalities or evidence of filter collapse. The ensemble was under-dispersive for some variables, but for others, the ensemble spread approximately matched the corresponding root-mean-square errors. Reducing the number of ensemble members led to slightly larger errors across all variables, due to the introduction of larger sampling errors into the system.

2.1 Introduction

Data assimilation (DA) methods can traditionally be classified into three categories: variational methods, which look for a maximum-a-posteriori (MAP) estimator, Kalman-based methods, which produce a minimum variance estimator (often in an ensemble implementation), and methods which attempt to estimate full probability

density functions (PDFs) without making any parametric assumptions (e.g., Markov Chain Monte Carlo and particle filters). For an introductory discussion the reader is referred to e.g., Asch et al. (2016). Each traditional DA method is subject to its own advantages and limitations, which determine the applicability in operational numerical weather prediction (NWP) systems. A wide spectrum of modern DA methods have been proposed in recent years, including hybrid ensemble-variational (hybrid-EnVar) methods which have gained significant traction. Within the category of hybrid-EnVar methods, there exist different flavours due to subtleties in the derivation and permutations arising from the usage of different variational or ensemble methods. One can modify, for instance, the elements of the problem or the solution algorithm, yielding different varieties of hybrid variants. Bannister (2017) provides a comprehensive review of the latest hybrid-EnVar methods used in modern DA.

In this article, we focus on the hybrid covariance ensemble-variational approach (Hamill and Snyder, 2000). This differs from the hybrid gain ensemble-variational DA approach (Penny, 2014), which is also commonly used. Most existing hybrid applications focus on the mid-latitudinal context and highlight the advantage of introducing flow-dependency in the error statistics. However, almost none have explored the hybrid application in the tropical context, where the characteristics of the error statistics are still poorly understood. Here, we introduce the hybrid-EnVar method to an existing convective-scale DA framework (Bannister, 2020) for a simplified non-hydrostatic fluid dynamics model (ABC model; Petrie et al., 2017), with the hope that the upgraded system can provide insights on the benefits and highlight potential issues that may arise using hybrid-EnVar methods in the tropical context. We note that this study is also the first to use a tropical configuration of the ABC-DA system.

The aims of this study are as follows:

(a) to document and test a hybrid-EnVar DA system for the ABC model, and

(b) to test generating an ensemble suitable for hybrid-EnVar DA to function.

Section 2.2 contains details of the existing system used in this study. Section 2.3 documents the development of an ABC ensemble system, necessary to generate a meaningful ensemble of ABC states, which feed into the hybrid-EnVar DA system along with the implementation of hybrid-EnVar DA system itself. Section 2.4 demonstrates the use of the hybrid-EnVar DA system within a tropical framework. Three appendices provide details that may be of interest to readers familiar with ensemble initialisation, and inter-variable localisation in ensemble DA.

2.2 The ABC-DA system

2.2.1 Model equations

The ABC model used in this study was originally developed by Petrie et al. (2017) and was designed as a simplified non-hydrostatic fluid dynamics model for use in convective-scale DA experiments. It comprises solving a set of simplified partial differential equations derived from the Euler equations. A vertical slice formulation containing only dry dynamics is used (two-dimensional *x*-*z* spatial grid). This section summarises the model equations and their properties. These are:

$$\frac{\partial u}{\partial t} + B\boldsymbol{u} \cdot \nabla u + C\frac{\partial \tilde{\rho}'}{\partial x} - fv = 0, \tag{2.1a}$$

$$\frac{\partial v}{\partial t} + B\boldsymbol{u} \cdot \nabla v + fu = 0, \tag{2.1b}$$

$$\frac{\partial w}{\partial t} + B\boldsymbol{u} \cdot \nabla w + C\frac{\partial \tilde{\rho}'}{\partial z} - b' = 0, \tag{2.1c}$$

$$\frac{\partial \tilde{\rho}'}{\partial t} + B\nabla \cdot (\tilde{\rho}\boldsymbol{u}) = 0, \tag{2.1d}$$

$$\frac{\partial b'}{\partial t} + B\boldsymbol{u} \cdot \nabla b' + A^2 w = 0, \tag{2.1e}$$

where $\boldsymbol{u} = (u, v, w)$ is the three-dimensional wind vector of zonal, meridional and vertical wind; $\tilde{\rho}'$ and b' are perturbation quantities from a reference state of scaled density ($\tilde{\rho}$) and buoyancy respectively (see Petrie et al., 2017). The coefficients A, B and C are tunable parameters which control the pure gravity wave frequency, the modulation of the advective and divergent terms, and the relationship between the pressure and density perturbations in the equation of state, respectively. The small-scale acoustic wave speed is given by $\sqrt{BC}$. Additionally, the Coriolis parameter f can be chosen depending on the desired latitudinal position of the vertical slice. Collectively the variables u, v, w, $\tilde{\rho}'$ and b' at every grid position in the domain are referred to as the state vector $\mathbf{x}$.

2.2.2 Variational data assimilation

Variational DA was subsequently implemented in the ABC model by Bannister (2020), termed as the ABC-DA system. As of version 1.4 (https://doi.org/10.5281/zenodo.3531926), 3DVar and 3DVar-FGAT (First Guess at Appropriate Time) are available in the ABC-DA system. The reader is directed to Bannister (2020) for the full details of this implementation, but here we summarise the key equations in the context of 3DVar-FGAT. The 3DVar-FGAT scheme is later in this article adapted into a hybrid scheme.

Incremental formulation

The objective of variational DA is to find an optimal state $\mathbf{x}^a$ which minimises a cost function $J(\mathbf{x})$ (e.g., Kalnay, 2003). This cost function usually comprises two terms: one for the departure of the state with respect to the background state $\mathbf{x}^b$, and one for the departure of the state (transformed to observation space) with respect to observations $\mathbf{y}$. A third term related to any model errors can be added in the so-called weak-constraint formulation, which is not needed in this work as we do not consider model errors in our set-up. Even though the terms in J are based on Mahalanobis distances, J can be non-quadratic (with respect to the state variable) due to the non-linearities of the (often) non-linear forecast model ($\mathcal{M}_{t-1\rightarrow t}$, used in the case of 4DVar) and observation operator ($\mathcal{H}_t$). Most variational systems implement an incremental formulation of the cost function (Courtier et al., 1994) which involves iteratively linearising $\mathcal{M}_{t-1\rightarrow t}$ and $\mathcal{H}_t$ around a reference state ($\mathbf{x}^r$) and framing the problem in terms of increments to $\mathbf{x}^r$ in a series of outer loops. This allows one to find an approximate solution of a complicated non-quadratic optimisation problem by tackling a series of easier quadratic ones. To illustrate, for a DA cycle with a window from $t = 0$ to T, the incremental form of the 3DVar-FGAT cost function is:

$$J(\delta\mathbf{x}) = \frac{1}{2}(\delta\mathbf{x} - \delta\mathbf{x}^b)^\top \mathbf{B}_c^{-1}(\delta\mathbf{x} - \delta\mathbf{x}^b) + \frac{1}{2}\sum_{t=0}^{T}(\mathbf{H}_t\delta\mathbf{x} - \mathbf{d}[t])^\top \mathbf{R}_t^{-1}(\mathbf{H}_t\delta\mathbf{x} - \mathbf{d}[t]) \quad (2.2)$$

where $\mathbf{x} = \mathbf{x}^r + \delta\mathbf{x}$, $\mathbf{x}^r$ is a reference state, $\delta\mathbf{x}$ is the state increment, and $\delta\mathbf{x}^b$ is the difference between the background and the reference. $\mathbf{B}_c$ is the background error covariance matrix, $\mathbf{R}_t$ is the observation error covariance matrix at time t, and $\mathbf{H}_t$ is the linearised observation operator at time t. We define the innovation:

$$\mathbf{d}[t] = \mathbf{y}[t] - \mathcal{H}_t[\mathcal{M}_{0\rightarrow t}[\mathbf{x}^r]]. \quad (2.3)$$

Note that Eq. (2.2) is the same as Eq. (7) of Bannister (2020), except that the linearised forecast model $\mathbf{M}_{t-1\rightarrow t}$ has been replaced here by the identity $\mathbf{I}$ (this replacement is what distinguishes 3DVar-FGAT from 4DVar). For the first outer loop, $\mathbf{x}^r$ is set as $\mathbf{x}^b$ (i.e., $\delta\mathbf{x}^b = 0$).

Estimation and modelling of $\mathbf{B}_c$

A vital component in variational DA is $\mathbf{B}_c$. It is the averaged (climatological) second moment of the PDF of forecast errors of the system (Bannister, 2008*a*). It determines the weighting between the use of observational and background information, and it

allows for the spreading of observational information spatially and between variables. We can disentangle the construction of $\mathbf{B}_c$ by considering how the background errors are first estimated, and then used in the modelling of $\mathbf{B}_c$.

In the original implementation by Bannister (2020), the estimation of the background error statistics was performed by the extraction of multiple longitude-height slices from one or more Met Office Unified Model outputs (since these were conveneiently available), which were processed to create an 'ensemble' of ABC states (and subsequently ABC forecasts). This set of forecast perturbations serve as proxies for the background errors used as training data for $\mathbf{B}_c$. The validity of this prescribed source of background error statistics has not been investigated, but the approach is convenient and practical. Another way to estimate the training data is to compute forecast differences (with different lead times and valid at the same time) over a climatological period (the National Meteorological Center method; Parrish and Derber, 1992), but as of version 1.4, this is not coded in the ABC-DA system. Instead, we introduce a different method to compute the ensemble forecasts for the training data (Section 2.3.1).

In many systems, $\mathbf{B}_c$ is too large to explicitly be computed using the training data. For instance, in operational models the size of the state variable can be $\mathcal{O}\left(10^9\right)$. Instead, $\mathbf{B}_c$ is often modelled through the use of a so-called control variable transform $\mathbf{U}$. Even though the ABC model is small enough for the explicit computation of $\mathbf{B}$ to be feasible, it is still far more practical to use a control variable. We introduce a control vector $\delta\boldsymbol{\chi}$ which is related to a state vector $\delta\mathbf{x}$ by:

$$\delta\mathbf{x} = \mathbf{U}\delta\boldsymbol{\chi}. \tag{2.4}$$

The choice of the control vector, $\delta\boldsymbol{\chi}$, and control variable transform $\mathbf{U}$ is flexible, but they dictate the eventual cross-covariances between model variables of $\delta\mathbf{x}$. In order to improve the conditioning of the incremental cost function (for more efficient minimisation), the control variables are chosen to be uncorrelated and have unit variance. Substituting Eq. (2.4) into Eq. (2.2) yields a new pre-conditioned incremental cost function:

$$J(\delta\boldsymbol{\chi}) = \frac{1}{2}(\delta\boldsymbol{\chi}-\delta\boldsymbol{\chi}^b)^\top(\delta\boldsymbol{\chi}-\delta\boldsymbol{\chi}^b)+\frac{1}{2}\sum_{t=0}^{T}(\mathbf{H}_t\mathbf{U}\delta\boldsymbol{\chi}-\mathbf{d}[t])^\top\mathbf{R}_t^{-1}(\mathbf{H}_t\mathbf{U}\delta\boldsymbol{\chi}-\mathbf{d}[t]) \tag{2.5}$$

where $\delta\mathbf{x}^b = \mathbf{U}\delta\boldsymbol{\chi}^b$. Since $\mathbf{B}_c$ is a symmetric and positive matrix, $\mathbf{U}$ may be chosen to be a lower triangular matrix (using a Cholesky decomposition). The implied $\mathbf{B}_c$ is

given by minimising Eq. (2.5) with the transform in Eq. (2.4); $\mathbf{B}_c = \mathbf{U}\mathbf{U}^\top$ (Bannister, 2008*b*). It is evident that the use of a carefully designed $\mathbf{U}$ removes the need to compute $\mathbf{B}_c^{-1}$ in order to minimise the cost function. By contrast, since observation errors are assumed to be uncorrelated in the ABC-DA system, $\mathbf{R}_t$ is diagonal. Hence, there is no requirement for a separate transform since $\mathbf{R}_t^{-1}$ can be easily computed.

The calibration of $\mathbf{U}$ (and thus the implied $\mathbf{B}_c$) is usually only performed once at the start of any cycling experiment using 'climatological' background error statistics, and then used for every DA cycle.

ABC-DA minimisation algorithm

In the ABC-DA system, a conjugate gradient algorithm is used to find the minimiser of the cost function. Differentiating Eq. (2.5) with respect to $\delta\boldsymbol{\chi}$ yields the gradient $\nabla_{\delta\boldsymbol{\chi}} J$, given by:

$$\nabla_{\delta\boldsymbol{\chi}} J = \delta\boldsymbol{\chi} - \delta\boldsymbol{\chi}^b + \mathbf{U}^\top \sum_{t=0}^{T} \mathbf{H}_t^\top \mathbf{R}_t^{-1} (\mathbf{H}_t \mathbf{U} \delta\boldsymbol{\chi} - \mathbf{d}[t]) \tag{2.6}$$

where $\mathbf{U}^\top$ and $\mathbf{H}_t^\top$ are the adjoints of $\mathbf{U}$ and $\mathbf{H}_t$ respectively. The reader is directed to Bannister (2020) for more details. The modifications required for specific steps in order to enable hybrid-EnVar DA are highlighted later in Section 2.3.2.

2.3 Technical implementation of the data assimilation and forecast framework

Hybrid-EnVar schemes stem from a combination of two approaches: ensemble methods and variational methods. For the former, the archetypical example is the Kalman filter (EnKF) in its different formulations. The reader is referred to e.g., Evensen (2006) for an introduction. The variational approach has been discussed in Section 2.2.2. Instead of a one-off retrieval of the background error statistics from a climatological source, the purpose of having an ensemble is to estimate time-dependent background error statistics from the ensemble forecasts valid at each cycle. As such, the background error statistics vary as the system evolves.

Accordingly, a parallel ensemble system that runs alongside the hybrid (single-trajectory) analysis is required in order to provide the background error statistics at each cycle. In this study, we explore the ensemble bred vectors (a variant of the bred

vector method; EBV) method to evolve the ensemble system, which will be described in Section 2.3.3. The following sections cover the step-by-step implementation of a cycling hybrid-EnVar DA system in the ABC model, in accordance with the schematic diagram (Fig. 2.1) which shows the coupling between the deterministic components and the parallel-run ensemble system using the two different ensemble propagation methods. Figure 2.1 is explained in the remainder of Section 2.3.

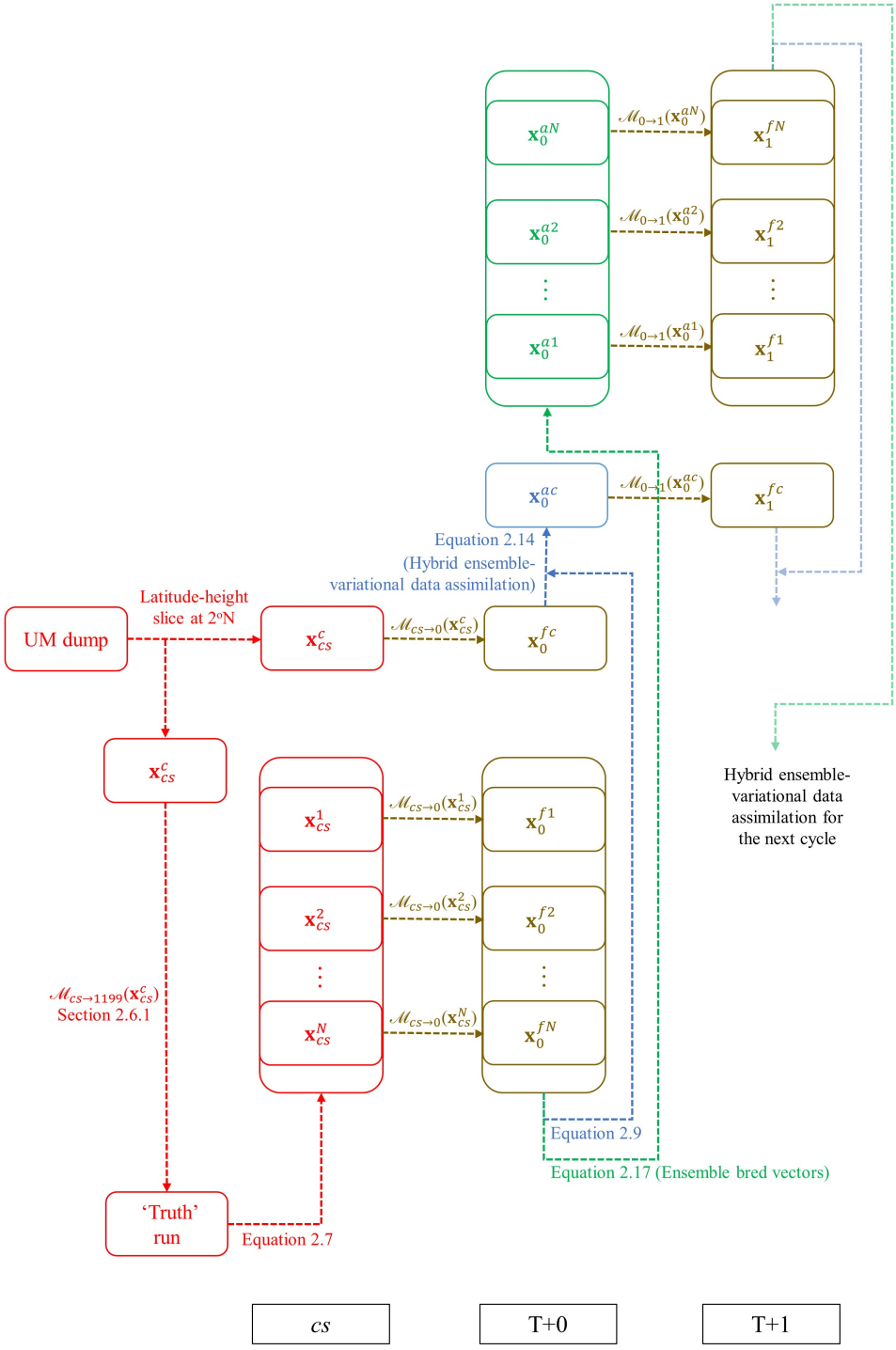

Figure 2.1: Schematic diagram of the ensemble and deterministic workflow for the hybrid-EnVar scheme in the ABC-DA system, illustrated for an hourly-cycling setup over the first cycle from a cold start. The subscripts refer to the validity time; cs refers to cold start. The superscripts fk and fc refer to the k^{th} member of the forecast ensemble and the control forecast respectively, ak and ac refer to the k^{th} member of the analysis ensemble and the hybrid control analysis respectively.

2.3.1 Generation of initial ensemble of states for ABC ensemble system

This section discusses the generation of an initial ensemble, the first step in Fig. 2.1 (red segments), which is needed in the case of a cold start. Subsequent propagation of the ensemble can then proceed after this problem is addressed.

In the ABC model, an initial two-dimensional state can be computed from a longitude-height slice of a Unified Model output, which is a convenient approach adopted by Petrie et al. (2017). In this light, the simplest method to generate an initial ensemble is to extract different longitude-height slices from the same Unified Model output, similar to how a population of training data is generated in Bannister (2020) for the calibration of the static background error covariances as mentioned in Section 2.2.2. Another method is to simply add statistical noise to the initial ABC model state, although it is not straightforward to determine the distributions for the noise sampling (which could vary for different variables and include multivariate correlations), so that the solutions are consistent with the underlying dynamics. The model evolution in the first few cycles may spuriously dampen or amplify the added statistical noise if it is drawn from an incorrectly-chosen distribution.

For this study, we adopt the random field perturbation method proposed by Magnusson et al. (2009) to generate the initial ensemble. The main idea relies on choosing two (assumed independent) ABC states and calculating their differences. The differences are treated as perturbations and can then be scaled to maintain a fixed amplitude between ensemble members and/or cycles, and are subsequently added to the initial ABC state computed above to generate an initial (arbitrary-sized) ensemble of states. Linear balances are approximately preserved in the resulting ensemble as only linear operations are performed on the fields (Magnusson et al., 2009).

Unlike in an operational NWP system where archived past analyses are available, here a long 'truth' simulation needs to be performed using the ABC model starting from a chosen initial state (the 'truth run' in Fig. 2.1). To generate each ensemble member, two states from the same 'truth' simulation are chosen at random. These need to be sufficiently separated in time for the assumption of independence to be valid. Following

the above steps, the initial ensemble of states is given by:

$$\mathbf{x}^k_{cs} = \mathbf{x}^c_{cs} + \frac{1}{\sqrt{2}} r^k (\mathbf{x}^{tr}_{kt1} - \mathbf{x}^{tr}_{kt2}), \tag{2.7a}$$

$$r^k = \frac{\epsilon^{rf}}{|\mathbf{x}^{tr}_{kt1} - \mathbf{x}^{tr}_{kt2}|_{E_{tot}}}, \quad \epsilon^{rf} = \overline{|\mathbf{x}^{tr}_{kt1} - \mathbf{x}^{tr}_{kt2}|_{E_{tot}}} \tag{2.7b}$$

where $\mathbf{x}^k_{cs}$ represents the k^{th} initial state of N ensemble members, $\mathbf{x}^c_{cs}$ is the initial unperturbed (hereafter referred to as control) state; the superscript 'cs' refers to cold start, $\mathbf{x}^{tr}_{kt1}$ and $\mathbf{x}^{tr}_{kt2}$ are the two random states drawn from the same 'truth' simulation at different times ($kt1$ and $kt2$), and r^k depends on the scaling factor ϵ^{rf} defined according to the total energy norm ($|\bullet|_{E_{tot}} = \sqrt{E_{tot}}$; see Eq. (2.19)) of the perturbations, to maintain a fixed amplitude (the ensemble mean $\overline{|\bullet|_{E_{tot}}}$) between ensemble members. The reason for the $\frac{1}{\sqrt{2}}$ factor is included because we are considering differences between two states so the variance of the difference is a reflection of the sum of their error variances, rather than considering differences between a state and a mean (see Appendix of Berre et al., 2006). More details and justification of the method are covered in Section 2.6.1.

After generating the initial ensemble, the cold start members are propagated to the analysis time of the first cycle at $T+0$ (from $\mathbf{x}^k_{cs}$ to $\mathbf{x}^{fk}_0$ and from $\mathbf{x}^c_{cs}$ to $\mathbf{x}^{fc}_0$ by $\mathcal{M}_{cs\rightarrow 0}$; Fig. 2.1, lower brown segments). Note that for subsequent cycles, the analysis ensemble (produced from Section 2.3.3) and control analysis are propagated instead of cold start members (i.e., from $\mathbf{x}^{ak}_t$ to $\mathbf{x}^{fk}_{t+1}$ and from $\mathbf{x}^{ac}_t$ to $\mathbf{x}^{fc}_{t+1}$ by $\mathcal{M}_{t\rightarrow t+1}$; Fig. 2.1, upper right brown segments). These ensemble forecasts are then used in the DA step, described in the next section.

2.3.2 Hybrid ensemble-variational data assimilation

The hybrid-EnVar approach seeks to implement a hybrid background error covariance $\mathbf{B}_h$ which is a linear combination of a climatological and an ensemble-derived background error covariance matrix ($\mathbf{B}_c$ and $\mathbf{B}_e$), in the form following Hamill and Snyder (2000):

$$\mathbf{B}_h = \beta_c^2 \mathbf{B}_c + \beta_e^2 \mathbf{B}_e \tag{2.8}$$

where β_c^2 and β_e^2 are (positive) scalar weights often determined empirically for the algorithm. These weights are often chosen to add to unity, but this need not be the case. This approach computes $\mathbf{B}_h$ explicitly, but is not practical in an NWP system. For the ABC-DA system, the alpha control variable approach of Lorenc (2003) is instead

implemented, which constructs an implied version of Eq. (2.8) using an alteration of the standard variational cost function and control variables. Wang et al. (2007) demonstrates the mathematical equivalence of both approaches.

Given the control background, which is a short-range forecast from the previous cycle ($\mathbf{x}^b = \mathbf{x}_t^{fc}$), the hybrid-EnVar DA yields the hybrid control analysis $\mathbf{x}_t^{ac}$ (Fig. 2.1, blue segments), which needs the ensemble members to implicitly construct the $\mathbf{B}_e$ part (recall that $\mathbf{B}_c$ in Eq. (2.8) is derived from the $\mathbf{U}$ transform, and $\mathbf{B}_e$ is derived from the ensemble). The steps to retrieve the hybrid control analysis are described below, but we first explain how $\mathbf{B}_e$ can be computed from the ensemble.

Computation of the ensemble-derived background error covariance matrix for the control analysis

At each cycle, one may compute a rectangular matrix $\mathbf{X}_t^f$ whose columns contain the scaled differences between the ensemble forecasts (i.e., $\mathbf{x}_t^{fk}$ for the k^{th} member forecast valid at time t) and the ensemble mean ($\bar{\mathbf{x}}_t^f$):

$$\mathbf{X}_t^f = \frac{1}{\sqrt{N-1}}(\mathbf{x}_t^{f1} - \bar{\mathbf{x}}_t^f, \mathbf{x}_t^{f2} - \bar{\mathbf{x}}_t^f, \ldots, \mathbf{x}_t^{fN} - \bar{\mathbf{x}}_t^f) = (\mathbf{x}_t'^1, \mathbf{x}_t'^2, \ldots, \mathbf{x}_t'^N) \qquad (2.9)$$

where $\mathbf{x}_t'^k$ are the scaled error modes valid at time t. The ensemble-derived background error covariance matrix (at time t) $\mathbf{P}_e^f[t]$ is explicitly given by the outer product:

$$\mathbf{P}_e^f[t] = \mathbf{X}_t^f \mathbf{X}_t^{f\top}. \qquad (2.10)$$

As we shall see, this matrix is not computed explicitly, although parts of it are computed explicitly for visualisation purposes later in this article.

In the limit where N tends to infinity, or where N is far greater than the degrees of freedom of the state n ($N \gg n$), $\mathbf{P}_e^f[t]$ may be full rank. In practice, however, a small number of ensemble members ($N \ll n$) will inevitably lead to sampling error and a rank-deficient matrix. Houtekamer and Mitchell (2001) proposed mitigating this problem by performing a Schur product of $\mathbf{P}_e^f$ with a correlation matrix (or localisation matrix) $\mathbf{L}$:

$$\mathbf{B}_e = \mathbf{L} \circ \mathbf{P}_e^f[t]. \qquad (2.11)$$

This seeks to address the sampling error by damping the long-range background error covariances, as well as effectively increasing the rank of $\mathbf{P}_e^f[t]$. The spatial and multivariate aspects of the localisation matrix are further discussed below, including how this

can be performed without constructing explicit matrices.

Alpha control variable transform

Following the approach of Lorenc (2003), we introduce an ensemble-related penalty in the variational cost function. This requires constructing so-called alpha fields $\boldsymbol{\alpha}^k$ (part of a new set of mutually uncorrelated control variables) associated with each ensemble member k, and constrained to have covariance $\mathbf{L}$ (the localisation matrix, as used in Eq. (2.11)). The number of elements in $\boldsymbol{\alpha}^k$ must be the same as the state vector of $\mathbf{x}_t'^k$ (number of model gridpoints N_g $\times$ number of model variables N_{var}). The modified cost function is:

$$J(\delta\boldsymbol{\chi}, \boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2, \ldots, \boldsymbol{\alpha}^N) = \overbrace{\frac{1}{2}(\delta\boldsymbol{\chi} - \delta\boldsymbol{\chi}^b)^\top(\delta\boldsymbol{\chi} - \delta\boldsymbol{\chi}^b)}^{J_b} + \overbrace{\frac{1}{2}\sum_{t=0}^{T}(\mathbf{H}_t\delta\mathbf{x} - \mathbf{d}[t])^\top\mathbf{R}_t^{-1}(\mathbf{H}_t\delta\mathbf{x} - \mathbf{d}[t])}^{J_o} + \overbrace{\frac{1}{2}\sum_{k=1}^{N}\boldsymbol{\alpha}^{k\top}\mathbf{L}^{-1}\boldsymbol{\alpha}^k}^{J_e} \tag{2.12a}$$

$$\text{with } \delta\mathbf{x} = \beta_c\mathbf{U}\delta\boldsymbol{\chi} + \beta_e\sum_{k=1}^{N}\mathbf{x}_t'^k \circ \boldsymbol{\alpha}^k, \tag{2.12b}$$

where J_b, J_o and J_e are the background, observation and ensemble penalties respectively. Equation (2.12a) is an extension of Eq. (2.5), and Eq. (2.12b) is an extension of Eq. (2.4), the hybrid control variable transform. Together these equations make up the hybrid scheme.

Similar to the way that $\mathbf{B}_c$ can be decomposed as $\mathbf{B}_c = \mathbf{U}\mathbf{U}^\top$, $\mathbf{L}$ can be decomposed in terms of the alpha control variable transform, $\mathbf{U}^\alpha$, i.e., $\mathbf{L} = \mathbf{U}^\alpha\mathbf{U}^{\alpha\top}$. Consider an alpha control vector $\boldsymbol{\chi}^{\alpha k}$ (again associated with ensemble member k) which is related to the alpha field $\boldsymbol{\alpha}^k$ via:

$$\boldsymbol{\alpha}^k = \mathbf{U}^\alpha\boldsymbol{\chi}^{\alpha k}. \tag{2.13}$$

Substituting Eq. (2.13) into Eq. (2.12a) yields:

$$J(\delta\boldsymbol{\chi}, \boldsymbol{\chi}^{\alpha 1}, \boldsymbol{\chi}^{\alpha 2}, \ldots, \boldsymbol{\chi}^{\alpha N}) = J_b + J_o + \frac{1}{2}\sum_{k=1}^{N}\boldsymbol{\chi}^{\alpha k\top}\boldsymbol{\chi}^{\alpha k} \tag{2.14a}$$

$$\text{with } \delta\mathbf{x} = \beta_c\mathbf{U}\delta\boldsymbol{\chi} + \beta_e\sum_{k=1}^{N}\mathbf{x}_t'^k \circ (\mathbf{U}^\alpha\boldsymbol{\chi}^{\alpha k}), \tag{2.14b}$$

$$\mathbf{x}_t^{ac} = \mathbf{x}^r + \delta\mathbf{x}^a. \tag{2.14c}$$

The variational problem (Eq. (2.14a)) is minimised with respect to the collective set of control vectors, comprising a part that is associated with $\mathbf{B}_c$ ($\delta\boldsymbol{\chi}$), and parts that are associated with $\mathbf{B}_e$ ($\boldsymbol{\chi}^{\alpha 1}, \boldsymbol{\chi}^{\alpha 2}, \ldots, \boldsymbol{\chi}^{\alpha N}$). Together, these are combined using the hybrid transform (Eq. (2.14b)) to give the particular $\delta\mathbf{x}$ that minimises Eq. (2.14a), namely $\delta\mathbf{x}^a$. This gives the analysis $\mathbf{x}_t^{ac}$ in Eq. (2.14c).

The total implied covariance matrix (that is effectively seen by the DA) is formally given by:

$$\mathbf{B}_h = \beta_c^2 \mathbf{U}\mathbf{U}^\top + \beta_e^2 (\mathbf{U}^\alpha \mathbf{U}^{\alpha\top}) \circ (\mathbf{X}_t^f \mathbf{X}_t^{f\top}), \tag{2.15}$$

which is a linear combination of the implied $\mathbf{B}_c$ and $\mathbf{B}_e$ (without explicitly constructing either), and is element-wise equivalent to the explicit hybrid covariance in Eq. (2.8).

Next, we reproduce the minimisation algorithm steps Section 3.5 of Bannister (2020), and highlight (underlined) the modifications required when the hybrid-EnVar scheme is enabled:

1. Set the reference state at $t = 0$ to the background state $\mathbf{x}^r = \mathbf{x}^b$. Decide values for N, β_c, and β_e.
2. Do the outer loop.
   (a) For the first outer loop, $\delta\boldsymbol{\chi}^b = 0$; otherwise, compute $\delta\boldsymbol{\chi}^b = \mathbf{U}^{-1}(\mathbf{x}^b - \mathbf{x}^r)$.
   (b) Compute $\mathbf{x}^r[t]$ over the time window, $1 \leq t \leq T$, with the non-linear model $\mathbf{x}^r[t] = \mathcal{M}_{t-1\rightarrow t}(\mathbf{x}^r[t-1])$.
   (c) Compute the reference state's observations: $\mathbf{y}^{mr}[t] = \mathcal{H}_t(\mathbf{x}^r[t])$.
   (d) Compute the differences: $\mathbf{d}[t] = \mathbf{y}[t] - \mathbf{y}^{mr}[t]$.
   (e) Set $\delta\boldsymbol{\chi} = 0$, $\delta\mathbf{x} = 0$, and $\boldsymbol{\chi}^{\alpha k} = 0$, $1 \leq k \leq N$.
   (f) Do the inner loop.
      i. Integrate the perturbation trajectory over the time window, $1 \leq t \leq T$, with the linear forecast model: $\delta\mathbf{x}[t] = \mathbf{M}_{t-1\rightarrow t}\delta\mathbf{x}[t-1]$.
      ii. Compute the perturbations to the model observations: $\delta\mathbf{y}^m[t] = \mathbf{H}_t\delta\mathbf{x}[t]$.
      iii. Compute $\boldsymbol{\Delta}[t]$ vectors as defined as $\boldsymbol{\Delta}[t] = \mathbf{H}_t^\top \mathbf{R}_t^{-1}(\delta\mathbf{y}^m[t] - \mathbf{d}[t])$.
      iv. Set the adjoint state $\boldsymbol{\lambda}[T+1] = 0$.
      v. Integrate the following adjoint equation backwards in time, $T \geq t \geq 0$: $\boldsymbol{\lambda}[t] = \boldsymbol{\Delta}[t] + \mathbf{M}_{t\rightarrow t+1}^\top \boldsymbol{\lambda}[t+1]$.

vi. Compute the gradient as follows: $\nabla_{\delta\boldsymbol{\chi}} J = \delta\boldsymbol{\chi} - \delta\boldsymbol{\chi}^b + \beta_c \mathbf{U}^\top \boldsymbol{\lambda}[0]$, and $\nabla_{\boldsymbol{\chi}^{\alpha k}} J = \boldsymbol{\chi}^{\alpha k} + \beta_e \mathbf{U}^{\alpha\top}(\mathbf{x}_t'^k \circ \boldsymbol{\lambda}[0])$. These are the gradients with respect to each control vector segment, $1 \leq k \leq N$.

vii. Use the conjugate gradient algorithm to adjust $\delta\boldsymbol{\chi}$ and $\boldsymbol{\chi}^{\alpha k}$ to reduce the value of J. Note that the cost function is $J = J_b + J_o + J_e$ (Eq. (2.14a)).

viii. Compute the new increment in model space using the control variable transform and alpha control variable transform:
$\delta\mathbf{x} = \beta_c \mathbf{U}\delta\boldsymbol{\chi} + \beta_e \sum_{k=1}^{N} \mathbf{x}_t'^k \circ (\mathbf{U}^\alpha \boldsymbol{\chi}^{\alpha k})$ (Eq. (2.14b)).

ix. Go to step 2fi until the inner-loop convergence criterion is satisfied.

(g) Update the reference state: $\mathbf{x}^r \rightarrow \mathbf{x}^r + \delta\mathbf{x}$.

(h) Go to step 2a until the outer-loop convergence criterion is satisfied. At convergence, set the hybrid control analysis $\mathbf{x}_t^{ac} = \mathbf{x}^r$.

3. Run a non-linear forecast from $\mathbf{x}_t^{ac}$ for the background of the next cycle and longer forecasts if required.

Inter-variable and spatial localisation

Localisation of the ensemble-derived background error covariance matrix, as in Eq. (2.11), is required to mitigate sampling error, which can dominate the computed covariance between distant points (Hamill et al., 2001). Localisation opens up a range of options and raises some pertinent questions: Should we localise only in space (and should these spatial localisation matrices depend on the variable), or should we additionally include localisation between different model variables? This depends on the design of $\boldsymbol{\chi}^{\alpha k}$ and $\mathbf{U}^\alpha$ in Eq. (2.13), and the implied $\mathbf{L}$. If $\boldsymbol{\chi}^{\alpha k}$ only depends on gridpoint location (i.e., it need only be of length N_g), then $\mathbf{U}^\alpha$ must be rectangular ($N_g N_{var} \times N_g$) so that $\boldsymbol{\alpha}^k$ has length $N_g N_{var}$ required for the Schur product in Eq. (2.12b). This approach was adopted by Wang et al. (2008*a*), except that $\boldsymbol{\chi}^{\alpha k}$ was only dependent on *horizontal* gridpoint locations. By design, $\mathbf{U}^\alpha$ functions to 'use the same $\boldsymbol{\chi}^{\alpha k}$ for each model variable and model level' (i.e., repeated rows in $\mathbf{U}^\alpha$) so the Schur product in Eq. (2.14b) can be computed. This point was not highlighted in the description of Eq. (1) of Wang et al. (2008*a*).

If $\boldsymbol{\chi}^{\alpha k}$ only has N_g elements, the implied $\mathbf{L} = \mathbf{U}^\alpha \mathbf{U}^{\alpha\top}$ can only involve spatial localisation, so full inter-variable covariances (as found from the raw ensemble) are retained (see Section 2.6.2). Alternatively, if $\boldsymbol{\chi}^{\alpha k}$ is full length (i.e., of length $N_g N_{var}$) with independent fields for each model variable, $\mathbf{U}^\alpha$ is square ($N_g N_{var} \times N_g N_{var}$) so

it is possible to use this transform to damp the ensemble-derived covariances between different variables and spatial locations. Nonetheless, there is flexibility to still retain the full inter-variable covariances in $\mathbf{L}$ depending on the design of $\mathbf{U}^\alpha$. For the record, a proof of the equivalence between this approach ($\boldsymbol{\chi}^{\alpha k}$ with length $N_g N_{var}$) with full inter-variable covariances retained, and the approach where $\boldsymbol{\chi}^{\alpha k}$ is of length N_g is included in Section 2.6.2. More complex designs of $\mathbf{U}^\alpha$ which allow the retention of inter-variable covariances only between certain model variables, or using different spatial localisation length-scales for different model variables are also possible[2]. In practice, these may be useful in convective-scale data assimilation, particularly when hydrometeor variables are involved (Xuguang Wang, personal communication).

In the ABC-DA system, $\mathbf{U}^\alpha$ is further decomposed into the horizontal $\mathbf{U}^\alpha_{horiz}$ and vertical $\mathbf{U}^\alpha_{vert}$ localisation transforms, similar to the decomposition of $\mathbf{U}$ in Bannister (2020). The series of transforms is given by:

$$\mathbf{U}^\alpha = \mathbf{U}^\alpha_{vert}\mathbf{U}^\alpha_{horiz} \tag{2.16}$$

thus treating the vertical and horizontal localisation separately.

Initial tests constructed $\mathbf{U}^\alpha_{horiz}$ using a Fourier decomposition (as is done in the standard horizontal transform of $\mathbf{U}$ — see Bannister (2020), but this yielded undesirable small negative correlations at longer localisation distances (presumably related to the Gibbs phenomenon, not shown). Thus a different approach was adopted as the basis for populating $\mathbf{U}^\alpha_{horiz}$, using the eigen-decomposition of a target horizontal localisation matrix $\mathbf{L}_{horiz} = \mathbf{U}^\alpha_{horiz}\mathbf{U}^{\alpha\top}_{horiz}$ (where $\mathbf{U}^\alpha_{horiz} = \mathbf{F}^\alpha_{horiz}(\mathbf{\Lambda}^\alpha_{horiz})^{1/2}$, and $\mathbf{F}^\alpha_{horiz}$ and $\mathbf{\Lambda}^\alpha_{horiz}$ are the eigenvectors and eigenvalues respectively of the imposed horizontal localisation matrix). We start by constructing $\mathbf{L}_{horiz}$ using the fifth-order piecewise Gaspari-Cohn function with a horizontal localisation length-scale h^α (equivalent to c, with $a = 1/2$, in Eq. (4.10) of Gaspari and Cohn, 1999). This function is approximately Gaussian over a compact support. As the ABC model uses periodic boundary conditions, $\mathbf{L}_{horiz}$ must be designed to be circulant and account for 'overlapping tails' of the Gaspari-Cohn function when h^α is larger than half the domain size. In the 'overlapping tails' regime, the correlation function does not satisfies the 'space-limited' requirement described in (Gaspari and Cohn, 1999). Thus, the resulting $\mathbf{L}_{horiz}$ is found to not be positive semi-definite when h^α is large and tends to infinity, so the horizontal eigenvectors associated with the negative eigenvalues need to be truncated. Offline

[2]This is explored in Chapter 4.

testing in idealised setups and within the ABC-DA system showed that with the remaining eigenvectors, $\mathbf{U}^{\alpha}_{horiz}\mathbf{U}^{\alpha\top}_{horiz}$ is a good approximation for $\mathbf{L}_{horiz}$. It is also possible to scale the remaining eigenvalues to restore the initial total variances for a better approximation. Figure 2.2 illustrates the implied correlation function ($h^{\alpha} =$ 250 km) with respect to longitudinal gridpoint 50, for an ABC-DA system with 364 longitudinal gridpoints and a 1.5 km horizontal grid, retrieved using the above steps. The original Gaspari-Cohn function with 'overlapping tails' is compared with the implied correlation function reconstructed from the eigenvectors and eigenvalues of the eigen-decomposition of $\mathbf{L}_{horiz}$ (Fig. 2.2a), the resulting implied correlation function when negative eigenvectors/eigenvalues are truncated (Fig. 2.2b), and the resulting implied correlation function after further restoration of the initial total variances by scaling (Fig. 2.2c). Note that in this example, the threshold for which negative eigenvalues appear is $h^{\alpha} \approx$ 138.33 km, found empirically. In the current version of the ABC-DA system, the scaling to restore initial total variances is not implemented yet.

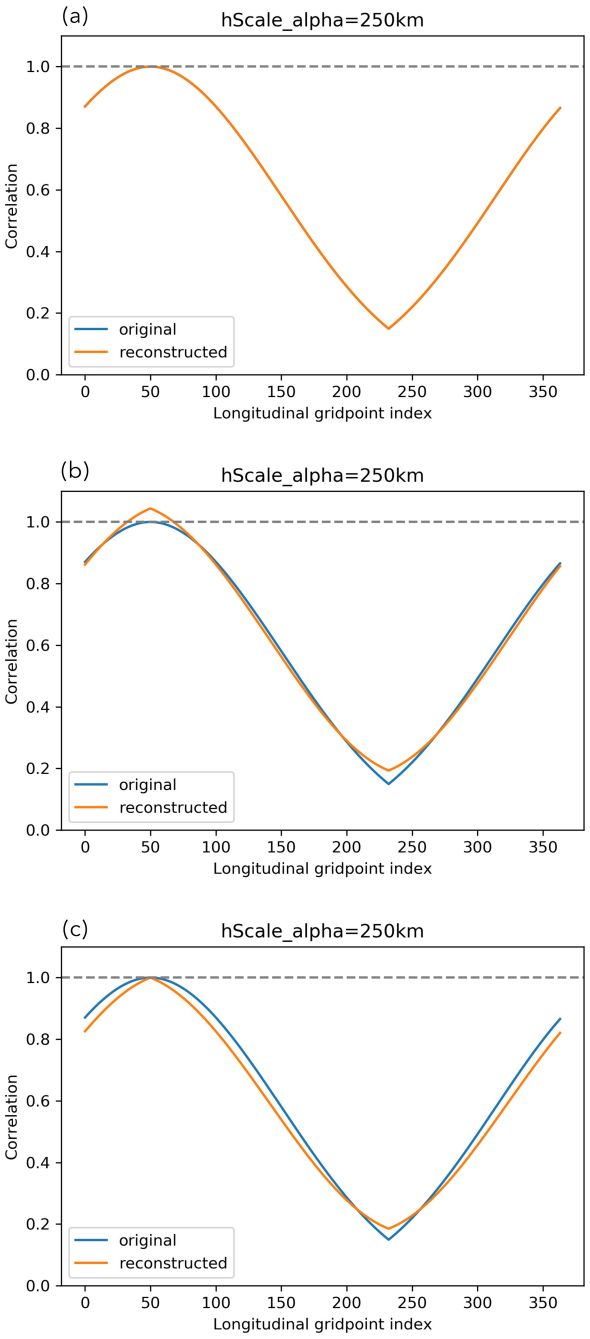

Figure 2.2: Correlation functions ($h^{\alpha} = 250$km) with respect to longitudinal gridpoint 50, for an ABC-DA system with 364 longitudinal gridpoints and 1.5 km horizontal grid. The implied correlation functions (orange) are reconstructed from (a) all eigenvectors and eigenvalues of the eigendecomposition of $\mathbf{L}_{horiz}$, (b) only eigenvectors with non-negative eigenvalues (c) only eigenvectors with non-negative eigenvalues that are scaled to restore initial total variance, and compared with the original Gaspari-Cohn function (blue).

To populate $\mathbf{U}^{\alpha}_{vert}$, a similar approach is adopted; a Gaspari-Cohn function is used with vertical localisation length-scale v^{α}. Note that the target vertical localisation matrix $\mathbf{L}_{vert}$ is a correlation matrix and so must be positive semi-definite, so truncation of eigenvectors is not required. Since $\mathbf{U}^{\alpha}_{horiz}$ and $\mathbf{U}^{\alpha}_{vert}$ are separate, it is possible to have a different $\mathbf{L}_{vert}$ for each horizontal eigenvector. However, for simplicity, the default setup in the ABC-DA system uses the same $\mathbf{L}_{vert}$ for each horizontal eigenvector. As for $\mathbf{L}_{horiz}$ the vertical eigenvectors are retrieved through the eigen-decomposition of $\mathbf{L}_{vert}$ and used to populate $\mathbf{U}^{\alpha}_{vert}$ such that $\mathbf{L}_{vert} = \mathbf{U}^{\alpha}_{vert}\mathbf{U}^{\alpha\top}_{vert}$ (where $\mathbf{U}^{\alpha}_{vert} = \mathbf{F}^{\alpha}_{vert}(\mathbf{\Lambda}^{\alpha}_{vert})^{1/2}$, and $\mathbf{F}^{\alpha}_{vert}$ and $\mathbf{\Lambda}^{\alpha}_{vert}$ are the eigenvectors and eigenvalues respectively of the imposed vertical localisation matrix).

