The effective use of anchor observations in variational bias correction in the presence of model bias

Article

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

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

RESEARCH ARTICLE

The effective use of anchor observations in variational bias correction in the presence of model bias

Devon J. Francis[1,2] ⊙ | **Alison M. Fowler[1,2]** ⊙ | **Amos S. Lawless[1,2]** ⊙ | **John Eyre[3]** ⊙ |
Stefano Migliorini[3] ⊙

[1]School of Mathematical, Physical and Computational Sciences, University of Reading, Reading, UK

[2]National Centre for Earth Observation, Reading, UK

[3]Met Office, Exeter, UK

Correspondence
Devon J. Francis, Department of Meteorology, Meteorology Building, University of Reading, Whiteknights Road, Earley Gate, Reading, RG6 6ET, Berkshire, UK.
Email: d.francis@pgr.reading.ac.uk

In numerical weather prediction, satellite radiance observations have a significant impact on forecast skill. However, radiance observations must usually be bias-corrected for the satellite data to have a positive impact. Many operational centres use variational bias correction (VarBC) to correct the observation biases, but VarBC assumes that there is no model bias within the system. As model biases are often non-negligible, unbiased observations (anchor observations) are known to play an important role in VarBC to reduce the contamination of model bias. However, more work is needed to understand what properties the network of anchor observations needs to have to reduce most contamination of model bias. We derive analytical expressions to show the sensitivity of the bias correction to the anchor observations and the expected value of the error in the analysed bias-correction coefficients. We find that the precision and location of the anchor observations are important in reducing the contamination of model bias in the estimate of observation bias. Anchor observations work best at reducing the effect of model bias when they observe the same state variables as the bias-corrected observations. When this is not the case, strong background-error correlations become more important, as they allow more information about the model bias to be passed from the anchor observations to the bias-corrected observations. The model bias observed by both the biased and anchor observations must be similar, otherwise the anchor observations cannot reduce the contamination of model bias in the observation-bias correction. These results show that, in operational systems, regions with sparse anchor observations could be more susceptible to model biases within the radiance observation-bias corrections. We demonstrate these results in a series of idealised numerical experiments that use the Lorenz 96 model as a simplified model of the atmosphere.

KEYWORDS
anchor observations, bias correction, data assimilation, model bias, unbiased observations, VarBC

1 | INTRODUCTION

In numerical weather prediction (NWP), the models that describe the evolution of the atmosphere are chaotic, which means that errors in the initial state will lead to larger errors in the forecast. To minimise errors in the initial state, it is important to incorporate observations, so that we have a more complete picture of the full system. Data assimilation (DA) is used to combine measurements from observations and previous forecasts (which we call the background state) to find the best estimate of the state of the atmosphere at the start of the forecast.

Observations from satellites are very important for NWP, as they have a large impact on improving NWP forecasts (English *et al.*, 2013; Eyre *et al.*, 2022). In particular, microwave and infrared hyperspectral radiances have a higher impact on NWP forecasts than other observation types (Candy *et al.*, 2021). However, most radiance observations need to be bias-corrected (Dee and Uppala, 2009), and this is now done in most operational centres such as the Met Office (Cameron and Bell, 2018) and European Centre for Medium-Range Weather Forecasts (ECMWF: (Han and Bormann, 2016) by variational bias correction (VarBC), a system that calculates the bias correction adaptively and can update the bias-correction coefficient at each cycle (Auligné *et al.*, 2007).

In addition to observation biases, biases can also be present as systematic errors in the model, into which the observations are assimilated. Model biases come from systematic errors in the formulation of the models, which approximate the physics and dynamics of the (real) atmosphere. Some operational centres—for example, ECMWF (Laloyaux *et al.*, 2020a)—use weak-constraint four-dimensional variational assimilation (WC4DVar) to reduce model biases in the stratosphere. This has recently had more success, due to the separation of scales between model bias (large-scale) and initial condition errors (smaller scale), as well as a large number of radio occultation observations available in the stratosphere, which can be assumed to be unbiased. WC4DVar works by including a forcing parameter within the model trajectory, which represents the model bias. The cost function is then minimised as a function of both the state and the new model bias parameter.

WC4DVar is not used in all NWP centres, due to its technical difficulty and the resources needed to implement it. Therefore model bias is often not accounted for explicitly within VarBC. If model biases are left within VarBC, then the observations will be corrected towards the model bias, as the system struggles to separate sources of bias (Talagrand, 1998). Therefore, as with WC4DVar, unbiased observations need to be used to anchor the system so that the magnitude of the analysis bias is smaller than that of

the forecast model. We call these observations that are not bias-corrected "anchor observations" (Eyre, 2016).

Currently, several types of observations are used as anchor observations within NWP systems. Radiosondes have been used as observations in NWP for over seven decades, measuring upper-air temperature, pressure, humidity, and wind. They are mainly located over land, but some are launched from ships to get some coverage over oceans. Most radiosondes have a bias due to solar and/or infrared radiation, but these are corrected prior to data assimilation so that the data in an assimilation system are considered to be unbiased (Sun *et al.*, 2013). Radio occultation (RO) measurements provide information about temperature in the stratosphere and upper troposphere with good spatial coverage across the Earth. They do not need to be bias-corrected, as any biases occurring in the bending angles are small in comparison with other biases within the DA system, hence RO observations work well as anchor observations (Healy and Thépaut, 2006; Cucurull *et al.*, 2014). Some radiance observations are not bias-corrected using VarBC, so that they can be used as anchor observations, as it is assumed that the model bias is much bigger than the observation bias in the variables that they observe, although these observations may still be bias-corrected using a static bias-correction prior to use in VarBC. For example, AMSU-A channel 14 is used to anchor the upper stratosphere (Di Tomaso and Bormann, 2011) in addition to RO; selected hyperspectral infrared window channels are used to anchor skin temperature at the Met Office; and selected ozone channels from hyperspectral infrared instruments are used to anchor the upper tropospheric/lower stratospheric ozone analysis at ECMWF (Han and McNally, 2010). However, the assumption that observation biases are small for these observations can be wrong: for example, biases in the observation operator may be large (Han and Bormann, 2016).

Over the past 30 years there has been a significant increase in the number of satellite observations, due to their large impact on NWP (Eyre *et al.*, 2020; Saunders, 2021; Eyre *et al.*, 2022). As a result, it is becoming more important to use unbiased observations more effectively in reducing the contamination of model bias in VarBC. Previous work from Eyre (2016) shows the importance of anchor observations in the presence of model bias for a scalar system with no explicit random error in the observations. Eyre (2016) shows that if there is model bias present in a VarBC system then, as the number of anchor observations reduces, the observation-bias correction is more affected by model bias, so the analysis bias will tend towards the model bias. In this study, we extend the work of Eyre (2016) by looking at a multivariate system with explicit random error to look at the role of the locations of

anchor observations and their uncertainty characteristics in reducing the effect of model bias on observation-bias correction.

The article has the following structure. In Section 2, the general method of VarBC is explained. In Section 3, we extend current VarBC theory to include two types of observation to demonstrate the role of anchor observations: in Section 3.1 we derive analysis equations for the state and bias coefficients when there are both observations that are bias-corrected and observations that are used as anchor observations; in Section 3.2 we look at the importance of the location of the anchor observations relative to the bias-corrected observations; in Section 3.3 we look at the importance of the location of the anchor observations relative to the locations of model bias; and in Section 3.4 we show how bias in the analysed bias-correction coefficients filters into the state in subsequent cycles. In Sections 4 and 5 we test the theory using an idealised 40-variable model of the atmosphere (Lorenz, 1996): Section 4 gives details about the setup of the toy data assimilation system used in the experiments and Section 5 presents idealised experiments to show the following: how the observation bias coefficients are affected by model bias, depending on the precision of the anchor observations; the effect of anchor and bias-corrected observations observing different parts of the state; and how the model bias contaminates both the observation-bias correction and the state estimation when VarBC is cycled. Finally, we present our conclusions and further discussion in Section 6.

2 | BIAS CORRECTION OF OBSERVATIONS

Auligné *et al.* (2007) describe three types of bias-correction techniques: static, "offline", and variational (VarBC). In the static scheme, a vector to describe the observation biases is calculated via the mean difference between the observations and the background state and this term is taken to be fixed until the bias correction is updated. As a result, this scheme cannot take into account any variation in the bias over different DA cycles since the last update. In the "offline" scheme, the bias-correction coefficients are calculated at every new cycle based on the background state from the analysis of the previous cycle. The bias-correction coefficients are calculated by minimising the cost function with respect to the bias-correction coefficients. Although this scheme can adapt more easily than the static scheme to changing biases, the bias coefficients are calculated via the background state, so the "offline" scheme will be more affected by model bias than VarBC would be (Eyre, 2016). VarBC, on the other

hand, is an adaptive scheme that updates the bias coefficients within the data assimilation system, taking advantage of the latest unbiased anchor observations. This is the method primarily used operationally, for example at the Met Office (Cameron and Bell, 2018) and ECMWF (Han and Bormann, 2016). In this article, we will only look at VarBC.

The bias correction for the kth observation (e.g., a measurement taken by an individual channel of a particular satellite) is given by

$$c_k = s_k + \sum_{i=1}^{r_k} \beta_{k,i} p_{k,i}(\mathbf{x}) \qquad (1)$$

where s_k is a constant; $p_{k,i}$ are the r_k predictors used for the kth observation; $\mathbf{x} \in \mathbb{R}^n$ is the NWP model state; and $\beta_{k,i}$ are the corresponding bias-correction coefficients that are updated at each cycle for each instrument that is bias-corrected. We define the vector $\boldsymbol{\beta} \in \mathbb{R}^r$ to hold all values of $\beta_{k,i}$, which for simplicity we will call the bias coefficient, where $r = \sum_k r_k$. We define the vector function $c(\mathbf{x}, \boldsymbol{\beta}) \in \mathbb{R}^{m_1}$ to be the bias correction for all observations that are bias-corrected, where m_1 is the number of biased observations.

The Met Office uses the following predictors: air-mass thicknesses, which are the differences in height between different pressure levels; Legendre polynomials to describe the spatial variation of bias across a scan; Fourier series to describe orbital bias (only used for SSMIS and MWRI); and a constant predictor that is not dependent on location (Cameron and Bell, 2018), which is given by $p_{k,0} = 1$ (Dee and Uppala, 2009). $p_{k,0}$ differs from the static bias correction s_k in Equation 1, as it is multiplied by the bias coefficient, $\beta_{k,0}$, so is updated within VarBC.