2.3.3 Generation of ABC analysis ensemble

After the initial ensemble has been generated (Section 2.3.1) using the method of Magnusson et al. (2009), and the initial hybrid control analysis has been retrieved (Section 2.3.2), the next step is to generate analysis ensembles (Fig. 2.1, green segments). The ensemble then proceed, via the forecast model (Fig. 2.1, upper right brown segments), as a forecast ensemble which is used in the next hybrid DA step. Various methods have been used in previous studies, such as singular vectors (Buizza et al., 1993), bred vectors (Toth and Kalnay, 1993; 1997), perturbed observations (Houtekamer and Derome, 1995), Ensemble Kalman filter (EnKF; Evensen, 1994), Ensemble Transform Kalman Filter (ETKF; Bishop et al., 2001), and other square-root filters. Here, we focus mainly on the EBV method. This method has useful information about the nature of dynamical error growth about the analysis state at each cycle, but is uninformed about the observation network.

The ensemble, which is run in parallel to the hybrid DA are important components in hybrid-EnVar since they provide the means to compute $\mathbf{X}^{f}_{t}$ in Eq. (2.9). The success of the scheme depends on the extent to which the ensemble forecasts can appropriately represent the background error statistics for the ABC-DA system, so proper design of the ensemble system is critical. We construct the analysis ensemble around the hybrid control analysis (i.e., adding ensemble perturbations to the hybrid control analysis; see below), so the ensemble is 'DA-centred'.

Ensemble bred vectors

In this approach, we consider a variant of the bred vectors method — the EBV method (Balci et al., 2012). The basic bred vectors method (Toth and Kalnay, 1993)

is generally simple to implement and has a cheap computational cost. The idea relies on breeding perturbations by running the non-linear forecast model for a fixed period for pairs of forecast ensemble members, taking the difference between the two forecasts, and then scaling the difference to have a specified and fixed amplitude. This process 'breeds' the fastest growing error modes. This is repeated to retrieve the required number of error modes, and the resulting perturbations are respectively added to the hybrid control analysis to generate an analysis ensemble. The intention is that these perturbations should adequately sample the space of possible analysis errors.

The main difference between the bred vectors and EBV methods lies in the scaling of the perturbations at each cycle. In the bred vectors method, the perturbations are scaled to maintain a fixed amplitude across cycles for each ensemble member. The scaling is independent for each ensemble member and there is therefore no mechanism to compare the dynamics with perturbations of the other members. The EBV method (Balci et al., 2012) on the other hand involves a global scaling factor, which depends on the amplitude of the largest perturbation, and offers better insights into the relative behaviour of nearby ensemble trajectories. Perturbations that have an amplitude smaller than the largest perturbation of the ensemble then play a smaller role after scaling; in other words, ensemble trajectories that are clustered around the control member trajectory are less important for identifying dominant directions of error growth. An in-depth comparison of the bred vectors and EBV methods is provided in Balci et al. (2012).

To generate the analysis ensemble, a target maximum amplitude ϵ_0 is required, but this opens the question on what to choose for ϵ_0. Here, we use $\epsilon_0 = \epsilon^{rf}$ (the mean total energy norm of the initial ensemble of states), although other choices are possible, such as running a series of experiments and finding the average analysis error to estimate the analysis uncertainty. This scaling factor is fixed across perturbations, so at each cycle the perturbations are scaled by the same ratio r_t^{ebv}, which is used and defined as follows:

$$\delta \mathbf{x}_t^{fk} = \frac{1}{\sqrt{2}} r_t^{ebv} (\mathbf{x}_t^{fk} - \mathbf{x}_t^{fc}), \quad r_t^{ebv} = \frac{\epsilon_0}{\mathsf{max}[|\mathbf{x}_t^{fk} - \mathbf{x}_t^{fc}|_{E_{tot}}]}, \tag{2.17a}$$

$$\mathbf{x}_t^{ak} = \mathbf{x}_t^{ac} + \delta \mathbf{x}_t^{fk} \tag{2.17b}$$

where $\mathbf{x}_t^{fk}$ and $\mathbf{x}_t^{fc}$ are the k^{th} ensemble and control forecast from the previous cycle respectively, $\delta \mathbf{x}_t^{fk}$ is the k^{th} scaled ensemble perturbation at time t. The k^{th} member of the analysis ensemble $\mathbf{x}_t^{ak}$ is centred on the hybrid control analysis $\mathbf{x}_t^{ac}$ produced by the DA step (Section 2.3.2). The $\frac{1}{\sqrt{2}}$ factor is not necessarily required because

we are computing differences between individual ensemble member forecasts and the same control member forecast, but we have included it as a deflation factor with our choice of ϵ_0. It is worth noting that the EBV method is not formally consistent with Kalman filter theory, but will not suffer from filter collapse as long as the ϵ_0 chosen is well-tuned. As also discussed in Kalnay et al. (2002), for non-linear atmospheric systems with enough physical space for several independent local instabilities (like the ABC model), bred vectors derived from different initial perturbations remain distinct from each other and do not collapse onto a single leading mode.

It is not uncommon to use such a set-up (i.e., separate hybrid deterministic and ensemble systems for, respectively, the first and second moments of the posterior). While the hybrid control analysis involves both the ensemble and climatological contributions to the background error covariance matrix Eq. (2.15), the computation of analysis perturbations involves only the forecast ensemble and neglects the climatological contributions. While this is a formal discrepancy, we assume that this setup is an adequate from a practical perspective.

2.4 Data assimilation experiments using the hybrid-EnVar scheme in a tropical setting

For this study, the Unified Model output is retrieved from a tropical convective-scale NWP system over the Maritime Continent (SINGV-DA; Heng et al., 2020). SINGV-DA operates on a 1.5 km core horizontal grid, with a model top height of 38.5 km. Longitude-height slices of fields u and v around 2°N are extracted from the SINGV-DA output by placing these fields onto the 1.5 km ABC model grid for the lowest 60 levels (up to around 18 km height), resulting in a 364 $\times$ 60 ABC model grid. These initial u and v fields are then modified to make them compatible with the ABC model's periodic boundary conditions, and the remaining fields, w, $\tilde{\rho}'$, and b' are derived following the procedure in Section 4.1 of Bannister (2020). For the ensemble system, 30 initial ensemble members are generated following Section 2.3.1, excluding the control state reconfigured from the longitude-height slice. This is a typical ensemble size used in operational NWP systems.

To represent a tropical setting of the ABC model, a value of $f = 10^{-5}$ s^{-1} is used. This corresponds approximately to a value of f at a latitude of 4°N in an NWP system. The other model parameters are set as follows: $A = 0.02$ s^{-1}, $B = 0.01$, $C = 10^4$ m^2

s^{-2}. A series of hourly-cycling multi-cycle DA observation system simulation experiments are conducted to demonstrate the incorporation of ensemble-derived background error covariances in hybrid-EnVar DA. The hybrid extension of 3DVar-FGAT may be termed hybrid-En3DVar-FGAT.

We run four experiments with the following configurations:

(a) 100% $\mathbf{B}_c$ (i.e., no flow-dependency, equivalent to 3DVar-FGAT),

(b) 50% $\mathbf{B}_c$, 50% $\mathbf{B}_e$; hybrid-En3DVar-FGAT (i.e., flow-dependency with an equal contribution from $\mathbf{B}_c$ and $\mathbf{B}_e$),

(c) 20% $\mathbf{B}_c$, 80% $\mathbf{B}_e$; hybrid-En3DVar-FGAT (i.e., flow-dependency with most contribution from $\mathbf{B}_e$),

(d) 100% $\mathbf{B}_e$; pure En3DVar-FGAT (i.e., no contribution from $\mathbf{B}_c$).

These experiments are referred to as EBV(a) to EBV(d) accordingly. Note that configuration (a) does not use ensemble information, but the experiment is named as EBV(a) for ease of reference.

2.4.1 Implied background error covariances

To show the workings of Eq. (2.8) (or equivalently Eq. (2.15)) and the localisation, we compute a selection of implied background error covariances, with the various weights assigned to $\mathbf{B}_c$ and $\mathbf{B}_e$ (as the above configurations (a) to (d)). This is similar to performing single observation experiments and retrieving the analysis increments. For $\mathbf{B}_e$, spatial localisation length-scales of $h^\alpha = 100$ km and $v^\alpha = 5$ km are set, and with no inter-variable localisation. The implied background error covariances are valid for the time of the first cycle after a cold start (i.e., at T+0) and use one-hour forecast ensemble perturbations. For $\mathbf{B}_c$, the same ensemble is used as training data to calibrate $\mathbf{U}$.

Figure 2.3 shows the implied background error covariances of $\tilde{\rho}'$, v and b' with respect to a fixed $\tilde{\rho}'$ point in the middle of the domain for the four configurations (four rows). Configuration (a) (top row) shows the implied background error covariances that are modelled purely by $\mathbf{U}$, and configuration (d) (bottom row) shows the purely ensemble-derived covariances with spatial localisation (implied by $\mathbf{U}^\alpha$). Configurations (b) and (c) are linear combinations with different weights, as demonstrated by Eq. (2.8).

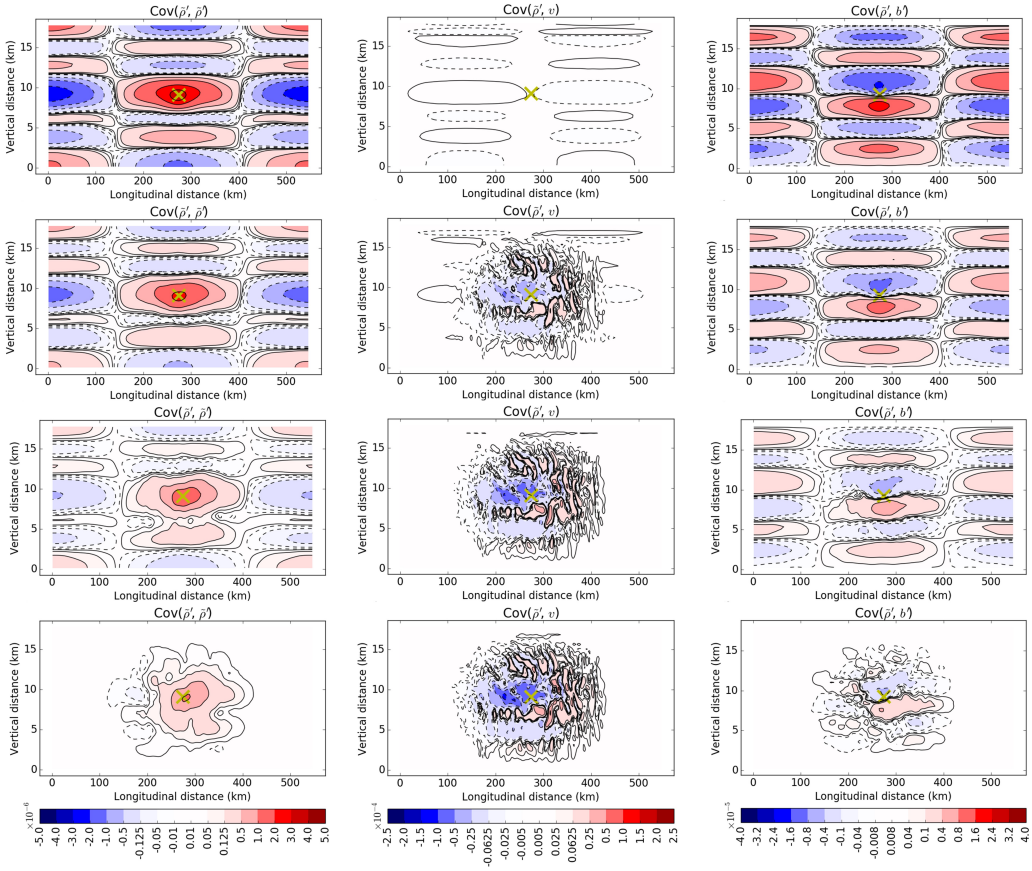

Figure 2.3: Implied background error covariances of $\tilde{\rho}'$ (leftmost column; Cov($\tilde{\rho}'$, $\tilde{\rho}'$)), v (middle column; Cov($\tilde{\rho}'$, v)) and b' (rightmost column; Cov($\tilde{\rho}'$, b')) with respect to a $\tilde{\rho}'$ point (yellow cross) near the centre of the domain for the first cycle after cold start. The rows represent configurations (a), (b), (c), and (d) respectively (see the list near the start of Section 2.4). Negative values have contours that are dashed.

For $\tilde{\rho}'$-$\tilde{\rho}'$ covariances in configuration (a) for pure 3DVar-FGAT, the central region of auto-correlation has horizontal and vertical length-scales of approximately 100 km and 2 km respectively, and is surrounded by oscillations, possibly reflecting the dominant gravity wave propagation. The vertical length-scale here is smaller than that found in the mid-latitude study of Bannister (2020), and such a contrast between low and higher latitudes is seen in other studies, e.g., Ingleby (2001). This can be compared with configuration (d) for pure (and localised) En3DVar-FGAT, which shows a narrower but taller region of auto-correlation. Most of the oscillations are beyond the localisation region so are not visible apart from small negative values to the west of the auto-correlation.

The $\tilde{\rho}'$-v covariances in configuration (a) follow from the use of geostrophic balance, which is manifested in $\mathbf{U}$ (Eq. (2a) in Bannister (2020)). The v pattern is consistent with an anti-cyclonic field around the source point (i.e., positive and negative v covariances west and east of the positive $\tilde{\rho}'$ source point respectively). Since f is small, these covariances are also small. Contrasting this with configuration (d), it appears that the ensemble-derived covariances are more substantial, and are of opposite sign, suggesting that there exists some (other) mass-wind relationship manifested in $\mathbf{B}_e$ (e.g., related to equatorial gravity wave processes not represented in $\mathbf{U}$).

The $\tilde{\rho}'$-b' covariances in configuration (a) follow from the use of hydrostatic balance, again manifested in $\mathbf{U}$ (Eq. (3) in Bannister (2020)). Hydrostatic balance relates b' increments with the vertical gradient of $\tilde{\rho}'$ increments, and the top-left and top-right panels are confirmed to be consistent in this way. In configuration (d), the $\tilde{\rho}'$-b' covariances have similar vertical patterns within the localisation region, although are weaker. As noted in Bannister (2020), the $\mathbf{B}_c$ covariances tend to be larger than their $\mathbf{B}_e$ counterparts even though both use the same training data for calibrating the transforms, but the broad structures are similar.

Note that the implied background error covariances between $\tilde{\rho}'$ and u and w are each zero in configuration (a) by definition of $\mathbf{U}$ (Bannister, 2020). By contrast, in configurations (b), (c) and (d), the implied background error covariances are prescribed directly between the associated model variables, implied by $\mathbf{U}^{\alpha}$ (not shown).

Even though the multivariate background error relationships relevant to the tropics are likely to be different from those at mid-latitudes, the same balance conditions designed for mid-latitudes are often used in $\mathbf{U}$ for tropical settings (as they are here). By

exploring ensemble-derived multivariate background error relationships, we may be able to identify alternative balances inherent in the dynamical fields. This will be explored in a separate study.

2.4.2 Details of observation system simulation experiments

In all experiments, 200 observations of each variable (u, v, w, $\tilde{\rho}'$ and b'), which are equally spaced throughout the domain, are assimilated at every hourly cycle. The observations are sampled from a 'truth' run, with added observation noise following a Gaussian distribution. The observation error standard deviations are chosen to be approximately 10% of the variable's root-mean-square value as seen in the 'truth' run. These are 0.2 m s^{-1}, 0.2 m s^{-1}, 0.01 m s^{-1}, 1.5 $\times$ 10^{-4} and 1.5 $\times$ 10^{-3} m s^{-2} for u, v, w, $\tilde{\rho}'$ and b' respectively. All generated observations are valid at the background/analysis time of each cycle, so there is no difference in the analysis between 3DVar and 3DVar-FGAT (and indeed 4DVar if it were implemented). The number of observations are $\approx 1\%$ of the degrees of freedom of the state (both spatial and multivariate), to mimic how observations are sparse in the tropical setting.

The initial background of the deterministic system is determined from the initial 'truth' plus a small background noise perturbation $\delta\mathbf{x} = \mathbf{U}\delta\boldsymbol{\chi}$, where $\delta\boldsymbol{\chi}$ is drawn randomly from $\mathcal{N}(0, \mathbf{I})$. In order to reduce the effect of random noise on the experiments, the ABC-DA system is first spun-up for 50 one-hour cycles, with the expectation that the DA-centred ensemble system and deterministic system will have lost memory of the particular way that the system was initialised from a cold start. The information from the 50^{th} cycle of spin-up is then used in the first cycle of all the actual experiments.

During the spin-up configuration testing, we noticed that the inclusion of vertical localisation in $\mathbf{B}_e$ was particularly detrimental to the evolution of the w field. Investigation revealed that this was due to the introduction of hydrostatic imbalance in the analysis increments (not shown). A similar well-known issue to do with horizontal localisation introducing geostrophic imbalance was discussed in Section 3c of Lorenc (2003). We include more comments on the hydrostatic imbalance issue in Section 2.6.3. For this reason, vertical localisation was excluded in $\mathbf{B}_e$ in the spin-up process and in all experiments. After inspecting the other fields during spin-up configuration testing, we found that in most configurations, the hybrid control analysis gradually converged around the 'truth' run as the observations were assimilated over the 50 spin-up cycles, which is logically expected. Particularly using the EBV(d) configuration, the evolution of the fields were reasonably in line with the 'truth' run, so this was the

chosen configuration which was run for 50 spin-up cycles, referred to as the spin-up run.

To ensure a fair comparison in the results, the spin-up run provides the same starting background (50^{th} cycle forecast), empirically tuned EBV ensemble and ensemble-derived error modes (if required) for the first cycle of all the actual experiments. Each experiment is run for 50 cycles and only differ in the DA algorithm configurations after spin-up. Where $\mathbf{B}_e$ is required, we use $h^{\alpha} = 20$ km for the horizontal localisation, while not performing any inter-variable localisation (see Section 2.6.2). The horizontal localisation length-scale was determined by comparing with the horizontal distance between adjacent observations ($\approx$ 23 km). For the minimisation, a total of 75 inner loops within a single outer loop is used. This was determined after testing to ensure that sufficient convergence was attained for all cycles in the experiments. This is demonstrated in Fig. 2.4, which shows the minimisation of the cost function for the first cycle of the EBV experiments. For this cycle, the cost function was minimised the fastest and slowest in EBV(d) and (a) respectively. The analysis misfit to assimilated observations was also the largest in EBV(d) with $J_o \approx 1500$ after minimisation. However, it is important to note that this metric is not a particularly useful indicator of analysis quality, but rather how each scheme draws the analysis towards the observations (Wang et al., 2008*b*). We would expect that the minimum of the cost function would approximate half the number of observations (i.e., $J_{min} \approx \frac{N_{obs}}{2} = 500$, the expected value of a chi-squared PDF), so EBV(c) neatly matches our expectations.

In addition to the experiments, a free background run, hereafter referred to as FreeBG, is performed starting from the same 50^{th} cycle forecast of the spin-up run. This is used as the control run to assess if the DA in the experiments is adding value by bringing the deterministic run trajectories closer to the 'truth' or if the trajectories are simply following the natural evolution of the system and neglecting the observational information.

2.4.3 Sensitivity to weighting of $\mathbf{B}_c$ and $\mathbf{B}_e$

Typically for the hybrid-EnVar scheme, tuning of the weights (β_c^2 and β_e^2) for $\mathbf{B}_c$ and $\mathbf{B}_e$ is performed empirically to assess the best configuration which combines the benefits from both sources of background error statistics. Figure 2.5 shows the comparison of domain-averaged analysis errors (root-mean-square errors; RMSE) with respect to the 'truth' for the EBV and FreeBG experiments.

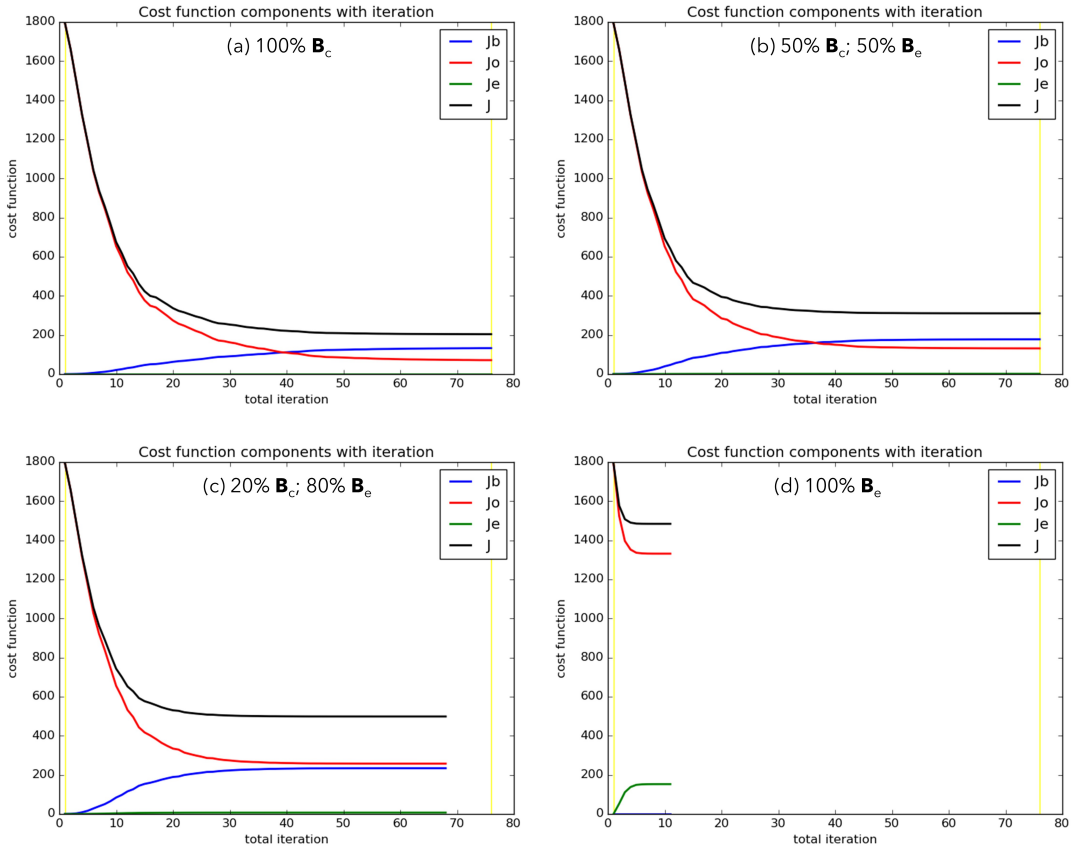

Figure 2.4: Total penalty (black) from the climatological background (blue), ensemble background (green) and observation (red) penalty contributions over the 75 inner loops for the first cycle of the EBV experiments, labelled (a) to (d) accordingly. Early termination of inner loops occurs when convergence criteria is satisfied, in (c) and (d). At convergence, ensemble penalty (green) in (b) and (c) is around 1.5 and 7 respectively.

The cycle-averaged analysis errors (Fig. 2.5, bottom right panel) for all prognostic variables except v are generally smaller for the EBV experiments compared to FreeBG, with an RMSE ratio less than 1. During the simulation, the w, $\tilde{\rho}'$ and b' errors were decreasing, suggesting that the deterministic run trajectories of the EBV experiments were converging around the 'truth' because of the availability of observational information. The u analysis errors were decreasing in EBV(c) and EBV(d), but were increasing in EBV(a) and EBV(b). Throughout the 50 cycles, the v analysis errors were generally increasing in the EBV experiments. This peculiar issue was exacerbated when the weighting towards $\mathbf{B}_c$ was increased, suggesting that the issue originates from $\mathbf{B}_c$. A feature of the u, $\tilde{\rho}'$, and b' RMSE time sequences is the eight-hour periodicity, which is also apparent in the basic dynamical root-mean-square fields. A normal mode analysis (not shown) and inspection of the basic dynamical fields reveals that there is a 16-hour period (local maxima to local maxima), which is within the period range of low-zonal-wavenumber gravity waves, suggesting that this feature is due to the dominant gravity waves in this system.

To test if the issue with the v analysis errors was due to the choice of training data, we repeated EBV(a) but with $\mathbf{B}_c$ calibrated using other training data (e.g., the ensemble perturbations from the 50^{th} cycle of the spin-up run instead of those from the initial forecast ensemble). Even with more time-appropriate training data (but same variances), the issue was only partially resolved (smaller increase in RMSE; not shown). Also, this issue does not appear to occur in mid-latitudinal experimental setups in Bannister (2021). From the implied background error covariances in Section 2.4.1, there exists some mass-wind relationship in $\mathbf{B}_e$ that is not well-represented in $\mathbf{B}_c$ through the geostrophic balance relationship since f is small in the tropical setting. Repeating EBV(a), but omitting the geostrophic balance constraint entirely in the calibration of $\mathbf{B}_c$ (i.e., treating $\tilde{\rho}'$ and v background errors univariately) also did not resolve the issue (not shown). We speculate that the issue could be due to the absence of a suitable balance constraint for prescribing the mass-wind relationship for $\mathbf{B}_c$, or a likely lack of tuning of the variances for $\mathbf{B}_c$. Early results with a tuned $\mathbf{B}_c$ showed that the issue with the v analysis errors could be resolved by reducing the variances of all variables substantially. The warrants further investigation in a separate study, to tune the variances for the system or assess the possibility of deriving a balance relationship between v and $\tilde{\rho}'$ for the tropical setting.

Comparing between the EBV experiments, the u analysis errors and v analysis errors are generally the smallest in EBV(d), indicating that allocating full weight to $\mathbf{B}_e$ in this

setup is ideal for minimising the horizontal wind-related analysis errors. The w, $\tilde{\rho}'$ and b' analysis errors are arguably the smallest in EBV(c), with the smallest cycle-averaged RMSE. The results presented here are not unsurprising given that previous studies evaluating hybrid-EnVar DA in simplified models (e.g., Hamill and Snyder, 2000) and NWP systems (e.g., Montmerle et al., 2018; Bédard et al., 2020) also show that the best configuration appears to rely on a combination of both $\mathbf{B}_e$ and $\mathbf{B}_c$, and not solely one or the other.

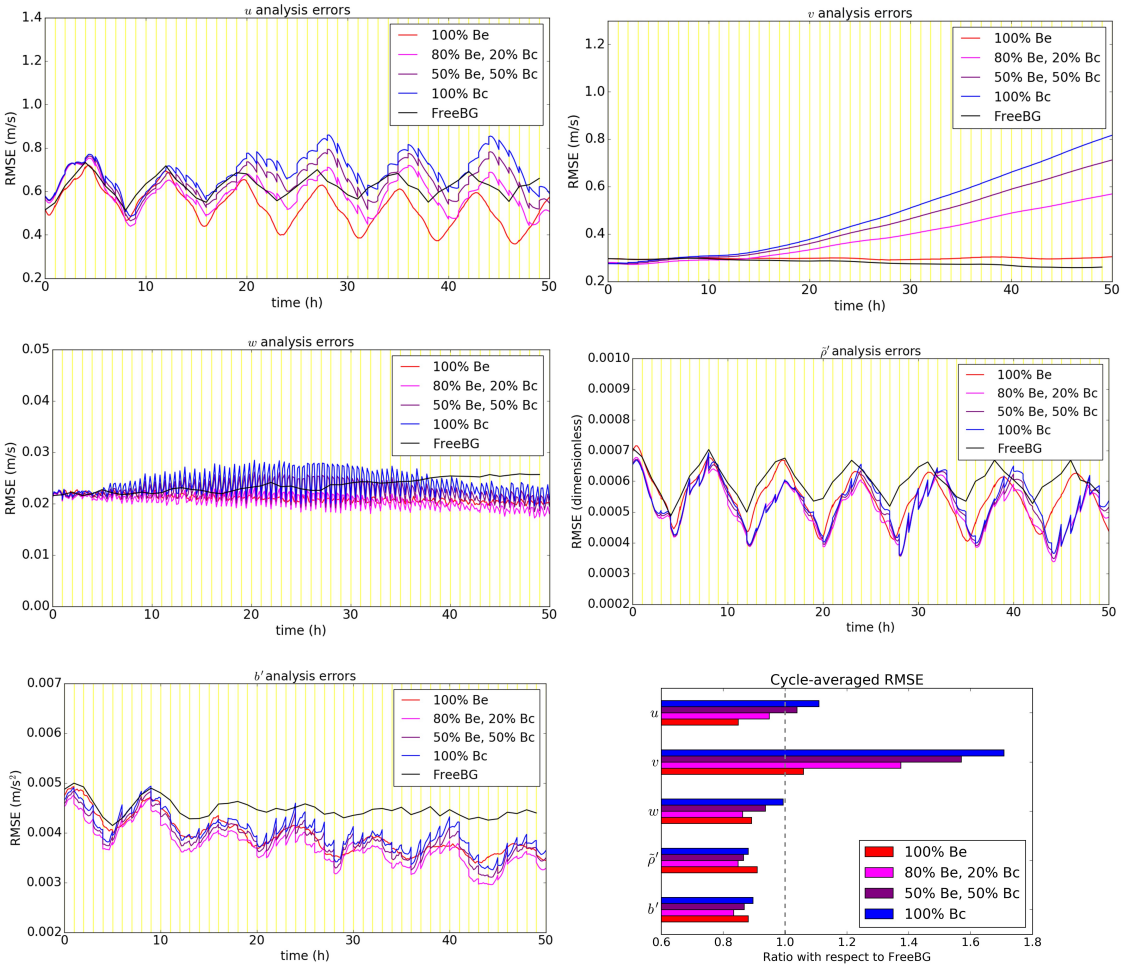

Figure 2.5: All panels except bottom right: time series of root-mean-square analysis errors for the EBV experiments (100% $\mathbf{B}_c$, configuration (a); 50% $\mathbf{B}_e$, 50% $\mathbf{B}_c$, (b); 80% $\mathbf{B}_e$, 20% $\mathbf{B}_c$, (c); 100% $\mathbf{B}_e$, (d)) and the free background run (FreeBG). The vertical yellow lines are the analysis times. Analysis errors are defined with respect to the 'truth' run, computed every 10 minutes within the respective assimilation windows for EBV experiments and every hour for FreeBG. Bottom right: the ratio of the cycle-averaged RMSE of the EBV experiments with respect to FreeBG for the five ABC model variables.

2.4.4 Ensemble trajectories and spread-error relationship

We can better appreciate the robustness of the ensemble by plotting the trajectories of the ensemble, its mean, the FreeBG, and the 'truth' (Fig. 2.6). To avoid over-smoothing the local spatial variations in the fields, the trajectories are computed by taking a gridpoint-averaged value of the fields for a subset of the full domain; a box located at the centre of the domain (model levels 25 to 35, longitudinal gridpoints 127 to 237). We have also investigated the trajectories using other subsets (boxes) distributed around the domain, but the main ideas are the same so we have excluded discussion on them.

In Fig. 2.6, the spread of the EBV(d) ensemble is centred around the ensemble mean throughout the 50 cycles. The 'truth' trajectory is also generally contained within the spread of the ensemble, particularly for u, $\tilde{\rho}'$ and b'. There were no evidence of filter collapse, nor bimodalities, which indicate that the DA-centred ensemble generated using the EBV method is healthy.

It is common practice to also compare the ensemble spread with the RMSE, which for a perfectly reliable large ensemble where observation density and errors are accounted for, the two quantities should be approximately the same (Leutbecher, 2009). In the EBV method, the ensemble spread is largely dependent on the choice of ϵ_0 since it does not account for the observation network, but this method is still worth considering as a 'control method' and comparing with future methods consistent with Kalman filter theory. In the computation of the ensemble spread, Fortin et al. (2014) also cautions against using the wrong metric. Following their recommended approach, we define the gridpoint-averaged ensemble spread $\bar{S}_t$ using the square-root gridpoint-averaged ensemble variance:

$$\bar{S}_t = \sqrt{\frac{1}{N_g}\sum_{i=1}^{N_g} S^2[i,t]} \tag{2.18}$$

where the ensemble spread S is computed using Eq. (4) of Whitaker and Loughe (1998). N_g in this case refers to the number of gridpoints over which the average is taken (i.e., the points within the same box used for Fig. 2.6), and i is the gridpoint index which represents points in the box. The RMSE is computed as before, except now over points within this box.

Figure 2.7 shows $\bar{S}_t$ for each model quantity in the EBV(d) ensemble. These are benchmarked against the RMSE and the (time-stationary) implied background error

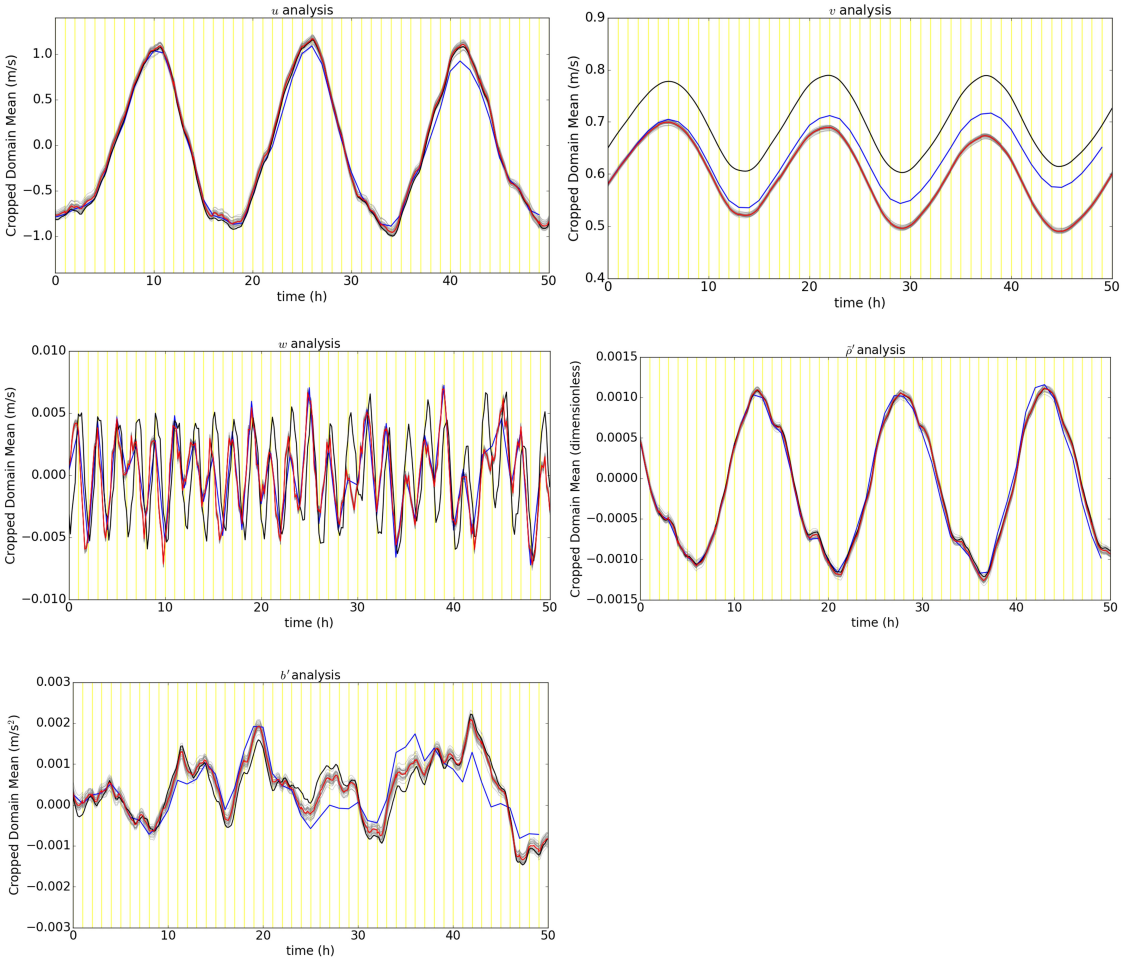

Figure 2.6: EBV(d) (100% $\mathbf{B}_e$) ensemble trajectories derived from gridpoint-averaged analysis fields and their forecasts over a subset of the full domain (a box located at the centre of the domain, model levels 25 to 35, longitudinal gridpoints 127 to 237). The corresponding ensemble mean (red), free background (blue) and 'truth' (black) trajectories for the same subset domain are plotted alongside the individual ensemble member (grey) trajectories. Values for the free background are indicated every hour, and every 10 minutes for the other trajectories.

standard deviations at model level 30 of $\mathbf{B}_c$, which are also plotted. For u and $\tilde{\rho}'$, the ensemble spread approximately matches the RMSE, particularly for later cycles as the hybrid control analysis converges around the 'truth'. For v, w and b', the ensemble is clearly under-dispersive. For all variables, the ensemble spread is also much smaller than the corresponding implied background error standard deviation at model level 30 of $\mathbf{B}_c$. This strongly suggests that the issue with v analysis errors highlighted in the previous section is due to lack of tuning of the variances of $\mathbf{B}_c$, which depends on the specific data assimilation setup. Note that the ensemble spread is computed with respect to the ensemble mean (Whitaker and Loughe, 1998), but the RMSE is computed between the hybrid control analysis (a surrogate to the ensemble mean) and the 'truth'. The spread-error relationship from this setup suggests that the DA-centring did not result in major statistical inconsistencies with the EBV ensemble. While the spread-error relationship is a useful diagnostic, it is not so straightforward to directly relate the ensemble spread to the eventual skill of the hybrid-EnVar DA system. Hence, it is not easy to determine whether to further inflate or deflate the EBV analysis perturbations by considering other choices of ϵ_0.

2.4.5 Sensitivity to number of ensemble members

As mentioned in Section 2.3.2, having a finite number of ensemble members will lead to sampling error in $\mathbf{P}_e^f[t]$. Logically, decreasing the number of ensemble members N used to compute $\mathbf{P}_e^f[t]$ should result in larger sampling errors. For a fixed $\mathbf{L}$ (as in Section 2.4.2), we demonstrate the sensitivity of the the skill of the hybrid-EnVar DA system to N in the ABC-DA system. We perform two additional experiments as variants of EBV(d) to maximise the impact of the ensemble size changes. The experiments follow the same configuration as EBV(d), but with only 20 and 10 ensemble members in the ensemble instead, referred to as EBV(d20) and EBV(d10) respectively, instead of the 30 members used until now.

Figure 2.8 shows the comparison of cycle-averaged analysis errors as N is varied. The RMSE is smallest in EBV(d) for almost all prognostic variables. Reducing the ensemble size from 30 to 20 in EBV(d20) leads to an increase in the RMSE, indicating poorer performance of the ABC-DA system. A further reduction of the ensemble size to 10 in EBV(d10) leads to the poorest performance overall. In this simple setup, these results are expected following the above argument that larger sampling errors are introduced into the system when N is smaller. For $\tilde{\rho}'$ and b', the RMSE ratio in EBV(d10) is even larger than in EBV(a), indicating that the pure EnVar setup may perform poorer than its 3DVar-FGAT counterpart when the ensemble size is too small

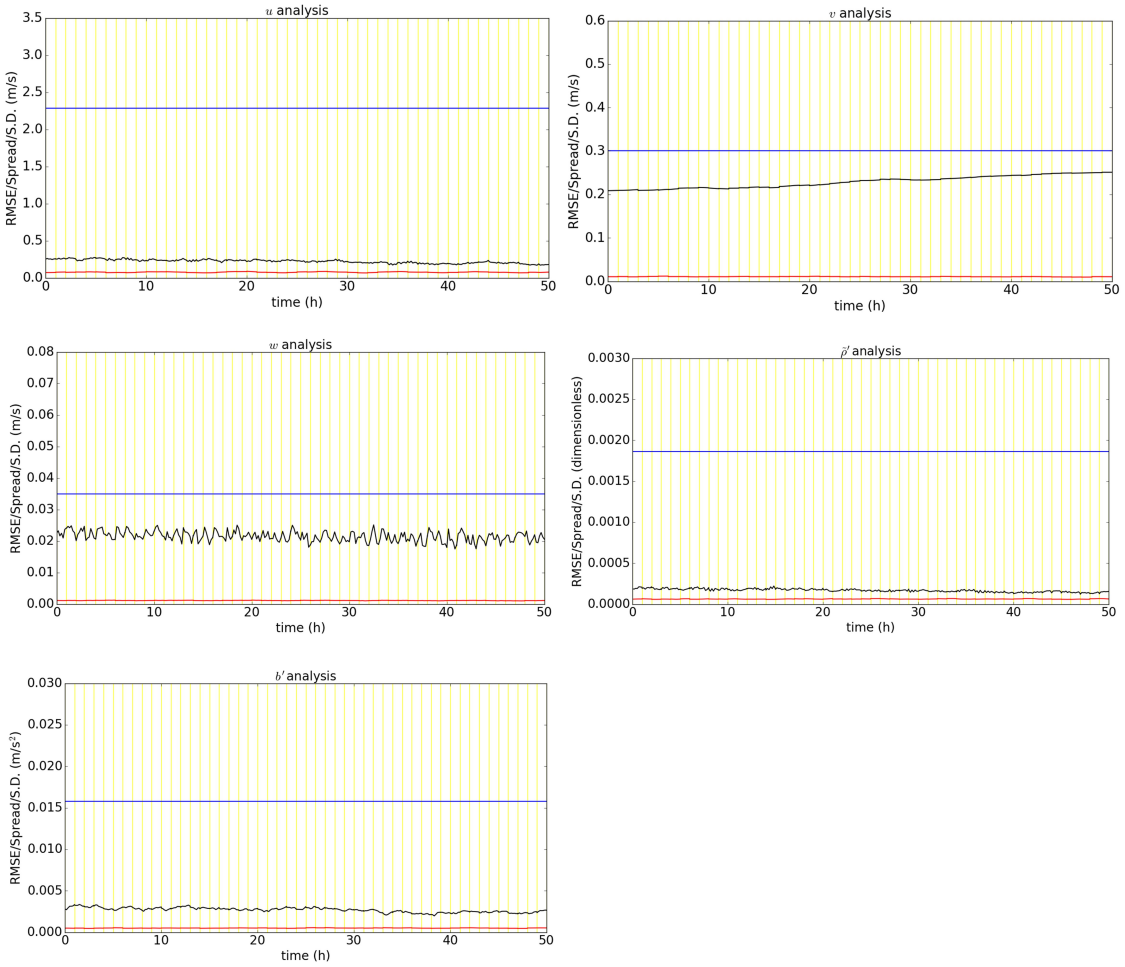

Figure 2.7: Time series of root-mean-square analysis errors (RMSE; black) and ensemble spread (Spread; red) for the EBV(d) (100% $\mathbf{B}_e$) ensemble, computed over a subset of the domain (a box located at the centre of the domain, model levels 25 to 35, longitudinal gridpoints 127 to 237). The implied (time-stationary) background error standard deviation at model level 30 is also included (S.D.; blue).

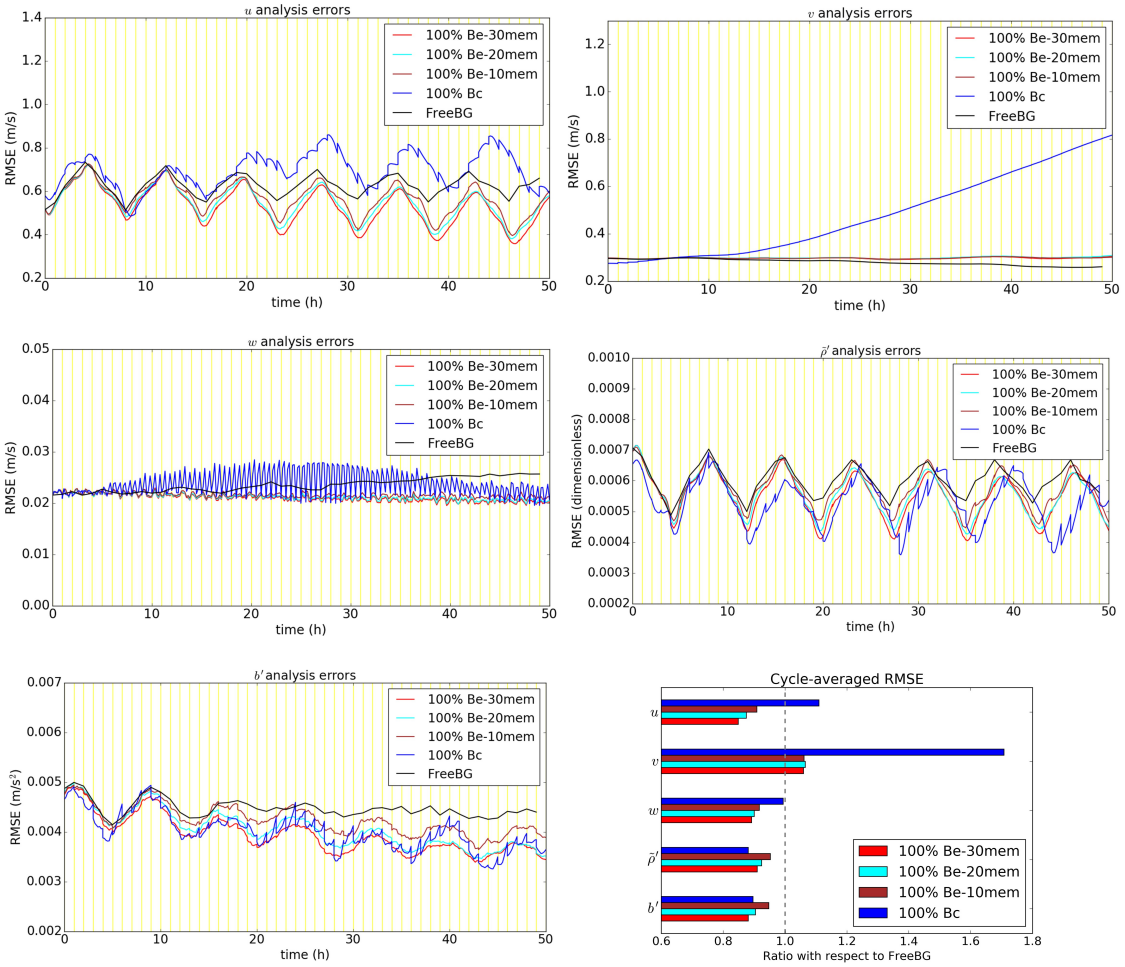

Figure 2.8: As in Fig. 2.5, but for EBV(d), EBV(d20) and EBV(d10) experiments (100% $\mathbf{B}_e$ with 30, 20, and 10 ensemble members respectively).

(Fig. 2.8; bottom right panel).

It is important to highlight that the results are specific to this ABC-DA setup where the localisation length-scales are kept fixed (and are arguably quite tight) across the experiments. For other setups where the localisation length-scales are broader, the optimal ensemble size would be expected to be larger. It would also be worth exploring if a further increase in the ensemble size by orders of magnitude would yield a 'saturation point' where there is little additional benefit to the system. However, one should also be aware of ensemble clustering (Amezcua et al., 2012) in very large ensembles. This issue has been shown to negatively impact hybrid-EnVar DA in simpler models such as the three variable Lorenz-63 model (Goodliff et al., 2015). In the case of the much larger ABC-DA system though, it is unlikely that N could practically be made large enough relative to n to be exposed to this handicap.