In VarBC the cost function, J, for three-dimensional variational assimilation (3DVar) is now dependent on $\mathbf{v} = \left(\mathbf{x}^{\mathrm{T}}, \ \boldsymbol{\beta}^{\mathrm{T}} \right)^{\mathrm{T}}$,

$$J(\mathbf{v}) = \frac{1}{2}(\mathbf{v} - \mathbf{v}^{\mathrm{b}})^{\mathrm{T}} \mathbf{B}_{\mathrm{v}}^{-1}(\mathbf{v} - \mathbf{v}^{\mathrm{b}})$$
$$+ \frac{1}{2}(\mathbf{y} - h_{\mathrm{v}}(\mathbf{v}))^{\mathrm{T}} \mathbf{R}^{-1}(\mathbf{y} - h_{\mathrm{v}}(\mathbf{v})), \qquad (2)$$

where $\mathbf{v}^{\mathrm{b}}$ is the background of the state and bias coefficients; $\mathbf{B}_{\mathrm{v}}$ is the background-error covariance matrix for the state and bias coefficients; $\mathbf{R}$ is the observation-error covariance matrix; $\mathbf{y}$ is the vector that contains all observations; $h_{\mathrm{v}}(\mathbf{v})$ is the observation operator given by a vector function that is the sum of the nonlinear observation operator transforming the state to the observation space, $h(\mathbf{x})$, and the bias correction, $c(\mathbf{x}, \boldsymbol{\beta})$ (Auligné *et al.*, 2007). $h_{\mathrm{v}}(\mathbf{v})$ will vary for each observation: for example, observations that are not bias-corrected will have $c(\mathbf{x}, \boldsymbol{\beta}) = \mathbf{0}$. In this article we refer to $\mathbf{y} - h_{\mathrm{v}}(\mathbf{v})$ as the innovation vector. The

cost function can be generalised to 4DVar by summing over the innovation vectors to include all observations in the assimilation window and incorporating the model into the observation operator. In the following we have looked only at 3DVar for simplicity.

Assuming the observation operator can be linearised around the background state $\mathbf{v}^b$, then the cost function in Equation 2 is minimised by the analysis equation,

$$\mathbf{v}^a = \mathbf{v}^b + \mathbf{K}_v(\mathbf{y} - h_v(\mathbf{v}^b)), \qquad (3)$$

where the superscripts a and b denote the analysis and background, respectively, and $\mathbf{K}_v$ is known as the Kalman gain matrix for $\mathbf{v}$, given by

$$\mathbf{K}_v = \mathbf{B}_v \mathbf{H}_v^T (\mathbf{H}_v \mathbf{B}_v \mathbf{H}_v^T + \mathbf{R})^{-1} \in \mathbb{R}^{(n+r) \times m}, \qquad (4)$$

where $\mathbf{H}_v$ is the Jacobian of $h_v(\mathbf{v}^b)$ and m is the total number of combined anchor and bias-corrected observations. We will refer to $\mathbf{K}_v$ as the sensitivity of the control vector to the observations.

3 | THE IMPORTANCE OF ANCHOR OBSERVATIONS IN VARBC: THEORETICAL RESULTS

Through experience at operational centres, anchor observations are known to be essential in VarBC. Their importance for constraining the system from diverging to the model bias was highlighted in the theoretical work of Eyre (2016) in a scalar system. In this section we derive general insight for VarBC that demonstrates the role of the anchor observations in limiting the contamination of model biases. We use this theory to understand how the role of anchor observations is sensitive to their error characteristics and their location within the spatial domain. We then look at the effect that model bias has on the bias coefficients and how this is translated into the state in subsequent cycles.

3.1 | Setup and derivation of theory for two observation types

In order to separate the roles of the bias-corrected and anchor observations, we write Equation 3 in terms of the state analysis and bias-coefficient analysis, by assuming that there are no correlations between the errors in the state and the bias coefficients, consistent with operational VarBC implementation (Dee, 2004). By also assuming that there are no correlations between the errors in the bias-corrected and anchor observations, we can further

extend Equation 3 to include two observation types, where $\mathbf{y}_{(1)} \in \mathbb{R}^{m_1}$ are bias-corrected within the assimilation (bias-corrected observations) and $\mathbf{y}_{(2)} \in \mathbb{R}^{m_2}$ are not bias-corrected (anchor observations):

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}_{xy_{(1)}} \mathbf{d}_{(1)} + \mathbf{K}_{xy_{(2)}} \mathbf{d}_{(2)}, \qquad (5)$$

$$\boldsymbol{\beta}^a = \boldsymbol{\beta}^b + \mathbf{K}_{\beta y_{(1)}} \mathbf{d}_{(1)} + \mathbf{K}_{\beta y_{(2)}} \mathbf{d}_{(2)}, \qquad (6)$$

where

$$\mathbf{d}_{(1)} = \mathbf{y}_{(1)} - h_{(1)}(\mathbf{x}^b) - c(\mathbf{x}^b, \boldsymbol{\beta}^b); \quad \mathbf{d}_{(2)} = \mathbf{y}_{(2)} - h_{(2)}(\mathbf{x}^b). \qquad (7)$$

We use $\mathbf{K}_{xy_{(1)}}$ and $\mathbf{K}_{xy_{(2)}}$ to denote the sensitivities of the state analysis to the bias-corrected and anchor observations, respectively, and $\mathbf{K}_{\beta y_{(1)}}$ and $\mathbf{K}_{\beta y_{(2)}}$ to denote the sensitivities of the bias-coefficient analysis to the bias-corrected and anchor observations, respectively. These four Kalman gain matrices are stored in the matrix $\mathbf{K}_v$:

$$\mathbf{K}_v = \mathbf{B}_v \mathbf{H}_v^T (\mathbf{H}_v \mathbf{B}_v \mathbf{H}_v^T + \mathbf{R})^{-1}$$
$$= \begin{pmatrix} \mathbf{K}_{xy_{(1)}} & \mathbf{K}_{xy_{(2)}} \\ \mathbf{K}_{\beta y_{(1)}} & \mathbf{K}_{\beta y_{(2)}} \end{pmatrix} \in \mathbb{R}^{(n+r) \times (m_1+m_2)}, \qquad (8)$$

where $\mathbf{H}_v$ is given by

$$\mathbf{H}_v(\mathbf{v}) = \begin{pmatrix} \mathbf{H}_{(1)} + \mathbf{C}_x & \mathbf{C}_\beta \\ \mathbf{H}_{(2)} & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{(m_1+m_2) \times (n+r)}. \qquad (9)$$

$\mathbf{H}_{(1)}$ is the Jacobian of the bias-corrected observation operator $h_{(1)}(\mathbf{x})$; $\mathbf{H}_{(2)}$ is the Jacobian of the anchor observation operator $h_{(2)}(\mathbf{x})$; $\mathbf{C}_x$ is the Jacobian of $c(\mathbf{x}, \boldsymbol{\beta})$ with respect to $\mathbf{x}$; and $\mathbf{C}_\beta$ is the Jacobian of $c(\mathbf{x}, \boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$. Operationally, the bias correction would be a function of the background state such that $c(\mathbf{x}^b, \boldsymbol{\beta})$, hence $\mathbf{C}_x$ would be zero, as c is dependent on the model state at the time of the background, not the variable x. Therefore in the remainder of this article we will set $\mathbf{C}_x$ to be zero. $\mathbf{B}_v$ is the background-error covariance matrix for the state and the bias coefficient, given by

$$\mathbf{B}_v = \begin{pmatrix} \mathbf{B}_x & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_\beta \end{pmatrix} \in \mathbb{R}^{(n+r) \times (n+r)}. \qquad (10)$$

$\mathbf{R}$ is the observation-error covariance matrix for the bias-corrected and anchor observations, given by

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_{(1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{(2)} \end{pmatrix} \in \mathbb{R}^{(m_1+m_2) \times (m_1+m_2)}. \qquad (11)$$

From Equation 6, the sensitivities of the bias-coefficient analysis to the bias-corrected observations,

anchor observations, state background, and bias-coefficient background respectively are given by

$$\frac{\partial \boldsymbol{\beta}^{a}}{\partial \mathbf{y}_{(1)}} = \mathbf{K}_{\beta y_{(1)}}, \tag{12}$$

$$\frac{\partial \boldsymbol{\beta}^{a}}{\partial \mathbf{y}_{(2)}} = \mathbf{K}_{\beta y_{(2)}}, \tag{13}$$

$$\frac{\partial \boldsymbol{\beta}^{a}}{\partial \mathbf{x}^{b}} = -\mathbf{K}_{\beta y_{(1)}} \mathbf{H}_{(1)} - \mathbf{K}_{\beta y_{(2)}} \mathbf{H}_{(2)}, \tag{14}$$

$$\frac{\partial \boldsymbol{\beta}^{a}}{\partial \boldsymbol{\beta}^{b}} = \mathbf{I} - \mathbf{K}_{\beta y_{(1)}} \mathbf{C}_{\beta}, \tag{15}$$

where we have used the convention that a partial derivative is a row vector. We will use Equation 14 to study the dependence of $\boldsymbol{\beta}^{a}$ on the state background, which is the mechanism for model bias to be passed into the bias-coefficient analysis.

In order to understand the sensitivity of $\boldsymbol{\beta}^{a}$ to $\mathbf{x}^{b}$ via Equation 14, we need to know what $\mathbf{K}_{\beta y_{(1)}}$ and $\mathbf{K}_{\beta y_{(2)}}$ are dependent on. Upon calculating $\mathbf{K}_{v}$ in terms of $\mathbf{K}_{xy_{(1)}}$, $\mathbf{K}_{xy_{(2)}}$, $\mathbf{K}_{\beta y_{(1)}}$, and $\mathbf{K}_{\beta y_{(2)}}$ (see Appendix A), we can rewrite $\mathbf{K}_{\beta y_{(2)}}$ in terms of $\mathbf{K}_{\beta y_{(1)}}$:

$$\mathbf{K}_{\beta y_{(2)}} = -\mathbf{K}_{\beta y_{(1)}} \mathbf{H}_{(1)} \mathbf{B}_{x} \mathbf{H}_{(2)}^{T} (\mathbf{H}_{(2)} \mathbf{B}_{x} \mathbf{H}_{(2)}^{T} + \mathbf{R}_{(2)})^{-1}. \tag{16}$$

Therefore this shows that the sensitivity of the bias-coefficient analysis to the anchor observations is dependent on the sensitivity of the bias-coefficient analysis to the bias-corrected observations and vice versa ($\mathbf{K}_{\beta y_{(1)}}$ is also dependent on $\mathbf{K}_{\beta y_{(2)}}$).