2.5 Summary

In this article, we document the development of the hybrid ensemble-variational data assimilation system for the ABC model (Petrie et al., 2017), built on the existing variational ABC-DA system (Bannister, 2020). The hybrid ensemble-variational algorithm that is introduced is based on the alpha control variable approach of Lorenc (2003). Key details related to the spatial and inter-variable localisation are discussed; the approach coded in the ABC-DA system allows flexibility in the localisation, for use in future exploratory studies. The hybrid ensemble-variational algorithm requires an ensemble system that is run parallel to the deterministic components to provide the flow-dependent error modes. To achieve this, the random field perturbations method is introduced in the ABC model for generating an initial ensemble. The ensemble bred vectors (EBV) method is also introduced in the ABC-DA system to propagate the ensemble, which is centred on the hybrid control analysis at each cycle.

Using a tropical setting of the ABC model, we test both ensemble propagation methods (30-member ensemble) in a series of hourly-cycling multi-cycle data assimilation observation system simulation experiments with hybrid ensemble-variational data assimilation. In the experiments, 3DVar-FGAT (First Guess at Appropriate Time) is employed together with EBV using different weightings assigned to the implied climatological (or static) background error covariance matrix ($\mathbf{B}_c$) and the implied ensemble-derived background error covariance matrix ($\mathbf{B}_e$); (a) 100% $\mathbf{B}_c$ (i.e., no flow-dependency, equivalent to 3DVar-FGAT), (b) 50% $\mathbf{B}_c$, 50% $\mathbf{B}_e$; hybrid-En3DVar-FGAT

(i.e., flow-dependency with an equal contribution from $\mathbf{B}_c$ and $\mathbf{B}_e$), (c) 20% $\mathbf{B}_c$, 80% $\mathbf{B}_e$; hybrid-En3DVar-FGAT (i.e., flow-dependency with most contribution from $\mathbf{B}_e$), and (d) 100% $\mathbf{B}_e$; pure En3DVar-FGAT (i.e., no contribution from $\mathbf{B}_c$).

The cycle-averaged analysis root-mean-square errors with respect to the 'truth' for all prognostic variables except v were generally smaller for the EBV experiments compared to the free background. All experiments that involved the ensemble outperformed pure $\mathbf{B}_c$ for all variables. EBV(c) was the best performing configuration for w, $\tilde{\rho}'$ and b', while EBV(d) was the best performing configuration for u and v. We also noted that the v field gradually diverged from the 'truth' during the simulations for experiments involving $\mathbf{B}_c$, even though fields of other variables were converging around the 'truth' as logically expected. Through further assessment of the implied background error covariances and sensitivity tests, it was found that for the tropical setting of the ABC model, there exists some mass-wind relationship that is captured in $\mathbf{B}_e$ which is not well-represented by the (weak) geostrophic balance constraint in $\mathbf{B}_c$. We speculate that the issue with v for configurations that involve $\mathbf{B}_c$ could be due to the absence of a suitable balance constraint for prescribing the mass-wind relationship which may exist in the tropical setting of the ABC model, warranting further investigation in a separate study since it is not trivial to derive one. The results demonstrate the advantages of employing hybrid ensemble-variational data assimilation in the ABC-DA system over traditional variational data assimilation.

An inspection of the EBV(d) ensemble trajectories showed that the ensemble was centred around the ensemble mean throughout the experiment, with the 'truth' trajectory generally contained within the spread of the ensemble. For v, w and b', the EBV ensemble was under-dispersive, but for u and $\tilde{\rho}'$, the ensemble spread approximately matched the corresponding RMSE. The EBV ensemble did not exhibit bimodalities or evidence of filter collapse, indicating that the DA-centred ensemble generated was healthy.

To illustrate the sensitivity to ensemble members, we performed two additional experiments as variants of EBV(d); EBV(d20) with 20 ensemble members and EBV(d10) with 10 ensemble members. The cycle-averaged analysis errors for almost all prognostic variables were smallest in EBV(d). Reducing the ensemble size from 30 to 20, and subsequently to 10 led to an increase in the RMSE, indicating poorer performance of the ABC-DA system. The results in this simple setup are consistent with the expectation that larger sampling errors are introduced into the system with a

smaller ensemble, thus resulting in larger RMSE.

During the testing and development of the hybrid ensemble-variational method, localisation-related issues like hydrostatic imbalance in the analysis increments also became apparent. Similar issues have been documented in previous studies, but we have included additional comments in this article. Given the rapid adoption and broad shift towards hybrid ensemble-variational methods in convective-scale numerical weather prediction, we hope that the ABC-DA system can prove useful in providing further insights and highlight other potential issues that may arise in such methods. Particularly for the tropics, further work is required to better understand the characteristics of the ensemble-derived background errors, such as disentangling its flow-dependency or designing the localisation to isolate or identify important multivariate relationships.

2.6 Supporting information

2.6.1 Details on the random field perturbations method

From Section 2.3.1, the random field perturbations method is used to generate the initial ensemble states for the ABC ensemble system. Equation (2.7a) describes the implementation where pairs of states are randomly chosen from a long 'truth' run.

In Magnusson et al. (2009), there are additional constraints placed on the choice of random fields. The dates must be from different years and must be from the same season in order to eliminate inter-annual correlations in the perturbations yet preserve the seasonal characteristics of the variability. In the ABC model, we have attempted to capture the essence of these constraints even though there are no seasons in the ABC model.

For the experiments, the long 'truth' run is generated with the initial control state $\mathbf{x}^c_{cs}$ as the initial condition, and is run for 50 days. ABC model dumps are produced every hour, resulting in a total of 1200 state dumps. A minimum threshold of 100 hours is set between the validity time of each random pair of states, for which they are assumed to be uncorrelated. In other words, pairs of states are selected randomly and are retained only if they are valid at least 100 hours apart. Additionally, Magnusson et al. (2009) did not indicate if the dates can be repeatedly selected (i.e., selection from a pool with replacement), so we have not imposed the additional constraint of selection from a pool without replacement. For the experiments, a total of 1200

state dumps is sufficiently large compared to the number of pairs required (number of ensemble members, 30 in most of this work).

One aspect that was highlighted in the implementation was the choice of fixed perturbation amplitude to scale the random field perturbations. It is not possible to follow the exact approach of Magnusson et al. (2009), using the average analysis error statistics, in the ABC model. However, we use the same metric (total energy norm) to gauge the initial fixed perturbation amplitude. As described in Section 2.3.1, the random field perturbations are scaled towards their mean total energy norm. This approach ensures that the random fields perturbations have the same fixed perturbation amplitude, but differ in directions of error growth. Further testing with the ABC model showed that reducing the fixed perturbation amplitude yielded smaller errors in the experiments, so a deflation factor of 5 was eventually adopted.

The total energy norm (E_{tot}) for the random field perturbations are computed using:

$$E_{tot} = E_k + E_b + E_e \tag{2.19a}$$

$$E_k = \int \frac{\tilde{\rho}(u^2 + v^2 + w^2)}{2} \rho_0 \, \mathrm{d}V \tag{2.19b}$$

$$E_b = \int \frac{\tilde{\rho} b'^2}{2A^2} \rho_0 \, \mathrm{d}V \tag{2.19c}$$

$$E_e = \int \frac{C\tilde{\rho}'^2}{2B} \rho_0 \, \mathrm{d}V \tag{2.19d}$$

where E_k, E_b, E_e are the kinetic, buoyant and elastic energy respectively, $\rho_0 = 1.225$ kg m^{-3} is a reference air density, and $\mathrm{d}V$ is the volume of a gridbox in the ABC model. Note that as mentioned in Section 2.3.3, we also use E_{tot} in the ensemble bred vectors method to scale the ensemble perturbations for subsequent cycles. Prior to the experiments, we performed some initial testing using the inner product norm instead of E_{tot}, which yielded similar results between the two norm choices. Since E_{tot} is a metric that is physically meaningful, it was eventually used for the scaling of the random field perturbations and ensemble perturbations from the ensemble bred vectors method for the experiments.

2.6.2 Accounting for inter-variable covariances — proof of equivalence of two approaches

As highlighted in Section 2.3.2, $\mathbf{L}$ (the localisation matrix) can be partitioned into a matrix $\mathbf{U}^{\alpha}$:

$$\mathbf{L} = \mathbf{U}^{\alpha}\mathbf{U}^{\alpha\top} \tag{2.20}$$

We seek to prove that two approaches used to code $\boldsymbol{\chi}^{\alpha k}$ and $\mathbf{U}^{\alpha}$ (described below) give the same result when $\mathbf{L}$ is applied on the same model space vector $\mathbf{v}$ (of length $N_g N_{var}$). Note that $\mathbf{L}$ is $N_g N_{var} \times N_g N_{var}$, which in principle means that the inter-variable localisation matrix can be set to have any correlation structure, including the limiting cases of full localisation between different variables (where the corresponding matrix elements are 0), and no localisation (matrix elements are 1). Recall that $\mathbf{L}$ is used in the DA via a Schur product with $\mathbf{P}_e^f[t]$ (Eq. (2.11)). While the number of rows in $\mathbf{U}^{\alpha}$ is constrained to be $N_g N_{var}$, the number of columns can be chosen. The fewer the columns, the smaller the corresponding size of the $\boldsymbol{\chi}^{\alpha k}$ vectors (Eq. (2.14b)), but the less flexible the implied localisation matrix. The first approach considered is based on Wang et al. (2008*a*) (N_g columns) and the second approach is coded in the ABC-DA system ($N_g N_{var}$ columns), inspired by Bannister (2017). The first approach requires less memory and computation, but has less flexibility than the second approach in terms of multivariate localisation choices.

For simplicity, the proof is demonstrated using pure EnVar with one non-zero element in $\mathbf{v}$. This is a similar procedure to computing a column of the implied $\mathbf{B}_c$ or $\mathbf{B}_e$, but now the implied $\mathbf{L}$ is being probed. It is easier to visualise the interactions of the matrix elements by partitioning $\mathbf{v}$ into segments of size $N_g \times 1$ based on the ABC prognostic variables, i.e., $\mathbf{v} = (\mathbf{v}_u, \mathbf{v}_v, \mathbf{v}_w, \mathbf{v}_{\tilde{\rho}'}, \mathbf{v}_{b'})^{\top}$. Similarly, we can consider blocks, each of size $N_g \times N_g$, used to construct $\mathbf{U}^{\alpha}$ (i.e., $\mathbf{U}^{\alpha}_u, \mathbf{U}^{\alpha}_v, \mathbf{U}^{\alpha}_w, \mathbf{U}^{\alpha}_{\tilde{\rho}'}$, and $\mathbf{U}^{\alpha}_{b'}$), which will determine the spatial localisations (horizontal and vertical) for each variable.

The main difference between the two approaches is in the design of $\mathbf{U}^{\alpha}$. In the first approach, based on Wang et al. (2008*a*), $\mathbf{U}^{\alpha}$ (denoted $\tilde{\mathbf{U}}^{\alpha}$) is rectangular ($N_g N_{var} \times N_g$,

and $\boldsymbol{\chi}^{\alpha k}$ has N_g elements), given by:

$$\tilde{\mathbf{U}}^{\alpha} = \begin{bmatrix} \mathbf{U}_u^{\alpha} \\ \mathbf{U}_v^{\alpha} \\ \mathbf{U}_w^{\alpha} \\ \mathbf{U}_{\tilde{\rho}'}^{\alpha} \\ \mathbf{U}_{b'}^{\alpha} \end{bmatrix}. \tag{2.21}$$

Applying $\mathbf{L}$ (denoted $\tilde{\mathbf{L}}$ for first approach) to $\mathbf{v}$ yields:

$$\tilde{\mathbf{L}}\mathbf{v} = \tilde{\mathbf{U}}^{\alpha}\tilde{\mathbf{U}}^{\alpha\top}\mathbf{v} = \begin{bmatrix} \mathbf{U}_u^{\alpha} \\ \mathbf{U}_v^{\alpha} \\ \mathbf{U}_w^{\alpha} \\ \mathbf{U}_{\tilde{\rho}'}^{\alpha} \\ \mathbf{U}_{b'}^{\alpha} \end{bmatrix} \begin{bmatrix} \mathbf{U}_u^{\alpha\top} & \mathbf{U}_v^{\alpha\top} & \mathbf{U}_w^{\alpha\top} & \mathbf{U}_{\tilde{\rho}'}^{\alpha\top} & \mathbf{U}_{b'}^{\alpha\top} \end{bmatrix} \mathbf{v} \tag{2.22a}$$

with the elements given by:

$$(\tilde{\mathbf{L}}\mathbf{v})_i = \sum_{j=1}^{N_g} \left\{ (\tilde{\mathbf{U}}^{\alpha})_{i,j} \sum_{i'=1}^{N_g N_{var}} (\tilde{\mathbf{U}}^{\alpha\top})_{j,i'} \mathbf{v}_{i'} \right\} \tag{2.22b}$$

If there is only one non-zero element (the q^{th} element of $\mathbf{v}$), this simplifies to:

$$(\tilde{\mathbf{L}}\mathbf{v})_i = \sum_{j=1}^{N_g} (\tilde{\mathbf{U}}^{\alpha})_{i,j} (\tilde{\mathbf{U}}^{\alpha\top})_{j,q} \mathbf{v}_q = \sum_{j=1}^{N_g} (\tilde{\mathbf{U}}^{\alpha})_{i,j} (\tilde{\mathbf{U}}^{\alpha})_{q,j} \mathbf{v}_q \tag{2.22c}$$

In the second approach, $\mathbf{U}^{\alpha}$ (denoted $\hat{\mathbf{U}}^{\alpha}$) is square ($N_g N_{var} \times N_g N_{var}$ and $\boldsymbol{\chi}^{\alpha k}$ has $N_g N_{var}$ elements), given by:

$$\hat{\mathbf{U}}^{\alpha} = \begin{bmatrix} \mathbf{U}_u^{\alpha} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_v^{\alpha} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{U}_w^{\alpha} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{U}_{\tilde{\rho}'}^{\alpha} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{U}_{b'}^{\alpha} \end{bmatrix} \tag{2.23}$$

where $\mathbf{0}$ is a $N_g \times N_g$ block containing zeroes. This is the default configuration that is coded in the ABC-DA system, which gives an implied $\mathbf{L}$ (denoted $\hat{\mathbf{L}}$ for the second

approach):

$$\hat{\mathbf{L}} = \hat{\mathbf{U}}^{\alpha}\hat{\mathbf{U}}^{\alpha\top} = \begin{bmatrix} \mathbf{U}_u^{\alpha}\mathbf{U}_u^{\alpha\top} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_v^{\alpha}\mathbf{U}_v^{\alpha\top} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{U}_w^{\alpha}\mathbf{U}_w^{\alpha\top} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{U}_{\tilde{\rho}'}^{\alpha}\mathbf{U}_{\tilde{\rho}'}^{\alpha\top} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{U}_{b'}^{\alpha}\mathbf{U}_{b'}^{\alpha\top} \end{bmatrix} \tag{2.24}$$

Notice that here, $\hat{\mathbf{L}}$ does a full inter-variable localisation, so that the Schur product of $\hat{\mathbf{L}}$ with $\mathbf{P}_e^f[t]$ will not retain any inter-variable covariances. This may be useful if N is small and sampling noise is problematic in $\mathbf{P}_e^f[t]$.

Next, we introduce a mapping matrix $\hat{\mathbf{I}}$, which consists of $N_p \times N_p$ blocks of identity matrices ($\mathbf{I}_{N_g}$, each of size $N_g \times N_g$):

$$\hat{\mathbf{I}} = \frac{1}{\sqrt{N_p}} \begin{bmatrix} \mathbf{I}_{N_g} & \mathbf{I}_{N_g} & \mathbf{I}_{N_g} & \mathbf{I}_{N_g} & \mathbf{I}_{N_g} \\ \mathbf{I}_{N_g} & \mathbf{I}_{N_g} & \mathbf{I}_{N_g} & \mathbf{I}_{N_g} & \mathbf{I}_{N_g} \\ \mathbf{I}_{N_g} & \mathbf{I}_{N_g} & \mathbf{I}_{N_g} & \mathbf{I}_{N_g} & \mathbf{I}_{N_g} \\ \mathbf{I}_{N_g} & \mathbf{I}_{N_g} & \mathbf{I}_{N_g} & \mathbf{I}_{N_g} & \mathbf{I}_{N_g} \\ \mathbf{I}_{N_g} & \mathbf{I}_{N_g} & \mathbf{I}_{N_g} & \mathbf{I}_{N_g} & \mathbf{I}_{N_g} \end{bmatrix}, \tag{2.25}$$

where N_p is the number of model variables whose inter-variable covariances are retained by the mapping matrix (i.e., $N_p = N_{var} = 5$ in the above). Note that other designs of $\hat{\mathbf{I}}$ (e.g., replacing some blocks with $\mathbf{0}$) will allow only the desired retention of specific covariances between certain model variables.

Using the second approach of coding $\mathbf{U}^{\alpha}$ and $\boldsymbol{\chi}^{\alpha k}$, it is possible to retain the full inter-variable covariances and achieve the exact same outcome as the first approach by defining $\mathbf{U}^{\alpha} = \hat{\mathbf{U}}^{\alpha}\hat{\mathbf{I}}$. The implied localisation matrix is thus $\mathbf{L} = \frac{1}{N_p}\hat{\mathbf{U}}^{\alpha}\hat{\mathbf{I}}\hat{\mathbf{I}}\hat{\mathbf{U}}^{\alpha\top}$. As

before, applying $\mathbf{L}$ to $\mathbf{v}$ yields:

$$\mathbf{L}\mathbf{v} = \frac{1}{N_p}\hat{\mathbf{U}}^{\alpha}\hat{\mathbf{I}}\hat{\mathbf{I}}\hat{\mathbf{U}}^{\alpha\top}\mathbf{v}$$

$$= \frac{1}{N_p}\begin{bmatrix} \mathbf{U}_u^{\alpha} & \mathbf{U}_u^{\alpha} & \mathbf{U}_u^{\alpha} & \mathbf{U}_u^{\alpha} & \mathbf{U}_u^{\alpha} \\ \mathbf{U}_v^{\alpha} & \mathbf{U}_v^{\alpha} & \mathbf{U}_v^{\alpha} & \mathbf{U}_v^{\alpha} & \mathbf{U}_v^{\alpha} \\ \mathbf{U}_w^{\alpha} & \mathbf{U}_w^{\alpha} & \mathbf{U}_w^{\alpha} & \mathbf{U}_w^{\alpha} & \mathbf{U}_w^{\alpha} \\ \mathbf{U}_{\tilde{\rho}'}^{\alpha} & \mathbf{U}_{\tilde{\rho}'}^{\alpha} & \mathbf{U}_{\tilde{\rho}'}^{\alpha} & \mathbf{U}_{\tilde{\rho}'}^{\alpha} & \mathbf{U}_{\tilde{\rho}'}^{\alpha} \\ \mathbf{U}_{b'}^{\alpha} & \mathbf{U}_{b'}^{\alpha} & \mathbf{U}_{b'}^{\alpha} & \mathbf{U}_{b'}^{\alpha} & \mathbf{U}_{b'}^{\alpha} \end{bmatrix} \begin{bmatrix} \mathbf{U}_u^{\alpha\top} & \mathbf{U}_v^{\alpha\top} & \mathbf{U}_w^{\alpha\top} & \mathbf{U}_{\tilde{\rho}'}^{\alpha\top} & \mathbf{U}_{b'}^{\alpha\top} \\ \mathbf{U}_u^{\alpha\top} & \mathbf{U}_v^{\alpha\top} & \mathbf{U}_w^{\alpha\top} & \mathbf{U}_{\tilde{\rho}'}^{\alpha\top} & \mathbf{U}_{b'}^{\alpha\top} \\ \mathbf{U}_u^{\alpha\top} & \mathbf{U}_v^{\alpha\top} & \mathbf{U}_w^{\alpha\top} & \mathbf{U}_{\tilde{\rho}'}^{\alpha\top} & \mathbf{U}_{b'}^{\alpha\top} \\ \mathbf{U}_u^{\alpha\top} & \mathbf{U}_v^{\alpha\top} & \mathbf{U}_w^{\alpha\top} & \mathbf{U}_{\tilde{\rho}'}^{\alpha\top} & \mathbf{U}_{b'}^{\alpha\top} \\ \mathbf{U}_u^{\alpha\top} & \mathbf{U}_v^{\alpha\top} & \mathbf{U}_w^{\alpha\top} & \mathbf{U}_{\tilde{\rho}'}^{\alpha\top} & \mathbf{U}_{b'}^{\alpha\top} \end{bmatrix} \mathbf{v}, \tag{2.26a}$$

with the elements given by:

$$(\mathbf{L}\mathbf{v})_i = \frac{1}{N_p}\sum_{j=1}^{N_g N_p}\left\{(\mathbf{U}^{\alpha})_{i,j}\sum_{i'=1}^{N_g N_p}(\mathbf{U}^{\alpha\top})_{j,i'}\mathbf{v}_{i'}\right\}. \tag{2.26b}$$

Note how in this case, the rows of $\mathbf{U}^{\alpha\top}$ are the same as in $\tilde{\mathbf{U}}^{\alpha\top}$ from the first approach (Eq. (2.21)), but repeated N_p times. If there is only one non-zero element (the q^{th} element of $\mathbf{v}$), then the computation simplifies to:

$$(\mathbf{L}\mathbf{v})_i = \frac{1}{N_p}\sum_{j=1}^{N_g N_p}(\mathbf{U}^{\alpha})_{i,j}(\mathbf{U}^{\alpha\top})_{j,q}\mathbf{v}_q = \frac{1}{N_p}\sum_{j=1}^{N_g} N_p(\tilde{\mathbf{U}}^{\alpha})_{i,j}(\tilde{\mathbf{U}}^{\alpha})_{q,j}\mathbf{v}_q = \sum_{j=1}^{N_g}(\tilde{\mathbf{U}}^{\alpha})_{i,j}(\tilde{\mathbf{U}}^{\alpha})_{q,j}\mathbf{v}_q, \tag{2.26c}$$

noting that when $N_p = N_{var}$, full inter-variable covariances are retained. In the computation of any inner products of $\boldsymbol{\chi}^{\alpha k}$ in the variational algorithm, such as for the minimisation, or in the computation of J_e, these also have to be scaled by $\frac{1}{N_p}$ accordingly.

The key thing to note here is that when using both approaches with $\tilde{\mathbf{U}}^{\alpha}$ and $\mathbf{U}^{\alpha}$ respectively, the implied localisation matrices are the same ($\tilde{\mathbf{L}} = \mathbf{L}$), as demonstrated by Eq. (2.22c) and Eq. (2.26c) being the same. Given the greater flexibility, the second approach was coded in the ABC-DA system.

2.6.3 Hydrostatic imbalance due to vertical localisation

According to Eq. (3) of Bannister (2020), the hydrostatic balance relation in the ABC model (also used in the control variable transform) is given by:

$$C\frac{\partial \tilde{\rho}'}{\partial z} = b'. \tag{2.27}$$

From Eq. (2.1), the prognostic w equation indicates that the change in w following an air parcel (i.e., a Lagrangian frame of reference) is given by the source/sink terms $C\frac{\partial \tilde{\rho}'}{\partial z}$ and b'. This neatly corresponds to hydrostatic balance. In other words, hydrostatic imbalance will lead to sources/sinks in w as the system evolves.

Applying vertical localisation directly to the ensemble-derived error modes via the Schur product results in alterations in the vertical gradient of the $\tilde{\rho}'$ field, depending on the kurtosis of the correlation curve applied. We can consider the following scenario: Assuming that the ensemble forecasts are hydrostatically balanced on the large scales, one could expect that assimilating a single $\tilde{\rho}'$ observation without vertical localisation would result in hydrostatically balanced $\tilde{\rho}'$ and b' increments. However, with vertical localisation, the sharpness of the correlation curve superimposes on the $\tilde{\rho}'$ fields in the ensemble-derived error modes and results in increments that decrease more rapidly with distance (sharper gradient) from the point of observation. Thus, a larger b' increment is required in order to maintain hydrostatic balance, but the actual b' increments are also reduced by the Schur product. In this scenario, the resulting b' increments would be sub-hydrostatic.

During the spin-up configuration testing with vertical localisation applied, it was noted that the root-mean-square value of the w field was gradually increasing throughout the earlier stages of the spin-up process. Since there exists a restoring A^2w source term in the prognostic b' equation, the root-mean-square value of the w field does not increase indefinitely because of corresponding induced changes in the b' field.

2.7 From the ABC-DA system to a full NWP system

The development of hybrid ensemble-variational data assimilation for the ABC-DA system allowed for the comparison of hybrid ensemble-variational data assimilation with traditional methods within a tropical framework. The results from Chapter 2 suggest that since the hybrid approach outperforms traditional methods in a simplified tropical convective-scale fluid dynamics model, it could also be beneficial for a full tropical NWP system. Notwithstanding this, no study so far has developed hybrid ensemble-variational data assimilation for a regional convective-scale NWP system over the Maritime Continent. Chapter 3 continues this exploration by documenting the development and comparison of hybrid ensemble-variational data assimilation with traditional methods, but this time using a full NWP system over the Maritime Continent.

In Chapter 2, the unique design of the hybrid ensemble-variational implementation also provided opportunities to explore less-known aspects of tropical data assimilation, such as the multivariate background error cross-covariances in the tropics and their impacts, and the dependence of localisation on prognostic variables (e.g., specific humidity, zonal/meridional wind). Chapter 4 continues this exploration by modifying the localisation design within the ensemble-variational approach to reveal insights about tropical data assimilation. These insights can then be leveraged to improve hybrid ensemble-variational data assimilation over the Maritime Continent.

Chapter 3

Development of a hybrid ensemble-variational data assimilation system over the western Maritime Continent

This chapter concerns RQ1 posed in Section 1.5 and has been published in *Weather and Forecasting* with the following reference:

Lee, J.C.K. and Barker, D.M., 2023. Development of a Hybrid Ensemble-Variational Data Assimilation System over the Western Maritime Continent. *Weather and Forecasting*, **38(3)**, pp. 425-444, https://doi.org/10.1175/WAF-D-22-0113.1.

It is unmodified from the published manuscript, other than being re-formatted in accordance with the thesis chapters and with minor typographical adjustments to maintain consistency throughout the thesis.

Abstract

A hybrid three-dimensional ensemble-variational (En3DVar) data assimilation system has been developed to explore incorporating information from an 11-member regional ensemble prediction system, which is dynamically downscaled from a global ensemble system, into a 3-hourly cycling convective-scale data assimilation system over the western Maritime Continent. From the ensemble, there exists small-scale ensemble perturbation structures associated with positional differences of tropical convection, but these structures are well represented only after the downscaled ensemble forecast has

evolved for at least 6 hours due to spinup. There was also a robust moderate negative correlation between total specific humidity and potential temperature background errors, presumably because of incorrect vertical motion in the presence of clouds. Time shifting of the ensemble perturbations, by using those available from adjacent cycles, helped to ameliorate the sampling error prevalent in their raw autocovariances. Monthlong hybrid-En3DVar trials were conducted using different weights assigned to the ensemble-derived and climatological background error covariances. The forecast fits to radiosonde relative humidity and wind observations were generally improved with hybrid-En3DVar, but in all experiments, the fits to surface observations were degraded compared to the baseline 3DVar configuration. Over the Singapore radar domain, there was a general improvement in the precipitation forecasts, especially when the weighting toward the climatological background error covariance was larger, and with the additional application of time-shifted ensemble perturbations. Future work involves consolidating the ensemble prediction and deterministic system, by centring the ensemble prediction system on the hybrid analysis, to better represent the analysis and forecast uncertainties.

3.1 Introduction

At the Meteorological Service Singapore (MSS), a hybrid ensemble-variational data assimilation system has been developed to explore incorporating information from an ensemble prediction system (SINGV-EPS; Porson et al., 2019) into a variational data assimilation system (SINGV-DA; Heng et al., 2020). Such hybrid ensemble-variational methods have recently gained traction, used in both global numerical weather prediction (NWP) systems (Buehner et al., 2013, Clayton et al., 2013, Kuhl et al., 2013, Wang et al., 2013, Bonavita et al., 2016, Kadowaki et al., 2020) and regional NWP systems (Zhang and Zhang, 2012, Gustafsson et al., 2014, Ito et al., 2016, Montmerle et al., 2018, Caron et al., 2019, Bédard et al., 2020) at leading operational NWP centres. Different centres employ their own flavor of hybrid ensemble-variational methods, due to many possible permutations in the design of the ensemble prediction system and/or variational data assimilation system. Bannister (2017) provides a thorough overview of hybrid ensemble-variational methods used in operational systems.

Apart from MSS, there are only a handful of research and operational centres that maintain a convective-scale NWP system over the western Maritime Continent (Centre for Climate Research Singapore, 2019). Very few centres have incorporated data assimilation, and these centres typical apply traditional variational methods partly due to the lack of a suitable high-resolution ensemble which is needed for the application

of hybrid ensemble-variational methods. In traditional variational data assimilation, the characterisation of the background errors often relies on the assumption of ergodicity by using climatological error statistics. Modeling the climatological background error covariance matrix also requires further assumptions of homogeneity, isotropy, and balance constraints in the background error statistics (Bannister, 2008*b*). These are necessary to prescribe a parameterised model of the climatological background error covariance matrix so that the variational data assimilation problem becomes computationally feasible. Over the western Maritime Continent, these assumptions may be often violated by the presence of nonlinear convective processes, land-sea interactions and local orographic effects (Lee and Huang, 2022). On the other hand, the characterisation of the ensemble-derived background error covariance matrix, usually within an ensemble Kalman filter (EnKF) framework (Evensen, 1994), does not require these assumptions as the background error statistics can be estimated directly from the model states. The error statistics may also contain meaningful flow-dependent error structures related to the short-lived tropical weather phenomena over the region. However, since the degrees of freedom for the model state is usually far greater than the ensemble size used to estimate the error statistics (resulting in a rank-deficient matrix), sampling noise and spurious long-range correlations may be present, particularly for smaller ensembles. To address sampling noise, Houtekamer and Mitchell (2001) proposed computing the Schur product of a localisation matrix (a correlation matrix) and the ensemble-derived background error covariance matrix. This effectively increases the rank of the ensemble-derived background error covariance matrix. However, the use of localisation can unintentionally introduce dynamical imbalances into the system (Lorenc, 2003), which may be detrimental.

The use of a weighted combination of both estimates of background error statistics in a hybrid ensemble-variational data assimilation framework was proposed by Hamill and Snyder (2000). The main idea is to alleviate the limitations and maximize the advantages offered by individual ensemble-based or variational methods themselves. The ensemble-derived background error covariance matrix can augment the climatological background error covariance matrix with its flow-dependent error structures, while the climatological background error covariance matrix can ameliorate the sampling noise issues associated with the ensemble-derived background error covariance matrix. Previous studies have provided evidence that the weighted combination results in improved verification scores over individual ensemble-based or variational methods. The optimal weighting combination varies between studies because of a multitude of factors including ensemble size (Hamill and Snyder, 2000), localisation

space and length-scales (Montmerle et al., 2018, Caron et al., 2019), design of the ensemble (e.g., EnKF or other deterministic analysis ensemble methods; Tippett et al., 2003), and variational method applied (e.g., three- or four-dimensional; 3DVar or 4DVar). Most operational centres rely on empirical tuning to find the optimal weighting.

In this article, we describe the development of a hybrid ensemble-variational data assimilation system over the western Maritime Continent, which is — to our knowledge — the first of its kind over this region. The initial implementation of SINGV-EPS is a simple 11-member ensemble prediction system dynamically downscaled from preselected European Centre for Medium-Range Weather Forecasts (ECMWF) global ensemble members every 12 hours over the western Maritime Continent. SINGV-DA is a 3-hourly cycling three-dimensional first guess at appropriate time variational system (3DVar-FGAT) over the same domain. There is also no centring of SINGV-EPS on the SINGV-DA analyses, so the information flow is one-way (i.e., from SINGV-EPS to SINGV-DA during the estimation of the ensemble-derived background error statistics). Section 3.2 describes the hybrid ensemble-variational formulation and implementation at MSS, along with more details on SINGV-EPS and SINGV-DA. We explore the structures of the ensemble-derived background error statistics and discuss their relevance in Section 3.3. Section 3.4 contains a description of the monthlong trials, which are conducted to seek a suitable configuration for operational implementation. We also discuss the justification of the tuning parameters for the western Maritime Continent context and include verification scores from the empirical tuning of the hybrid three-dimensional ensemble-variational (En3DVar) system (see Section 3.2.3 for nomenclature) at MSS.

3.2 Hybrid-En3DVar formulation

3.2.1 Traditional 3DVar-FGAT in SINGV-DA

SINGV-DA is a convective-scale regional NWP data assimilation system that uses a horizontal grid spacing of approximately 1.5 km, with 80 vertical levels up to 38.5 km over the western Maritime Continent (see Fig. 3.2 for the domain). The lateral boundary conditions (LBCs) are provided by ECMWF analyses and forecasts every 6 hours (0000, 0600, 1200, and 1800 UTC). SINGV-DA is based on the Met Office (UKMO) Unified Model framework (Tang et al., 2013), and is designed as a 3-hourly cycling 3DVar-FGAT system. Assimilated observations are retrieved from WMO's Global Telecommunication System. These include radiosondes, surface and aircraft observations, all-sky and clear-sky satellite radiance observations, satellite-

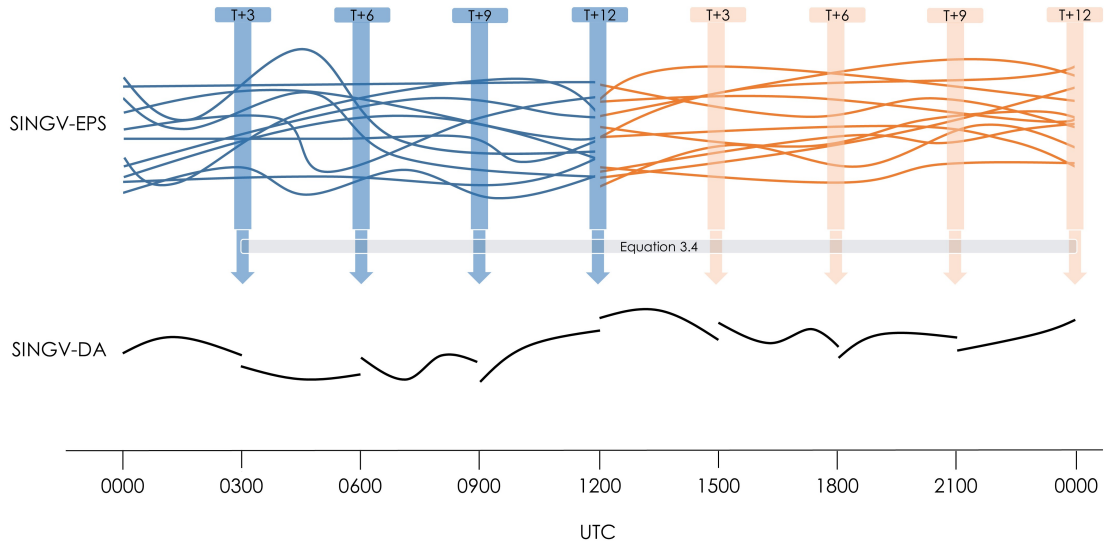

Figure 3.1: Schematic of the information flow between SINGV-EPS and SINGV-DA. The blue and orange lines represent the ensemble trajectories of SINGV-EPS (excluding control member) for forecasts initialised at 0000 and 1200 UTC, respectively. The black lines represent the trajectory of the SINGV-DA forecasts in a 3-hourly cycling set-up. The ensemble perturbations from SINGV-EPS are computed using the 3- ($T + 3$) to 12-h ($T + 12$) forecasts depending on the time of the day.

derived wind observations, and satellite-derived pseudo-cloud observations (see Heng et al., 2020 for the full list). The observation error profiles are retrieved from the UKMO.

At each SINGV-DA cycle, we seek an 'optimal' state $\mathbf{x}^a$ that minimises a cost function $J(\mathbf{x})$ (e.g., Kalnay, 2003). The cost function is usually nonquadratic because of the often nonlinear forecast model (M) and nonlinear observation operator (H), which appear in one of its components: the departure of the 'optimal' state with respect to observations; observation penalty. Like most variational systems, SINGV-DA implements an incremental formulation of the cost function (Courtier et al., 1994), which requires a linearisation of M and H around a reference state ($\mathbf{x}^r$) and formulating the problems in terms of increments $\delta\mathbf{x}$ to $\mathbf{x}^r$ in a series of outer loops. In SINGV-DA, we do not include an imbalance penalty (e.g., as included in Clayton et al., 2013, Milan et al., 2020) and assume a perfect model. Since FGAT is used, the linearised forecast model $\mathbf{M} = \mathbf{I}$, an identity matrix. The (strong-constraint) incremental form of the 3DVar-FGAT cost function is thus given by

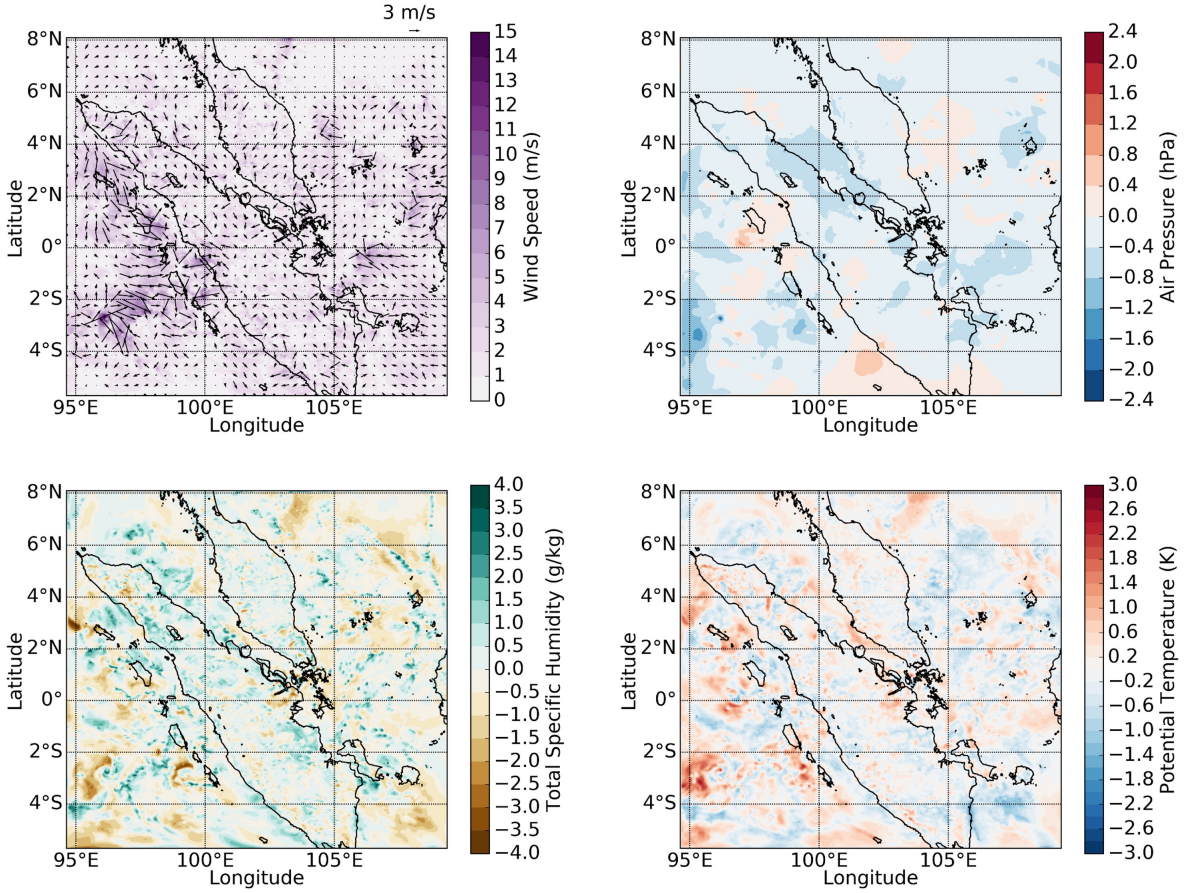

Figure 3.2: Ensemble perturbation fields of the horizontal wind (top left), pressure (top right), total specific humidity (bottom left), and potential temperature (bottom right) for ensemble member 1 at model level 15 (∼1-km height AGL), valid at 0600 UTC 1 Jun 2019 (6-h forecast from 0000 UTC 1 Jun 2019). The vector represents the horizontal wind deviation from the ensemble mean.

$$J(\delta\mathbf{x}) = J_b + J_o,$$
$$= \frac{1}{2}(\delta\mathbf{x} - \delta\mathbf{x}^r)^\top \mathbf{B}^{-1}(\delta\mathbf{x} - \delta\mathbf{x}^r) + \frac{1}{2}(\mathbf{H}\delta\mathbf{x} - \mathbf{d})^\top \mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x} - \mathbf{d}), \tag{3.1}$$

where J_b and J_o are the background and observation penalties, respectively; $\delta\mathbf{x}^r$ is the difference between the background and $\mathbf{x}^r$; $\mathbf{B}$ and $\mathbf{R}$ are the background and observation error covariance matrices; and $\mathbf{H}$ is the linearised observation operator. The innovation $\mathbf{d}$ is defined as

$$\mathbf{d} = \mathbf{y}^o - H[M(\mathbf{x}^r], \tag{3.2}$$

where $\mathbf{y}^o$ is the observations vector. As SINGV-DA uses the observations processing and data assimilation framework from the UKMO, the computation of the observation penalty is the same as in Eq. (2) of Clayton et al. (2013), except that the linear perturbation forecast model (analogous to $\mathbf{M}$) is not required in 3DVar-FGAT (hereafter referred to as 3DVar for brevity). These details are also discussed in Lorenc et al. (2000).

The conditioning of the cost function minimisation (Eq. (3.1)) can be improved by introducing a control variable transform $\mathbf{U}$ (a 'square root' of $\mathbf{B}$; i.e., $\mathbf{B} = \mathbf{U}\mathbf{U}^\top$). This avoids the need to compute $\mathbf{B}^{-1}$. The increment $\delta\mathbf{x}$ can be expressed as

$$\delta\mathbf{x} = \mathbf{U}\delta\boldsymbol{\chi}, \tag{3.3}$$

where $\delta\boldsymbol{\chi}$ is the control vector. SINGV-DA adopts the same control variable transform (for full transforms, see Lorenc et al., 2000) as the UKMO limited-area model (LAM; UKV; Tang et al., 2013) prior to July 2017. The control variables used in $\mathbf{U}$ are streamfunction, velocity potential, unbalanced pressure and a nonlinear transformed humidity (Ingleby et al., 2013). Linear balance is used as a balance constraint, so for the western Maritime Continent the geostrophically balanced pressure is small and pressure is near-wholly unbalanced (i.e., horizontal wind and pressure background errors are assumed to be almost uncorrelated). The same vertical modes in the vertical transform are used for each horizontal mode in the horizontal transform, so $\mathbf{B}$ is homogeneous in the domain. Note that the additional vertical adaptive grid transform (e.g., used in Milan et al., 2020) is not applied. The same $\mathbf{B}$ (strictly speaking the same $\mathbf{U}$) is used in the variational minimisation for each cycle, thus $\mathbf{B}$ is often referred to as the climatological background covariance matrix $\mathbf{B}_c$. Time stationarity of $\mathbf{B}_c$ is assumed even though the true forecast errors of the system (and their probability

distribution function) may vary according to the flow conditions associated with various tropical weather phenomena.

The training data for calibrating $\mathbf{B}_c$ for SINGV-DA was generated using the lagged National Meteorological Center (NMC) method (Široká et al., 2003), which is an extension of the original NMC method (Parrish and Derber, 1992). The main difference is that the lagged NMC method excludes error sources from the driving model in regional NWP, by ensuring that the forecast differences along the boundaries are zero. The same set of LBCs are used in forecast pairs that are valid at the same time. This method is suitable for SINGV-DA as it retains the mesoscale and small-scale background error structures and does not reanalyse the scales already treated by the global model (Široká et al., 2003). Other studies with LAMs have also opted for the lagged NMC method (Sadiki and Fischer, 2005, Montmerle et al., 2006, Stanesic et al., 2019). The training data were based on the differences between 12- and 6-h forecasts valid at the same time, over the period 22 January–6 March 2013. However, while SINGV-DA uses a 1.5-km horizontal grid spacing, the training data were generated using a preliminary forecast model that uses a 4.5-km grid spacing instead. This discrepancy is further discussed in Section 3.4.5.

The characteristics of $\mathbf{B}_c$ that is used in SINGV-DA have been explored in Heng et al. (2020) and Lee and Huang (2020). We remark that the variances in $\mathbf{B}_c$ are simply the in-sample variances from the training data of lagged forecast differences. However, since the inception of SINGV-DA, the original $\mathbf{B}_c$ has not been replaced due to poorer verification scores with other calibrated covariances. Heng et al. (2020) also describes other aspects of SINGV-DA, including the observational coverage, the application of an incremental analysis update (IAU) scheme (Bloom et al., 1996), satellite bias corrections, and other relevant SINGV-DA details.

3.2.2 Ensemble-derived background error covariances from SINGV-EPS

SINGV-EPS is a convective-scale 11-member ensemble prediction system that uses a horizontal grid spacing of 4.5 km. Each ensemble member is dynamically downscaled over the western Maritime Continent (near-identical domain and with the same model settings in each ensemble member as SINGV-DA, but at a lower resolution) using the corresponding global ECMWF ensemble member analysis and forecast. As ECMWF offers 51 global ensemble members (at $\sim$18-km grid spacing; see Buizza and

Richardson, 2017) and SINGV-EPS only requires 11, the initial conditions and LBCs for SINGV-EPS are retrieved from the same members (simple fixed preselection, based on the first 11 odd number indexes of the 51 members) for reinitialisation every 12 hours. Since SINGV-EPS does not incorporate data assimilation, it is uninformed of the SINGV-DA observation network, unlike in other ensembles within the EnKF framework that can account for observation uncertainty and network. Only the nature of the dynamical error growth about the ensemble mean is represented during the estimation of the ensemble-derived background error covariance matrix. This is a limitation of the initial implementation, and future work on SINGV-EPS should address it.

An ensemble prediction system like SINGV-EPS allows for the representation of flow-dependent forecast errors due to varying flow conditions, which can be estimated using the ensemble forecast trajectories. Only the necessary ensemble forecast fields (zonal and meridional wind, potential temperature, a density term, pressure, total specific humidity) are required for reconfiguration onto the 3DVar assimilation grid, usually at a coarser resolution and smaller than the forecast domain. One may compute a rectangular matrix $\mathbf{X}^f$ whose columns contain the scaled differences between the ensemble forecasts and the ensemble mean:

$$\begin{aligned}\mathbf{X}^f &= \frac{1}{\sqrt{N-1}}(\mathbf{x}_t^1 - \bar{\mathbf{x}}_t, \mathbf{x}_t^2 - \bar{\mathbf{x}}_t, \ldots, \mathbf{x}_t^N - \bar{\mathbf{x}}_t) \\ &= (\mathbf{x}_t^{1'}, \mathbf{x}_t^{2'}, \ldots, \mathbf{x}_t^{N'}),\end{aligned} \tag{3.4}$$

where N is the number of ensemble members, $\mathbf{x}_t^k$ is the k^{th} member forecast and $\bar{\mathbf{x}}_t$ is ensemble mean valid at time t, and $\mathbf{x}_t^{k'}$ is the k^{th} scaled ensemble perturbation. The discrete validity time of the ensemble perturbations are chosen to correspond to the relevant cycle times (eight cycles a day) in SINGV-DA. Since SINGV-EPS is not coupled to SINGV-DA and does not require the deterministic analysis (no centring) to reinitialise the ensemble, these ensemble perturbations can be generated prior to running SINGV-DA.

The raw ensemble-derived background error covariance $\mathbf{P}_t^f$ is explicitly given by the outer product:

$$\mathbf{P}_t^f = \mathbf{X}_t^f \mathbf{X}_t^{f\top}. \tag{3.5}$$

The number of ensemble members only has N degrees of freedom to fit the observations, so $\mathbf{P}_t^f$ is rank-deficient and is contaminated by sampling noise. Typically, a Schur product

(Houtekamer and Mitchell, 2001) with a localisation matrix ($\mathbf{L}$) is used to damp any possible spurious long-range correlations:

$$\mathbf{B}_e = \mathbf{L} \circ \mathbf{P}_t^f, \tag{3.6}$$

where $\mathbf{B}_e$ is the ensemble-derived background error covariance matrix after localisation and the $\circ$ operator denotes the Schur product (or Hadamard product), which conducts an element-wise product of two same-sized matrices. The design of $\mathbf{L}$ is further discussed in Section 3.2.4.

3.2.3 Hybrid background error covariance

The hybrid background error covariance matrix $\mathbf{B}_h$ is a linear combination of $\mathbf{B}_c$ and $\mathbf{B}_e$ (Hamill and Snyder, 2000), in the following form:

$$\mathbf{B}_h = \beta_c^2 \mathbf{B}_c + \beta_e^2 \mathbf{B}_e, \tag{3.7}$$

where β_c and β_e are (positive) scalar weights that are usually determined empirically. These weights are often chosen to sum to unity, although it is not mandatory. It is not feasible to explicitly compute $\mathbf{B}_h$ from individual components for an NWP system like SINGV-DA, so the alpha control variable approach of Lorenc (2003) is employed which constructs an implied version of Eq. (3.6) using an modified version of Eq. (3.1). Wang et al. (2007) demonstrates how both approaches yield equivalent results. The modified cost function (extension of Eq. (3.1)) is given by

$$\begin{aligned} J(\delta\mathbf{x}, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_N) &= J_b + J_o + J_e \\ = \frac{1}{2}(\delta\mathbf{x} - \delta\mathbf{x}^r)^\top \mathbf{B}_c^{-1}(\delta\mathbf{x} - \delta\mathbf{x}^r) &+ \frac{1}{2}(\mathbf{H}\delta\mathbf{x} - \mathbf{d})^\top \mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x} - \mathbf{d}) + \frac{1}{2}\sum_{k=1}^{N} \boldsymbol{\alpha}_k^\top \mathbf{L}^{-1} \boldsymbol{\alpha}_k^\top, \end{aligned} \tag{3.8}$$

where J_e is the ensemble penalty, and $\boldsymbol{\alpha}_k$ is an alpha field with dimensions the same as the state size, associated with the k^{th} ensemble member. Correspondingly, a modification of Eq. (3.3) is also required to include the ensemble contribution from a linear combination of ensemble perturbations:

$$\delta\mathbf{x} = \beta_c \mathbf{U}\delta\boldsymbol{\chi} + \beta_e \sum_{k=1}^{N} \mathbf{x}_t^{k'} \circ \boldsymbol{\alpha}_k. \tag{3.9}$$

The alpha fields essentially control the contribution of each ensemble perturbation to the analysis increment. Like $\mathbf{B}_c$, $\mathbf{L}$ can be partitioned into a 'square-root' matrix $\mathbf{U}^\alpha$ (i.e., $\mathbf{L} = \mathbf{U}^\alpha \mathbf{U}^{\alpha\top}$), the alpha control variable transform, which can be applied to an alpha control vector associated with the k^{th} ensemble member ($\boldsymbol{\chi}_k^\alpha$), so Eq. (3.9) becomes

$$\delta\mathbf{x} = \beta_c \mathbf{U}\delta\boldsymbol{\chi} + \beta_e \sum_{k=1}^{N} \mathbf{x}_t^{k'} \circ (\mathbf{U}^\alpha \boldsymbol{\chi}_k^\alpha). \tag{3.10}$$

Note that using $\mathbf{U}^\alpha$ avoids the need to compute $\mathbf{L}^{-1}$ in Eq. (3.8). Following the naming convention of Bannister (2017), this approach is termed as hybrid-En3DVar. Note that this is simply 3DVar-FGAT using $\mathbf{B}_h$, or hybrid-3DVar-Ben following the nomenclature of Lorenc (2013).

Since SINGV-DA is a 3-hourly cycling system, $\mathbf{B}_h$ is required at each cycle so there needs to be valid ensemble perturbations from SINGV-EPS every 3 hours. However, SINGV-EPS is initialised once every 12 hours, so the ensemble forecast trajectories downscaled from a given set of ECMWF driving data have to provide ensemble perturbations for multiple SINGV-DA cycles that fall within the 12-h window until the next set of ECMWF driving data is available. Figure 3.1 shows the schematic diagram illustrating the design of the hybrid-En3DVar system and flow of information from SINGV-EPS to SINGV-DA. As an example, the 12-h ensemble forecasts are used to compute the ensemble perturbations (Eq. (3.5)) for the 0000 and 1200 UTC SINGV-DA cycles, and the 9-h forecasts are used for the 0900 and 2100 UTC SINGV-DA cycles. Evidently, a key limitation is that the ensemble statistics are calculated from a longer forecast range (for 6–12-h forecasts) than required for a 3-hourly cycling SINGV-DA system requiring 3-h forecast error statistics. Since SINGV-EPS is also downscaled from global driving data, spinup may also affect the ensemble statistics so the variances and length-scales in $\mathbf{B}_e$ may vary across different forecast ranges. Dipankar et al. (2020) noted that the spinup duration in similar downscaled simulations over the same domain is around 6–9 hours. Consequently, the ensemble statistics computed from the 3-h forecasts may only be estimating the large-scale forecast errors. In the global implementation, Clayton et al. (2013) noted only a small difference in the ensemble statistics between various lead times and expected only minor impacts on the verification scores. We investigate these issues further and discuss their relevance in Section 3.3.

3.2.4 Localisation and weightings

A key aspect of the hybrid-En3DVar algorithm is the design of the localisation applied on $\mathbf{B}_e$. In many respects, the localisation approach is very similar to Clayton et al. (2013) since SINGV-DA also uses the same data assimilation framework of the UKMO, but for the LAM implementation. Here, we focus on the key differences and other important aspects for consideration in the SINGV-DA implementation.

Lorenc (2003) discusses how localisation directly in the space of the model variables (hereafter referred to as model space) can result in the generation of subgeostrophic wind increments when a single height observation is assimilated because the Schur product alters the kurtosis of the covariance curve and its gradient. In this light, Clayton et al. (2013) applied balance-preserving localisation by first transforming the ensemble perturbations from model space into control variable space (using balance constraints) in the algorithm. Over the western Maritime Continent (deep tropics), we expect that adhering to geostrophic balance is relatively unimportant. Furthermore, for a convective-scale system, transient dynamical imbalances inevitably occur because of nonlinear convective processes. We have thus opted to perform localisation on the model prognostic variables in SINGV-DA, despite possible resulting dynamical imbalances in the analysis increments. In this manner, multivariate background error relationships are directly prescribed between model variables by the cross covariances from the ensemble perturbations (Eq. (3.5)). Note that we have also disabled inter-variable localisation, which removes cross covariances between model variables. Therefore, we preserve the inherent background error relationships captured by the ensemble (see Sections 3.3.3 and 3.3.4).

We can define the 'square-root' $\mathbf{U}^{\alpha}$ using separate horizontal and vertical transforms for the spatial localisation:

$$\mathbf{U}^{\alpha} = \mathbf{U}_v^{\alpha}\mathbf{U}_h^{\alpha}, \tag{3.11}$$

where $\mathbf{U}_h^{\alpha}$ and $\mathbf{U}_v^{\alpha}$ are the horizontal and vertical transforms that apply the localisation, respectively. For the horizontal localisation in the LAM implementation, a homogeneous and isotropic Gaussian correlation C is modeled using a spectral representation of

$$C(r) = exp\left[-\frac{1}{2}\left(\frac{r}{2s}\right)^2\right], \tag{3.12}$$

where s is the horizontal length-scale localisation parameter, and r is the horizontal

distance between two grid points. Note that Eq. (3.12) (the default Gaussian expression used in the LAM implementation by the UKMO) differs from Clayton et al. (2013) by a factor of 2 for s in the denominator, so the resulting Gaussian correlation function is broader for the same value of s. Additionally, a boundary relaxation is applied using half-cosine functions on the ensemble perturbations so that the values are zero at the lateral boundaries. This ensures that $\mathbf{B}_e$ satisfies the same boundary conditions that are built into the design of $\mathbf{B}_c$.

A side effect of the horizontal localisation is the aliasing of the length-scales of the analysis increments onto the localisation length-scale (due to the Schur product), which may degrade the quality of the analysis. To address this, Clayton et al. (2013) introduced an antialiasing 'high-pass' horizontal filter to remove the power from the lower wavenumbers (threshold determined by the localisation length-scale) and spurious gravity wave activity. However, they noted that the justification for its application depends on the length-scales present in the ensemble, since it is somewhat an ad hoc modification. As SINGV-EPS is dynamically downscaled and likely contains significant power in the scales larger than the localisation length-scale, we have chosen to apply this filter in SINGV-DA.

For the vertical localisation, an eigendecomposition of a target vertical localisation matrix is used. Only a fixed number of leading eigenvectors are retained to reduce the computational costs. In SINGV-DA, the target vertical localisation is constructed using a Gaspari-Cohn correlation function (Eq. (4.10) of Gaspari and Cohn (1999)) with ln(p) as a coordinate, where p is the level-mean pressure stored in $\mathbf{B}_c$. A ln(p) separation parameter controls the vertical localisation, in a similar manner as s for the horizontal case. This approach follows Buehner (2005) and uses a vertical coordinate such that the same vertical correlation length-scales can be used regardless of model level, even with the varying vertical mesh spacing.

Instead of using a uniform weight between $\mathbf{B}_e$ and $\mathbf{B}_c$ for all model levels, it is possible to use a vertically dependent weighting between $\mathbf{B}_e$ and $\mathbf{B}_c$ (Buehner et al., 2013, Clayton et al., 2013). Specifically, above a certain height above ground level (AGL), they introduce a transition zone where the weighting toward $\mathbf{B}_c$ is increased approximately linearly to full weight. This allows the background error correlation length-scales to gradually adjust to the climatological value due to practical model lid constraints in their set-up. Gradually weighting toward $\mathbf{B}_c$ also allows the horizontal correlation length-scales to vary in the upper model levels (e.g., stratosphere), instead of using a

single localisation length-scale in $\mathbf{B}_e$, which would likely be shorter than appropriate for the stratosphere (Clayton et al., 2013) since the horizontal correlation length-scales tend to be larger in that region (e.g., using observed residuals, Lönnberg and Hollingsworth, 1986, Bartello and Mitchell, 1992, and using forecast difference statistics, Ingleby, 2001). Over the Maritime Continent, the tropopause is expected to be around 16 km in height AGL, so we have chosen the transition zone to be 16–21 km (i.e., increasing weighting toward $\mathbf{B}_c$ starting from 16 km).

3.3 Structures in the ensemble-derived background error statistics

3.3.1 Analysis of ensemble perturbation structures

It is helpful to examine the spatial variation of the ensemble perturbations since the analysis increment computation involves a linear combination of the ensemble perturbations and is constrained to be within the subspace spanned by them (Lorenc, 2003). To illustrate the spatial variation present in the ensemble perturbations, we plot the ensemble member 1 perturbation fields ($\mathbf{x}_t^{1'}$) of the horizontal wind vector, pressure, total specific humidity and potential temperature at model level 15 ($\sim$1-km height AGL), valid at 0600 UTC 1 June 2019 (6-h forecast from 0000 UTC 1 June 2019; Fig. 3.2). This corresponds to the early afternoon in local time, with development of scattered thunderstorms over most parts of the domain, particularly off the western coast of Sumatra (not shown). There is a mix of large- and small-scale structures within the fields, related to positional differences of tropical convection between ensemble members and their mean. This is particularly pertinent in the horizontal wind vector, total specific humidity and potential temperature fields, especially around the western coast of Sumatra. Notably, the potential temperature and total specific humidity fields also exhibit similar spatial structures — regions of positive values in the total specific humidity field tend to correspond with regions of negative values in the potential temperature field. We also note a larger ensemble perturbation magnitude over the adjacent sea than land, which is common across all ensemble members (not shown). This is related to the diurnal cycle over land, compared with over sea. In general, the SINGV systems are well able to capture the diurnal cycle of precipitation over land, but do not represent the offshore migration of precipitation (tends to underestimate) during the night, through to the early afternoon (Dipankar et al., 2020, Lee et al., 2021) during intermonsoon seasons. Therefore, the forecast errors tend to be larger over the ocean during the diurnal cycle peak (and immediately after the peak, at 0600 UTC).

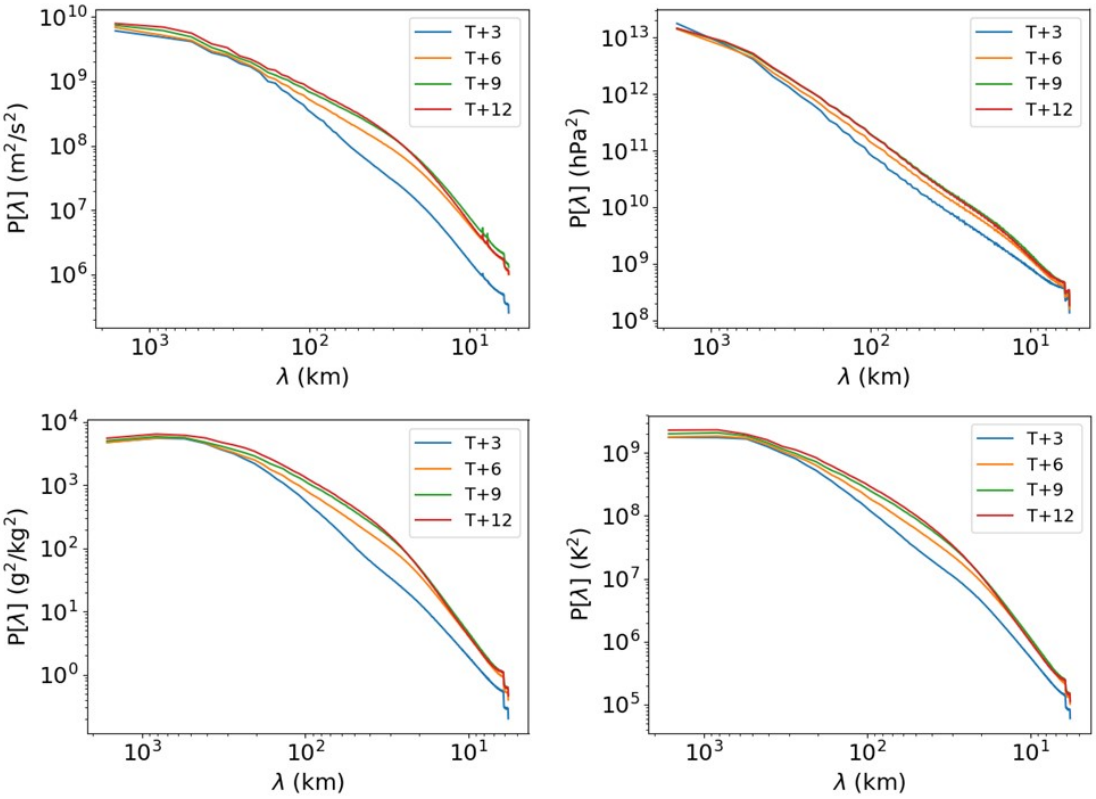

Figure 3.3: Mean power spectrum of the ensemble perturbations, for the (top left) horizontal wind, (top right) pressure, (bottom left) total specific humidity, and (bottom right) potential temperature fields at model level 15. The power P(λ) is binned according to total horizontal wavelength (λ) for different lead times (3–12-h forecasts; from $T + 3$ to $T + 12$). Each spectrum is averaged across all ensemble members and cycles in June 2019 (660 samples).

Next, we compute the mean power spectrum of the ensemble perturbations across all ensemble members and cycles in June 2019 (11 ensemble members and 2 cycles per day, total of 660 samples) for each lead time (Fig. 3.3). The mean power spectrum for all four variables shows that for shorter lead times (3-h forecasts), the high-resolution structures in the ensemble perturbations (smaller scales) do not contain the same power as for longer lead times (6-, 9-, and 12-h forecasts). This indicates that the ensemble statistics computed from the 3-h forecasts are estimating mainly the large-scale forecast errors. The differences are much smaller in the spectrum of the ensemble perturbations computed using the 6-, 9-, and 12-h forecasts, which suggests that their forecast error variances do not differ substantially, especially after 9 hours. This aligns with the notion that the spinup duration of 6–9 h — following Dipankar et al. (2020) — is sufficient for the high-resolution structures in the forecasts to develop. The mean power

spectrum of the ensemble perturbations is also closely related to the mean ensemble spread by definition. A higher power is derived from having larger perturbations from the ensemble mean, and this implies a larger ensemble spread. Having similarities in the mean power spectrum for longer lead times indicate that the increase in ensemble spread after 9 hours is fairly muted. Clayton et al. (2013) commented that using an ensemble constantly recentred around a deterministic analysis, they had only small differences in the variances and length-scales of the ensemble perturbations, even when comparing across different lead times.

For the mean power spectrum for pressure, the ensemble perturbation fields contain mainly large-scale structures (highest power at the largest total horizontal wavelength of 1600 km). However, for total specific humidity and potential temperature, the highest power occurs mainly at around a total horizontal wavelength of 800 km, which indicates that there are smaller scale ensemble perturbation structures in these fields which are more dominant. Consequently, following the schematic (Fig. 3.1), it is reasonable to expect that the analyses for some SINGV-DA cycles (0300 and 1500 UTC) would be disadvantaged by the lack of representation or underrepresentation of small-scale forecast errors in the ensemble perturbations, unless time shifting of ensemble perturbations (i.e., using ensemble perturbations that are valid prior and after the target analysis time; Lorenc, 2007) is considered.

3.3.2 Selection of localisation length-scales from auto-covariance structures

To better understand the localisation scales suitable for this En3DVar set-up, we compute the raw $\mathbf{P}_t^f$ (valid 0600 UTC 1 June 2019) with respect to a point near the centre of the SINGV-DA domain, focusing on the model level 15 autocovariance structures for each of the four variables. One would expect larger positive covariances near the point of interest and negligible covariances distant from the point of interest, assuming the absence of large-scale phenomena causing long-range spatial correlations. Figure 3.4 shows that when using only 11 ensemble perturbations, the autocovariances around the point of interest are generally dominated by sampling noise.

We explore recomputing the autocovariances using time-shifted ensemble perturbations, by including the ensemble perturbations valid 3 hours prior and after the target analysis time. This also serves to artificially boost the size to a total of 33 ensemble perturbations. Figure 3.5 shows that the inclusion of time-shifted ensemble

perturbations helps to alleviate some of the sampling noise, particularly in the potential temperature and horizontal wind fields. There are larger positive autocovariances tightly centred on the point of interest for the pressure ($\sim$250-km radius), potential temperature and total specific humidity fields ($\sim$50-km radius). However, even when using 33 ensemble perturbations, some long-range covariances still exist. Similar to Fig. 3.2, note how the autocovariance structures between total specific humidity and potential temperature are broadly similar (regions of positive/negative covariances tend to coincide), suggesting that the similarities in spatial patterns between these variables are also present across all the ensemble perturbations.

We repeat the computation of the autocovariances (33 ensemble perturbations) with respect to 10 other (land, ocean and coastal) points in the domain. These had similar qualitative takeaways on the localisation radius and the similarities in the autocovariance structures between total specific humidity and potential temperature (not shown). Note that in most cases, the sampling noise is drastically reduced, but still prevalent.

In the SINGV-DA implementation, the same spatial localisation is applied to all variables, regardless of horizontal position or model level. The specific humidity field is chosen as a benchmark for determining the localisation length-scales, since it was statistically the noisiest. However, one can argue that from Fig. 3.5, the localisation length-scales suitable for other variables (e.g., potential temperature) are also comparable. Over the western Maritime Continent, the variation in the total specific humidity background errors are partly governed by localised hydrometeor-related processes. Destouches et al. (2021) estimated the optimal horizontal localisation length-scales for hydrometeor variables to be around 20–80 km and around 120 km for specific humidity. Together with Figs. 3.4 and 3.5 (and considering that we use a smaller ensemble), one can hazard an educated guess on the appropriate horizontal localisation length-scale, of about 50 km, to eliminate the detrimental impacts of sampling noise yet retain meaningful spatial information. This value is also similar to the climatological background error correlation length-scales for specific humidity. For vertical localisation, we also take reference from Destouches et al. (2021) and climatological background error statistics — the optimal vertical localisation length-scale, in units of $\ln(p)$, for specific humidity using a Gaussian function is around 0.5. A comparable separation parameter for a Gaspari-Cohn correlation function is around 1.5 (i.e., no correlation beyond a $\ln(p)$ separation of 1.5). The selection of localisation length-scales based on climatological background error statistics is not new; Clayton et al. (2013) also selected the horizontal localisation

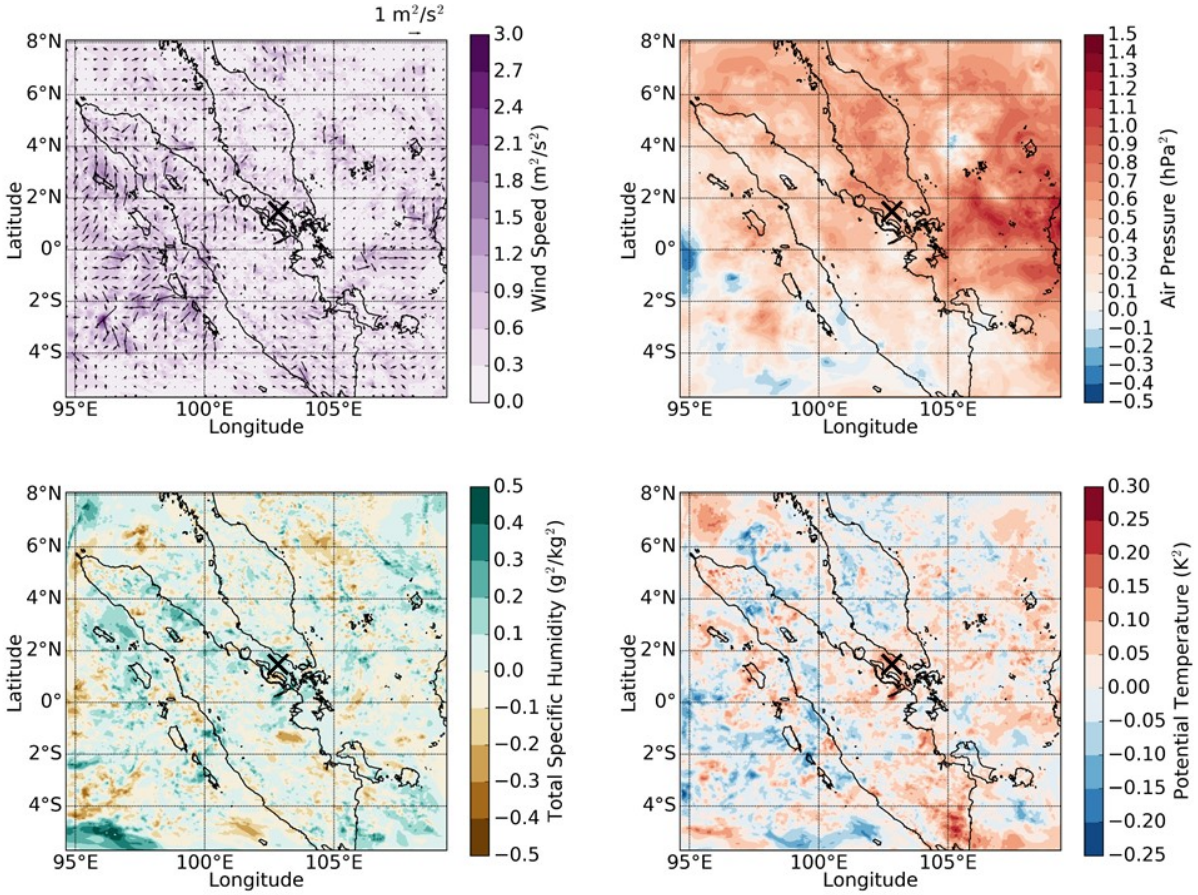

Figure 3.4: Raw ensemble-derived autocovariances of horizontal wind (with respect to a southwesterly wind; top left), pressure (top right), total specific humidity (bottom left), and potential temperature (bottom right) at model level 15 (∼1-km height AGL), computed from the 11 ensemble perturbations valid at 0600 UTC 1 Jun 2019 (6-h forecast from 0000 UTC 1 Jun 2019), with respect to a point in the centre of the domain (black cross). The vector represents the horizontal wind covariances (i.e., positive covariances in both the zonal and meridional components are represented by a vector pointing northeast).

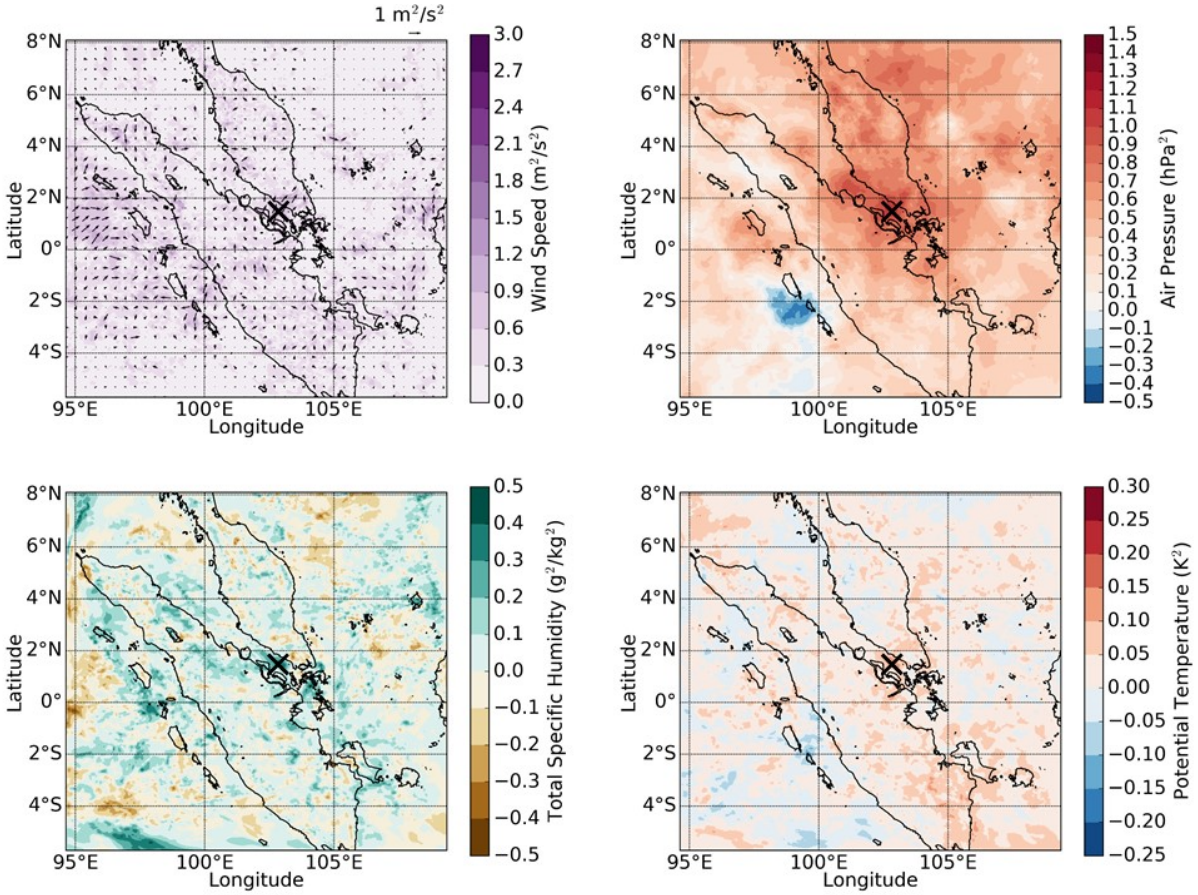

Figure 3.5: Raw ensemble-derived autocovariances of horizontal wind (with respect to a southwesterly wind; top left), pressure (top right), total specific humidity (bottom left), and potential temperature (bottom right) at model level 15 (∼1-km height AGL), computed from the 33 ensemble perturbations (by inclusion of time-shifted ensemble perturbations) valid from 0300 to 0900 UTC 1 Jun 2019 (3–9-h forecast from 0000 UTC 1 Jun 2019), with respect to a point in the centre of the domain (black cross). The vector represents the horizontal wind covariances (i.e., positive covariances in both the zonal and meridional components are represented by a vector pointing northeast).

length-scales based on error correlation length-scales for streamfunction. A more robust and objective approach has previously been developed (Ménétrier et al., 2015*a*,*b*), using sample estimated quantities and a linear filter. It is worth considering to reselect the length-scales using such advanced system-specific methods once the development of SINGV-EPS becomes more mature, or to use the 51-member ECMWF global ensemble perturbations, albeit at coarser resolution. We provide further comments in Section 3.4.5.

3.3.3 Potential temperature pseudo-observation

To illustrate the effect of the chosen localisation parameters, we insert a single pseudo-observation of potential temperature (1 K above the background) near the centre of the domain at different model levels and assess the resulting analysis increments. This variable choice also allows us to further investigate the existence of a multivariate relationship between the background errors of total specific humidity and potential temperature highlighted in Sections 3.3.1 and 3.3.2.

Figure 3.6 shows the vertical cross section of the potential temperature and total specific humidity responses to the single pseudo-observation of potential temperature, using traditional 3DVar and pure EnVar (full weight on $\mathbf{B}_e$ computed using 33 ensemble perturbations from 0300 to 0900 UTC 1 June 2019) with and without vertical localisation (columns) for different model levels (rows). One can observe how the vertical correlation length-scales are relatively constant ($\sim$7 km) when using ln(p) as the vertical coordinate for localisation (Fig. 3.6; right column). We note that the inclusion of vertical localisation helps to remove the apparent sampling error due to the small SINGV-EPS ensemble size.

Using 3DVar, there is a strong drying associated with the positive potential temperature pseudo-observation when it is inserted at the lower model levels (i.e., 15 and 29). When inserted at higher levels, the total specific humidity increments are negligible. This relationship is largely dependent on the calibrated nonlinear transformed humidity control variable (Ingleby et al., 2013) which for SINGV-DA, prescribes strong negative background error cross covariances between potential temperature and total specific humidity when the background is relatively far from saturation in the lower troposphere. This relationship is also captured when using pure EnVar, although the covariance is much smaller and more localised. Moisture is occasionally added at locations adjacent to its removal. When the pseudo-observation is inserted at higher model levels, moisture may be added or removed depending on insertion height AGL, which could be a

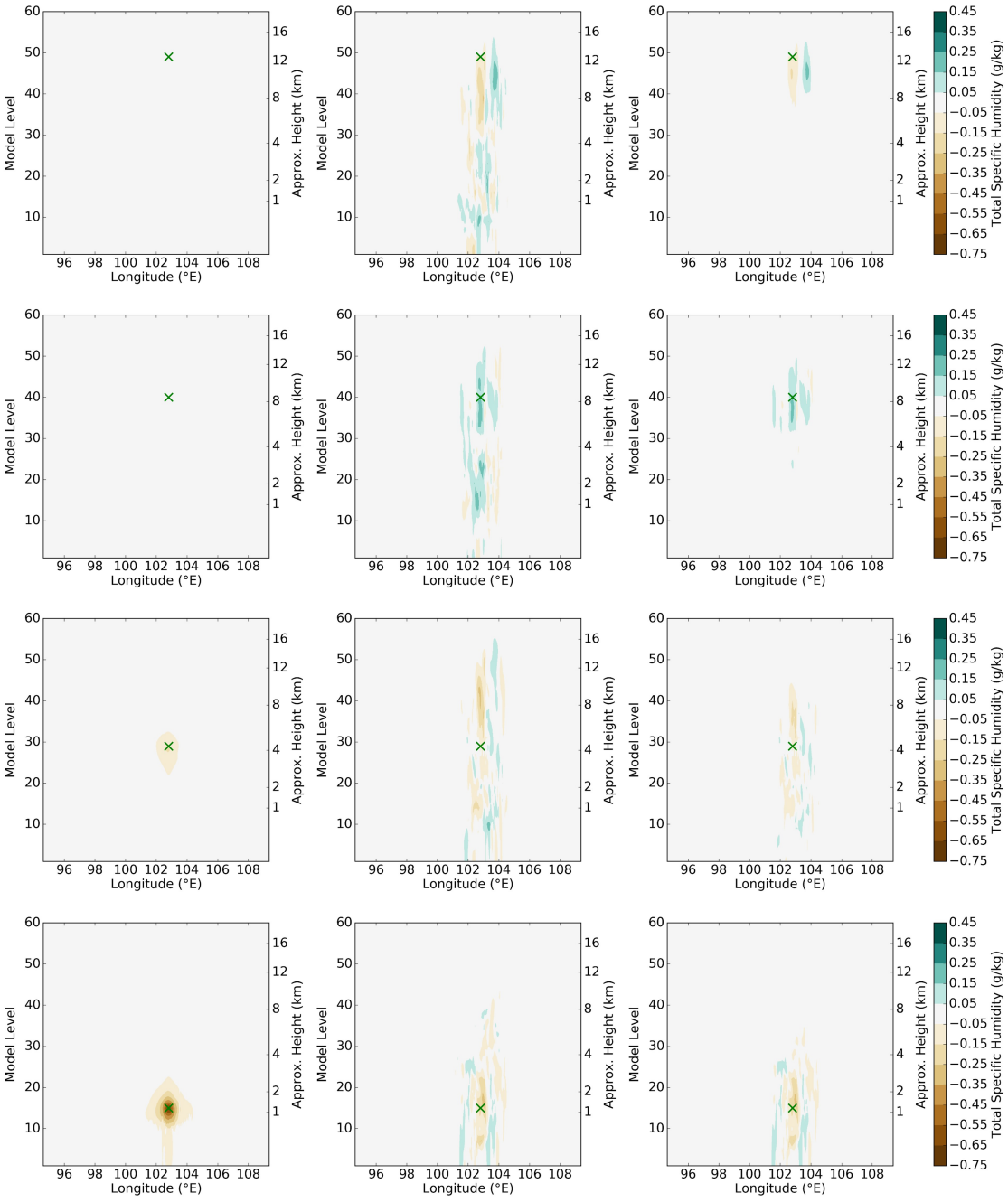

Figure 3.6: Total specific humidity analysis increment response to a pseudo-single observation of potential temperature 1 K above the background (observation error of 0.5 K) inserted near the centre of the domain (green cross) at model levels 15, 29, 40, and 49 (shown in rows from top to bottom, corresponding to 1-, 4-, 8-, and 12-km height AGL, respectively) for (left) 3DVar, (centre) pure En3DVar without vertical localisation, and (right) pure En3DVar with full spatial localisation. See text for details on ensemble perturbations used in pure En3DVar.

reflection of the actual day-to-day variability of the background error statistics.

3.3.4 Time-averaged cross-correlation with potential temperature

We also compute the raw cross correlation between the total specific humidity and potential temperature, using ensemble perturbations from 6-h forecasts sampled from all ensemble members and cycles in June 2019, for different model levels and locations (Fig. 3.7). It is evident from Fig. 3.7 that there exists a moderate negative correlation ($\sim$0.4) between the two variables at the lower levels, subject to small spatial variations. This correlation weakens higher up in the troposphere. At model level 49 (12 km), the correlation is occasionally weakly positive ($\sim$0.2) instead, albeit extremely localised. This was also noted for cross correlations with respect to other points in the domain. Our findings on the negative correlation in the lower troposphere is consistent with Lorenc (2007), who analysed radiosonde innovation statistics in the UKMO global model. They also found a weak negative correlation between errors of specific humidity and temperature in the lower troposphere and postulated that this was associated with the presence or absence of clouds, highlighting incorrect vertical motion (e.g., excessive descent leading to warming and drying) as a possible root cause. Over the Maritime Continent, where tropical convection and hydrostatic imbalance is prevalent, incorrect vertical motion (including positional errors in convection) is likely a dominant source of background error. Additionally, moisture convergence leading to vertical motion is usually restricted to the lower troposphere, which explains why the negative correlation is mainly confined below 8 km. The results in Sections 3.3.3 and 3.3.4 indicate that there exist potentially meaningful multivariate background error correlations. While enabling inter-variable localisation removes sampling noise between variables, it also inadvertently removes these background error correlations. Therefore, it seems reasonable to disable inter-variable localisation (Section 3.2.4) even though the ensemble size is relatively small.

3.4 Experimental set-up and trials

3.4.1 Description of trials

The impact of hybrid-En3DVar on SINGV-DA forecasts was evaluated over June 2019, which featured both localised thunderstorm and large-scale convective occurrences throughout the domain. The initial development work trialed hybrid-En3DVar in

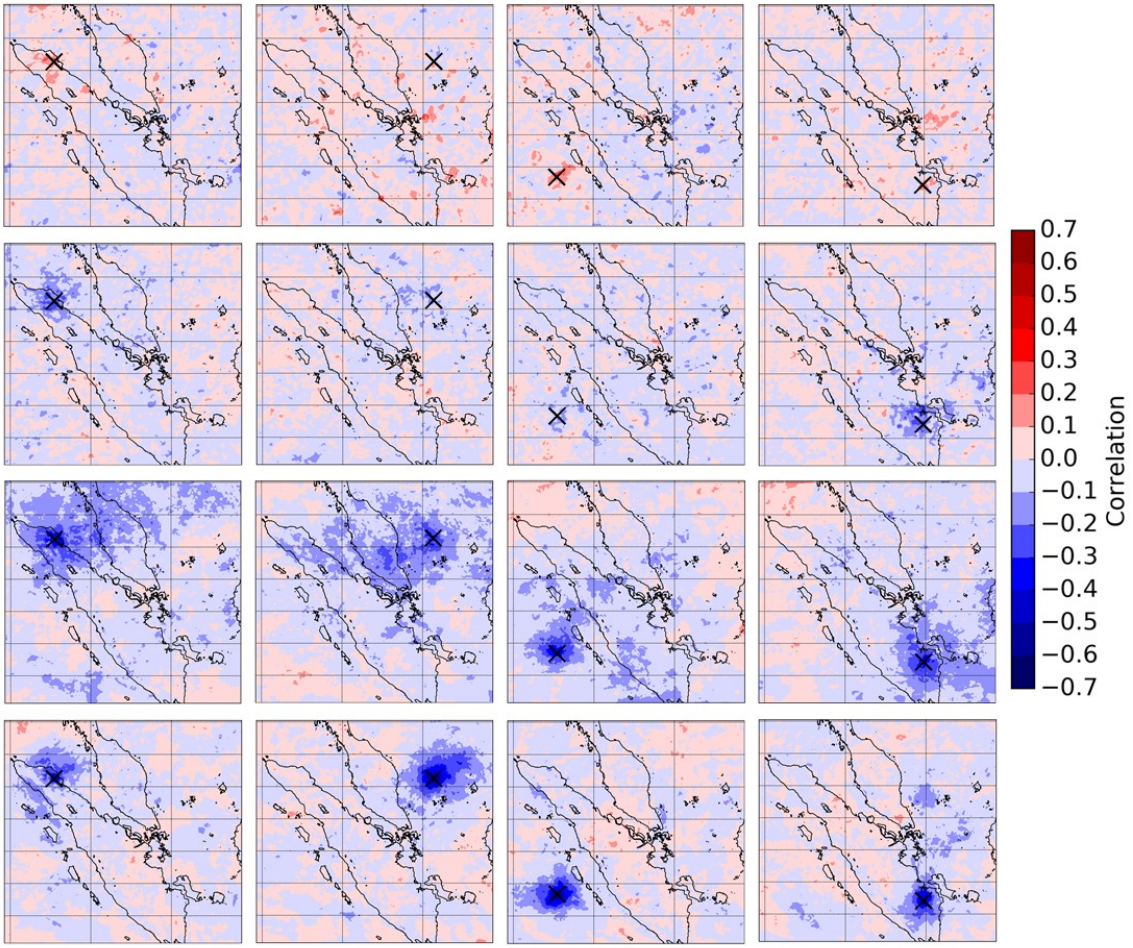

Figure 3.7: Cross correlation of total specific humidity with respect to potential temperature at four different corner locations (black cross) in the domain (columns) at different model levels [(shown from bottom to top) model levels 15, 29, 40, and 49 corresponding to 1-, 4-, 8-, and 12-km height AGL, respectively], using ensemble perturbations from 6-h forecasts ($T + 6$) across all ensemble members and cycles in June 2019 (660 samples).

Table 3.1: Summary of SINGV-DA configurations testing hybrid-En3DVar with different weightings to climatological and ensemble components, and application of time-shifted ensemble perturbations.

Experiment name	Weightings	Time-shifted ensemble perturbations
CTRL (3DVar)	100% $\mathbf{B}_c$	No (11 ensemble perturbations)
EXPT-80C-20E	80% $\mathbf{B}_c$, 20% $\mathbf{B}_e$	No (11 ensemble perturbations)
EXPT-50C-50E	50% $\mathbf{B}_c$, 50% $\mathbf{B}_e$	No (11 ensemble perturbations)
EXPT-20C-80E	20% $\mathbf{B}_c$, 80% $\mathbf{B}_e$	No (11 ensemble perturbations)
EXPT-0C-100E (Pure EnVar)	100% $\mathbf{B}_e$	No (11 ensemble perturbations)
EXPT-100C-80E	100% $\mathbf{B}_c$, 80% $\mathbf{B}_e$	No (11 ensemble perturbations)
EXPT-80C-20E-TS	80% $\mathbf{B}_c$, 20% $\mathbf{B}_e$	Yes (33 ensemble perturbations)
EXPT-50C-50E-TS	50% $\mathbf{B}_c$, 50% $\mathbf{B}_e$	Yes (33 ensemble perturbations)

SINGV-DA using the 11-member SINGV-EPS (Fig. 3.1) with different weightings between $\mathbf{B}_c$ and $\mathbf{B}_e$ (Table 3.1). An additional experiment (EXPT-100C-80E) was included to trial using substantial inflation of the background error statistics, augmenting a fully weighted $\mathbf{B}_c$ with flow-dependent information from $\mathbf{B}_e$. Separate tests showed that SINGV-EPS can be under-dispersive, especially during the peak of the diurnal cycle. Additionally, $\mathbf{B}_c$ variances were originally obtained from the lagged NMC method without any tuning. Thus, EXPT-100C-80E helps to preliminarily assess if larger background error variances in general are desirable for SINGV-DA. We also performed two additional experiments incorporating time-shifted ensemble perturbations based on two of the weighting combinations. In all experiments, where required, the localisation settings follow the description and justification in Sections 3.2.4 and 3.3.2.

3.4.2 Impact on analysis increments

To illustrate the effect of varying the weighting, Fig. 3.8 shows the model level 29 ($\sim$4 km) horizontal cross section of the potential temperature, total specific humidity, and wind analysis increments for the first cycle of the monthlong trials (0300 UTC 1 June 2019). The same first guess has been used in all experiments.

The analysis increments structures in general appear as a combination of the experiments with 100% $\mathbf{B}_c$ or 100% $\mathbf{B}_e$ (i.e., CTRL or EXPT-0C-100E, respectively), which is logically expected. We note that for experiments with higher weightings of $\mathbf{B}_e$, the analysis increments contain smaller scale structures. This was also previously highlighted in Montmerle et al. (2018). Localised values of potential temperature increments, for example, are slightly larger over certain regions (e.g., off the east coast of the Malaysian Peninsula), associated with larger local variances reflecting the ensemble forecast uncertainty over those regions. Most of the analysis increments at this level are

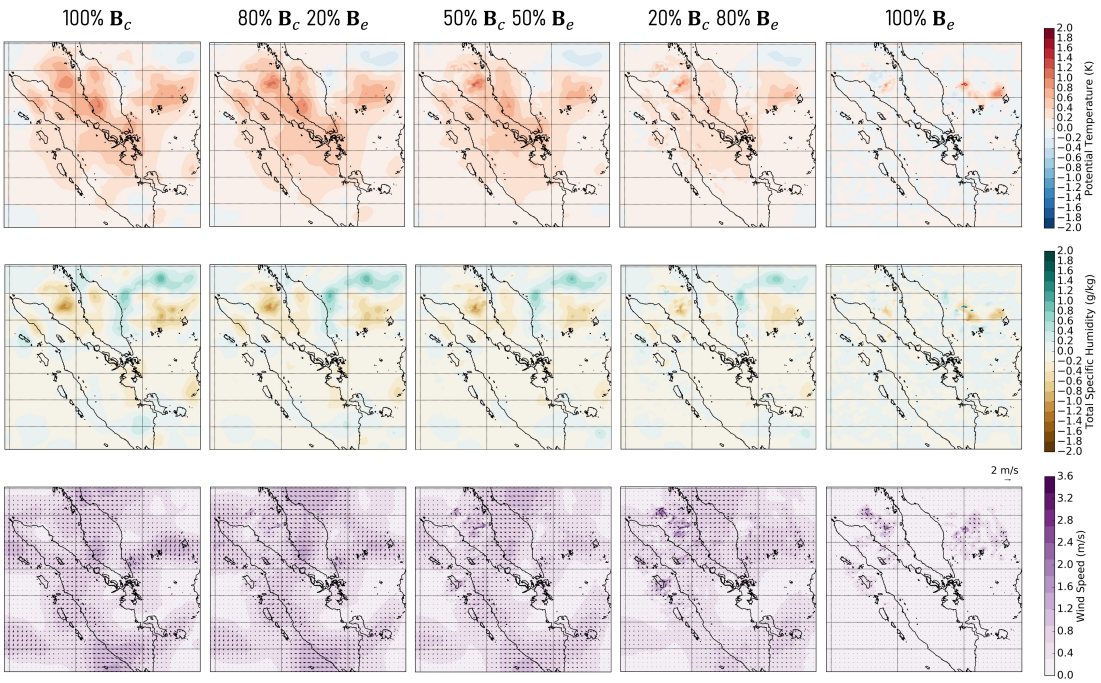

Figure 3.8: Analysis increments of (top) potential temperature, (middle) total specific humidity, and (bottom) horizontal wind at model level 29 (∼4 km) for the first cycle of monthlong trials (0300 UTC 1 Jun 2019), for different weightings between $\mathbf{B}_c$ and $\mathbf{B}_e$ (corresponding to first five experiments in Table 3.1 The same first guess has been used for all experiments. The vector represents the direction and magnitude of the horizontal wind analysis increments.