As $\mathbf{K}_{\beta y_{(2)}}$ can be written in terms of $\mathbf{K}_{\beta y_{(1)}}$, we can rewrite the sensitivity of the bias-coefficient analysis to the state background, Equation 14, so that it is written in terms of $\mathbf{K}_{\beta y_{(1)}}$:

$$\frac{\partial \boldsymbol{\beta}^{a}}{\partial \mathbf{x}^{b}} = -\mathbf{K}_{\beta y_{(1)}} \mathbf{H}_{(1)} (\mathbf{I} - \mathbf{D}), \tag{17}$$

where $\mathbf{D}$ is given by

$$\mathbf{D} = \mathbf{B}_{x} \mathbf{H}_{(2)}^{T} (\mathbf{H}_{(2)} \mathbf{B}_{x} \mathbf{H}_{(2)}^{T} + \mathbf{R}_{(2)})^{-1} \mathbf{H}_{(2)}. \tag{18}$$

In order to understand how biases in the model are passed into the bias-coefficient analysis, we define the errors in the analysis, background, and observations for both the state and the bias coefficients as follows:

$$\boldsymbol{\epsilon}_{x}^{a} = \mathbf{x}^{a} - \mathbf{x}^{t}, \quad \boldsymbol{\epsilon}_{\beta}^{a} = \boldsymbol{\beta}^{a} - \boldsymbol{\beta}^{t}, \tag{19}$$

$$\boldsymbol{\epsilon}_{x}^{b} = \mathbf{x}^{b} - \mathbf{x}^{t}, \quad \boldsymbol{\epsilon}_{\beta}^{b} = \boldsymbol{\beta}^{b} - \boldsymbol{\beta}^{t}, \quad \boldsymbol{\epsilon}_{v}^{b} = \mathbf{v}^{b} - \mathbf{v}^{t}, \tag{20}$$

$$\boldsymbol{\epsilon}_{1}^{o} = \mathbf{y}_{(1)} - h_{(1)}(\mathbf{x}^{t}) - c(\mathbf{x}^{t}, \boldsymbol{\beta}^{t}), \quad \boldsymbol{\epsilon}_{2}^{o} = \mathbf{y}_{(2)} - h_{(2)}(\mathbf{x}^{t}), \tag{21}$$

where the "t" superscript denotes the true value of the state, bias coefficient, or control vector, respectively. The true bias coefficient is the theoretical vector that would correct the observations in the bias correction perfectly if the biased observations had no random error.

Taking the expected value of the analysis errors in the state and bias coefficient (see Appendix B), we find

$$\langle \boldsymbol{\epsilon}_{x}^{a} \rangle = \langle \boldsymbol{\epsilon}_{x}^{b} \rangle + \mathbf{K}_{xy_{(1)}} \langle \boldsymbol{\epsilon}_{1}^{o} \rangle - \mathbf{K}_{xy_{(1)}} \mathbf{H}_{v_{(1)}} \langle \boldsymbol{\epsilon}_{v}^{b} \rangle \\ + \mathbf{K}_{xy_{(2)}} \langle \boldsymbol{\epsilon}_{2}^{o} \rangle - \mathbf{K}_{xy_{(2)}} \mathbf{H}_{v_{(2)}} \langle \boldsymbol{\epsilon}_{v}^{b} \rangle, \tag{22}$$

$$\langle \boldsymbol{\epsilon}_{\beta}^{a} \rangle = \langle \boldsymbol{\epsilon}_{\beta}^{b} \rangle + \mathbf{K}_{\beta y_{(1)}} \langle \boldsymbol{\epsilon}_{1}^{o} \rangle - \mathbf{K}_{\beta y_{(1)}} \mathbf{H}_{v_{(1)}} \langle \boldsymbol{\epsilon}_{v}^{b} \rangle \\ + \mathbf{K}_{\beta y_{(2)}} \langle \boldsymbol{\epsilon}_{2}^{o} \rangle - \mathbf{K}_{\beta y_{(2)}} \mathbf{H}_{v_{(2)}} \langle \boldsymbol{\epsilon}_{v}^{b} \rangle, \tag{23}$$

where $\mathbf{H}_{v_{(1)}} = \begin{pmatrix} \mathbf{H}_{(1)}, & \mathbf{C}_{\beta} \end{pmatrix}$ and $\mathbf{H}_{v_{(2)}} = \begin{pmatrix} \mathbf{H}_{(2)}, & \mathbf{0} \end{pmatrix}$, see Equation 9. If the right-hand side of Equation 22 is nonzero then the state analysis is biased. If the right-hand side of Equation 23 is nonzero then the bias-coefficient analysis has a bias. Biases in $\boldsymbol{\beta}^{a}$ will filter into $\langle \boldsymbol{\epsilon}_{x}^{a} \rangle$ when $\boldsymbol{\beta}^{a}$ becomes $\boldsymbol{\beta}^{b}$ in the next cycle, as will be shown in more depth in Section 3.4.