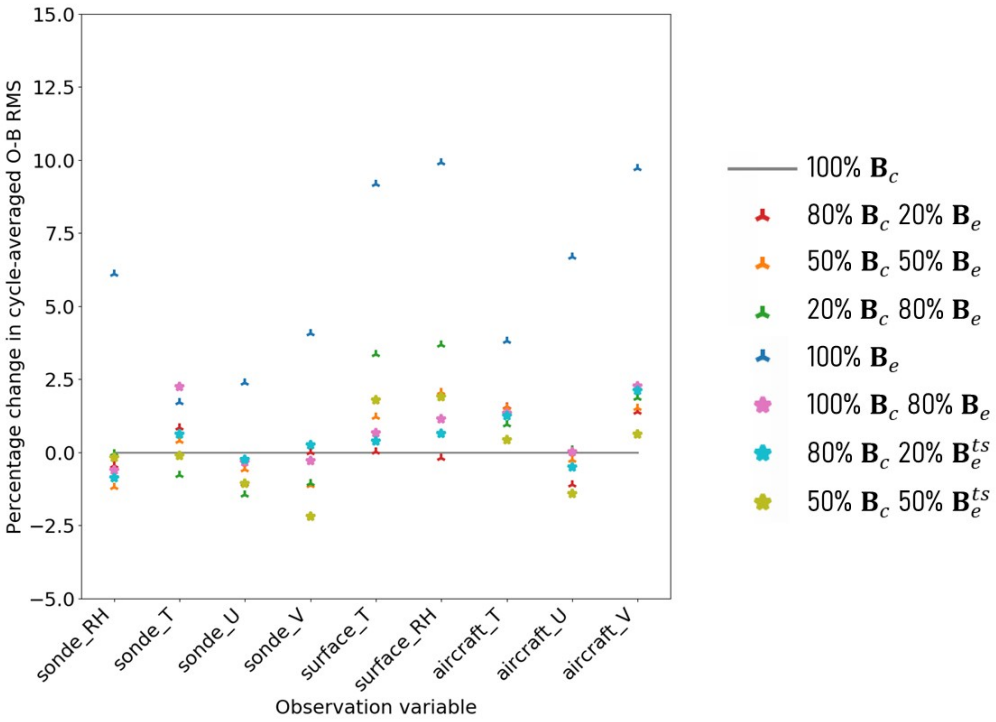

Figure 3.9: Observation minus background root-mean-square (O-B RMS) statistics averaged over the trial period for all experiments as a percentage change in cycle-averaged O-B RMS with respect to CTRL for conventional radiosonde (sonde; vertically averaged), surface, and aircraft observations. The experiment names and weightings are described in Table 3.1. Statistics are computed for relative humidity (RH), temperature (T), zonal wind (U), and meridional wind (V).

due to satellite and aircraft observations, since no radiosonde information is available for this particular cycle.

3.4.3 Verification against conventional observations

Next, we assess the background fit to conventional observations by computing the root-mean-square differences between the observation and background (O-B RMS), averaged over all cycles during the monthlong trials. In most of the experiments, there was a reduction in the O-B RMS for radiosonde relative humidity, zonal and meridional wind, and aircraft zonal wind compared to CTRL (Fig. 3.9). The vertical profiles of differences in O-B RMS compared to CTRL for radiosonde relative humidity, zonal and meridional wind (Fig. 3.10) also highlight this reduction, particularly below model level 29 ($\sim$4 km). For EXPT-0C-100E, the forecast fit to all conventional observations was poorer. This is likely a consequence of applying strict localisation in a data-sparse

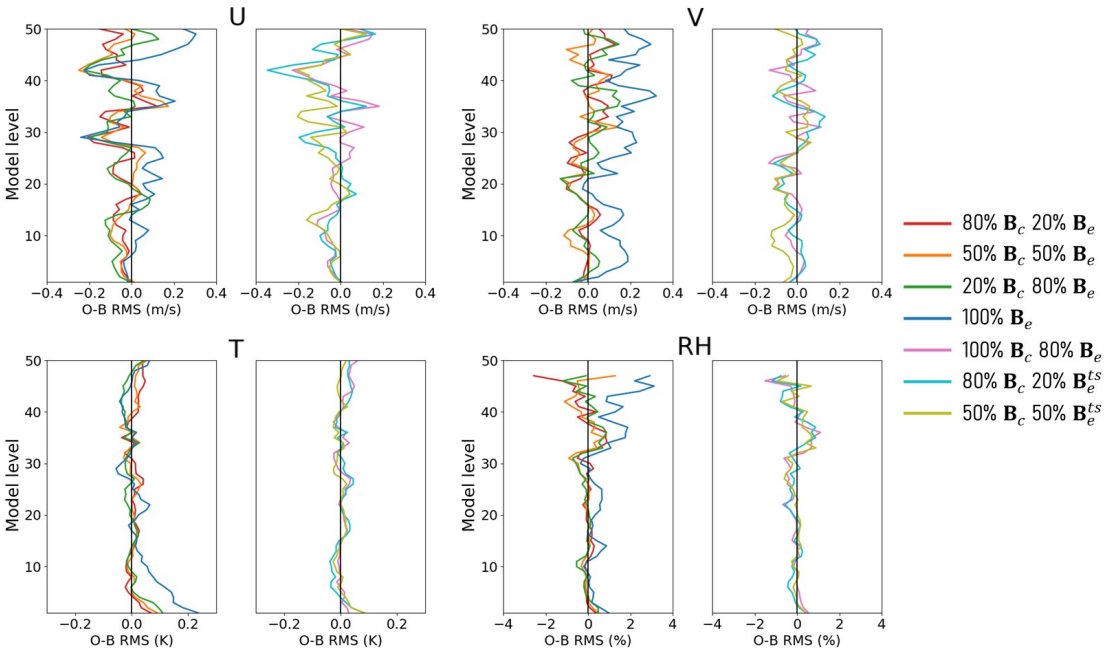

Figure 3.10: Vertical profiles of differences in O-B RMS compared to CTRL for radiosonde variables for all experiments. The experiment names and weightings are described in Table 3.1. Statistics are computed for relative humidity (RH), temperature (T), zonal wind (U), and meridional wind (V).

region such as over the western Maritime Continent, so the analysis increments are very localised and observational information is not well spread throughout the domain by $\mathbf{B}_e$ (Fig. 3.8; rightmost panel).

In all the experiments, there was also an increase in the O-B RMS for surface temperature and relative humidity compared to CTRL. An increase in the weighting toward $\mathbf{B}_e$ appears to result in larger O-B RMS values. The vertical profiles of O-B RMS for radiosonde temperature and relative humidity also show that the O-B RMS is larger near the surface compared to CTRL. Note that this result is not seen in the wind variables; we do not assimilate wind information from surface observations because the prevailing wind in the tropics is weak and noisy. One possible reason for the poorer fit to observations near the surface may be related to the inconsistencies in the ECMWF soil moisture data and the Unified Model soil moisture scheme used in SINGV-EPS, which would impact the ensemble perturbations used in hybrid-En3DVar. This technical change will be explored in future studies.

3.4.4 Verification against satellite-derived precipitation

To assess the impact of hybrid-En3DVar on SINGV-DA precipitation forecasts, we compute the fractions skill score (FSS; Roberts and Lean, 2008) statistics from hourly accumulated precipitation of 60 forecasts (in June 2019) initialised at 0300 and 1500 UTC over the SINGV-DA domain. We have followed the routine verification procedure of the SINGV-DA system, which focuses on 0300 and 1500 UTC because the latest sets of ECMWF LBCs are available only for these two cycles (prior to availability four times a day). We also first focus on the Singapore radar domain since we expect that the impact of hybrid-En3DVar will be more pronounced in regions with more observations (such as in the vicinity of Singapore), given the strict localisation. The FSS are computed using a neighborhood size of 50 km, as a function of eight precipitation thresholds (0.125, 0.25, 0.5, 1, 2, 4, 8, and 16 mm). The verification is performed against the Global Precipitation Mission (GPM) data created with the Integrated Multi-satellitE Retrievals for GPM, Final Precipitation product (GPM_3IMERGHH v06B, Huffman et al., 2019), which is available at $0.1^\circ \times 0.1^\circ$ spatial resolution, and a 30-min temporal resolution.

Over the Singapore radar domain, there is generally an improvement in the precipitation forecasts compared to CTRL in EXPT-80C-20E, EXPT-50C-50E, and EXPT-100C-80E, up to a lead time of 21 hours, particularly for thresholds above 2 mm (Fig. 3.11). Increasing the weighting toward $\mathbf{B}_e$ (including full weight in EXPT-0C-100E) led to larger degradations in the very short-range forecasts, reflecting the immediate effect of the data assimilation. This is likely due to the poorer fits to conventional observations, particularly near the surface, as seen in Fig. 3.9. Incorporating substantial inflation, as done in EXPT-100C-80E in Fig. 3.11, also improved the 9–21-h forecasts, but degraded the forecasts at longer lead times. Overall, EXPT-80C-20E yielded the largest forecast improvements, which is unsurprising, following Hamill and Snyder (2000) who suggested that the optimal combination should be weighted toward $\mathbf{B}_c$ if the ensemble size is small. Previous studies (e.g., Lorenc, 2003, Ménétrier et al., 2015*a*, 2015*b*) have also noted that the suitability of localisation length-scales depends on ensemble size. We found that in this case, our choice of the strict localisation length-scales appears to be appropriate given the small SINGV-EPS ensemble size.

Over the full domain, the precipitation forecasts are improved in EXPT-50C-50E and EXPT-20C-80E compared to CTRL for thresholds above 4 mm, but degraded at very short lead times for thresholds below 4 mm, as also seen in Fig. 3.11. The results for EXPT-80C-20E and EXPT-100C-80E are mixed, with very small degradation

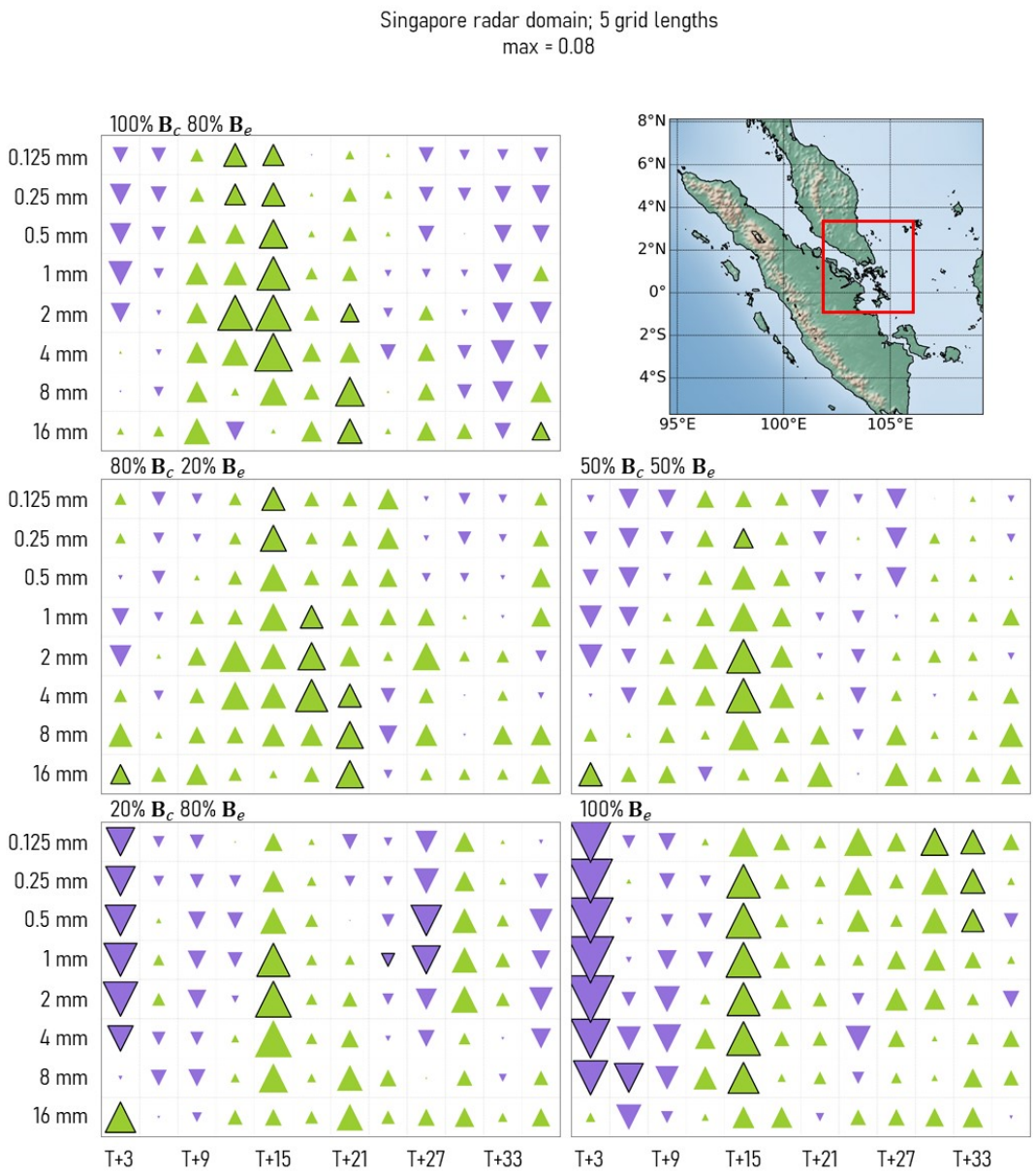

Figure 3.11: Hinton diagrams of fraction skill scores (FSS) computed over the Singapore radar domain (red rectangle in top-right panel) for all experiments (without time-shifted ensemble perturbations) with respect to CTRL, verified against GPM data. A green (purple) triangle indicates that the forecasts are improved (degraded). A larger triangle indicates a greater improvement or degradation, by up to 0.08 (the same size as the bounding box). Significance is determined using the nonparametric two-sided Wilcoxon signed-rank test at the 90% confidence level, indicated using bold triangles.

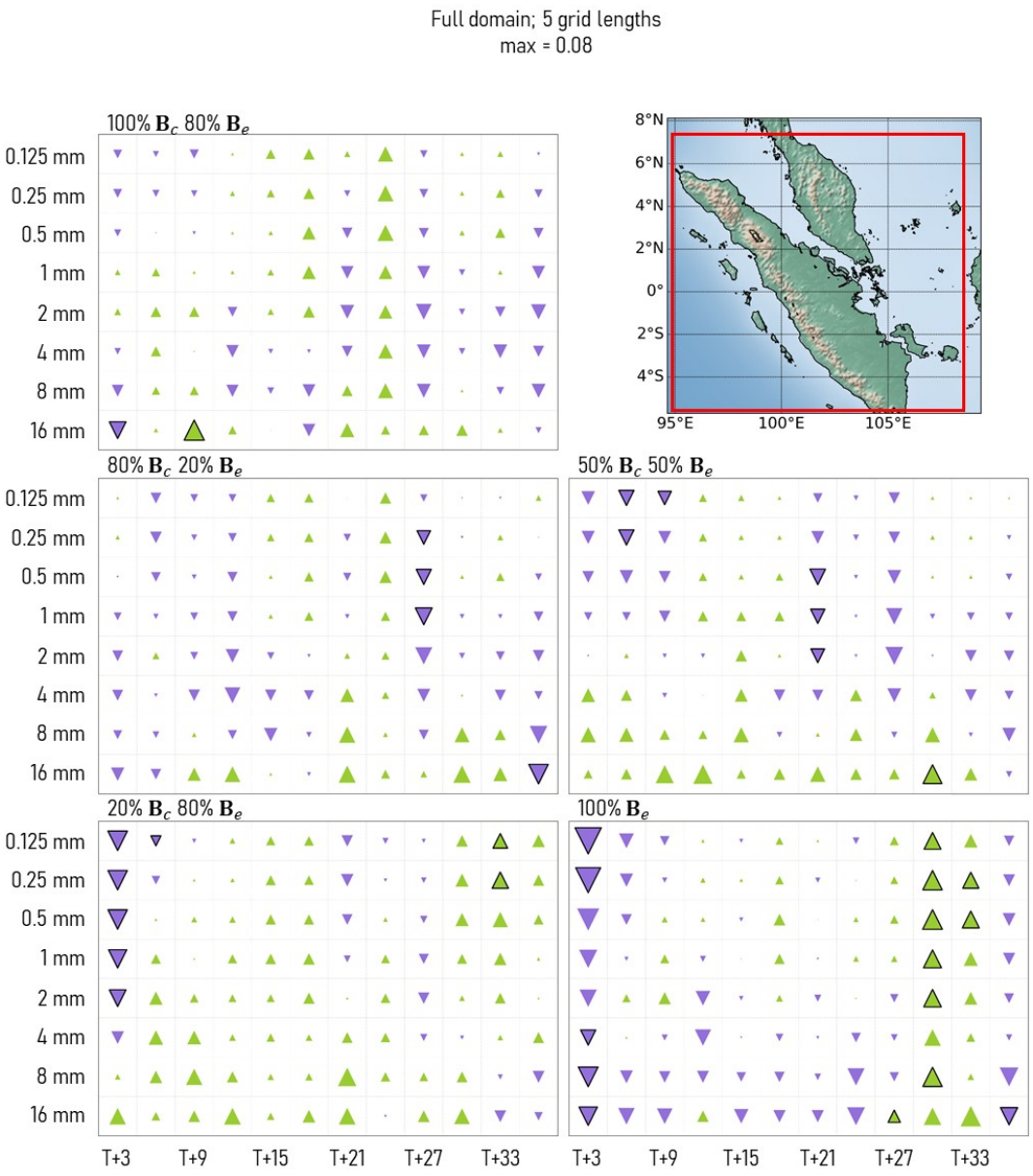

Figure 3.12: As in Fig. 3.11, but over the full domain (red rectangle in top-right panel).

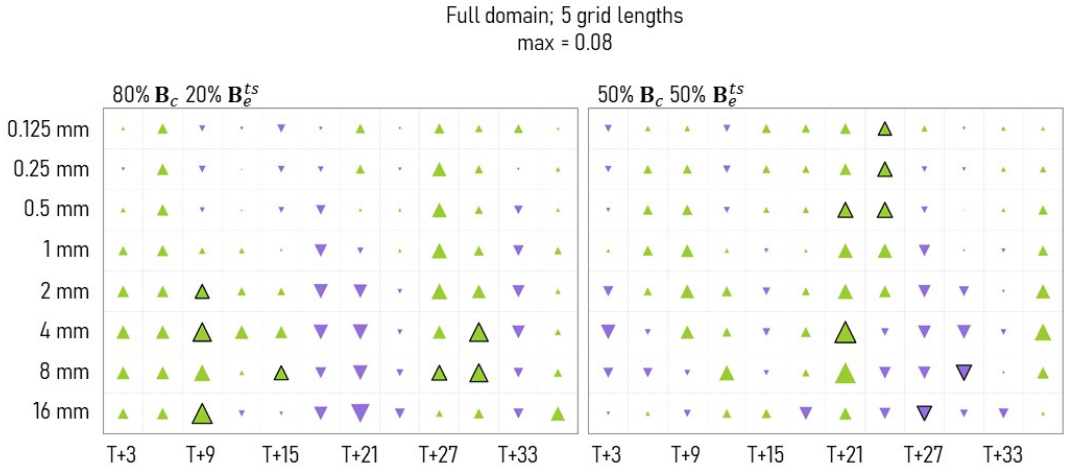

Figure 3.13: As in Fig. 3.11, but over the full domain, and compared with EXPT-80C-20E and EXPT-50C-50E (respective experiment counterparts without time-shifted ensemble perturbations) instead of CTRL.

or improvements in the forecasts across different thresholds and lead times. The differences between Figs. 3.11 and 3.12 suggest that while the forecasts are improved over the Singapore radar domain, the forecasts may be degraded over other parts of the domain, where sampling noise may be more prevalent (especially around Sumatra; not shown) despite the strict localisation.

Sections 3.3.2 and 3.3.3 showed that using time-shifted ensemble perturbations reduces the sampling noise in $\mathbf{B}_e$ and highlighted a robust negative background error correlation between total specific humidity and potential temperature. To assess the impact of using time-shifted ensemble perturbations on the precipitation forecasts, we compare the FSS for EXPT-80C-20E-TS and EXPT-50C-50E-TS with EXPT-80C-20E and EXPT-50C-50E, respectively. Figure 3.13 shows that when time-shifted ensemble perturbations are used, the precipitation forecasts over the whole domain are generally improved. Over the Singapore radar domain, the impact is neutral; EXPT-80C-20E-TS and EXPT-50C-50E-TS yield relatively equal improvements as their experiment counterparts without time-shifted ensemble perturbations, with respect to CTRL (not shown). The improvements when using time-shifted ensemble perturbations agree with the results in Section 6.2.2 of Gustafsson et al. (2014). They commented that the use of time-shifted ensemble perturbations allows for timing errors and spatial phase errors to be represented, which perhaps is critical over convection-dominated regions, such as the western Maritime Continent.

3.4.5 Other experiments and discussion

We have also conducted further experiments that are not included in the results above. As alluded to in Section 3.2.1, the training data for calibrating $\mathbf{B}_c$ was generated using a 4.5-km forecast system, whereas SINGV-DA now uses a 1.5-km horizontal grid spacing. One would expect the forecast errors (and their variances) to be smaller in a higher resolution system, hence we tested reducing the standard deviation of streamfunction and velocity potential in $\mathbf{B}_c$ by 0.8, together with hybrid-En3DVar. This led to further improvements in the precipitation verification scores and forecast fits to conventional observations, and was thus included in subsequent package upgrade trials.

In Section 3.3.2, we noted that there were other objective methods for selecting localisation length-scales, although the approach by trial-and-error is often adopted by many operational centres. We have hence conducted further experiments using different horizontal localisation length-scales of up to 200 km with various weightings. The details are omitted in this article given the large number of possible permutations. None of the other experiments outperformed EXPT-80C-20E-TS and in general, given the small SINGV-EPS ensemble size, stricter localisation yielded better precipitation verification scores.

A common criticism of using time-shifted ensemble perturbations is that the ensemble perturbations are, strictly speaking, not independent samples. One would expect that by omitting cross covariances between ensemble perturbations, the full ensemble-derived covariances would be underestimated, and inflation of the resulting covariances may be required. We have nonetheless opted to test the time-shifting approach for our system, following Gustafsson et al. (2014), Huang and Wang (2018) and Gasperoni et al. (2022). Huang and Wang (2018) had also previously shown that using time-shifted ensemble perturbations improved the Gaussianity of the background ensemble distribution, because time-shifting produces a temporal smoothing effect on the ensemble-derived covariances. They focused on a tropical cyclone case, but for the western Maritime Continent the time-shifted ensemble perturbations may contain diurnal convection signals at different locations, and thus may lead to non-Gaussianity.

3.5 Conclusions

At the Meteorological Service Singapore (MSS), a hybrid ensemble-variational (En3DVar) data assimilation system has been developed to explore incorporating information from an ensemble prediction system into a variational data assimilation system

over the western Maritime Continent. In this initial implementation, the 11-member ensemble prediction system is dynamically downscaled from global ensemble members every 12 hours, with ensemble-derived background error statistics used in the 3-hourly cycling variational data assimilation system.

To understand the ensemble-derived background error statistics, we analysed the structures and raw covariances of the flow-dependent ensemble perturbations from the ensemble prediction system. There exists small-scale error structures associated with positional differences of tropical convection, but these structures are well represented only after the downscaled ensemble forecast has evolved for at least 6 hours due to spinup. This result highlighted a limitation of the initial implementation of the hybrid-En3DVar system, where certain cycles were disadvantaged by the underrepresentation of small scale forecast errors in the ensemble perturbations. Sampling noise was prevalent in the raw autocovariances computed using such a small ensemble, with an estimated horizontal localisation scale of around 50 km suitable for this set-up. We found that time shifting of the ensemble perturbations, by using those available from adjacent cycles, helped to ameliorate sampling error.

We also discovered a robust and moderate negative correlation between total specific humidity and potential temperature background errors which was confined to the lower troposphere. The negative covariance was also captured by the control variable transform in 3DVar, but when using the alpha control variable transform with the ensemble-derived covariance, the relationship was weaker and more localised. We postulate that this robust relationship is associated with incorrect vertical motion in the presence of clouds.

Monthlong trials in June 2019 were conducted to assess the impact of hybrid-En3DVar on the analysis increments, forecast fits to observations and precipitation forecasts. Multiple trials were conducted using different weights assigned to the ensemble-derived and climatological background error covariances, respectively. The analysis increments generally contained smaller scale structures and had larger localised values reflecting the larger forecast uncertainty over certain regions as the weighting toward the ensemble-derived background error covariances increased. The forecast fits to radiosonde relative humidity and wind observations were generally improved with hybrid-En3DVar, but in all experiments, the forecast fits to surface temperature and relative humidity observations were degraded compared to the baseline 3DVar configuration. Over the Singapore radar domain, there was a general improvement in

the precipitation forecasts, particularly for thresholds above 2 mm, compared to the baseline 3DVar configuration, especially when the weighting toward the climatological background error covariance was larger (e.g., 50% or 80% weight). However, the results were mixed over the full domain, possibly because sampling noise was more prevalent over other parts of the domain (e.g., Sumatra). This issue was mediated by the application of time-shifted ensemble perturbations, which then led to an improvement in the precipitation forecasts instead. Overall, the experiment using a weighting of 80% climatological, 20% ensemble-derived background error covariances, with time-shifted ensemble perturbations, yielded the best verification scores. These results are encouraging, given the simple initial implementation where SINGV-EPS is not centred on the SINGV-DA analysis and is uninformed of the SINGV-DA observation network.

Future work involves consolidating the ensemble prediction system and the deterministic system by centring the ensemble prediction system on the hybrid analysis. This should avoid spinup issues and better represent the analysis and forecast uncertainties since in the consolidated system, the 3-h ensemble forecasts centred on the SINGV-DA analysis are available for all cycles. The ensemble analysis perturbations can also be generated using various ensemble approaches (e.g., bred vectors, ensemble of 3DVar).

It would also be interesting to explore the error structures in the ensemble perturbations during other seasons, such as the northeast monsoon. Previous investigations have identified other error structures reminiscent of a sea breeze off the coast of Sumatra, but these were present only after full onset of the southwest monsoon (in July; not shown, and in September; Lee and Huang, 2022). It would be interesting to see if capturing this flow-dependent information provides the same benefit on precipitation forecasts.

We are also considering reducing the ensemble horizontal grid spacing from 4.5 to 2.2 km in the future. This brings the current SINGV-EPS horizontal grid spacing closer to that in SINGV-DA. Currently, ensemble forecasts are interpolated from 4.5- to the 2.83-km SINGV-DA variational assimilation grid, which may introduce unintended smoothing effects. However, there is also no guarantee that reducing the ensemble horizontal grid spacing will drastically improve the analysis. Feng and Wang (2021) previously showed that there was a larger positive impact on the analysis when the horizontal grid spacing of the first guess (from the deterministic system) compared to the ensemble forecasts is reduced. Therefore, we may instead choose to reduce the horizontal grid spacing of the first guess in further trials with hybrid-En3DVar.

3.6 Improving the localisation design within the ensemble-variational approach

The results from Chapter 2 highlighted potential challenges in representing the mass-wind error cross-covariances — whether through the geostrophic balance constraint or directly from the ensemble — in the tropics. The geostrophic balance constraint did not appear to be helpful for tropical data assimilation. However, it was unclear if the ensemble-derived error cross-covariances contained useful balances that were beneficial for convective-scale data assimilation. One crude approach was to simply remove all the error cross-covariances through the design of localisation, but this could introduce imbalances in the analysis.

Chapter 3 also showed that there were certain robust ensemble-derived multivariate error cross-covariances present over the Maritime Continent, which one might not want to knock-out through localisation. This suggests that a selection of multivariate error cross-covariances to retain or knock-out could be helpful for ensemble-variational data assimilation over the Maritime Continent. Chapter 3 also showed how each variable had a different power spectrum, indicating that the localisation length-scales for each variable should differ over the Maritime Continent.

In this light, further research on the modifications to improve the localisation design was conducted. This is covered in Chapter 4.

Chapter 4

Variable-dependent and selective multivariate localisation for ensemble-variational data assimilation in the tropics

This chapter concerns RQ2 posed in Section 1.5 and has been published in *Monthly Weather Review* with the following reference:

Lee, J.C.K., Amezcua, J. and Bannister, R.N., 2024. Variable-dependent and selective multivariate localisation for ensemble-variational data assimilation in the tropics. *Monthly Weather Review*, **152(4)**, pp. 1097-1118, https://doi.org/10.1175/MWR-D-23-0201.1.

It is unmodified from the published manuscript, other than being re-formatted in accordance with the thesis chapters and with minor typographical adjustments to maintain consistency throughout the thesis.

Abstract

Two aspects of ensemble localisation for data assimilation are explored using the simplified non-hydrostatic ABC model in a tropical setting. The first aspect (i) is the ability to prescribe different localisation length-scales for different variables (variable-dependent localisation). The second aspect (ii) is the ability to control (i.e., to knock-out by localisation) multivariate error covariances (selective multivariate localisation). These aspects are explored in order to shed light on the cross-covariances that are important in the

tropics and to help determine the most appropriate localisation configuration for a tropical ensemble-variational (EnVar) data assimilation system. Two localisation schemes are implemented within the EnVar framework to achieve (i) and (ii). One is called the isolated variable-dependent localisation scheme (IVDL) and the other is called the symmetric variable-dependent localisation (SVDL) scheme. Multi-cycle Observation System Simulation Experiments are conducted using IVDL or SVDL mainly with a 100-member ensemble, although other ensemble sizes are studied (between 10 and 1000 members). The results reveal that selective multivariate localisation can reduce the cycle-averaged root-mean-square error (RMSE) in the experiments when cross-covariances associated with hydrostatic balance are retained and when zonal wind/mass error cross-covariances are knocked-out. When variable-dependent horizontal and vertical localisation are incrementally introduced, the cycle-averaged RMSE is further reduced. Overall, the best performing experiment using both variable-dependent and selective multivariate localisation leads to a 3-4% reduction in cycle-averaged RMSE compared to the traditional EnVar experiment. These results may inform the possible improvements to existing tropical numerical weather prediction systems which use EnVar data assimilation.

4.1 Introduction

Ensemble-variational (EnVar) data assimilation methods have recently gained traction and have been widely tested in several operational global and regional numerical weather prediction (NWP) systems (Buehner et al., 2013, Clayton et al., 2013, Wang et al., 2013, Gustafsson et al., 2014, Hu et al., 2017, Montmerle et al., 2018, Singh and Prasad, 2019, Bédard et al., 2020, Kadowaki et al., 2020), and in research case studies focusing on extreme weather events (Schwartz et al., 2013, Shen et al., 2016, Lu et al., 2017, Gao et al., 2019, Kutty et al., 2020). The main idea relies on using ensemble-derived background error statistics to replace the climatological error statistics used in the variational approach. Often, a hybrid-EnVar approach is adopted by weighting the ensemble-derived and climatological error statistics with respective weights that depend on the ensemble size (Hamill and Snyder, 2000). Where the ensemble is sufficiently large, one might solely rely on ensemble-derived error statistics by placing full weight on it in the variational algorithm.

Most studies reported a benefit from using EnVar data assimilation, either in the hybrid or pure EnVar form, as opposed to traditional three-dimensional or four-dimensional variational (3D-Var or 4D-Var) approaches. They attributed the benefit broadly to the flow-dependency introduced by the ensemble-derived error statistics.

This flow-dependency generically encompasses time-appropriateness of error variances, flow-consistency of spatial covariances, as well as flow-consistency of multivariate error relationships (i.e., cross-covariances between different variables in the model). The benefit stemming from each component has yet to be clearly distinguished. Shen et al. (2016) and Lu et al. (2017) highlighted through tropical cyclone case studies that using ensemble-derived error statistics yielded more realistic multivariate error cross-covariances, particularly in the vicinity of the cyclone vortex. Johnson et al. (2015) and Gao et al. (2019) found, through case studies over the United States and China, that using ensemble-derived error statistics resulted in more dynamically coherent analyses of mesoscale convective systems. In these case studies, the flow-consistency of multivariate error relationships from the ensemble-derived error statistics was found to be a key contributor leading to the improvements in the quality of the analysis and subsequent forecasts.

While capturing appropriate multivariate error relationships may help to retrieve a dynamically consistent and balanced analysis, sampling noise may also contaminate the error covariances. This is because the ensemble size is usually far smaller than the degrees of freedom of the state, and therefore the estimated error covariance matrix will be rank-deficient. Houtekamer and Mitchell (2001) suggested using a Schur product of a correlation matrix (referred to as a localisation matrix) with the ensemble-derived background error covariance matrix to apply spatial covariance localisation and mitigate spurious long-range correlations. This is widely adopted in traditional EnVar implementations (e.g., in Wang et al. 2008*a*), especially in most operational weather prediction centres adopting EnVar techniques. However, there are two potential limitations with current traditional approaches. Firstly, the same spatial localisation is usually applied to all variables, irrespective of their characteristic length-scales associated with the system dynamics. For example, Huang et al. (2021) and Caron and Buehner (2022) specified variable-independent localisation scales. This assumption of the same spatial localisation was shown to be rather unrealistic (Lei et al., 2015), especially at convective scales (Destouches et al., 2021, Necker et al., 2023). The error cross-covariance localisation should ideally also reflect a mix of the characteristic length-scales of the variables involved, but again this is not usually done in traditional EnVar implementations. Secondly, in traditional EnVar schemes, the ability to do multivariate localisation (the knocking-out of correlations between variables) is not currently implemented, so multivariate error relationships between all variables are determined by the spatially localised ensemble.

In the absence of a reasonable physically-based constraint, some of the ensemble-derived relationships may be useful, as seen in tropical cyclone case studies (Shen et al., 2016, Lu et al., 2017) and over Southeast Asia (Lee and Barker, 2023; Chapter 3). However, other cross-covariances and their characteristic length-scales may not be well-represented by a limited-size ensemble, and are likely to be dominated by sampling noise. It is possible that these cross-covariances are non-informative and may predominantly be introducing noise to the analysis, yet cannot be removed using traditional localisation frameworks.

In the tropics, the disadvantages of the above mentioned two potential limitations may become more apparent given the nature of convective weather and less balanced flow in the region. One would desire, for instance, the flexibility to prescribe smaller localisation length-scales for convection-related variables, which may typically involve vertical wind and hydrometeors, following Destouches et al. (2021). Additionally, appropriate multivariate localisation may also be required since it is not trivial to specify multivariate error relationships for the tropics, particularly between mass (e.g., temperature, pressure) and wind variables. Some operational systems prescribe geostrophic balance or linear balance in their background error covariance model (e.g., Lorenc et al. 2000), but in the tropics, this procedure effectively treats the mass and wind variables univariately since the Coriolis parameter is small there. It remains to be seen if all multivariate error relationships estimated by an ensemble are physically meaningful or even required in the tropics. In this light, an improved EnVar implementation for the tropics should allow for *variable-dependent localisation* (a concept suggested by Necker et al. 2020) and a way to constrain the multivariate error relationships (keeping some cross-covariances and not others), which we term as *selective multivariate localisation* (this concept is similar to that in Kang et al. 2011 for ensemble Kalman filters). Notwithstanding this, neither have been explored in the tropics yet.

To this end, one possible modification to traditional EnVar is to use scale-dependent localisation (Buehner and Shlyaeva 2015; Huang et al. 2021; Caron and Buehner 2022), which allows the localisation length-scales at different scales to be specified independently. Therefore, the large-scale and small-scale errors are allowed to have different error characteristics. However, in these studies, for each scale, all variables still share the same localisation length-scales. Wang and Wang (2023) further extended this to include both variable-dependent localisation and scale-dependent localisation, for a few tornadic supercell case studies over the United States. They

introduced an approach to modify traditional EnVar and found that further applying variable-dependent localisation was beneficial to see storm maintenance. Another possible modification was proposed by Stanley et al. (2021) to construct separate localisation functions for the multivariate error cross-covariances (within a bivariate Lorenz 96 system with coupled data assimilation), but this has yet to be applied to the EnVar framework. Additionally, the approach does not allow one to select specific cross-covariances to be retained.

In this study, we implement and explore two approaches to grant the ability to apply variable-dependent and selective multivariate localisation within the EnVar data assimilation framework. The first approach is termed as the *isolated variable-dependent localisation scheme (IVDL)*. The second approach is termed as the *symmetric variable-dependent localisation scheme (SVDL)*. Details are given in Section 4.2, along with their similarities or novelties vis-à-vis existing schemes. Both schemes allow for variable-dependent localisation; IVDL is more computationally efficient, but also less flexible than SVDL. By design, the IVDL scheme implicitly determines the multivariate localisation, while the SVDL scheme explicitly prescribes the multivariate localisation on top of spatial localisation. This study aims to answer the following questions:

1. How many ensemble members are sufficient to show a significant degree of 'signal' in the covariances, but still benefit from localisation of sampling noise?

2. Which multivariate error relationships in the ensemble-derived error covariances are important/beneficial for EnVar data assimilation in the tropics?

3. Is variable-dependent spatial localisation beneficial for EnVar data assimilation in the tropics?

Section 4.2 describes the design and implementation of IVDL and SVDL schemes. A simplified non-hydrostatic convective-scale model, the ABC model (Petrie et al., 2017), is used for this study. A tropical configuration of the ABC model (a longitude/height dry model, without diabatic processes) with data assimilation is set up to demonstrate the two schemes. Section 4.3 provides further details on the model and data assimilation framework. Section 4.4 describes the experiments and gives guidance on the ensemble size (Question 1). When it comes to deciding on a suitable ensemble size, we consider linear independence of the ensemble members and sampling error. Section 4.5 evaluates the two schemes through empirical experiments. Due to the cheaper computational cost, we use only the IVDL scheme to explore selective multivariate localisation by controlling which multivariate error relationships are retained (Question 2), but we use both schemes

to explore variable-dependent spatial localisation (Question 3). Section 4.6 discusses and concludes the results of this study.

4.2 Variable-dependent and selective multivariate localisation applied with IVDL and SVDL

This work builds on the implementation in Lee et al. (2022), hereafter L22 (Chapter 2), who introduced hybrid-EnVar data assimilation via the alpha control variable approach (Lorenc, 2003) for the ABC model (Section 4.3). We shall follow their notation for consistency. For the pure EnVar approach, a given alpha control variable transform $\mathbf{U}^{\alpha}$ acts on an alpha control vector $\boldsymbol{\chi}^{\alpha k}$ to give an alpha field (i.e., $\boldsymbol{\alpha}^{k} = \mathbf{U}^{\alpha}\boldsymbol{\chi}^{\alpha k}$), which controls the linear combination of ensemble perturbations $\mathbf{x}_t'^{k}$. There is one alpha control variable (and hence one alpha field) per ensemble perturbation member. The analysis increment $\delta\mathbf{x}$ at time t is then:

$$\delta\mathbf{x} = \sum_{k=1}^{N} \mathbf{x}_t'^{k} \circ \boldsymbol{\alpha}^{k}, \tag{4.1}$$

where N is the number of ensemble members, k is the ensemble member index, and $\circ$ is the Schur product. The implied localisation matrix $\mathbf{L}$ in the variational algorithm is $\mathbf{L} = \mathbf{U}^{\alpha}\mathbf{U}^{\alpha\top}$, so changing the design of $\mathbf{U}^{\alpha}$ is key to controlling the application of variable-dependent and selective multivariate localisation. Following Eq. (15) from L22 for the pure EnVar approach, the implied background error covariance matrix is therefore:

$$\mathbf{B}_e = (\mathbf{U}^{\alpha}\mathbf{U}^{\alpha\top}) \circ (\mathbf{X}_t^{f}\mathbf{X}_t^{f\top}), \tag{4.2}$$

where $\mathbf{X}_t^{f}$ is the matrix whose columns contain the ensemble perturbations $\mathbf{x}_t'^{k}$ divided by $\sqrt{N-1}$.

4.2.1 The isolated variable-dependent localisation scheme (IVDL)

We first describe the implementation of IVDL. In Section 2.6.2, a proof of equivalence between their approach in designing $\mathbf{U}^{\alpha}$ and the traditional EnVar approach of Wang et al. (2008*a*) — who presented it slightly differently — is provided. L22 further showed that their choice of $\mathbf{U}^{\alpha}$ does full inter-variable localisation (where no multivariate error cross-covariances are retained). In Wang et al. (2008*a*), the length of $\boldsymbol{\chi}^{\alpha k}$ for each ensemble member k is given by the number of horizontal gridpoints (N_{g}), whereas in

L22, the length of $\boldsymbol{\chi}^{\alpha k}$ is further multiplied by the number of prognostic variables (N_{var}). This obviously influences the dimensions of $\mathbf{U}^{\alpha}$ (the number of columns of $\mathbf{U}^{\alpha}$ must match the length of $\boldsymbol{\chi}^{\alpha k}$ and the number of rows must match the length of the state perturbations $\mathbf{x}_t'^k$ for the Schur product in Eq. (4.1)). Here, we reproduce parts of the L22 proof to illustrate how variable-dependent and selective multivariate localisation can be implemented. For an arbitrary state with three one-dimensional (horizontal) physical variables (e.g., p, q, r) per gridpoint ($\mathbf{x} \in \mathbb{R}^{3N_{\mathrm{g}}}$), the alpha control variable transform implied by Wang et al. (2008*a*) can be mathematically represented by the choice $\mathbf{U}^{\alpha} = \tilde{\mathbf{U}}^{\alpha}$:

$$\tilde{\mathbf{U}}^{\alpha} = \begin{bmatrix} \mathbf{U}_p^{\alpha} \\ \mathbf{U}_q^{\alpha} \\ \mathbf{U}_r^{\alpha} \end{bmatrix}, \tag{4.3}$$

in contrast to the approach in L22, which takes the choice $\mathbf{U}^{\alpha} = \hat{\mathbf{U}}^{\alpha}$:

$$\hat{\mathbf{U}}^{\alpha} = \begin{bmatrix} \mathbf{U}_p^{\alpha} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_q^{\alpha} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{U}_r^{\alpha} \end{bmatrix}. \tag{4.4}$$

Here, $\mathbf{0}$ is an $N_{\mathrm{g}} \times N_{\mathrm{g}}$ null matrix, $\mathbf{U}_p^{\alpha}$, $\mathbf{U}_q^{\alpha}$, and $\mathbf{U}_r^{\alpha}$ can each have the form of an eigenvector matrix scaled by the square-root of the eigenvalue matrix associated with the eigendecomposition of a spatial correlation (localisation) matrix for a specified length-scale (h^{α}). For example, $\mathbf{U}_r^{\alpha} = \mathbf{F}_r \boldsymbol{\Lambda}_r^{1/2}$ where $\mathbf{F}_r$ contains the eigenvectors and $\boldsymbol{\Lambda}_r$ contain the eigenvalues for variable r. For this study, we have used a Gaspari-Cohn localisation function (Gaspari and Cohn 1999; see L22 for details) to prescribe the correlation matrix.

Note that Eq. (4.3) presents a mathematically consistent interpretation of the approach in Wang et al. (2008*a*). In their implementation, they use recursive filters instead of the eigen-approach mentioned above and apply the same transform to each variable, i.e., $\mathbf{U}_p^{\alpha} = \mathbf{U}_q^{\alpha} = \mathbf{U}_r^{\alpha}$. This precludes the possibility of using a different length-scale for each variable (although this can be relaxed if required). In the L22 approach however, a different spatial correlation matrix for each variable is used to achieve variable-dependent localisation, but forces full multivariate localisation. Note that extra memory cost is required to store the eigenvectors and eigenvalues for each variable. In our implementation, this is the case even if, e.g., two variables share the same h^{α}; the same eigenvectors and eigenvalues are stored twice — once for each variable.

Next, we show how selective multivariate localisation can be achieved in the IVDL scheme. Equation (4.4) shows the most primitive form of $\mathbf{U}^{\alpha}$. This can be considered one limiting/extreme case where full inter-variable localisation is implied because of its design. In L22, they further highlighted that it was possible to extend Eq. (4.4) to include selective multivariate localisation by introducing a mapping matrix $\hat{\mathbf{I}}$ comprising scaled blocks of either null or identity matrices ($\hat{\mathbf{U}}^{\alpha}\hat{\mathbf{I}}$, see below). We refer to this extension using variants of $\hat{\mathbf{I}}$ as selective multivariate localisation because we can select which block matrices of $\hat{\mathbf{I}}$ are identity matrices, and which are null matrices. If all block matrices in $\hat{\mathbf{I}}$ are identity matrices, we get the other limiting/extreme case where all error cross-covariances are retained (no inter-variable localisation). Here, $\hat{\mathbf{I}}$ is given in this case by:

$$\hat{\mathbf{I}} = \frac{1}{\sqrt{3}} \begin{bmatrix} \mathbf{I}_{N_{\mathrm{g}}} & \mathbf{I}_{N_{\mathrm{g}}} & \mathbf{I}_{N_{\mathrm{g}}} \\ \mathbf{I}_{N_{\mathrm{g}}} & \mathbf{I}_{N_{\mathrm{g}}} & \mathbf{I}_{N_{\mathrm{g}}} \\ \mathbf{I}_{N_{\mathrm{g}}} & \mathbf{I}_{N_{\mathrm{g}}} & \mathbf{I}_{N_{\mathrm{g}}} \end{bmatrix}, \tag{4.5}$$

where $\mathbf{I}_{N_{\mathrm{g}}}$ is the $N_{\mathrm{g}} \times N_{\mathrm{g}}$ identity matrix, and 3 is the number of variables whose cross-covariances are retained (all $N_{\mathrm{var}} = 3$ variables in this case). We can then choose $\mathbf{U}^{\alpha}$ to be given by:

$$\mathbf{U}^{\alpha} = \hat{\mathbf{U}}^{\alpha}\hat{\mathbf{I}} = \frac{1}{\sqrt{3}} \begin{bmatrix} \mathbf{U}^{\alpha}_{p} & \mathbf{U}^{\alpha}_{p} & \mathbf{U}^{\alpha}_{p} \\ \mathbf{U}^{\alpha}_{q} & \mathbf{U}^{\alpha}_{q} & \mathbf{U}^{\alpha}_{q} \\ \mathbf{U}^{\alpha}_{r} & \mathbf{U}^{\alpha}_{r} & \mathbf{U}^{\alpha}_{r} \end{bmatrix}. \tag{4.6}$$

One can compute the implied localisation matrices $\mathbf{L}$ using $\tilde{\mathbf{U}}^{\alpha}$ and $\hat{\mathbf{U}}^{\alpha}\hat{\mathbf{I}}$, from Wang et al. (2008*a*) and L22 ($\tilde{\mathbf{U}}^{\alpha}\tilde{\mathbf{U}}^{\alpha\top}$ and $\hat{\mathbf{U}}^{\alpha}\hat{\mathbf{I}}\hat{\mathbf{I}}\hat{\mathbf{U}}^{\alpha\top}$ respectively) to see that they are equivalent (proven element-wise in Section 2.6.2).

Now consider a variant of $\hat{\mathbf{I}}$ where only p and q cross-covariances are retained in the localisation scheme, $\hat{\mathbf{I}}$ is then given by:

$$\hat{\mathbf{I}} = \begin{bmatrix} \frac{1}{\sqrt{2}}\mathbf{I}_{N_{\mathrm{g}}} & \frac{1}{\sqrt{2}}\mathbf{I}_{N_{\mathrm{g}}} & \mathbf{0} \\ \frac{1}{\sqrt{2}}\mathbf{I}_{N_{\mathrm{g}}} & \frac{1}{\sqrt{2}}\mathbf{I}_{N_{\mathrm{g}}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{1}{\sqrt{1}}\mathbf{I}_{N_{\mathrm{g}}} \end{bmatrix}. \tag{4.7}$$

Here, we have partitioned the variables into two groups with two and one parameters respectively, where variables p and q are allowed to be correlated in the assimilation, but each is uncorrelated with r. With this setup, many permutations of selective multivariate localisation are possible, depending on how the sets are determined. Each set of variables is treated independently in $\boldsymbol{\chi}^{\alpha k}$ (i.e., as a partition) by knocking out selected multivariate

error cross-covariances. We can verify that the implied localisation matrix $\mathbf{L}$ of Eq. (4.7) is given by:

$$\mathbf{L} = \hat{\mathbf{U}}^{\alpha}\hat{\mathbf{I}}\hat{\mathbf{I}}\hat{\mathbf{U}}^{\alpha\top} = \begin{bmatrix} \mathbf{U}_p^{\alpha}\mathbf{U}_p^{\alpha\top} & \mathbf{U}_p^{\alpha}\mathbf{U}_q^{\alpha\top} & \mathbf{0} \\ \mathbf{U}_q^{\alpha}\mathbf{U}_p^{\alpha\top} & \mathbf{U}_q^{\alpha}\mathbf{U}_q^{\alpha\top} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{U}_r^{\alpha}\mathbf{U}_r^{\alpha\top} \end{bmatrix}, \tag{4.8}$$

demonstrating how selective multivariate localisation can be achieved. This approach can be extended in an obvious way to grouping/isolating any number of variables.

One should also note that with the current approach — using eigenvectors decomposed from correlation matrices — special care must be taken when dealing with periodic domains since the correlation matrix is circulant and may not be positive semi-definite. The implication is that if the localisation length-scales for p and q are different and any eigenvectors associated with negative eigenvalues are truncated, $\mathbf{U}_p^{\alpha}\mathbf{U}_q^{\alpha\top}$ and $\mathbf{U}_q^{\alpha}\mathbf{U}_p^{\alpha\top}$ (the associated off-diagonal blocks) will not strictly be cross-correlation matrices (not shown). The off-diagonal block matrices with this extra symmetry is explored with SVDL in the next section.

Here, we have described IVDL in detail to provide clarity on how one might technically implement it in an NWP system. This approach is outlined in Wang and Wang (2023), and is referred to as Basic-SDLVDL[3] (with scale-dependent aspects in their case), but they implemented and tested a variant MinorE-SDLVDL[4] instead. Mathematically, Basic-SDLVDL and MinorE-SDLVDL are equivalent (Wang and Wang, 2023). Prior to this study, L22 had already discussed the technical implementation of IVDL (although not named IVDL then) and the approach to apply alpha fields to one or all variables, along with the proof of equivalence. Another more computationally efficient approach has since been proposed by Menetrier (https://doi.org/10.5281/zenodo.7547230).

4.2.2 The symmetric variable-dependent localisation scheme (SVDL)

We note that while the full localisation matrix is always symmetric, the off-diagonal blocks of $\mathbf{L}$ are not themselves symmetric if the localisation length-scales (and hence transforms) for p and q are different (i.e., $\mathbf{U}_p^{\alpha}\mathbf{U}_q^{\alpha\top} \neq \mathbf{U}_q^{\alpha}\mathbf{U}_p^{\alpha\top}$). Buehner and Shlyaeva (2015) also found the asymmetry in their between-scale cross-covariances when

[3]This naming is adopted by Wang and Wang (2023) and stands for Basic Scale-Dependent localisation Variable-Dependent localisation.

[4]This stands for Minor Extension Scale-Dependent localisation Variable-Dependent localisation.

applying scale-dependent localisation. This is not necessarily a criticism of the existing approaches, but is rather a prompt to pose the question if imposing extra symmetry in $\mathbf{U}^{\alpha}$ could improve the performance of a multivariate localisation scheme.

In SVDL, the full localisation matrix is specified explicitly:

$$\mathbf{L} = \begin{bmatrix} \mathbf{L}_{p,p} & \mathbf{L}_{p,q} & \mathbf{L}_{p,r} \\ \mathbf{L}_{q,p} & \mathbf{L}_{q,q} & \mathbf{L}_{q,r} \\ \mathbf{L}_{r,p} & \mathbf{L}_{r,q} & \mathbf{L}_{r,r} \end{bmatrix}, \tag{4.9}$$

and this is made, by construction, to have all block correlation matrices symmetric, e.g., $\mathbf{L}_{p,q} = \mathbf{L}_{p,q}^{\top}$. SVDL can be considered a 'brute force' approach — all block correlation matrices (including off-diagonal ones) within Eq. (4.9) are fully prescribed. The eigendecomposition is then performed on the full localisation matrix (instead of blocks of it like in IVDL) to retrieve $\mathbf{U}^{\alpha} = \mathbf{L}^{\frac{1}{2}}$ to use in the variational algorithm. Due to the application of the eigendecomposition on the full localisation matrix, the computational cost of the SVDL approach is estimated to be $\mathcal{O}([N_{\mathrm{g}}N_{\mathrm{var}}]^3)$ compared to $N_{\mathrm{var}}\mathcal{O}(N_{\mathrm{g}}^3)$ for IVDL if based solely on the computational complexity of eigendecomposition. For a small N_{var}, this may still be acceptable even for a full NWP system, although this needs further testing.

To prescribe the off-diagonal correlation matrices (e.g., $\mathbf{L}_{p,q}$), the average correlation length-scale, $\bar{h}^{\alpha}$, of two associated variables (p and q in this example) is computed, which is then used as the length-scale in the Gaspari-Cohn localisation function to construct $\mathbf{L}_{p,q}$. Other approaches may also be considered instead of using $\bar{h}^{\alpha}$, e.g., computing localisation functions separately and taking their average. As the off-diagonal matrices are constructed like autocorrelation matrices, they are symmetric, unlike in IVDL. Additionally, each off-diagonal matrix pair (e.g., $\mathbf{L}_{p,q}$ and $\mathbf{L}_{q,p}$) uses exactly the same $\bar{h}^{\alpha}$ to construct the localisation function, so the full localisation matrix is automatically symmetric. Furthermore, to apply selective multivariate localisation, one could set selective off-diagonal correlation matrices of $\mathbf{L}$ to $\mathbf{0}$, similar to Eq. (4.8).

It is also important to note that in SVDL, $\mathbf{L}$ is constructed with block correlation matrices (or with $\mathbf{0}$ in selected off-diagonal blocks), but may not be a correlation matrix as a whole. The implication is that without further adjustment it is not possible to guarantee positive semi-definiteness. Stanley et al. (2021) proposed how one might prescribe the off-diagonal correlation matrices such that the implied $\mathbf{L}$ is positive semi-definite. SVDL does not use their approach; this is to maintain flexibility

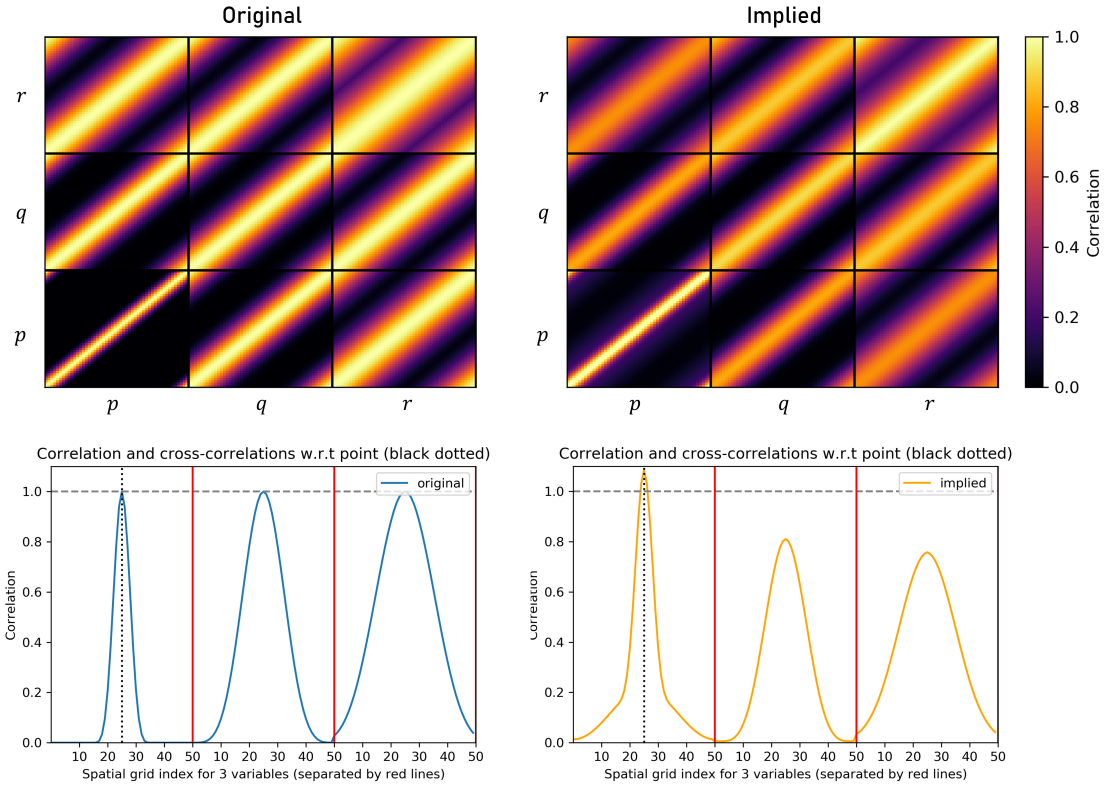

Figure 4.1: Full localisation matrix (Eq. (4.9)) with variable-dependent localisation using SVDL for a one-dimensional periodic domain of 50 points, and three variables (p, q, and r; localisation length-scale of 5, 10 and 15 points respectively). The original matrix (top left) is prescribed explicitly, while the implied matrix (top right) is re-constructed from eigenvectors after truncating negative eigenvalues and re-scaling. Auto and cross-correlations with respect to the midpoint (index 25) of variable p are shown for the original matrix (bottom left) and for the implied matrix (bottom right). The black dotted lines in the bottom panels are at value 1, which is the desired value of the peak correlations.

to prescribe some of the off-diagonal blocks to $\mathbf{0}$. This also means that an extra step is required to guarantee positive semi-definiteness. Following L22, any negative eigenvalues are truncated and the remaining eigenvalues are re-scaled (e.g., by a uniform factor given by the ratio of the original sum of eigenvalues to the sum of eigenvalues after truncation) to restore the original total variance. Figure 4.1 shows how variable-dependent localisation can be explicitly prescribed in SVDL, ensuring that the off-diagonal block matrices are symmetric. After decomposing the original localisation matrix and re-constructing, the implied localisation matrix is not identical to the original. Even with re-scaling, the truncation of negative eigenvalues has the effect of damping, particularly on the cross-correlations. The kurtosis of the correlations is also slightly altered. This effect is more severe for periodic domains where circulant matrices may be involved. There may be alternatives to using a uniform factor to re-scale, but they are not investigated here.

SVDL allows full control of the localisation, including multivariate error cross-covariances, but is computationally expensive. As mentioned, it does enforce symmetry in the off-diagonal correlation matrices unlike previous approaches (IVDL or in Wang and Wang 2023), and allows specification of null matrices on certain cross-correlation components unlike in Stanley et al. (2021). Nevertheless, it remains to be seen if symmetry is beneficial as there may not be a physical justification (see Section 4.3.2 for figures illustrating selective multivariate localisation with IVDL and the differences between IVDL and SVDL). Due to the expensive — but flexible — formulation of SVDL, one could also easily apply a cross-localisation weight factor to diminish the error cross-covariances, similar to that discussed in Stanley et al. (2021), but this is not investigated here.

4.3 Model and data assimilation framework

4.3.1 Development of the ABC-DA system

To evaluate variable-dependent and selective multivariate localisation for the tropics, we use the ABC model (Petrie et al., 2017), which solves a modified set of the compressible Euler equations. This model uses a vertical slice formulation (a two-dimensional longitude-height plane) and contains only dry dynamics. It is named after its key parameters: the pure gravity wave frequency A, the controller of acoustic wave speed B, and the constant of proportionality between pressure and density perturbations C. Additionally, a Coriolis parameter f can be set based on the desired latitudinal position

of the chosen longitude-height plane. This allows for a deep tropical environment to be mimicked by selecting a very small value for f. In this configuration, a value of $f = 10^{-5}$ s^{-1} is used. This corresponds approximately to a value of f at a latitude of 4°N. The other model parameters are also set as $A = 0.02$ s^{-1}, $B = 0.01$, and $C = 10^4$ m^2 s^{-2}. There are five prognostic variables: zonal wind u, meridional wind v, vertical wind w, scaled density perturbation $\tilde{\rho}'$ (a pressure-like variable), and buoyancy perturbation b' (a potential temperature-like variable), which govern the model dynamics. The ABC model is thus sufficiently complex as a multivariate dynamical system, while retaining simplicity to expedite research and development.

The associated data assimilation was introduced in Bannister (2020), supporting incremental 3DVar and 3DVar-FGAT (First Guess at Appropriate Time). The system is solely based on variational data assimilation. Initial implementation had an arguably crude form of generating an ensemble by considering multiple latitudinal slices from a three-dimensional operational model's output file (a version of the Unified Model). This was used for calibrating the background error covariance matrix and no ensemble-based methods (e.g., ensemble Kalman Filter or square-root filters) were implemented. Further development of the ABC-DA system by L22 introduced hybrid ensemble-variational data assimilation via the alpha control variable transform approach (Lorenc, 2003). Concurrently, L22 also introduced other ensemble generation and propagation approaches. The random field perturbations method (Magnusson et al., 2009) was used to cold start an ensemble and the ensemble bred vectors method (EBV; Balci et al. 2012) was introduced to propagate the ensemble at each data assimilation cycle — this was computationally cheaper than traditional ensemble Kalman Filter or square-root filters and did not suffer from filter collapse (see L22 for details). This parallel-run ensemble was necessary to support hybrid 3DVar and hybrid 3DVar-FGAT in the ABC-DA system. From the ensemble forecasts, the error modes $\mathbf{x}_t'^k$ can be computed and used with $\mathbf{U}^\alpha$ as in Eq. (4.1). Using these newly implemented features in the ABC-DA system, L22 showed that hybrid 3DVar outperformed 3DVar and pure EnVar methods in the ABC-DA configured for the tropical environment. However, for the purpose of this study, we will focus on the pure EnVar framework.

4.3.2 Illustration of IVDL and SVDL

Before exploring the research questions using variable-dependent and selective multivariate localisation in assimilation experiments, we illustrate how IVDL and SVDL can control the localisation with ABC model variables. First, selective multivariate localisation is illustrated using IVDL. For demonstration, the state variables have been

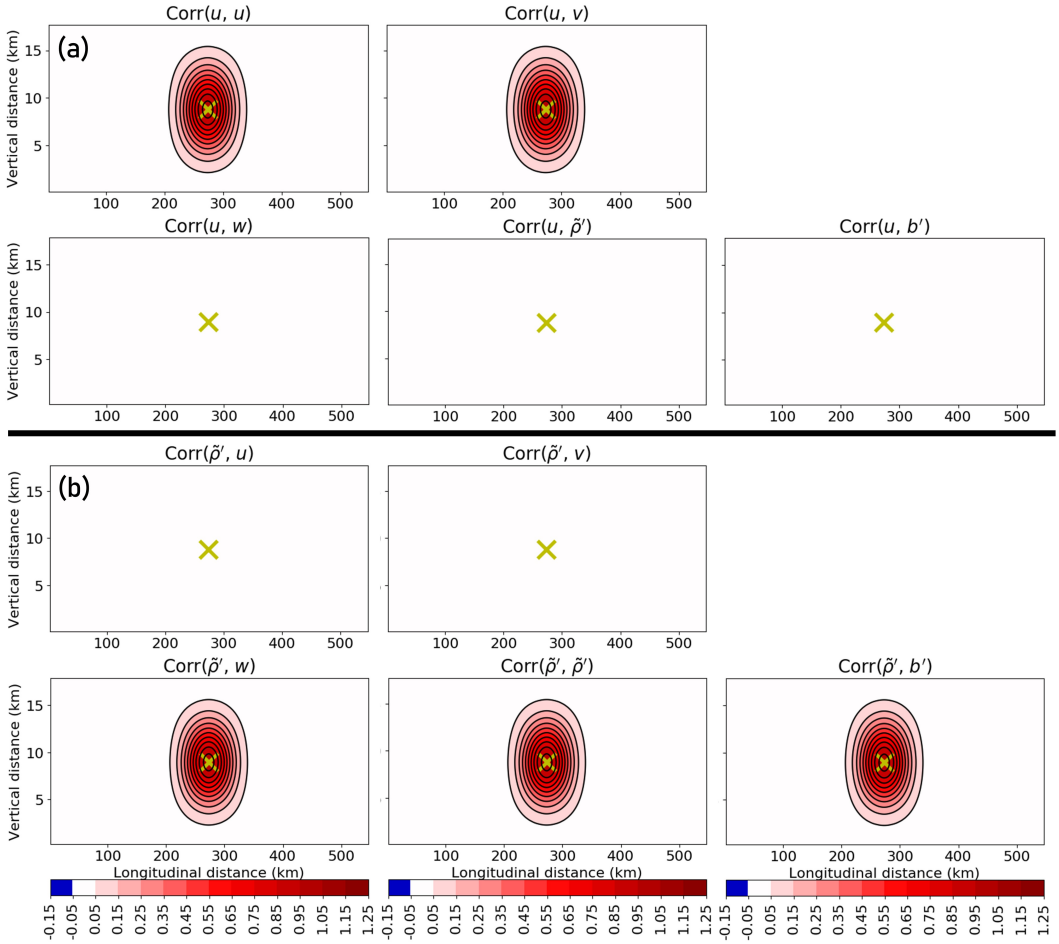

Figure 4.2: Implied localisation functions with respect to a point (yellow cross) using IVDL; for (a) cross-correlations of all variables with respect to u, (b) cross-correlations of all variables with respect to $\tilde{\rho}'$. The state variables have been grouped into two sets: (i) u and v; (ii) w, $\tilde{\rho}'$ and b' to illustrate selective multivariate localisation.

grouped into two sets: (i) u and v; (ii) w, $\tilde{\rho}'$ and b'. This means that variables in the same set retain their cross-correlations (and thus cross-covariances after Schur product with the ensemble-derived error covariances), but variables in different sets have cross-correlations knocked-out by localisation. Figure 4.2 shows the implied localisation functions with respect to u and $\tilde{\rho}'$ points. It is clear that only u and v cross-correlations are retained in the first set, and w, $\tilde{\rho}'$ and b' cross-correlations are retained in the second set. Between variables of different sets, no cross-correlations are retained by the design of $\mathbf{U}^{\alpha}$. Other grouping options have been implemented in the ABC-DA system, which we use to isolate important multivariate error relationships (see Section 4.5.1). This will enable us to explore Question 2 on which multivariate error relationships are beneficial for EnVar data assimilation in the tropics.

Next, variable-dependent localisation is illustrated using both IVDL and SVDL. For demonstration, only the vertical localisation length-scale is changed between variables. Figure 4.3 shows the comparison of IVDL and SVDL implied localisation functions with respect to u and b' points. There are subtle differences in Fig. 4.3a, b due to truncation of negative eigenvalues in SVDL. Additionally, note how there are differences in the u-b' and b'-u cross-correlations using IVDL (Figure 4.3c, d; left), which is due to the asymmetry in the off-diagonal block matrices. Using SVDL on the other hand (Figure 4.3c, d; right), the u-b' and b'-u cross-correlations are identical. As discussed in Section 4.2.2, it is not known a-priori whether the extra symmetry imposed by SVDL is beneficial to data assimilation, but SVDL is certainly more flexible than IVDL. This will enable us to explore Question 3 on whether variable-dependent spatial localisation is beneficial for EnVar data assimilation in the tropics.

4.4 Description of the experiments

4.4.1 Setup for the ABC-DA system

To evaluate the performance of IVDL and SVDL in data assimilation to learn about tropical multivariate covariances and the best localisation settings, we conduct a series of hourly-cycling Observation System Simulation Experiments (OSSEs) similar to those in L22. To represent the incompleteness of the observation network in an NWP system, only u, v and $\tilde{\rho}'$ are observed at a set of points in a sub-domain (longitudinal distance between 50 km to 500 km; height between 9 km to 14 km, i.e., the upper portion of vertical slice of 546 km length by 16 km height). The observation operator used is bi-linear interpolation. This setup is more akin to an NWP system with only satellite-related point observations (e.g., satellite-derived wind) that are available in the upper troposphere and stratosphere. This setup may also accentuate the impact of selective multivariate localisation because unobserved variable fields are updated solely on localised multivariate error cross-covariances. At each cycle, 100 observations of each of the abovementioned variables are assimilated. The observations are sampled from a 'truth' run (using a timestep of 4 seconds) with the same values of A, B, C, but with added Gaussian noise based on the observation error standard deviation. For this setup, the observation error standard deviations are 0.1 m s^{-1}, 0.1 m s^{-1} and 1.5 $\times$ 10^{-4} respectively. All observations are valid at the analysis time of each cycle and all experiments use the same observations.

The random field perturbations method (Section 3.1 of L22) is first used to generate

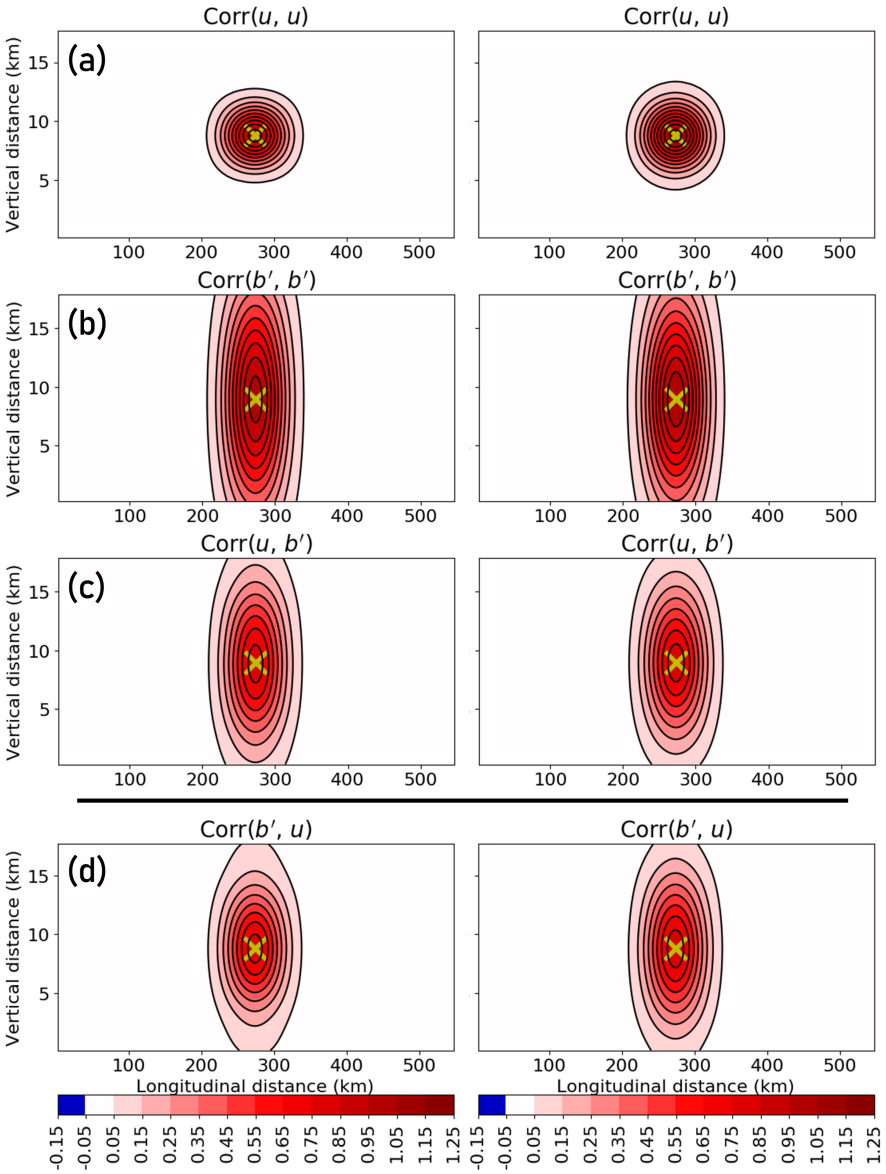

Figure 4.3: Implied localisation functions with respect to a point (yellow cross) using IVDL (left) and SVDL (right); for (a) autocorrelations of u, (b) autocorrelations of b', (c) cross-correlations of b' with respect to u, (d) cross-correlations of u with respect to b'. The vertical localisation length-scales are larger in b' to illustrate variable-dependence.

a 1000-member ensemble set (i.e., pairs of states are sampled from a 150-day 'truth' run with hourly outputs; 24 $\times$150 = 3600 available states). The number of available states have to be sufficiently large to avoid repetition in the pairs of states drawn (with replacement back into pool) by the random field perturbations method. The new 1000-member ensemble is centred around one randomly chosen ensemble analysis from the last cycle of the experiment labelled EBVd in L22 (cold start) and allowed to spin-up for 75 cycles (75 hours) so that the system would have lost memory of the cold start initialisation using random field perturbations. At each cycle, the ensemble perturbations are updated using the EBV method, similar to L22. No inflation is used, but by definition of the EBV method, a scaling based on a fixed factor divided by the max norm of the perturbations from the previous cycle (see Balci et al. 2012 or L22 for details) is used to get the updated perturbations. After spinning up for 75 cycles, the ensemble perturbations from the last spin-up cycle are used for the start of the experiments, computed according to the number of ensemble members chosen for a particular experiment (these are subsets of the 1000-member ensemble; see Section 4.4.2). For example, for a 1000-member ensemble, 999 perturbations are computed using the member-minus-mean approach; for a 100-member ensemble, the first 100 members are retained and 99 perturbations are computed. Each experiment in Section 4.5 is run for 100 cycles.

4.4.2 Guidance on ensemble size for experiments

Since the pure EnVar approach is used for the experiments (i.e., no climatological error statistics are used), we conduct further analysis to provide guidance on the ensemble size that is suitable for the tropical ABC-DA system. Firstly we assess the degree of orthogonality of the ensemble, by computing the linear independence of each successive ensemble perturbation with respect to previous perturbations at the start of the experiment (i.e., only the perturbations from the first cycle, even though they get updated throughout the experiment run), following Bannister et al. (2017). We do this separately for each variable, and so the results can differ for each one. Given the full set of 999 perturbations and ignoring time indices for brevity, $\mathbf{x}'^k$, the Gram-Schmidt procedure is used to successively compute a set of ortho-normalised vectors, $\hat{\mathbf{x}}'^k$,

$$\hat{\mathbf{x}}'^k = \frac{1}{\hat{N}_k}\left(\frac{\mathbf{x}'^k}{|\mathbf{x}'^k|} - \sum_{j=1}^{k-1}\left\langle \frac{\mathbf{x}'^k}{|\mathbf{x}'^k|}, \hat{\mathbf{x}}'^j \right\rangle \hat{\mathbf{x}}'^j\right) \tag{4.10}$$

where $|.|$ denotes the inner product, $\langle \mathbf{a}, \mathbf{b}\rangle = \mathbf{a}^\top\mathbf{b}$, and $\hat{N}_k$ is chosen to ensure that each successive perturbation ($\hat{\mathbf{x}}'^k$) has unit length. If $\hat{N}_k$ is small then the newly

orthogonalised vector $\hat{\mathbf{x}}'^k$ has only a small component that is linearly independent from the previously considered vectors of index $1 \dots k-1$. The magnitude of $\hat{N}_k$ is thus a measure of the degree of linear independence of the ensemble perturbation $\hat{\mathbf{x}}'^k$.

Figure 4.4 shows the degree of linear independence for each successive ensemble perturbation for the five prognostic variables at the start of the experiment. It reflects the capability of each new random field perturbation to sufficiently explore the additional direction in the sub-space. This is more likely if the system dynamics develop with strong non-linearities within the first hour (since we are using 1-hour forecast ensemble perturbations from a randomly generated analysis ensemble). Based on Fig. 4.4, and using the 20-member rolling average in red, we note that the degree of independence of each successive ensemble perturbation varies for each variable. We use the $\hat{N}_k = 0.3$ level to help decide on whether a sufficient level of linear independence is reached. For w, the degree of independence remains above 0.3 threshold until after the 400$^{\text{th}}$ member. On the other hand, for v, this occurs at about the 40$^{\text{th}}$ member. For u, $\tilde{\rho}'$, the threshold is met after the 100$^{\text{th}}$ member; and for b', the 120$^{\text{th}}$ member. This suggests that non-linearities largely develop in the w field when computing the 'truth' run, which are captured by the random field perturbations method, and/or are evolved within the first hour. The non-linearities captured by the random field perturbations method in other variables are weaker. The results suggest that if the ensemble size is larger than about 100, most of the ensemble perturbations would virtually be linear combinations of others, and thus have limited impacts on EnVar data assimilation (since the analysis increment is a linear combination of perturbations, Eq. (4.1)). This method does not however give any indication that sampling error is sufficiently small for localisation to be unnecessary. The results do highlight though how each field has its own characteristics based on the system dynamics, so variable-dependent localisation in particular should be worth exploring.

We further conducted a sensitivity test to the number of ensemble members using the OSSE framework described above. We try experiments with 10, 50, 100, 200, and 1000 members. Horizontal and vertical localisation are not applied in this sensitivity test to reveal the impacts of sampling noise on the (raw) ensemble-derived error covariances. The runs are evaluated using the root-mean-square error (RMSE) with respect to the 'truth' run. This was the approach taken in Bannister (2020), Bannister (2021) and L22 to assess the performance of their experiments. It is also a straightforward metric to measure the deviation of the forecasts from the 'truth' run.

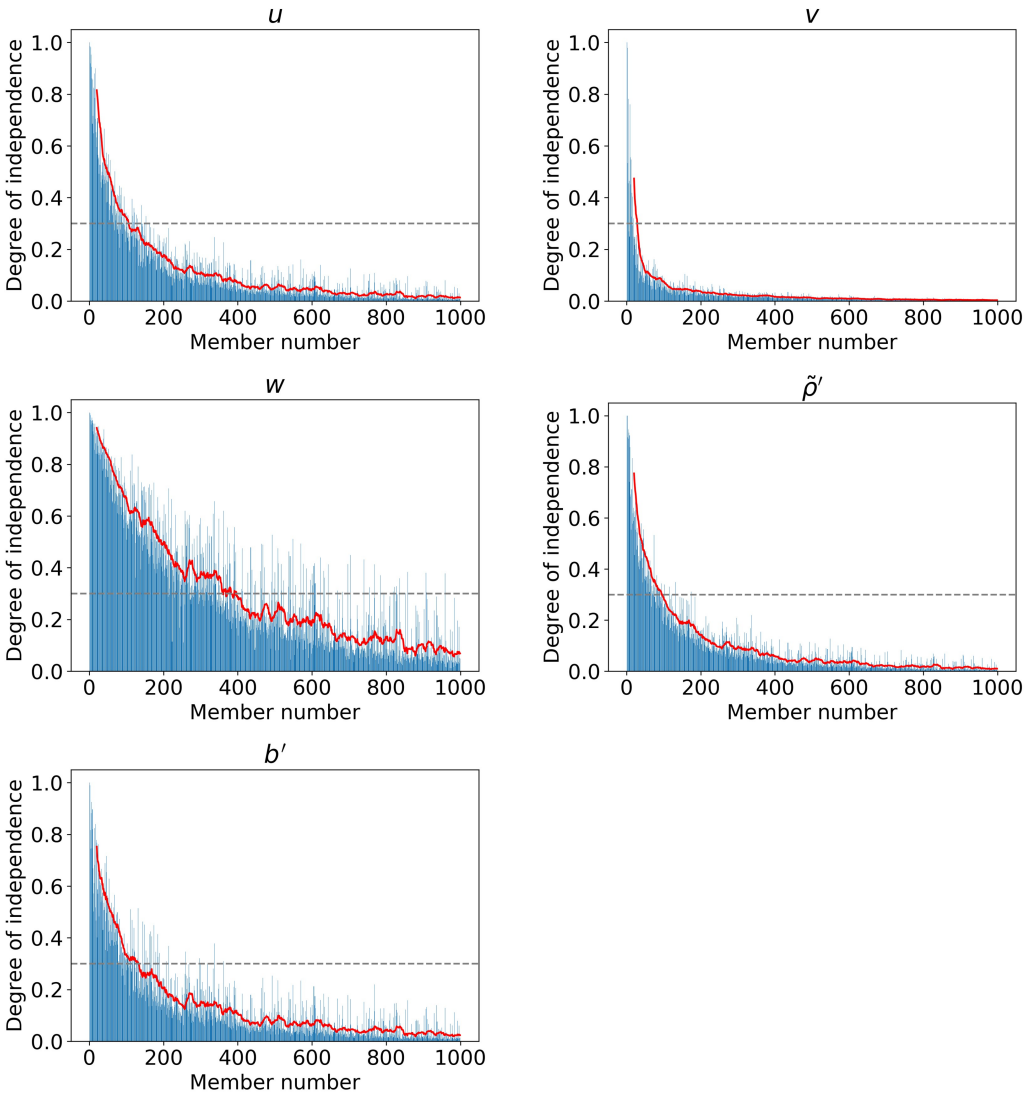

Figure 4.4: Degree of linear independence, $\hat{N}_k$, of each successive ensemble perturbation for all five prognostic variables. Perturbations are valid at the start of the experiments. An arbitrary threshold of 0.3 indicated by gray dotted line. The 20-member rolling averages are indicated in red.

Figure 4.5 contains the timeseries of RMSE for each variable over the 100 cycles, and a summary of the cycle-averaged RMSE for each variable compared to the free background run (FreeBG, which is the reference forecast without any data assimilation starting from the cold start background state). It shows that in general, the cycle-averaged RMSE decreases as the number of ensemble members increases, particularly for u and $\tilde{\rho}'$. The error reduction is about 2-4% depending on the variable when the ensemble size is increased from 10 to 1000. The highest cycle-averaged RMSE is seen in the runs with 10 and 50 members respectively, as expected due to the lack of localisation to address sampling error. Also, the reduction in cycle-averaged RMSE between the runs with 200 and 1000 members are small compared to that between the runs with 10 and 50 members, for almost all variables. For w, the decrease in RMSE with an increasing number of ensemble members is less pronounced beyond 50 members. As w is highly non-linear, it is unsurprising that FreeBG deviates from the 'truth' run more rapidly than the other variables and so the error increases over time.

Also, note the different cycle-averaged RMSE patterns for u and v. This difference may be due to the two-dimensional nature of the ABC model (no latitude dependence). In this set-up, vorticity ($\partial v/\partial x$) is associated with v and divergence ($\partial u/\partial x$) is associated with u. We would expect vorticity and divergence to have different timescales, which are indeed observed in the different cycle-averaged RMSE patterns for u and v.

Next, we examine the raw ensemble-derived error covariances between u and $\tilde{\rho}'$ using the ensemble perturbations drawn from the first cycle of the sensitivity test. Figure 4.6 shows how a selection of covariance structures change as the number of ensemble members is increased. It shows that the ensemble-derived error covariances become less noisy, particularly in the u-u and $\tilde{\rho}'$-$\tilde{\rho}'$ autocovariances. Notably, the ensemble size threshold at which the covariance structures start to appear consistent is 100 members. Above the threshold (i.e., 200 and 1000 members), the structures do not differ substantially. We also note that with 1000 members, the u-u autocovariances contain wave-like patterns which result in non-negligible longer-range spatial covariances. These patterns have previously been seen in Bannister (2020) and L22, albeit for $\tilde{\rho}'$ autocovariances instead. The wave-like patterns suggest that the main error modes in the ABC-DA system could be strongly influenced by periodic waves resonating in the domain, as also noted by L22 (i.e., they are real features rather than artifacts of sampling error).

Given the results in Figs. 4.5 and 4.6, and weighing the computational costs of

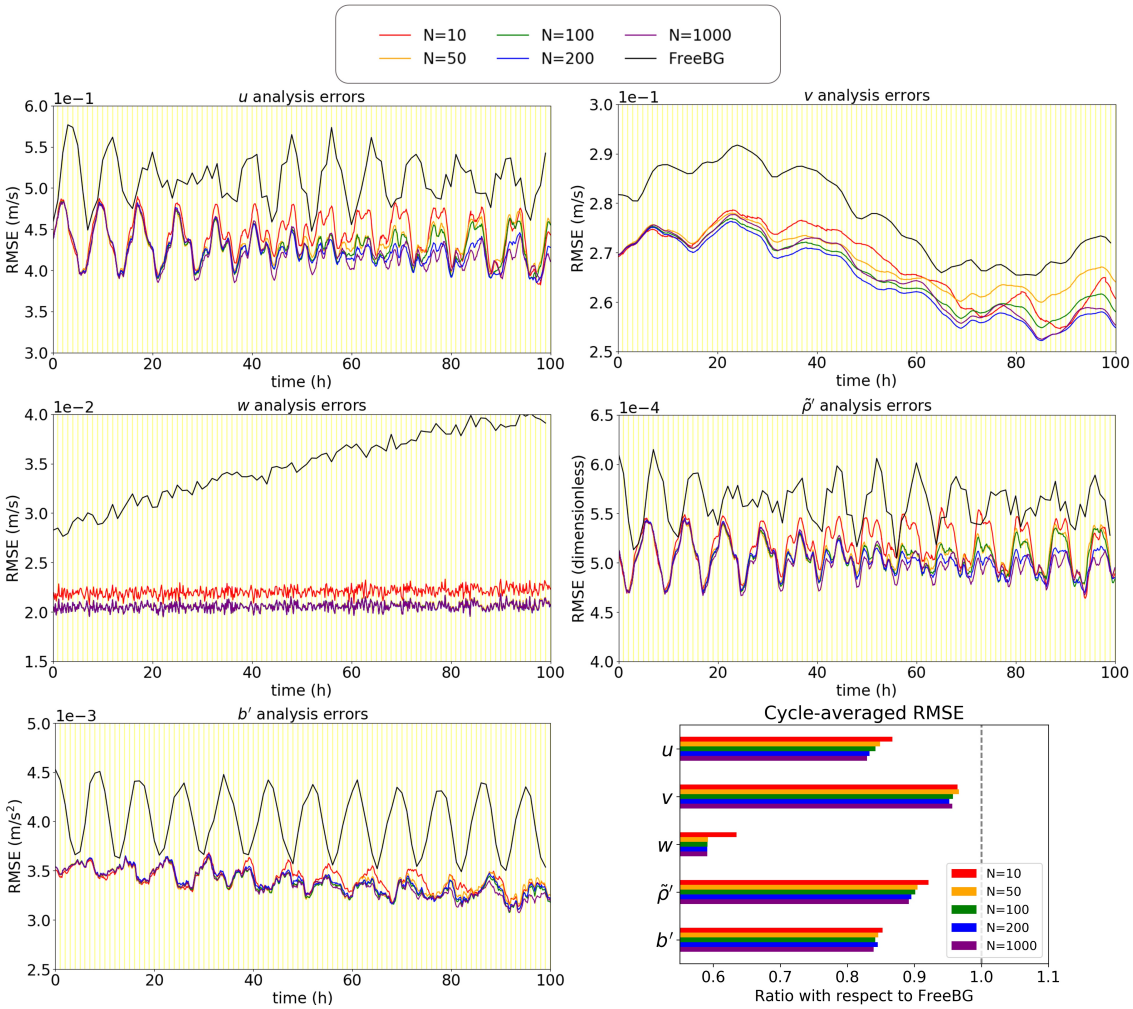

Figure 4.5: All panels except bottom right: time series of root-mean-square analysis errors for the ensemble sensitivity experiments (10, 50, 100, 200, and 1000 members) and the free background run (FreeBG). No localisation is used in these experiments. The vertical yellow lines are the analysis times. Analysis errors are defined with respect to the 'truth' run, computed every 10 minutes within the respective assimilation windows for experiments and every hour for FreeBG. Bottom right: the ratio of the cycle-averaged RMSE for each experiment with respect to FreeBG for the five ABC model variables.

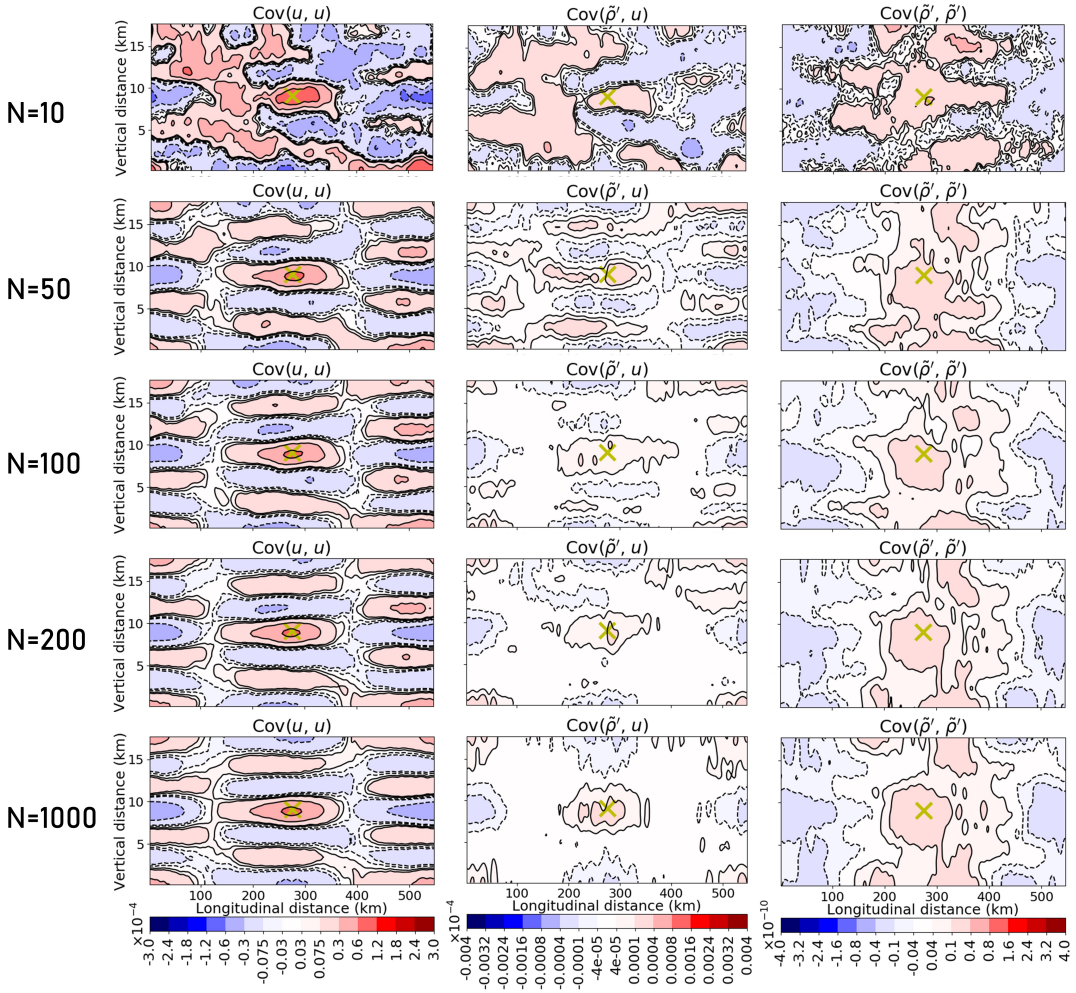

Figure 4.6: Raw ensemble-derived error autocovariances of u (leftmost column), cross-covariances of u with respect to $\tilde{\rho}'$ (middle column), and autocovariances of $\tilde{\rho}'$ (rightmost column) as a function of number of ensemble members N (increasing from top to bottom). Negative values have contours that are dashed and contour intervals are non-uniform to elucidate any features. The covariances are computed with respect to a point (yellow cross) near the centre of the domain. The ensemble perturbations are drawn from the first cycle of the sensitivity test.

running very large (1000-member) ensembles, a suitable ensemble size of 100 is used in further experiments to explore the possibilities of variable-dependent and selective multivariate localisation. This number of members is large enough to show coherent structures in the unlocalised covariances, but still shows evidence of sampling errors.