Assuming that $c(\mathbf{x}^{t}, \boldsymbol{\beta}^{t})$ is the true bias correction, then $\langle \boldsymbol{\epsilon}_{1}^{o} \rangle = \mathbf{0}$, as $\boldsymbol{\epsilon}_{(1)}^{o}$ is not dependent on the state background. We assume the anchor observations are unbiased ($\langle \boldsymbol{\epsilon}_{2}^{o} \rangle = \mathbf{0}$). To isolate the effect of model bias on the bias-coefficient analysis, we make the theoretical assumption that the bias-coefficient background has no bias ($\langle \boldsymbol{\epsilon}_{\beta}^{b} \rangle = \mathbf{0}$). In subsequent cycles, $\langle \boldsymbol{\epsilon}_{\beta}^{b} \rangle$ will be propagated from the previous cycle, so this assumption can only be possible for the first cycle. We will demonstrate the impact that a biased background bias coefficient has on the state and bias coefficient analyses in Sections 3.4 and 5.2.1. Using Appendix C, the expected value of the bias-coefficient analysis error, Equation 23, simplifies to

$$\langle \boldsymbol{\epsilon}_{\beta}^{a} \rangle \Big|_{\langle \boldsymbol{\epsilon}_{1}^{o} \rangle = \mathbf{0}, \langle \boldsymbol{\epsilon}_{2}^{o} \rangle = \mathbf{0}, \langle \boldsymbol{\epsilon}_{\beta}^{b} \rangle = \mathbf{0}} = -\mathbf{K}_{\beta y_{(1)}} \mathbf{H}_{(1)} (\mathbf{I} - \mathbf{D}) \langle \boldsymbol{\epsilon}_{x}^{b} \rangle, \tag{24}$$

where $\mathbf{D}$ is as defined in Equation 18 and we denote $\langle \boldsymbol{\epsilon}_{\beta}^{a} \rangle$ with the above assumptions as $\langle \boldsymbol{\epsilon}_{\beta}^{a} \rangle_{t=0}$ to highlight that this is only valid for the initial cycle. In the next few sections we will study Equation 24 to see how $\langle \boldsymbol{\epsilon}_{\beta}^{a} \rangle_{t=0}$ is affected by model bias when we vary the anchor observation parameters.

The effect of the model bias on $\boldsymbol{\beta}^{a}$ will be small if at least one term in the product $\mathbf{K}_{\beta y_{(1)}} \mathbf{H}_{(1)} (\mathbf{I} - \mathbf{D})$ is also small. If $\mathbf{K}_{\beta y_{(1)}}$ is small, then the sensitivity of the bias-coefficient analysis to the bias-corrected observations, Equation 12, would be small. If $\mathbf{H}_{(1)}$ is small, then the bias-corrected observations would not be used to determine the state analysis. Therefore, as these are both uninteresting cases for determining the bias-coefficient analysis, in this study

we will focus instead on when the magnitude of $\mathbf{I} - \mathbf{D}$ is small depending on the parameters given in the system.

Note that $\langle \epsilon_\beta^a \rangle_{t=0}$ can also be rewritten in terms of the sensitivity of the bias-coefficient analysis to the state background, Equation 17, such that Equation 24 becomes

$$\langle \epsilon_\beta^a \rangle_{t=0} = \frac{\partial \boldsymbol{\beta}^a}{\partial \mathbf{x}^b} \langle \epsilon_x^b \rangle. \quad (25)$$

Therefore if the sensitivity of the bias-coefficient analysis to the state background reduces, less model bias will be able to contaminate the analysed bias coefficient.

In subsequent sections, we look at the importance of the location of the anchor observations relative to the bias-corrected observations for reducing $\mathbf{I} - \mathbf{D}$ and thus reducing the contamination of model bias in the observation-bias correction.

3.2 | Position of anchor observations relative to bias-corrected observations

In this section, we want to understand how anchor observations can reduce the contamination of model bias in the bias-coefficient analysis depending on whether the bias-corrected and anchor observations observe the same state variables.

3.2.1 | Anchor observations fully observe the domain

In order to understand the role of anchor observations in reducing the contamination of model bias in VarBC, we first consider an almost perfect case where we have anchor observations everywhere in the domain. If $\mathbf{H}_{(1)}$ and $\mathbf{H}_{(2)}$ are both equal to the identity, that is, the state is fully observed directly by both bias-corrected and anchor observations, then $\mathbf{D}$, Equation 18, is given by

$$\mathbf{D}\Big|_{\mathbf{H}_{(2)}=\mathbf{I}} = \mathbf{B}_x(\mathbf{B}_x + \mathbf{R}_{(2)})^{-1} := \mathbf{D_I}. \quad (26)$$

Equation 24 is then given by

$$\langle \epsilon_\beta^a \rangle_{t=0}\Big|_{\mathbf{D}=\mathbf{D_I}} = -\mathbf{K}_{\beta y_{(1)}}\Big|_{\mathbf{H}_{(1)}=\mathbf{H}_{(2)}=\mathbf{I}}(\mathbf{I} - \mathbf{D_I})\langle \epsilon_x^b \rangle := \langle \epsilon_\beta^a \rangle_{t=0,\mathbf{D_I}}. \quad (27)$$

Equation 27 shows that, even with complete observation coverage, $\langle \epsilon_\beta^a \rangle_{t=0,\mathbf{D_I}}$ is still nonzero, as it is still a function of the state model bias. However, if $\mathbf{D_I}$ tends to the identity, then the right-hand side of Equation 27 would tend to zero, such that the observation-bias correction would no longer be contaminated by the model bias. $\mathbf{D_I}$ would tend to the

identity when the elements of $\mathbf{R}_{(2)}$ are smaller than the elements of $\mathbf{B}_x$. This would occur when the anchor observations are more precise than the state backgrounds that they observe.