4.5 Results from data assimilation experiments using localisation

4.5.1 Exploring the important/beneficial multivariate error relationships

In this section, we run EnVar data assimilation experiments to test different multivariate localisation options. All data assimilation experiments start with the same initial background and ensemble perturbations as the 100-member experiment in Section 4.4.2. The observations are also the same. The details of the selective multivariate localisation experiment variants are listed here. As a reminder, variables that are in different groups do not retain cross covariances with variables in other groups. For example, in experiment 1c below, localisation is used to ensure that u and v errors are made to be completely uncorrelated with w, $\tilde{\rho}'$, and b' errors.

(1a) One set: limiting case where all multivariate error cross-covariances are retained, as in traditional EnVar implementations.

(1b) Five sets: limiting case where no multivariate error cross-covariances are retained (full inter-variable localisation).

(1c) Two sets: (i) u, v and (ii) w, $\tilde{\rho}'$, b'.

(1d) Two sets: (i) v, w, $\tilde{\rho}'$, b' and (ii) u.

(1e) Two sets: (i) u, v, w, b' and (ii) $\tilde{\rho}'$.

(1f) Two sets: (i) u, w, $\tilde{\rho}'$, b' and (ii) v.

(1g) Three sets: (i) u, $\tilde{\rho}'$, b' and (ii) v and (iii) w.

For this sub-section, all variables use the same horizontal localisation length-scales of 20km and use the IVDL scheme (see first row of Table 4.1 and Section 4.5.2 for justification of choice). No vertical localisation is used, following Section 2.6.3 which showed that vertical localisation of $\tilde{\rho}'$ and b' could result in hydrostatic imbalances.

Here, IVDL is first used to enable selective multivariate localisation; variable-dependence is incorporated in the next sub-section.

To demonstrate the impact of selective multivariate localisation on the analysis, the analysis increments for the first cycle of experiments 1a, 1b, 1c, 1d, and 1g are plotted in Fig. 4.7. These experiments represent the diverse possibilities and the implications of choosing a different number of sets, and/or different number of variables in each set (experiments 1e and 1f are not shown as they are similar to 1d; partitioning into two sets: four and one variables). A comparison of successive experiment pairs allows for the impact of specific multivariate error relationships to be disentangled. For example, comparing experiments 1a and 1d reveals the impact of isolating u. Note that since EBV is used to propagate the ensemble (without data assimilation), the analysis increments shown are for the control member which assimilates the observations. Here, the impact of u observations on all other variables through the error cross-covariances is substantial; the analysis increments in experiment 1d are less widespread than in 1a, especially over unobserved regions. For the full inter-variable localisation limiting case (experiment 1b), there are analysis increments for observed variables only, as expected. The $\tilde{\rho}'$ analysis increment patterns are broadly similar to those from experiments where $\tilde{\rho}'$ is not influenced by u observations (i.e., experiments 1c, 1d, and 1e (latter not shown)). Similarly, the v increment patterns are broadly similar to those from experiments 1f (not shown) and 1g. When v is isolated in experiment 1g, the impact on the analysis increments of other variables is subtle, due to relatively small error cross-covariances associated with v. Note that the widespread analysis increments in experiment 1a do not yet reveal if the multivariate error relationships associated with u are meaningful or undesirable. Unlike the mid-latitudes, where the multivariate error covariances may be explained largely by geostrophic and hydrostatic theory, the essential multivariate error covariances in the tropics are not well-known.

To identify important/beneficial multivariate error relationships in the ensemble-derived error covariances, we analyse the performance of each experiment over the course of 100 cycles, benchmarked against two limiting cases: (i) where all multivariate error cross-covariances are retained (experiment 1a), and (ii) where no multivariate error cross-covariances are retained (experiment 1b). Experiments that have smaller cycle-averaged RMSE than benchmark (i) suggest that certain multivariate error cross-covariances are not important and may be introducing more noise into the analysis than signal. Likewise, experiments that perform better than benchmark (ii) suggest that certain multivariate error cross-covariances are important/beneficial for

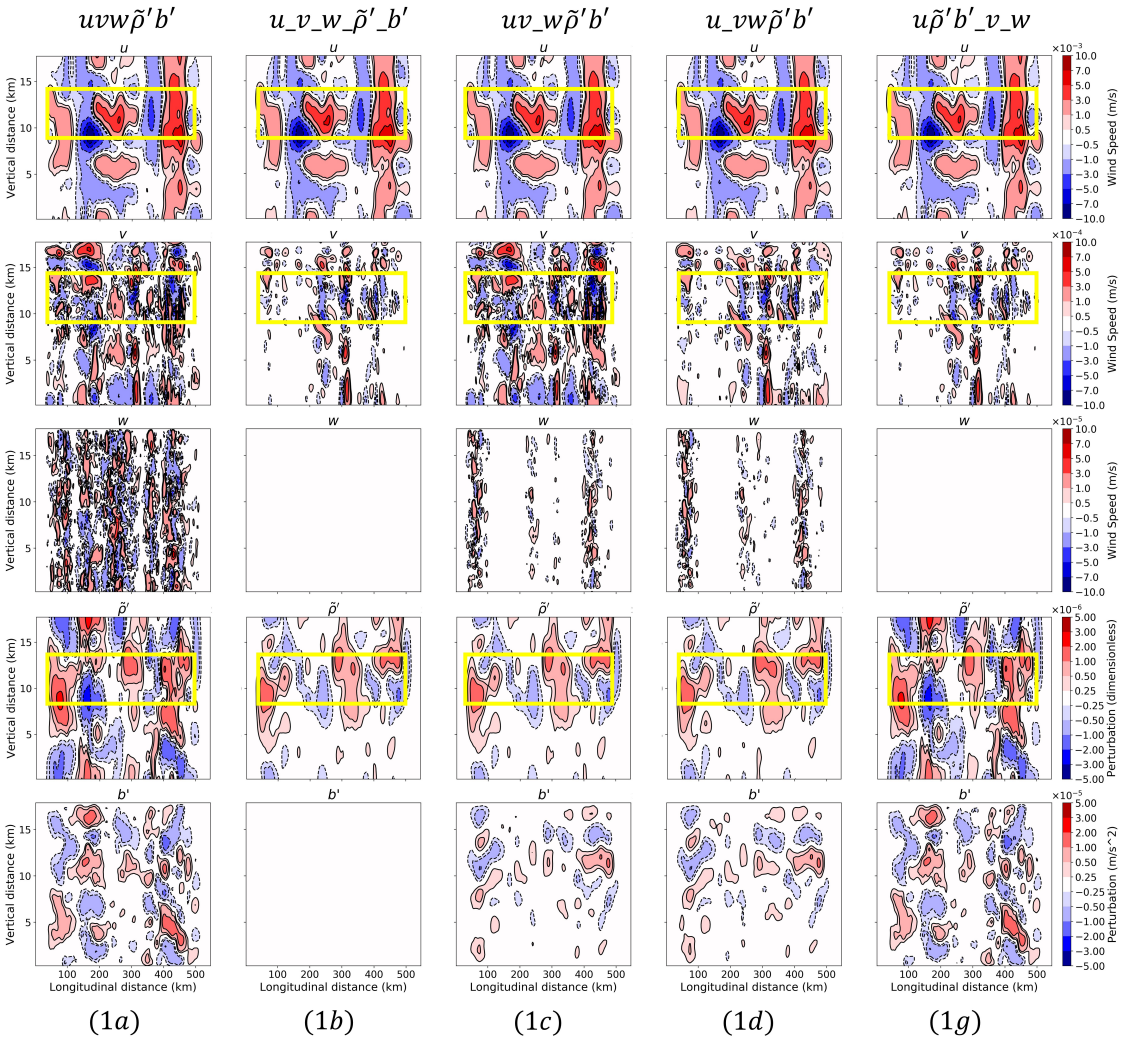

Figure 4.7: Analysis increments from the first cycle of experiments 1a, 1b, 1c, 1d and 1g (left to right, see row 1 of Table 4.1), for the five prognostic variables (top to bottom). All experiments start with the same background ensemble, assimilate the same observations (of u, v, and $\tilde{\rho}'$, which are equally spaced within the yellow box) and use the same spatial localisation. Selective multivariate localisation is applied with IVDL; variables are partitioned into sets (see text for description), demarcated by underscores (e.g., $u\_wv\tilde{\rho}'b'$ refers to experiment 1d).

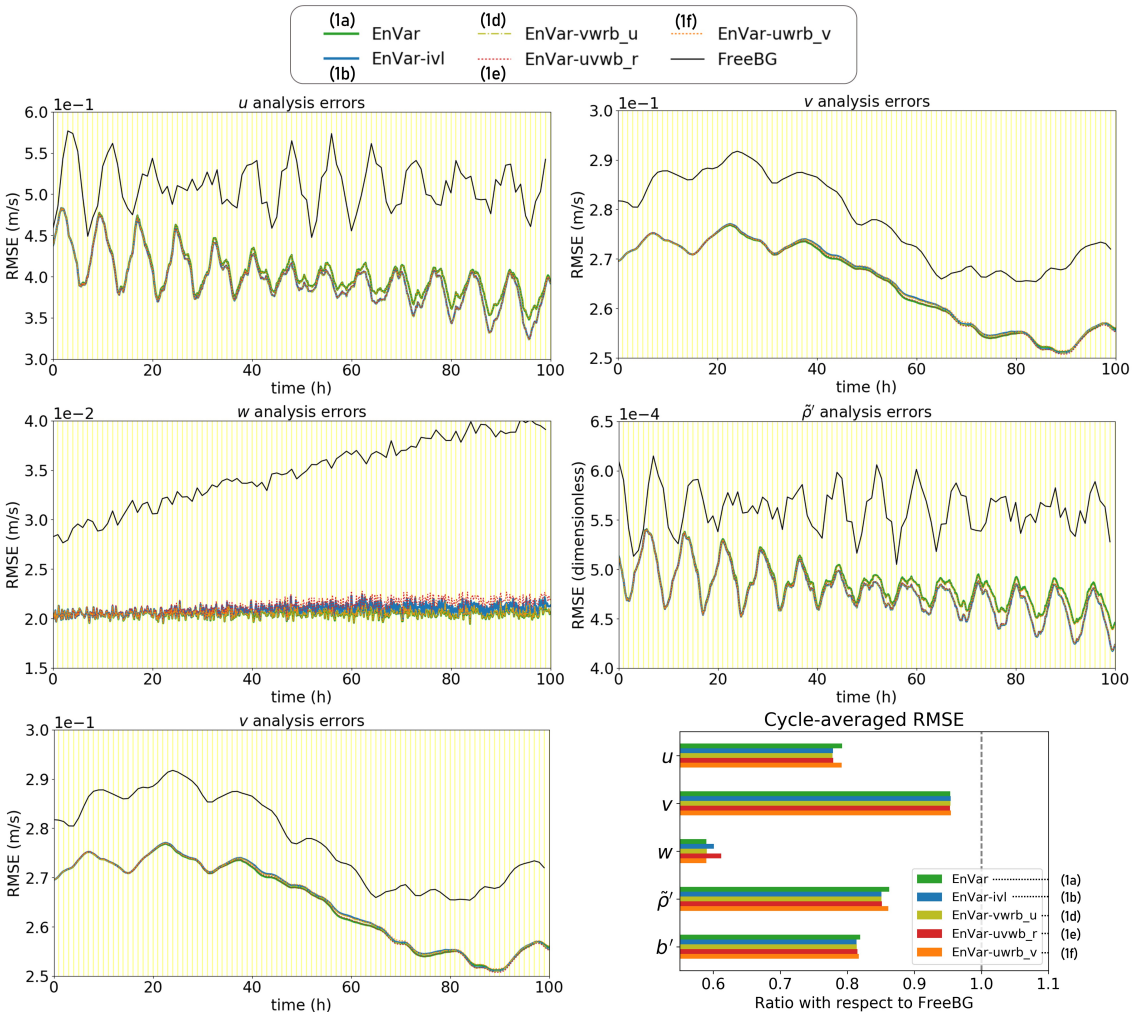

Figure 4.8: As in Fig. 4.5, but for experiments 1a (EnVar; limiting case), 1b (EnVar-ivl; limiting case), 1d (EnVar-vwrb_u), 1e (EnVar-uvwb_r) and 1f (EnVar-uwrb_v) compared to the free background run (FreeBG).

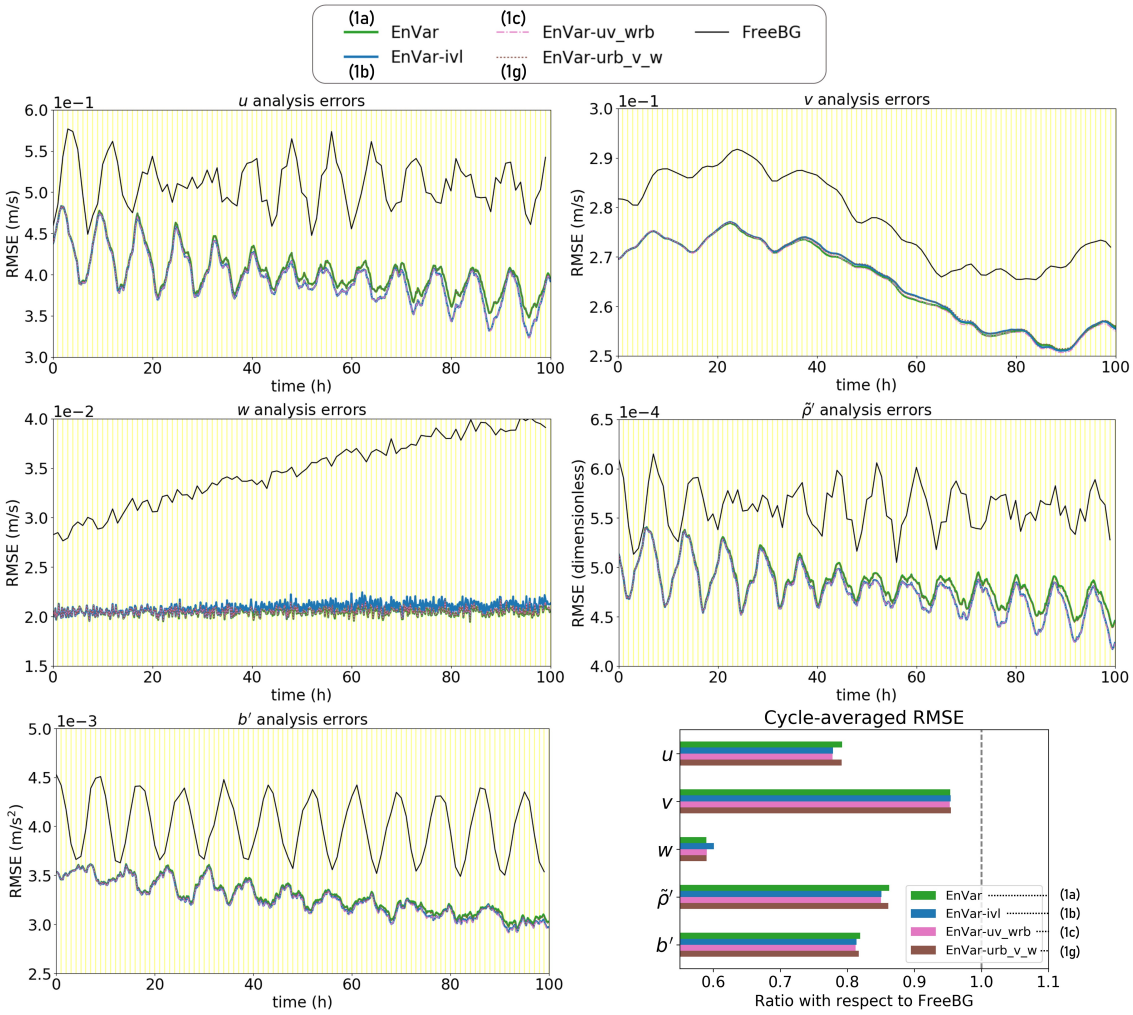

Figure 4.9: As in Fig. 4.5, but for experiments 1a (EnVar; limiting case), 1b (EnVar-ivl; limiting case), 1c (EnVar-uv_wrb) and 1g (EnVar-urb_v_w) compared to the free background run (FreeBG).

EnVar data assimilation in the tropics. The experiments which isolate individual variables (1d, 1e and 1f) are plotted in Fig. 4.8. The remaining experiments (1c and 1g) are plotted in Fig. 4.9.

The most salient feature from both figures is that the experiments have similar cycle-averaged RMSE to either experiments 1a or 1b (the two limiting cases), for all variables except w. This may be unsurprising based on the comparison of analysis increments in Fig. 4.7. Experiments 1c, 1d and 1e have similar values as experiment 1b (except w), while experiments 1f and 1g have similar values as 1a. Note that experiment 1b has smaller cycle-averaged RMSE than 1a for all variables except w. The key points from the seven experiments are as follows.

- Experiments that remove the $\tilde{\rho}'$-b' error cross-covariances (1b and 1e) lead to a substantially larger RMSE for w.
- Experiments that remove the u-$\tilde{\rho}'$ error cross-covariances (1b, 1c, 1d and 1e) lead to a notably smaller RMSE for two observed variables u and $\tilde{\rho}'$, and one unobserved variable b'.

The first point relates to hydrostatic imbalances in the analysis. From the prognostic equations for w, there are source and sink terms which relate to b' and the vertical gradient of $\tilde{\rho}'$. Any imbalance between the two will result in changes to w as the system evolves. Clearly, this imbalance is undesirable, as seen from the experiments. Lorenc (2003) previously discussed this issue in the context of mass-wind balance. Here, we find that maintaining signals of hydrostatic balance where the ensemble forecasts are hydrostatically balanced — by retaining the $\tilde{\rho}'$-b' error cross-covariances — is important/beneficial for the tropical ABC-DA system. Vetra-Carvalho et al. (2012) previously showed that for regions (in the United Kingdom) where convection was weak, hydrostatic balance holds very well. However, in regions where moist convection was involved, hydrostatic balance was not preserved. To generalise our results for the tropics, moist processes would need to be considered in the ABC model (Zhu and Bannister, 2023), but that version does not yet have data assimilation incorporated. Notwithstanding the limitations, the tropical dry dynamics representing vertical wind (dry convection) and the mass-wind interactions are still relevant to explore within the ABC-DA system.

The second point relates to mass-wind sampling errors in the analysis. As discussed previously, there is no clear mass-wind balance relationship for the tropics and many centres implicitly treat mass and wind variables univarately in their climatological

background error covariance matrices. Here, we find that explicitly treating the mass and wind variables univariately in EnVar data assimilation (by knocking-out covariances between u and $\tilde{\rho}'$) is beneficial for the tropical ABC-DA system. From Fig. 4.6, it appears that even with 100 members, the mass-wind error relationship still contains some sampling noise. Referencing earlier results from Fig. 4.7, the impact of u observations on other mass variables ($\tilde{\rho}'$ and b') is likely negative when u-related error cross-covariances are retained. The results suggest that the mass-wind error relationship prescribed directly from a 100-member ensemble is not beneficial for the tropical ABC-DA system. It remains a challenge to find a scale-dependent balance between mass and wind errors for the tropics (e.g., handling the larger scales based on large-scale balances, but handling the smaller scales based on convective-scale balances, if they exist at all). Until then, our results suggest that they should be treated entirely univariately.

We further examined if other multivariate error relationships are important/beneficial in experiments 1f and 1g, but isolating v and w does not have a substantial impact on the analysis increments nor cycle-averaged RMSE. We further repeated experiments 1a, 1c, 1d and 1e two more times with different random seeds for the observations (not shown), arriving at the same conclusions. There is limited additional benefit of repeating the other experiments, since they would yield similar results (as seen above) which would lead to the same conclusions. The abovementioned two points therefore suggest that (i) where dry dynamics are concerned, hydrostatic balance is important for EnVar data assimilation in the tropics; and (ii) treating mass and wind errors univariately is also beneficial for EnVar data assimilation in the tropics.

4.5.2 Exploring the benefits from variable-dependent spatial localisation

Next, we explore the benefits of variable-dependent localisation for EnVar data assimilation in the tropics by assimilating with different localisation length-scale values for different model variables. Thus far, all experiments have included only horizontal localisation of uniform length-scales. For some of the subsequent experiments, we use both horizontal and vertical localisation of length-scales $h^{\alpha}_{\mathrm{horiz}}$ and $h^{\alpha}_{\mathrm{vert}}$, respectively (see Section 4.2 for localisation function details, but applied separately for each spatial dimension). As before, the details of the variable-dependent localisation experiment variants are listed here.

(2a) As in experiment 1a, but applying SVDL and changing horizontal localisation for

Table 4.1: Horizontal and vertical localisation length-scales, $h^{\alpha}_{\text{horiz}}$ and h^{α}_{vert} for the experiments to evaluate variable-dependent localisation. 'NIL' means no localisation is used in the relevant direction (horizontal, vertical). IVDL refers to the isolated variable-dependent localisation scheme (Section 4.2.1) and SVDL refers to the symmetric variable-dependent localisation scheme (Section 4.2.2).

Experiment	Multivariate localisation	Scheme	u	v	w	$\tilde{\rho}'$	b'
1a to 1g	See text	IVDL	20km, NIL	20km, NIL	20km, NIL	20km, NIL	20km, NIL
2a	One set: (i) u, v, w, $\tilde{\rho}'$, b'	SVDL	50km, NIL	50km, NIL	20km, NIL	20km, NIL	20km, NIL
2b	One set: (i) u, v, w, $\tilde{\rho}'$, b'	SVDL	20km, 5km	10km, 2km	10km, 2km	20km, NIL	20km, NIL
2c	Two sets: (i) u, v (ii) w, $\tilde{\rho}'$, b'	IVDL	50km, NIL	50km, NIL	20km, NIL	20km, NIL	20km, NIL
2d	Two sets: (i) u, v (ii) w, $\tilde{\rho}'$, b'	IVDL	50km, 5km	50km, 5km	20km, NIL	20km, NIL	20km, NIL

horizontal wind variables only.

(2b) As in experiment 1a, but applying SVDL and varying horizontal and vertical localisation for wind variables.

(2c) As in experiment 1c, applying IVDL but changing horizontal localisation for horizontal wind variables only.

(2d) As in experiment 1c, applying IVDL but changing horizontal and vertical localisation for horizontal wind variables only.

For reference, Table 4.1 shows a summary of $h^{\alpha}_{\text{horiz}}$ and h^{α}_{vert} used for the experiments. The default length-scales for experiments 1a to 1g were determined based on the horizontal distance between adjacent observations (≈23 km), following L22. Experiments 2a and 2c implement variable-dependent horizontal localisation only. This is to assess if the flexibility granted by horizontal localisation alone is beneficial. Experiments 2b and 2d further implement variable-dependent vertical localisation to assess if it is also beneficial. Experiments using SVDL (2a and 2b) retain all multivariate error relationships, so they are compared to experiment 1a as the benchmark. Experiments using IVDL (2c and 2d) require that variables in the same set use the same $h^{\alpha}_{\text{horiz}}$ (see issues with periodic domains highlighted in Section 4.2.1), and they are compared to experiment 1c as the benchmark. Experiment 1c is also suitable to be the benchmark since it was the best performing out of the selective multivariate localisation experiments.

Figure 4.10 shows how variable-dependent localisation changes the analysis increments for the first cycle for experiments 1a, 2a, 2b, 2c and 2d. As expected, the u and v analysis increments in experiments 2a and 2c have broader structures than before since $h^{\alpha}_{\text{horiz}}$ is larger. Similarly, when vertical localisation is further

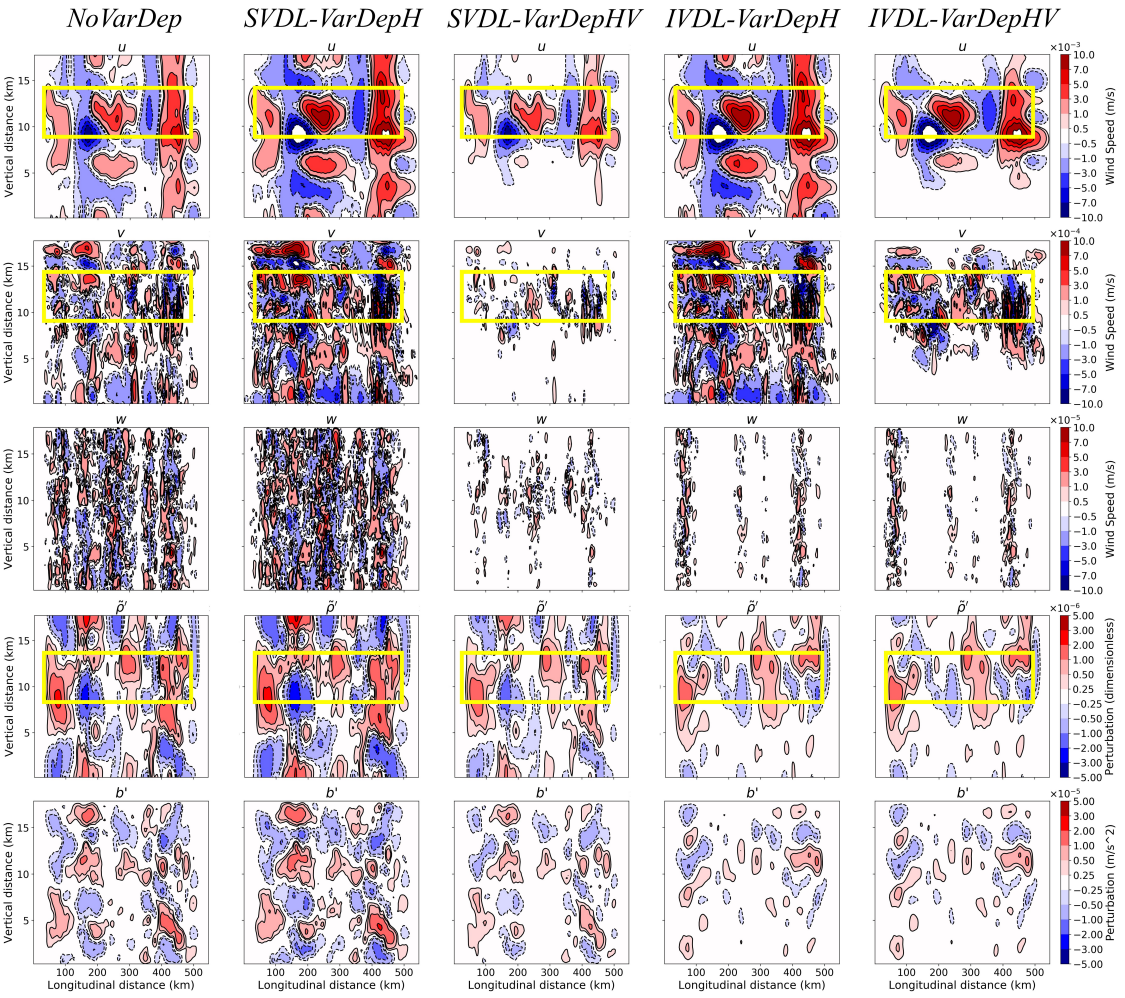

Figure 4.10: As in Fig. 4.7 but for experiments 1a (NoVarDep), 2a (SVDL-VarDepH), 2b (SVDL-VarDepHV), 2c (IVDL-VarDepH) and 2d (IVDL-VarDepHV).

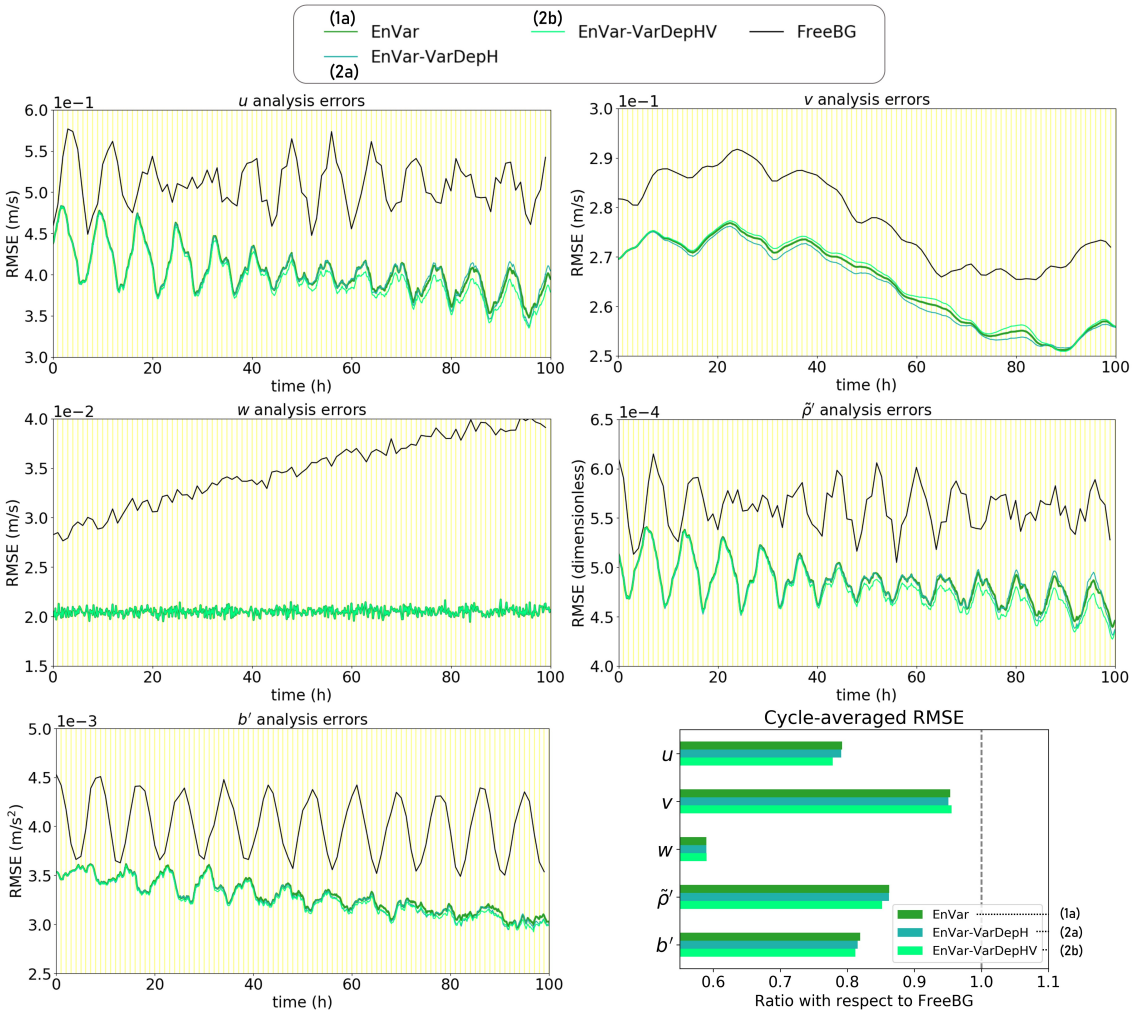

Figure 4.11: As in Fig. 4.5, but for experiments 1a (EnVar; limiting case), 2a (EnVar-VarDepH) and 2b (EnVar-VarDepHV) compared to the free background run (FreeBG).

introduced in experiments 2b and 2d, the u and v analysis increments are confined to the observed regions. Note how the w, $\tilde{\rho}'$ and b' analysis increments are similar between experiments 2c and 2d despite changes to the vertical localisation length-scales of the wind variables. This highlights how both variable-dependent and selective multivariate localisation can be simultaneously applied to the ensemble-derived error covariances (using IVDL) to constrain the observation impact on specific variables.

Next, we examine the impacts of variable-dependent localisation on the cycle-averaged RMSE. Figure 4.11 shows that the cycle-averaged RMSE for experiment 2a (when u and v use different $h^{\alpha}_{\text{horiz}}$ from the rest of the variables) is marginally smaller than 1a for most variables. The oscillations in the u and $\tilde{\rho}'$ RMSE evolution are also marginally more pronounced, likely related to the gravity waves (and their associated

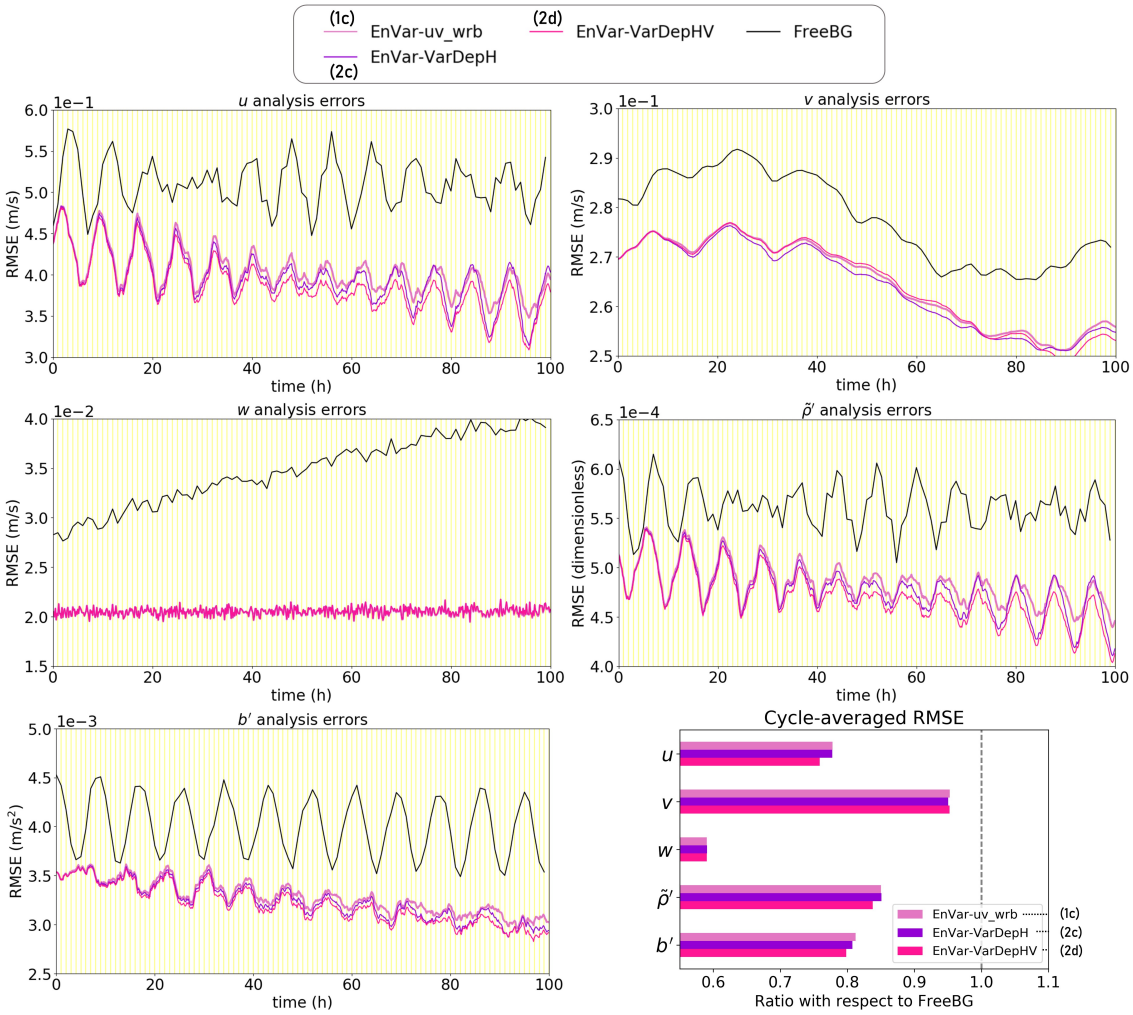

Figure 4.12: As in Fig. 4.5, but for experiments 1c (EnVar-uv_wrb), 2c (EnVar-VarDepH) and 2d (EnVar-VarDepHV) compared to the free background run (FreeBG).

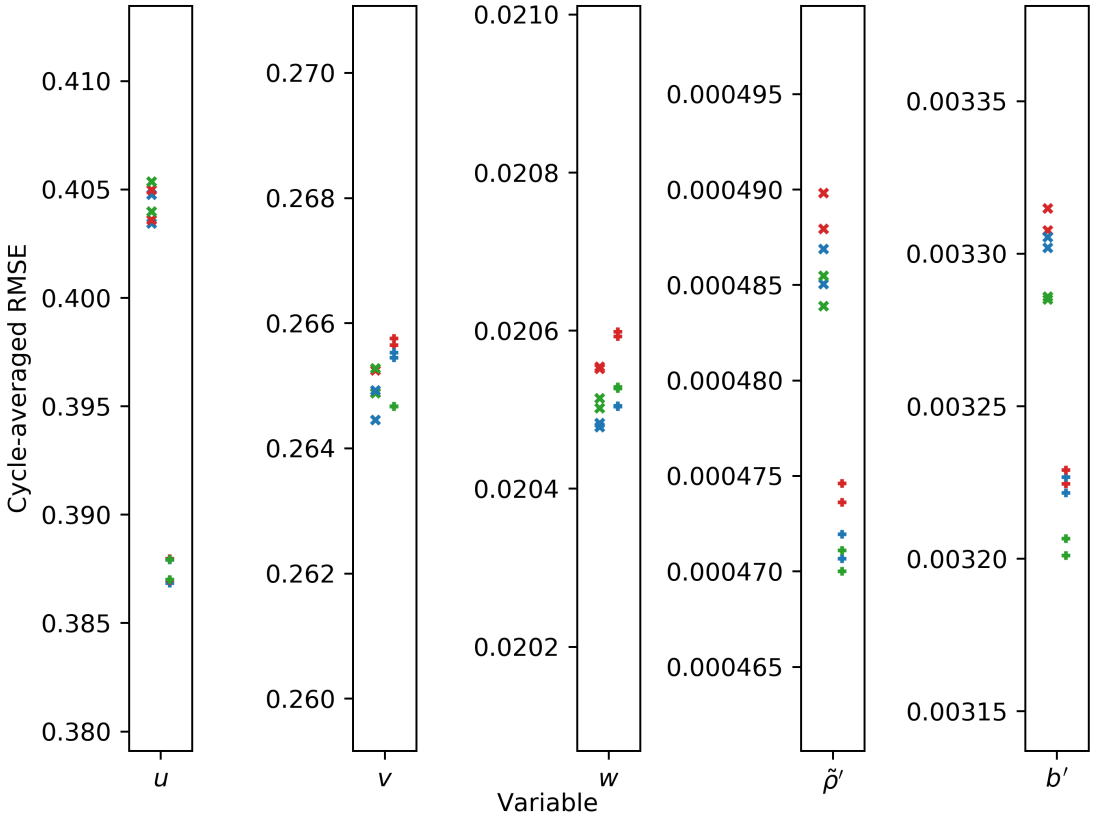

Figure 4.13: Comparison of the cycle-averaged RMSE for experiments 1a (crosses) and 2d (pluses) for all five prognostic variables. Each experiment is run three times with different random seeds for the observations (blue, red, green) and using a 50- or 100-member ensemble (total of six runs).

frequencies) within the domain. Similar results are seen comparing experiment 2c with 1c (Figure 4.12); using different $h^{\alpha}_{\text{horiz}}$ for both u and v leads to marginally improved forecasts, despite both experiments omitting error cross-covariances between u and $\tilde{\rho}'$ (unlike in experiment 2a, which uses SVDL). When variable-dependent vertical localisation is further introduced in experiments 2b and 2d, the cycle-averaged RMSEs in general are further reduced, even for b' which is unobserved. Across all experiments, we find that 2d leads to the smallest cycle-averaged RMSE, about 3-4% smaller (for u, $\tilde{\rho}'$ and b') compared to experiment 1a (the benchmark) which uses the widely adopted traditional localisation approach. We have further repeated experiments 1a and 2d two more times with different random seeds for the observations. We have also repeated the experiments using a 50-member ensemble instead of a 100-member ensemble to ensure the conclusions are not heavily dependent on ensemble size. Figure 4.13 shows that the cycle-averaged RMSE differences of u, $\tilde{\rho}'$ and b' within groups of experiment 1a or 2d configurations are much smaller than the differences between runs of experiment 1a and their respective 2d configurations.

We use a paired t-test and a Kolmogorov-Smirnov test to quantify the statistical significance. The two tests compare the distributions of the pairs of experiments 1a and 2d using a 100-member ensemble (3 samples). The paired t-test assumes that the cycle-averaged RMSE follow a normal distribution, while the Kolmogorov-Smirnov test does not. For a small sample size, it is useful to use both tests. Using the paired t-test, the p-values for u, $\tilde{\rho}'$ and b' are much less than 0.01, while the p-values for v and w are 0.301 and 0.0163 respectively. Using the Kolmogorov-Smirnov test, the Kolmogorov-Smirnov statistic for u, $\tilde{\rho}'$ and b' are each 1, with a p-value of 0.1. For v and w, the Kolmogorov-Smirnov statistic are each 0.667, with a p-value of 0.6. A p-value of less than 0.1 indicates that the result is statistically significant at the 90% confidence level. Note that experiments with 50-member ensemble are omitted from the statistical significance tests as they are repeats of the 100-member ensemble experiments (same corresponding observations) and are therefore not independent samples from the 100-member experiments. Overall, the results show that experiment 2d produces a smaller cycle-averaged RMSE than 1a, which is statistically significant at the 90% confidence level (at least) for u, $\tilde{\rho}'$ and b', but not statistically significant for v and w.

4.6 Conclusions

4.6.1 Summary and key results

In this study, the benefits of retaining or rejecting multivariate error relationships, and controlling the localisation length-scales separately for each variable in ensemble-variational (EnVar) data assimilation in the tropics are explored. This is conducted using a simplified non-hydrostatic model, the ABC model. Two approaches are implemented within the EnVar framework of the ABC model, which we refer to as the isolated variable-dependent localisation (IVDL) and symmetric variable-dependent localisation (SVDL) schemes. These grant the ability to (i) prescribe different spatial localisation length-scales for different variables; and (ii) control (i.e., knock-out by localisation) multivariate error cross-covariances. The IVDL determines the multivariate localisation in an implicit way, while the SVDL prescribes the multivariate localisation in an explicit way.

Using IVDL and SVDL in multi-cycle Observation System Simulation Experiments (OSSEs) with the ABC-DA system, we explore which multivariate relationships in the ensemble-derived error covariances are important/beneficial, and the benefits of variable-dependent spatial localisation for EnVar data assimilation in the tropics.

Using up to 1000 ensemble members (generated at each cycle as ensemble bred vectors), we first decide on a suitable ensemble size to test the localisation methods (Question 1 posed in the introduction). We compute the degree of linear independence of each successive ensemble perturbation. This provides an indication of whether the ensemble bred vectors suitably span the subspace. This analysis shows that each variable has different characteristics due to the system dynamics. The fact that the degree of independence is different for each variable suggests that variable-dependent spatial localisation is justifiable. We further show that for the ABC-DA system, increasing the number of ensemble members results in a decrease in the cycle-averaged root-mean-square errors (RMSE) in the OSSEs. We find that the threshold at which covariance structures start to appear consistent (but still have appreciable sampling error) is about 100 members for the ABC model. This ensemble size is used for most of the remaining experiments.

Further experiments reveal the following:

- Using selective multivariate localisation is beneficial, particularly when covariances associated with hydrostatic balance are retained and when the zonal wind errors are decoupled from the mass (scaled density perturbation) errors in the background

error covariances (Question 2). These lead to reduced cycle-averaged RMSE in corresponding experiments. The results suggest that when tropical dry dynamics is concerned (as with the ABC model), hydrostatic balance can still be important even at convective scales. There is also little basis for retaining the mass-wind error covariances that are presented by the ensemble, which we believe could introduce more sampling noise than useful signal in the error cross-covariances, even with as many as 100 ensemble members.

- Using variable-dependent localisation is beneficial for EnVar data assimilation in the tropics, albeit to a smaller extent compared to selective multivariate localisation (Question 3). For this particular setup, we show that using different horizontal length-scales for wind and mass variables (longer length-scales for wind) reduces the cycle-averaged RMSE, perhaps because they are more optimal for this system.

- The best performing experiment uses both variable-dependent localisation and selective multivariate localisation (retaining covariances associated with hydrostatic balance and omitting covariances associated with mass-wind error relationships). It leads to a 3-4% smaller cycle-averaged RMSE than the experiment using a traditional EnVar setup, where there is often little control over multivariate localisation nor over separate localisation length-scale for each variable. This is of the same order of magnitude improvement that is typical of upgrades to, for instance, the EnVar system of Environment and Climate Change Canada, e.g., (Caron and Buehner, 2022).

4.6.2 Discussion and future work

In Section 4.2.2, we highlighted how SVDL allows for extra symmetry to be enforced in the off-diagonal block matrices. Although there is no mathematical need for such extra symmetry in $\mathbf{L}$ (only that $\mathbf{L}$ as a whole is symmetric), we thought that this could be nonetheless investigated as a plausible option. Comparing experiments 2a with 1a and 2c with 1c shows that variable-dependent spatial localisation using SVDL and IVDL respectively reduces the RMSE of respective benchmarks by about the same percentages. From other further experiments (not shown), there also does not appear to be additional benefits of having symmetric off-diagonal block matrices using SVDL compared to IVDL. For computational efficiency, it may be more prudent to implement IVDL over SVDL, especially over non-periodic domains (see Section 4.2.1 for this caveat). This is because the potential lack of positive semi-definiteness in circulant block matrices only arise in IVDL because of periodic boundary conditions, but they

frequently arise in SVDL regardless of boundary conditions.

A more complete study of the optimal horizontal and vertical length-scales, and the correlation shapes for variable-dependent localisation, is beyond the scope of this paper. In the above experiments, the specific length-scale and correlation function choices for wind variables led to improved forecasts, but this could be because the initial choices for experiments 1a and 1c were severely sub-optimal. To briefly explore this further, we apply the empirical optimal localisation approach of Necker et al. (2023) to find the optimal tapering factors — the factors for populating the optimal localisation matrix for a given ensemble size. The optimal tapering factors are computed using the mean of middle five vertical levels, 28 to 32, and with respect to the middle horizontal gridpoint as a simple test. The 1000-member background ensemble and 100-member subsamples from the first cycle of the 1000-member experiment (Section 4.4.2) is used. The cutoff length of the optimal tapering factor curve with respect to the middle horizontal gridpoint is about 150 km for u, 80 km for $\tilde{\rho}'$ and 40 km for b' (not shown), corresponding to horizontal localisation length-scales of 75 km, 40 km and 20 km respectively (half of cutoff length) based on the commonly used Gaspari-Cohn localisation function (Necker et al., 2023). The optimal tapering factor curves for v and w were too noisy, owing to the limited number of vertical levels used in this simple test. This simple test suggests that the initial choices were moderately sub-optimal for some variables. Nonetheless, the key point from these experiments is that variable-dependent localisation can be beneficial — the impacts are sizeable even in a simplified multivariate system. These results support the findings by Necker et al. (2023), Lei et al. (2015), Wang and Wang (2023), who found that the localisation length-scales depend on the variable. Future work may focus on identifying suitable correlation functions and length-scales, as above, to implement with variable-dependent localisation to achieve further improvements in the forecasts.