Overall, Equation 26 shows that, even with anchor observations at every model grid point, we still have model bias contaminating the bias-coefficient analysis. This can only be reduced by having anchor observations that are more precise than the state backgrounds that they observe.

3.2.2 | Anchor observations partially observe the domain

Anchor observations and observations to be bias-corrected could observe different parts of the state. Therefore in this section we derive equations for the expected value of the bias-coefficient analysis error and the sensitivity of the bias-coefficient analysis to the state background when the anchor observations do not observe the whole domain.

Let the state $\mathbf{x}$ be separated into two parts: $\mathbf{x}_\phi$ and $\mathbf{x}_\psi$, such that $\mathbf{x} = (\mathbf{x}_\phi^T, \quad \mathbf{x}_\psi^T)^T$. Let the anchor observations only observe a subset of the state, such that they only observe variables in $\mathbf{x}_\psi$. Bias-corrected observations could observe variables in $\mathbf{x}_\phi$ and $\mathbf{x}_\psi$. Then the linearised bias-corrected and anchor observation operators will be given by

$$\mathbf{H}_{(1)} = \left(\mathbf{H}_{(1)_\phi}, \quad \mathbf{H}_{(1)_\psi}\right), \quad (28)$$

$$\mathbf{H}_{(2)}\Big|_{\mathbf{H}_{(2)_\phi}=0} = \left(\mathbf{0}, \quad \mathbf{H}_{(2)_\psi}\right) := \mathbf{H}_{(2)p}, \quad (29)$$

where $\mathbf{H}_{(1)_\phi}$ is related to observations of $\mathbf{x}_\phi$ and $\mathbf{H}_{(1)_\psi}$ and $\mathbf{H}_{(2)_\psi}$ are related to observations of $\mathbf{x}_\psi$. We have denoted $\mathbf{H}_{(2)p}$ to be the Jacobian of the anchor observation operator when the state is only partially observed by anchor observations, that is, anchor observations only observe variables in $\mathbf{x}_\psi$.

The background-error covariance matrix that describes the relationships between $\mathbf{x}_\phi^b$ and $\mathbf{x}_\psi^b$ is

$$\mathbf{B}_x = \begin{pmatrix} \mathbf{B}_{x_\phi} & \mathbf{B}_{x_{\phi\psi}} \\ \mathbf{B}_{x_{\phi\psi}}^T & \mathbf{B}_{x_\psi} \end{pmatrix}. \quad (30)$$

The magnitude of $\mathbf{B}_{x_{\phi\psi}}$ determines how much information about the observations will be shared between the state variables (Bannister, 2008). For example, if the elements in $\mathbf{B}_{x_{\phi\psi}}$ are small then there are weak correlations between the errors in $\mathbf{x}_\phi^b$ and $\mathbf{x}_\psi^b$.

We are interested in how the value of $\mathbf{D}$, Equation 18, affects the sensitivity of the bias-coefficient analysis to

the state background. When the anchor observations only observe a subset of the state, we denote $\mathbf{D}$ by

$$\mathbf{D}\Big|_{\mathbf{H}_{(2)}=\mathbf{H}_{(2)p}} := \mathbf{D}_p = \begin{pmatrix} \mathbf{D}_{p\phi} & \mathbf{D}_{p\phi\psi} \\ \mathbf{D}_{p\psi\phi} & \mathbf{D}_{p\psi} \end{pmatrix}.$$

The blocks of $\mathbf{D}_p$ are given by (see Appendix D)

$$\mathbf{D}_{p\phi} = \mathbf{0}, \tag{31}$$

$$\mathbf{D}_{p\phi\psi} = \mathbf{B}_{\mathbf{x}_{\phi\psi}} \mathbf{H}_{(2)_\psi}^T (\mathbf{H}_{(2)_\psi} \mathbf{B}_{\mathbf{x}_\psi} \mathbf{H}_{(2)_\psi}^T + \mathbf{R}_{(2)})^{-1} \mathbf{H}_{(2)_\psi}, \tag{32}$$

$$\mathbf{D}_{p\psi\phi} = \mathbf{0}, \tag{33}$$

$$\mathbf{D}_{p\psi} = \mathbf{B}_{\mathbf{x}_\psi} \mathbf{H}_{(2)_\psi}^T (\mathbf{H}_{(2)_\psi} \mathbf{B}_{\mathbf{x}_\psi} \mathbf{H}_{(2)_\psi}^T + \mathbf{R}_{(2)})^{-1} \mathbf{H}_{(2)_\psi}. \tag{34}$$