There are still many avenues to explore with regard to localisation. It has been shown that scale-dependent localisation is useful in general (Caron and Buehner, 2022, Wang and Wang, 2023), although its performance in the tropics is still questionable (Caron and Buehner, 2022), and so requires further study. The localisation function is not necessarily best specified as a fixed function of distance. Anderson (2012) for instance showed that the localisation function is a function of the sample correlation itself in addition to the ensemble size. This issue may be dealt with using adaptive localisation schemes such as those proposed by Bishop and Hodyss (2007, 2009). These schemes define the localisation function from the ensemble itself, but these

schemes are quite expensive to apply.

One limitation of this study is that the ABC model is a dry dynamics model and so does not represent moist processes, even though moist processes are obviously important for convective-scale data assimilation. However, as mentioned in Section 4.5.1, the tropical dry dynamics representing vertical wind (dry convection) and the mass-wind interactions are still interesting and relevant to explore as these are captured by the ABC model. Another limitation relates to the lower dimensionality of the ABC model (two-dimensional plane), so divergence and vorticity is only related to either zonal or meridional wind. Despite the limitations, this study could still pave the way for further work in testing IVDL and SVDL in full NWP systems, especially in the tropics, to assess their feasibility and benefits in applying variable-dependent or selective multivariate localisation to improve EnVar data assimilation. Should computational costs permit, one might even combine this work with that of Buehner and Shlyaeva (2015) to have scale-dependent, variable-dependent and selective multivariate localisation altogether.

Chapter 5

Conclusions

Accurate NWP relies on data assimilation to produce a well-chosen starting point to initialise the forecasts. While data assimilation approaches have been explored in many regions, there is a dearth of literature focusing on convective-scale data assimilation over the Maritime Continent. In particular, while hybrid data assimilation has gained traction in recent years, no study until now has explored its application over the Maritime Continent.

This thesis focused on exploring convective-scale hybrid data assimilation over the Maritime Continent. The hybrid ensemble-variational approach was developed and implemented for a simplified fluid dynamics model (ABC-DA system) and a full NWP model (SINGV-DA system) to assess its impact. Further improvements to the approach were proposed, namely amending the design of the alpha control variable transform to allow for variable-dependent localisation (prescribing different localisation length-scales for different variables) and selective-multivariate localisation (knocking-out by localisation certain multivariate error covariances), which suits the tropical application and addresses limitations of traditional ensemble-variational approaches.

The main results of this thesis and how they answer the two research questions outlined in Chapter 1 are summarised below. The implications and limitations of the results are also discussed alongside other avenues for future work.

5.1 Summary

5.1.1 How does the performance of hybrid ensemble-variational data assimilation compare with traditional variational data assimilation over the Maritime Continent?

We first explored this question using a simplified tropical fluid dynamics model, before extending it to a full NWP model over the Maritime Continent.

In Chapter 2, the hybrid ensemble-variational data assimilation approach was implemented in the ABC-DA system. In general, hybrid-En3DVar-FGAT outperformed both pure EnVar and 3DVar-FGAT using the ABC-DA system within a tropical framework. Sensitivity tests showed that the design of the ensemble (i.e., a sufficiently large ensemble and an ensemble propagation method that does not suffer from filter divergence) was important to ensure that the ensemble-derived background errors were helpful for hybrid data assimilation. The results also highlighted how a sub-optimal background error covariance model for the tropics in 3DVar-FGAT, which uses geostrophic balance as a balance constraint, led to erroneous analysis increments in meridional wind and negatively impacted the forecasts.

In Chapter 3, the hybrid ensemble-variational data assimilation approach was implemented in the SINGV-DA system. Without proper tuning, the initial hybrid-En3DVar-FGAT setup had a relatively neutral impact compared to the bare 3DVar-FGAT. However, tuning the weighting between the ensemble-derived background errors and climatological background errors and introducing time-shifting (taking ensemble-derived ensemble perturbations from adjacent cycles) was vital to reap the benefits from hybrid-En3DVar-FGAT, leading to improved precipitation forecasts and forecast fits to radiosonde humidity and wind observations compared to 3DVar-FGAT. Time-shifting was required likely because the parallel-run ensemble size of 11 members was too small, so the sampling noise affected the SINGV-DA analysis, even when localisation was used to mitigate for some sampling error. The results also highlighted how different variables had different autocovariance structures, and that there were some robust ensemble-derived background error cross-correlation structures between the moisture and temperature-related variables which could have physical significance over the Maritime Continent.

5.1.2 How can traditional ensemble-variational data assimilation approaches, in particular the localisation, be better designed to improve data assimilation and NWP over the tropics?

In Chapter 4, the localisation aspect of ensemble-variational data assimilation approaches was modified to allow for variable-dependent localisation and selective multivariate localisation in the ABC-DA system using the pure EnVar approach. This was to address two limitations of traditional ensemble-variational data assimilation approaches, which were especially pertinent over the tropics. Using selective multivariate localisation was beneficial, particularly when covariances associated with hydrostatic balance were retained and when zonal wind errors were decoupled from the mass (scaled density perturbation) errors in the cross-covariances. Even though the main purpose of localisation is to reduce the impact of sampling errors, its use here also served as a tool to help show which multivariate ensemble-derived background error covariances were helpful or harmful in the tropics. The results also highlighted that using variable-dependent localisation could be beneficial if the localisation length-scales were well-tuned for each variable. Together, these two enhancements led to improved forecasts in the ABC-DA system. The improvements using the pure EnVar approach is expected to carry over to the hybrid approach (En3DVar-FGAT), although this was not tested.

5.2 Relevance to adjacent research and future work

The most relevant outcome from this thesis for operational NWP was the development and exploration of hybrid data assimilation over the Maritime Continent. Most of the modifications to the underpinning ensemble-variational approach had to be performed using a simplified model within a tropical framework to expedite the development, with the expectation that the lessons would translate to a full NWP system over the Maritime Continent. Naturally, the immediate follow-on work from Chapter 4 would be to implement variable-dependent localisation and selective multivariate localisation in an operational NWP model, like SINGV-DA. Recent studies implementing variable-dependent localisation in a regional NWP model over the United States (Wang and Wang, 2023) have suggested similar potential benefits.

Apart from the translation of research outcomes to operational NWP, other underpinning research building on this thesis could also be conducted in the future. For example, repeating the ABC-DA experiments with the hydro-ABC model — a version of

the ABC model which includes moist dynamics (Zhu and Bannister, 2023) — once data assimilation has been incorporated, would allow for a deeper understanding of the relevance of the moisture-related background error autocovariances and cross-covariances, especially within the tropical framework. Together with the results from Chapter 4 showing that decoupling the wind and mass errors at convective scales was beneficial, this would set the premise for the development of new climatological (non-ensemble) background error covariance models for convective-scale data assimilation over the tropics.

The benefits of hybrid data assimilation have often been attributed to the flow-dependency of the background error covariances. This flow-dependency generically encompasses time-appropriateness of error variances, flow-consistency of spatial covariances, as well as flow-consistency of multivariate error cross-covariances. However, it remains an open question if the benefit of hybrid data assimilation stems from the time-appropriateness (i.e., using different ensemble perturbations at each cycle instead of climatological background error statistics) or the flow-consistency (i.e., having multivariate error cross-covariances that are represented by the ensemble directly instead of using balance constraints in the control variable transform which may not hold in the tropics). The disentangling of components was not addressed in this thesis. Previous studies have attempted to explore the growth of forecast errors using initial condition perturbations from climatological or ensemble sources (Hamill and Whitaker, 2011, Piccolo, 2011). One could further explore disentangling the time-appropriateness and flow-consistency by using climatological perturbations to supplant the ensemble perturbations for pure EnVar, or by using the ensemble perturbations to calibrate the climatological background error covariances for 3DVar. This would shed light on the relative importance of time-appropriateness or flow-consistency.

All of the parallel-run ensembles that were developed to enable hybrid data assimilation in this thesis had not taken into account observations directly when propagating the ensemble across data assimilation cycles — an ensemble bred vectors approach was used in Chapters 2 and 4, and a downscaler ensemble approach was used in Chapter 3. Nevertheless, the ensemble-derived background error covariances at each cycle were still sufficient to represent the directions of error growth that led to improvements via hybrid data assimilation. It would be interesting to repeat the experiments with a parallel-run ensemble that uses ensemble Kalman-based methods. Preliminary work was performed using an deterministic ensemble Kalman filter (Sakov and Oke, 2008) in the ABC-DA system, but in early tests the ensemble

suffered severely from filter divergence, so the results were omitted from this thesis. Perhaps it would be reasonable to explore this within a full NWP system directly instead.

As highlighted in Section 4.1, selective multivariate localisation was conceptually similar to the localisation approach in Kang et al. (2011) for ensemble Kalman-based methods. Their logic for knocking-out specific multivariate error covariances was that some of the variables were not physically related. For both variational and ensemble Kalman-based methods, it would be interesting to further explore this notion of selecting cross-covariances to knock-out when the physical relationship between variables is indirect, but still plausible. The findings would likely differ based on region and application. For example, one could extend this notion for strongly coupled data assimilation, where indirect physical relationships are plausible between atmosphere and ocean variables at different spatial and temporal timescales, which are not straightforward to describe. Likewise, for aerosol data assimilation, this approach may reveal new insights on the importance of physical relationships between atmospheric and aerosol variables.

All the work in this thesis focused on ensemble-variational methods, but localisation is also a fundamental part of ensemble Kalman-based methods. As highlighted in Section 3.2.4, the localisation space can affect the degree of imbalance in the analysis. All the work in this thesis has relied on localisation in model prognostic variable state space. However, for ensemble Kalman-based methods, localisation in observation space is more common. Campbell et al. (2010) and Lei et al. (2015) discussed the benefits and drawbacks of conducting localisation in observation or model space for satellite radiance assimilation. In the same vein as how we have introduced variable-dependent and selective multivariate localisation in model space in Chapter 4, one could in principle apply some form of observation-dependent localisation in observation space. This is technically feasible if the localisation matrix is explicitly specified (e.g., similar to the 'brute force' approach mentioned in Chapter 4). Another possible follow-up could be to explore variable-dependent localisation in control variable space, as done in Clayton et al. (2013) for some variables, to address issues with imbalances. In control variable space, selective multivariate localisation may be less useful since the background errors between control variables are already assumed to be uncorrelated.

Finally, the results in this thesis relate to a broader theme of how creating additional flexibility in the localisation can be beneficial — in this case for convective-scale hybrid data assimilation over the Maritime Continent — at the cost of being more memory

intensive and computationally expensive. This trade-off needs to be balanced, so the incremental benefit must be assessed in future work for different data assimilation approaches, systems and applications. Furthermore, with the rise of artificial intelligence NWP, running super large ensembles may be permitted. This then raises the question on whether localisation is even required in the future since sampling error may no longer plague ensemble-based data assimilation methods.

References

Amezcua, J., Ide, K., Bishop, C. H. and Kalnay, E. (2012), 'Ensemble clustering in deterministic ensemble Kalman filters', *Tellus A: Dynamic Meteorology and Oceanography* **64**(1), 18039.

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.

Asch, M., Bocquet, M. and Nodet, M. (2016), *Data assimilation: Methods, algorithms, and applications*, Fundamentals of Algorithms, SIAM, Society for Industrial and Applied Mathematics.
URL: *https://books.google.co.uk/books?id=A3Q6vgAACAAJ*

Augros, C., Caumont, O., Ducrocq, V., Gaussiat, N. and Tabary, P. (2016), 'Comparisons between S-, C-and X-band polarimetric radar observations and convective-scale simulations of the HyMeX first special observing period', *Quarterly Journal of the Royal Meteorological Society* **142**, 347–362.

Balci, N., Mazzucato, A. L., Restrepo, J. M. and Sell, G. R. (2012), 'Ensemble dynamics and bred vectors', *Monthly Weather Review* **140**(7), 2308–2334.

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.

Bannister, R. N. (2008*a*), 'A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances', *Quarterly Journal of the Royal Meteorological Society* **134**(637), 1951–1970.

Bannister, R. N. (2008*b*), 'A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance statistics', *Quarterly Journal of the Royal Meteorological Society* **134**(637), 1971–1996.

Bannister, R. N. (2020), 'The ABC-DA system (v1.4): A variational data assimilation system for convective-scale assimilation research with a study of the impact of a balance constraint', *Geoscientific Model Development* **13**(8), 3789–3816.

Bannister, R. N. (2021), 'Balance conditions in variational data assimilation for a high-resolution forecast model', *Quarterly Journal of the Royal Meteorological Society* .

Bannister, R. N., Migliorini, S., Rudd, A. C. and Baker, L. H. (2017), 'Methods of investigating forecast error sensitivity to ensemble size in a limited-area convection-permitting ensemble', *Geoscientific Model Development Discussions* pp. 1–38.

Barker, D. (2005), 'Southern high-latitude ensemble data assimilation in the Antarctic Mesoscale Prediction System', *Monthly Weather Review* **133**(12), 3431–3449.

Bartello, P. and Mitchell, H. L. (1992), 'A continuous three-dimensional model of short-range forecast error covariances', *Tellus A: Dynamic Meteorology and Oceanography* **44**(3), 217–235.

Bauer, P., Thorpe, A. and Brunet, G. (2015), 'The quiet revolution of numerical weather prediction', *Nature* **525**(7567), 47–55.

Baxter, G., Dance, S., Lawless, A. and Nichols, N. (2011), 'Four-dimensional variational data assimilation for high resolution nested models', *Computers & Fluids* **46**(1), 137–141.

Bédard, J., Caron, J.-F., Buehner, M., Baek, S.-J. and Fillion, L. (2020), 'Hybrid background error covariances for a limited-area deterministic weather prediction system', *Weather and Forecasting* **35**(3), 1051–1066.

Berre, L., Ştefaănescu, S. E. and Pereira, M. B. (2006), 'The representation of the analysis effect in three error simulation techniques', *Tellus A: Dynamic Meteorology and Oceanography* **58**(2), 196–209.

Bierman, G. J. (1977), *Factorization methods for discrete sequential estimation*, Courier Corporation.

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.

Bishop, C. H. and Hodyss, D. (2007), 'Flow-adaptive moderation of spurious ensemble correlations and its use in ensemble-based data assimilation', *Quarterly Journal of the Royal Meteorological Society* **133**(629), 2029–2044.

Bishop, C. H. and Hodyss, D. (2009), 'Ensemble covariances adaptively localized with ECO-RAP. Part 1: Tests on simple error models', *Tellus A: Dynamic Meteorology and Oceanography* **61**(1), 84–96.

Bloom, S., Takacs, L., Da Silva, A. and Ledvina, D. (1996), 'Data assimilation using incremental analysis updates', *Monthly Weather Review* **124**(6), 1256–1271.

Bocquet, M., Pires, C. A. and Wu, L. (2010), 'Beyond Gaussian statistical modeling in geophysical data assimilation', *Monthly Weather Review* **138**(8), 2997–3023.

Bonavita, M., Hólm, E., Isaksen, L. and Fisher, M. (2016), 'The evolution of the ECMWF hybrid data assimilation system', *Quarterly Journal of the Royal Meteorological Society* **142**(694), 287–303.

Bouttier, F. and Courtier, P. (1999), 'Data assimilation concepts and methods, March 1999', *Meteorological training course lecture series. ECMWF* **718**, 59.

Bouyssel, F., Berre, L., Bénichou, H., Chambon, P., Girardot, N., Guidard, V., Loo, C., Mahfouf, J.-F., Moll, P., Payan, C. et al. (2022), 'The 2020 global operational NWP data assimilation system at Météo-France', *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. IV)* pp. 645–664.

Buehner, M. (2005), 'Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP setting', *Quarterly Journal of the Royal Meteorological Society* **131**(607), 1013–1043.

Buehner, M., Morneau, J. and Charette, C. (2013), 'Four-dimensional ensemble-variational data assimilation for global deterministic weather prediction', *Nonlinear Processes in Geophysics* **20**(5), 669–682.

Buehner, M. and Shlyaeva, A. (2015), 'Scale-dependent background-error covariance localisation', *Tellus A: Dynamic Meteorology and Oceanography* **67**(1), 28027.

Buizza, R. and Richardson, D. (2017), '25 years of ensemble forecasting at ECMWF', *ECMWF Newsletter* **153**, 20–31.

Buizza, R., Tribbia, J., Molteni, F. and Palmer, T. (1993), 'Computation of optimal unstable structures for a numerical weather prediction model', *Tellus A: Dynamic Meteorology and Oceanography* **45**(5), 388–407.

Burgers, G., Jan van Leeuwen, P. and Evensen, G. (1998), 'Analysis scheme in the ensemble Kalman filter', *Monthly Weather Review* **126**(6), 1719–1724.

Campbell, W. F., Bishop, C. H. and Hodyss, D. (2010), 'Vertical covariance localization for satellite radiances in ensemble kalman filters', *Monthly Weather Review* **138**(1), 282–290.

Caron, J.-F. and Buehner, M. (2022), 'Implementation of scale-dependent background-error covariance localization in the Canadian Global Deterministic Prediction System', *Weather and Forecasting* **37**(9), 1567–1580.

Caron, J.-F., Michel, Y., Montmerle, T. and Arbogast, É. (2019), 'Improving background error covariances in a 3D ensemble–variational data assimilation system for regional NWP', *Monthly Weather Review* **147**(1), 135–151.

Centre for Climate Research Singapore (2019), 'Weather Prediction by Numerical Methods Module 1 (WPNM-M1)', *Workshop Report* pp. 1–23.

Chan, M.-Y., Chen, X. and Anderson, J. L. (2023), 'The potential benefits of handling mixture statistics via a bi-Gaussian EnKF: Tests with all-sky satellite infrared radiances', *Journal of Advances in Modeling Earth Systems* **15**(2), e2022MS003357.

Chen, K., Bai, L., Ling, F., Ye, P., Chen, T., Chen, K., Han, T. and Ouyang, W. (2023), 'Towards an end-to-end artificial intelligence driven global weather forecasting system', *arXiv preprint arXiv:2312.12462* .

Chen, Y., Rizvi, S. R., Huang, X.-Y., Min, J. and Zhang, X. (2013), 'Balance characteristics of multivariate background error covariances and their impact on analyses and forecasts in tropical and Arctic regions', *Meteorology and Atmospheric Physics* **121**, 79–98.

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.

Cohn, S. E. (1997), 'An introduction to estimation theory', *Journal of the Meteorological Society of Japan. Ser. II* **75**(1B), 257–288.

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, *in* 'Annales geophysicae (1988)', Vol. 11, pp. 634–647.

Destouches, M., Montmerle, T., Michel, Y. and Caron, J.-F. (2023), 'Impact of hydrometeor control variables in a convective-scale 3DEnVar data assimilation scheme', *Quarterly Journal of the Royal Meteorological Society* .

Destouches, M., Montmerle, T., Michel, Y. and Ménétrier, B. (2021), 'Estimating optimal localization for sampled background-error covariances of hydrometeor variables', *Quarterly Journal of the Royal Meteorological Society* **147**(734), 74–93.

Dillon, M. E., Skabar, Y. G., Ruiz, J., Kalnay, E., Collini, E. A., Echevarría, P., Saucedo, M., Miyoshi, T. and Kunii, M. (2016), 'Application of the WRF-LETKF data assimilation system over southern South America: Sensitivity to model physics', *Weather and Forecasting* **31**(1), 217–236.

Dipankar, A., Webster, S., Sun, X., Sanchez, C., North, R., Furtado, K., Wilkinson, J., Lock, A., Vosper, S., Huang, X.-Y. et al. (2020), 'SINGV: A convective-scale weather forecast model for Singapore', *Quarterly Journal of the Royal Meteorological Society* **146**(733), 4131–4146.

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.

Evensen, G. (2006), *Data assimilation: The ensemble Kalman filter*, Springer Berlin Heidelberg.
URL: *https://books.google.co.uk/books?id=VJ2oOecHhOYC*

Feng, J. and Wang, X. (2021), 'Impact of increasing horizontal and vertical resolution during the HWRF hybrid EnVar data assimilation on the analysis and prediction of Hurricane Patricia (2015)', *Monthly Weather Review* **149**(2), 419–441.

Fillion, L., Tanguay, M., Lapalme, E., Denis, B., Desgagne, M., Lee, V., Ek, N., Liu, Z., Lajoie, M., Caron, J.-F. et al. (2010), 'The Canadian regional data assimilation and forecasting system', *Weather and Forecasting* **25**(6), 1645–1669.

Fortin, V., Abaza, M., Anctil, F. and Turcotte, R. (2014), 'Why should ensemble spread match the RMSE of the ensemble mean?', *Journal of Hydrometeorology* **15**(4), 1708–1713.

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.

Gao, S., Min, J., Liu, L. and Ren, C. (2019), 'The development of a hybrid EnSRF-En3DVar system for convective-scale data assimilation', *Atmospheric Research* **229**, 208–223.

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.

Gasperoni, N. A., Wang, X. and Wang, Y. (2022), 'Using a cost-effective approach to increase background ensemble member size within the GSI-based EnVar system for improved radar analyses and forecasts of convective systems', *Monthly Weather Review* **150**(3), 667–689.

Goodliff, M., Amezcua, J. and Van Leeuwen, P. J. (2015), 'Comparing hybrid data assimilation methods on the Lorenz 1963 model with increasing non-linearity', *Tellus A: Dynamic Meteorology and Oceanography* **67**(1), 26928.

Gustafsson, N., Bojarova, J. and Vignes, O. (2014), 'A hybrid variational ensemble data assimilation for the HIgh Resolution Limited Area Model (HIRLAM)', *Nonlinear Processes in Geophysics* **21**(1), 303–323.

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.

Hamill, T. M. and Snyder, C. (2000), 'A hybrid ensemble Kalman filter–3D variational analysis scheme', *Monthly Weather Review* **128**(8), 2905–2919.

Hamill, T. M. and Whitaker, J. S. (2011), 'What constrains spread growth in forecasts initialized from ensemble Kalman filters?', *Monthly Weather Review* **139**(1), 117–131.

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.

Hawkness-Smith, L. and Simonin, D. (2021), 'Radar reflectivity assimilation using hourly cycling 4D-Var in the Met Office Unified Model', *Quarterly Journal of the Royal Meteorological Society* **147**(736), 1516–1538.

Heng, B. P., Tubbs, R., Huang, X.-Y., Macpherson, B., Barker, D. M., Boyd, D. F., Kelly, G., North, R., Stewart, L., Webster, S. et al. (2020), 'SINGV-DA: A data assimilation system for convective-scale numerical weather prediction over Singapore', *Quarterly Journal of the Royal Meteorological Society* **146**(729), 1923–1938.

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. (2019), 'Collecting and utilising crowd-sourced data for numerical weather prediction: Propositions from the meeting held in Copenhagen, 4–5 December 2018', *Atmospheric Science Letters* **20**(7), e921.

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: Dynamic Meteorology and Oceanography* **38**(2), 111–136.

Holton, J. R. (1973), 'An introduction to dynamic meteorology', *American Journal of Physics* **41**(5), 752–754.

Honda, T., Miyoshi, T., Lien, G.-Y., Nishizawa, S., Yoshida, R., Adachi, S. A., Terasaki, K., Okamoto, K., Tomita, H. and Bessho, K. (2018), 'Assimilating all-sky Himawari-8 satellite infrared radiances: A case of Typhoon Soudelor (2015)', *Monthly Weather Review* **146**(1), 213–229.

Houtekamer, P. and Derome, J. (1995), 'Methods for ensemble prediction', *Monthly Weather Review* **123**(7), 2181–2196.

Houtekamer, P. L. and Mitchell, H. L. (1998), 'Data assimilation using an ensemble Kalman filter technique', *Monthly Weather Review* **126**(3), 796–811.

Houtekamer, P. L. and Mitchell, H. L. (2001), 'A sequential ensemble Kalman filter for atmospheric data assimilation', *Monthly Weather Review* **129**(1), 123–137.

Hu, G., Dance, S. L., Bannister, R. N., Chipilski, H. G., Guillet, O., Macpherson, B., Weissmann, M. and Yussouf, N. (2023), 'Progress, challenges, and future steps in data assimilation for convection-permitting numerical weather prediction: Report on the virtual meeting held on 10 and 12 november 2021', *Atmospheric Science Letters* **24**(1), e1130.

Hu, M., Benjamin, S. G., Ladwig, T. T., Dowell, D. C., Weygandt, S. S., Alexander, C. R. and Whitaker, J. S. (2017), 'GSI three-dimensional ensemble–variational hybrid data assimilation using a global ensemble for the regional Rapid Refresh model', *Monthly Weather Review* **145**(10), 4205–4225.

Huang, B. and Wang, X. (2018), 'On the use of cost-effective valid-time-shifting (VTS) method to increase ensemble size in the GFS hybrid 4DEnVar system', *Monthly Weather Review* **146**(9), 2973–2998.

Huang, B., Wang, X., Kleist, D. T. and Lei, T. (2021), 'A simultaneous multiscale data assimilation using scale-dependent localization in GSI-based hybrid 4DEnVar for NCEP FV3-based GFS', *Monthly Weather Review* **149**(2), 479–501.

Huffman, G. J., Stocker, E., Bolvin, D., Nelkin, E. and Tan, J. (2019), 'GPM IMERG Final Precipitation L3 1 day 0.1 degree x 0.1 degree V06 (GPM_3IMERGDF)', `https://doi.org/10.5067/GPM/IMERGDF/DAY/06`. Accessed: 2022-11-11.

Ikuta, Y., Fujita, T., Ota, Y. and Honda, Y. (2021), 'Variational data assimilation system for operational regional models at Japan Meteorological Agency', *Journal of the Meteorological Society of Japan. Ser. II* **99**(6), 1563–1592.

Ingleby, N. B. (2001), 'The statistical structure of forecast errors and its representation in the Met. Office global 3-D variational data assimilation scheme', *Quarterly Journal of the Royal Meteorological Society* **127**(571), 209–231.

Ingleby, N., Lorenc, A. C., Ngan, K., Rawlins, F. and Jackson, D. (2013), 'Improved variational analyses using a nonlinear humidity control variable', *Quarterly Journal of the Royal Meteorological Society* **139**(676), 1875–1887.

Ito, K., Kunii, M., Kawabata, T., Saito, K., Aonashi, K. and Duc, L. (2016), 'Mesoscale hybrid data assimilation system based on JMA nonhydrostatic model', *Monthly Weather Review* **144**(9), 3417–3439.

Johnson, A., Wang, X., Carley, J. R., Wicker, L. J. and Karstens, C. (2015), 'A comparison of multiscale GSI-based EnKF and 3DVar data assimilation using radar and conventional observations for midlatitude convective-scale precipitation forecasts', *Monthly Weather Review* **143**(8), 3087–3108.

Jones, T. A., Skinner, P., Yussouf, N., Knopfmeier, K., Reinhart, A., Wang, X., Bedka, K., Smith Jr, W. and Palikonda, R. (2020), 'Assimilation of GOES-16 radiances and retrievals into the Warn-on-Forecast System', *Monthly Weather Review* **148**(5), 1829–1859.

Kadowaki, T., Ota, Y. and Yokota, S. (2020), 'Introduction of a new hybrid data assimilation system for the JMA Global Spectral Model', *Research activities in Earth system modelling. Working Group on Numerical Experimentation* pp. 1–09.

Kalman, R. E. (1960), 'A new approach to linear filtering and prediction problems'.

Kalnay, E. (2003), *Atmospheric modeling, data assimilation and predictability*, Cambridge University Press.

Kalnay, E., Corazza, M. and Cai, M. (2002), Are bred vectors the same as lyapunov vectors?, *in* 'EGS general assembly conference abstracts', p. 6820.

Kang, J.-S., Kalnay, E., Liu, J., Fung, I., Miyoshi, T. and Ide, K. (2011), '"variable localization" in an ensemble kalman filter: Application to the carbon cycle data assimilation', *Journal of Geophysical Research: Atmospheres* **116**(D9).

Kotsuki, S. and Bishop, C. H. (2022), 'Implementing hybrid background error covariance into the LETKF with attenuation-based localization: Experiments with a simplified AGCM', *Monthly Weather Review* **150**(1), 283–302.

Kretschmer, M., Hunt, B. R. and Ott, E. (2015), 'Data assimilation using a climatologically augmented local ensemble transform Kalman filter', *Tellus A: Dynamic Meteorology and Oceanography* **67**(1), 26617.

Kuhl, D. D., Rosmond, T. E., Bishop, C. H., McLay, J. and Baker, N. L. (2013), 'Comparison of hybrid ensemble/4DVar and 4DVar within the NAVDAS-AR data assimilation framework', *Monthly Weather Review* **141**(8), 2740–2758.

Kutty, G., Gogoi, R., Rakesh, V. and Pateria, M. (2020), 'Comparison of the performance of hybrid ETKF-3DVAR and 3DVAR data assimilation scheme on the forecast of tropical cyclones formed over the Bay of Bengal', *Journal of Earth System Science* **129**, 1–14.

Lee, J. C. and Huang, X.-Y. (2020), 'Background error statistics in the tropics: Structures and impact in a convective-scale numerical weather prediction system', *Quarterly Journal of the Royal Meteorological Society* **146**(730), 2154–2173.

Lee, J. C. K., Amezcua, J. and Bannister, R. N. (2022), 'Hybrid ensemble-variational data assimilation in ABC-DA within a tropical framework', *Geoscientific Model Development* **15**(15), 6197–6219.

Lee, J. C. K., Amezcua, J. and Bannister, R. N. (2024), 'Variable-dependent and selective multivariate localization for ensemble-variational data assimilation in the tropics', *Monthly Weather Review* **152**(4), 1097–1118.

Lee, J. C. K. and Barker, D. M. (2023), 'Development of a hybrid ensemble–variational data assimilation system over the western Maritime Continent', *Weather and Forecasting* **38**(3), 425–444.

Lee, J. C. K., Dipankar, A. and Huang, X.-Y. (2021), 'On the sensitivity of the simulated diurnal cycle of precipitation to 3-hourly radiosonde assimilation: A case study over the western Maritime Continent', *Monthly Weather Review* **149**(10), 3449–3468.

Lee, J. C. K. and Huang, X.-Y. (2022), 'Modelling the background error covariance matrix: Applicability over the Maritime Continent', *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. IV)* pp. 599–627.

Lei, L., Anderson, J. L. and Romine, G. S. (2015), 'Empirical localization functions for ensemble Kalman filter data assimilation in regions with and without precipitation', *Monthly Weather Review* **143**(9), 3664–3679.

Leuenberger, D., Haefele, A., Omanovic, N., Fengler, M., Martucci, G., Calpini, B., Fuhrer, O. and Rossa, A. (2020), 'Improving high-impact numerical weather prediction with lidar and drone observations', *Bulletin of the American Meteorological Society* **101**(7), E1036–E1051.

Leutbecher, M. (2009), Diagnosis of ensemble forecasting systems, *in* 'Seminar on Diagnosis of Forecasting and Data Assimilation Systems', pp. 235–266.

Lewis, J. M., Lakshmivarahan, S. and Dhall, S. (2006), *Dynamic data assimilation: A least squares approach*, Vol. 13, Cambridge University Press.

Liu, C., Xiao, Q. and Wang, B. (2008), 'An ensemble-based four-dimensional variational data assimilation scheme. Part I: Technical formulation and preliminary test', *Monthly Weather Review* **136**(9), 3363–3373.

Lönnberg, P. and Hollingsworth, A. (1986), 'The statistical structure of short-range forecast errors as determined from radiosonde data. Part II: The covariance of height and wind errors', *Tellus A: Dynamic Meteorology and Oceanography* **38**(2), 137–161.

Lorenc, A. (2007), 'A study of ob monitoring statistics from radiosondes, composited for low-level cloud layers', *Met Office NWP Forecasting Research Technical Report* **504**, 1–32.

Lorenc, A., Ballard, S., Bell, R., Ingleby, N., Andrews, P., Barker, D., Bray, J., Clayton, A., Dalby, T., Li, D. et al. (2000), 'The Met. Office global three-dimensional variational data assimilation scheme', *Quarterly Journal of the Royal Meteorological Society* **126**(570), 2991–3012.

Lorenc, A. C. (1986), 'Analysis methods for numerical weather prediction', *Quarterly Journal of the Royal Meteorological Society* **112**(474), 1177–1194.

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. (2013), 'Recommended nomenclature for EnVar data assimilation methods'.

Lorenz, E. N. (1963), 'Deterministic nonperiodic flow', *Journal of Atmospheric Sciences* **20**(2), 130–141.

Lorenz, E. N. (1996), Predictability: A problem partly solved, *in* 'Proc. Seminar on predictability', Vol. 1, Reading.

Lu, X., Wang, X., Li, Y., Tong, M. and Ma, X. (2017), 'GSI-based ensemble-variational hybrid data assimilation for HWRF for hurricane initialization and prediction: Impact of various error covariances for airborne radar observation assimilation', *Quarterly Journal of the Royal Meteorological Society* **143**(702), 223–239.

Magnusson, L., Nycander, J. and Källén, E. (2009), 'Flow-dependent versus flow-independent initial perturbations for ensemble prediction', *Tellus A: Dynamic Meteorology and Oceanography* **61**(2), 194–209.

Ménétrier, B., Montmerle, T., Michel, Y. and Berre, L. (2015*a*), 'Linear filtering of sample covariances for ensemble-based data assimilation. Part I: Optimality criteria and application to variance filtering and covariance localization', *Monthly Weather Review* **143**(5), 1622–1643.

Ménétrier, B., Montmerle, T., Michel, Y. and Berre, L. (2015*b*), 'Linear filtering of sample covariances for ensemble-based data assimilation. Part II: Application to a convective-scale NWP model', *Monthly Weather Review* **143**(5), 1644–1664.

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.

Montmerle, T., Lafore, J.-P., Berre, L. and Fischer, C. (2006), 'Limited-area model error statistics over Western Africa: Comparisons with midlatitude results', *Quarterly Journal of the Royal Meteorological Society* **132**(614), 213–230.

Montmerle, T., Michel, Y., Arbogast, E., Ménétrier, B. and Brousseau, P. (2018), 'A 3D ensemble variational data assimilation scheme for the limited-area AROME model: Formulation and preliminary results', *Quarterly Journal of the Royal Meteorological Society* **144**(716), 2196–2215.

Müller, M., Homleid, M., Ivarsson, K.-I., Køltzow, M. A., Lindskog, M., Midtbø, K. H., Andrae, U., Aspelien, T., Berggren, L., Bjørge, D. et al. (2017), 'AROME-MetCoOp: A Nordic convective-scale operational weather prediction model', *Weather and Forecasting* **32**(2), 609–627.

Necker, T., Geiss, S., Weissmann, M., Ruiz, J., Miyoshi, T. and Lien, G.-Y. (2020), 'A convective-scale 1,000-member ensemble simulation and potential applications', *Quarterly Journal of the Royal Meteorological Society* **146**(728), 1423–1442.

Necker, T., Hinger, D., Griewank, P. J., Miyoshi, T. and Weissmann, M. (2023), 'Guidance on how to improve vertical covariance localization based on a 1000-member ensemble', *Nonlinear Processes in Geophysics* **30**(1), 13–29.

Neyestani, A., Gustafsson, N., Ghader, S., Mohebalhojeh, A. R. and Körnich, H. (2021), 'Operational convective-scale data assimilation over Iran: A comparison between WRF and HARMONIE-AROME', *Dynamics of Atmospheres and Oceans* **95**, 101242.

Nuryanto, D. E., Pawitan, H., Hidayat, R. and Aldrian, E. (2017), Propagation of convective complex systems triggering potential flooding rainfall of Greater Jakarta using satellite data, *in* 'IOP Conference Series: Earth and Environmental Science', Vol. 54, IOP Publishing, p. 012028.

Parrish, D. F. and Derber, J. C. (1992), 'The National Meteorological Center's spectral statistical-interpolation analysis system', *Monthly Weather Review* **120**(8), 1747–1763.

Penny, S. G. (2014), 'The hybrid local ensemble transform Kalman filter', *Monthly Weather Review* **142**(6), 2139–2149.

Petrie, R. E., Bannister, R. N. and Cullen, M. J. P. (2017), 'The ABC model: A non-hydrostatic toy model for use in convective-scale data assimilation investigations', *Geoscientific Model Development* **10**(12), 4419–4441.

Piccolo, C. (2011), 'Growth of forecast errors from covariances modeled by 4DVAR and ETKF methods', *Monthly Weather Review* **139**(5), 1505–1518.

Porson, A. N., Hagelin, S., Boyd, D. F., Roberts, N. M., North, R., Webster, S. and Lo, J. C.-F. (2019), 'Extreme rainfall sensitivity in convective-scale ensemble modelling over Singapore', *Quarterly Journal of the Royal Meteorological Society* **145**(724), 3004–3022.

Posselt, D. J. and Bishop, C. H. (2018), 'Nonlinear data assimilation for clouds and precipitation using a gamma inverse-gamma ensemble filter', *Quarterly Journal of the Royal Meteorological Society* **144**(716), 2331–2349.

Poterjoy, J. (2022), 'Implications of multivariate non-Gaussian data assimilation for multiscale weather prediction', *Monthly Weather Review* **150**(6), 1475–1493.

Poterjoy, J., Sobash, R. A. and Anderson, J. L. (2017), 'Convective-scale data assimilation for the weather research and forecasting model using the local particle filter', *Monthly Weather Review* **145**(5), 1897–1918.

Roberts, N. M. and Lean, H. W. (2008), 'Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events', *Monthly Weather Review* **136**(1), 78–97.

Sadiki, W. and Fischer, C. (2005), 'A posteriori validation applied to the 3D-VAR Arpège and Aladin data assimilation systems', *Tellus A: Dynamic Meteorology and Oceanography* **57**(1), 21–34.

Sakov, P. and Oke, P. R. (2008), 'A deterministic formulation of the ensemble Kalman filter: An alternative to ensemble square root filters', *Tellus A: Dynamic Meteorology and Oceanography* **60**(2), 361–371.

Schwartz, C. S., Liu, Z., Huang, X.-Y., Kuo, Y.-H. and Fong, C.-T. (2013), 'Comparing limited-area 3DVAR and hybrid variational-ensemble data assimilation methods for typhoon track forecasts: Sensitivity to outer loops and vortex relocation', *Monthly Weather Review* **141**(12), 4350–4372.

Seity, Y., Brousseau, P., Malardel, S., Hello, G., Bénard, P., Bouttier, F., Lac, C. and Masson, V. (2011), 'The AROME-France convective-scale operational model', *Monthly Weather Review* **139**(3), 976–991.

Shen, F., Min, J. and Xu, D. (2016), 'Assimilation of radar radial velocity data with the WRF hybrid ETKF–3DVAR system for the prediction of Hurricane Ike (2008)', *Atmospheric Research* **169**, 127–138.

Simonin, D., Ballard, S. and Li, Z. (2014), 'Doppler radar radial wind assimilation using an hourly cycling 3D-Var with a 1.5 km resolution version of the Met Office Unified Model for nowcasting', *Quarterly Journal of the Royal Meteorological Society* **140**(684), 2298–2314.

Singh, S. K. and Prasad, V. (2019), 'Evaluation of precipitation forecasts from 3D-Var and hybrid GSI-based system during Indian summer monsoon 2015', *Meteorology and Atmospheric Physics* **131**, 455–465.

Široká, M., Fischer, C., Cassé, V., Brožková, R. and Geleyn, J.-F. (2003), 'The definition of mesoscale selective forecast error covariances for a limited area variational analysis', *Meteorology and Atmospheric Physics* **82**(1), 227–244.

Sobel, A. H., Nilsson, J. and Polvani, L. M. (2001), 'The weak temperature gradient approximation and balanced tropical moisture waves', *Journal of the Atmospheric Sciences* **58**(23), 3650–3665.

Stanesic, A., Horvath, K. and Keresturi, E. (2019), 'Comparison of NMC and ensemble-based climatological background-error covariances in an operational limited-area data assimilation system', *Atmosphere* **10**(10), 570.

Stanley, Z., Grooms, I. and Kleiber, W. (2021), 'Multivariate localization functions for strongly coupled data assimilation in the bivariate Lorenz 96 system', *Nonlinear Processes in Geophysics* **28**(4), 565–583.

Sun, J., Wang, H., Tong, W., Zhang, Y., Lin, C.-Y. and Xu, D. (2016), 'Comparison of the impacts of momentum control variables on high-resolution variational data assimilation and precipitation forecasting', *Monthly Weather Review* **144**(1), 149–169.

Sun, T., Chen, Y., Meng, D. and Chen, H. (2021), 'Background error covariance statistics of hydrometeor control variables based on Gaussian transform', *Advances in Atmospheric Sciences* **38**, 831–844.

Tang, Y., Lean, H. W. and Bornemann, J. (2013), 'The benefits of the Met Office variable resolution NWP model for forecasting convection', *Meteorological Applications* **20**(4), 417–426.

Tangang, F. T., Juneng, L., Salimun, E., Vinayachandran, P., Seng, Y. K., Reason, C., Behera, S. K. and Yasunari, T. (2008), 'On the roles of the northeast cold surge, the Borneo vortex, the Madden-Julian Oscillation, and the Indian Ocean Dipole during

the extreme 2006/2007 flood in southern Peninsular Malaysia', *Geophysical Research Letters* **35**(14).

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.

Toth, Z. and Kalnay, E. (1993), 'Ensemble forecasting at NMC: The generation of perturbations', *Bulletin of the American Meteorological Society* **74**(12), 2317–2330.

Toth, Z. and Kalnay, E. (1997), 'Ensemble forecasting at NCEP and the breeding method', *Monthly Weather Review* **125**(12), 3297–3319.

Ullrich, P. A., Jablonowski, C., Kent, J., Lauritzen, P. H., Nair, R., Reed, K. A., Zarzycki, C. M., Hall, D. M., Dazlich, D., Heikes, R. et al. (2017), 'DCMIP2016: A review of non-hydrostatic dynamical core design and intercomparison of participating models', *Geoscientific Model Development* **10**(12), 4477–4509.

Vaughan, A., Markou, S., Tebbutt, W., Requeima, J., Bruinsma, W. P., Andersson, T. R., Herzog, M., Lane, N. D., Hosking, J. S. and Turner, R. E. (2024), 'Aardvark weather: end-to-end data-driven weather forecasting', *arXiv preprint arXiv:2404.00411* .

Vetra-Carvalho, S., Dixon, M., Migliorini, S., Nichols, N. K. and Ballard, S. P. (2012), 'Breakdown of hydrostatic balance at convective scales in the forecast errors in the Met Office Unified Model', *Quarterly Journal of the Royal Meteorological Society* **138**(668), 1709–1720.

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.

Wang, H., Liu, Y., Duan, J., Shi, Y., Lou, X. and Li, J. (2022), 'Assimilation of radar reflectivity using a time-lagged ensemble based ensemble Kalman filter with the "cloud-dependent" background error covariances', *Journal of Geophysical Research: Atmospheres* **127**(10), e2021JD036207.

Wang, X., Barker, D. M., Snyder, C. and Hamill, T. M. (2008*a*), 'A hybrid ETKF–3DVAR data assimilation scheme for the WRF model. Part I: Observing system simulation experiment', *Monthly Weather Review* **136**(12), 5116–5131.

Wang, X., Barker, D. M., Snyder, C. and Hamill, T. M. (2008*b*), 'A hybrid ETKF–3DVAR data assimilation scheme for the WRF model. Part II: Real observation experiments', *Monthly Weather Review* **136**(12), 5132–5147.

Wang, X., Parrish, D., Kleist, D. and Whitaker, J. (2013), 'GSI 3DVar-based ensemble–variational hybrid data assimilation for NCEP Global Forecast System: Single-resolution experiments', *Monthly Weather Review* **141**(11), 4098–4117.

Wang, X., Snyder, C. and Hamill, T. M. (2007), 'On the theoretical equivalence of differently proposed ensemble–3DVAR hybrid analysis schemes', *Monthly Weather Review* **135**(1), 222–227.

Wang, Y. and Wang, X. (2021), 'Development of convective-scale static background error covariance within GSI-based hybrid EnVar system for direct radar reflectivity data assimilation', *Monthly Weather Review* **149**(8), 2713–2736.

Wang, Y. and Wang, X. (2023), 'Simultaneous multiscale data assimilation using scale- and variable-dependent localization in EnVar for convection allowing analyses and forecasts: Methodology and experiments for a tornadic supercell', *Journal of Advances in Modeling Earth Systems* **15**(5), e2022MS003430.

Wattrelot, E., Montmerle, T. and Mahfouf, J. (2016), Higher density radar assimilation in the operational AROME model at 1.3 km horizontal resolution, *in* '6th Workshop on the Impact of Various Observing Systems on NWP', pp. 10–13.

Whitaker, J. S. and Hamill, T. M. (2002), 'Ensemble data assimilation without perturbed observations', *Monthly Weather Review* **130**(7), 1913–1924.

Whitaker, J. S. and Loughe, A. F. (1998), 'The relationship between ensemble spread and ensemble mean skill', *Monthly Weather Review* **126**(12), 3292–3302.

WMO (2022), 'Observing Systems Capability Analysis and Review (OSCAR)', https://space.oscar.wmo.int.

Würsch, M. and Craig, G. C. (2014), 'A simple dynamical model of cumulus convection for data assimilation research', *Meteorologische Zeitschrift* (5), 483–490.

Yang, G.-Y. and Slingo, J. (2001), 'The diurnal cycle in the tropics', *Monthly Weather Review* **129**(4), 784–801.

Yano, J.-I. and Bonazzola, M. (2009), 'Scale analysis for large-scale tropical atmospheric dynamics', *Journal of the Atmospheric Sciences* **66**(1), 159–172.

Žagar, N., Gustafsson, N. and Källén, E. (2004), 'Variational data assimilation in the tropics: The impact of a background-error constraint', *Quarterly Journal of the Royal Meteorological Society* **130**(596), 103–125.

Zhang, F., Sun, Y. Q., Magnusson, L., Buizza, R., Lin, S.-J., Chen, J.-H. and Emanuel, K. (2019), 'What is the predictability limit of midlatitude weather?', *Journal of the Atmospheric Sciences* **76**(4), 1077–1091.

Zhang, M. and Zhang, F. (2012), 'E4DVar: Coupling an ensemble Kalman filter with four-dimensional variational data assimilation in a limited-area weather prediction model', *Monthly Weather Review* **140**(2), 587–600.

Zhu, J. and Bannister, R. N. (2023), 'The Hydro-ABC model (Version 2.0): A simplified convective-scale model with moist dynamics', *Geoscientific Model Development* **16**(21), 6067–6085.