Then the sensitivity of $\boldsymbol{\beta}^a$ to $\mathbf{x}^b$ (Equation 17), when $\mathbf{H}_{(2)} = \mathbf{H}_{(2)p}$, is given by

$$\frac{\partial \boldsymbol{\beta}^a}{\partial \mathbf{x}^b}\Big|_{\mathbf{H}_{(2)}=\mathbf{H}_{(2)p}}$$
$$= -\mathbf{K}_{\beta \mathbf{y}_{(1)}}\Big|_{\mathbf{H}_{(2)}=\mathbf{H}_{(2)p}} \left(\mathbf{H}_{(1)_\phi}, -\mathbf{H}_{(1)_\phi} \mathbf{D}_{p\phi\psi} + \mathbf{H}_{(1)_\psi} (\mathbf{I} - \mathbf{D}_{p\psi}) \right).$$

The first block element of the matrix in Equation 35 (i.e., $-\mathbf{K}_{\beta \mathbf{y}_{(1)}}\Big|_{\mathbf{H}_{(2)}=\mathbf{H}_{(2)p}} \mathbf{H}_{(1)_\phi}$) gives the sensitivity of $\boldsymbol{\beta}^a$ to $\mathbf{x}_\phi^b$ and the second block element of the matrix in Equation 35 (i.e., $-\mathbf{K}_{\beta \mathbf{y}_{(1)}}\Big|_{\mathbf{H}_{(2)}=\mathbf{H}_{(2)p}} [-\mathbf{H}_{(1)_\phi} \mathbf{D}_{p\phi\psi} + \mathbf{H}_{(1)_\psi} (\mathbf{I} - \mathbf{D}_{p\psi})]$) gives the sensitivity of $\boldsymbol{\beta}^a$ to $\mathbf{x}_\psi^b$. $\mathbf{D}_{p\phi}$ and $\mathbf{D}_{p\psi\phi}$ are zero (from Equations 31 and 33) so do not appear in this equation, but their role would be to vary the sensitivity of $\boldsymbol{\beta}^a$ on $\mathbf{x}_\phi^b$. Therefore, as $\mathbf{D}_{p\phi}$ and $\mathbf{D}_{p\psi\phi}$ are zero, the sensitivity of $\boldsymbol{\beta}^a$ to $\mathbf{x}_\phi^b$ is not explicitly dependent on anchor observations, when anchor observations do not observe state variables in $\mathbf{x}_\phi$. However, there is some implicit dependence of anchor observations in $\mathbf{K}_{\beta \mathbf{y}_{(1)}}$. If $\mathbf{D}_{p\phi\psi}$ and $\mathbf{D}_{p\psi}$ change magnitude, then this will vary the sensitivity of $\boldsymbol{\beta}^a$ to $\mathbf{x}_\psi^b$. Therefore, as $\mathbf{D}_{p\phi\psi}$ and $\mathbf{D}_{p\psi}$ are explicitly dependent on the anchor observations, the sensitivity of $\boldsymbol{\beta}^a$ to $\mathbf{x}_\psi$ is also explicitly dependent on the anchor observations, as in this case the anchor observations observe state variables in $\mathbf{x}_\psi$.

Putting $\mathbf{D}_p$ into Equation 24 gives $\langle \epsilon_\beta^a \rangle_{t=0}$ when the anchor observations observe part of the state:

$$\langle \epsilon_\beta^a \rangle = -\mathbf{K}_{\beta \mathbf{y}_{(1)}} \bigg(\mathbf{H}_{(1)_\phi} \langle \epsilon_{\mathbf{x}_\phi}^b \rangle$$
$$- \left[\mathbf{H}_{(1)_\phi} \mathbf{D}_{p\phi\psi} - \mathbf{H}_{(1)_\psi} (\mathbf{I} - \mathbf{D}_{p\psi}) \right] \langle \epsilon_{\mathbf{x}_\psi}^b \rangle \bigg)$$
$$:= \langle \epsilon_\beta^a \rangle_{t=0,\mathbf{D}_p}, \tag{36}$$

where $\langle \epsilon_{\mathbf{x}_\phi}^b \rangle$ and $\langle \epsilon_{\mathbf{x}_\psi}^b \rangle$ are the biases in the background state variables $\mathbf{x}_\phi^b$ and $\mathbf{x}_\psi^b$ respectively. We will use this equation further in Sections 3.2.3 and 3.3 to show how the sensitivity of the bias in $\boldsymbol{\beta}^a$ varies depending on the location of the model bias in relation to the anchor observations.

3.2.3 | Special case: anchor observations and bias-corrected observations observe different parts of the state

Next we look at the case where anchor and bias-corrected observations observe different parts of the state to understand the importance of the background-error covariance matrix. Let the bias-corrected observations only observe $\mathbf{x}_\phi$ and anchor observations only observe $\mathbf{x}_\psi$. Then the observation operators of both observation types will be given by

$$\left(\mathbf{H}_{(1)_\phi}, \quad \mathbf{0} \right) := \mathbf{H}_{(1)p}, \quad \left(\mathbf{0}, \quad \mathbf{H}_{(2)_\psi} \right) := \mathbf{H}_{(2)p}, \tag{37}$$

where we have denoted $\mathbf{H}_{(1)p}$ and $\mathbf{H}_{(2)p}$ to be the Jacobians of the observation operators when both bias-corrected and anchor observations observe the state only partially.

Substituting Equation 37 into Equation 36 gives the expected value of the bias-coefficient analysis when anchor and bias-corrected observations observe different parts of the state,

$$\langle \epsilon_\beta^a \rangle_{t=0,\mathbf{D}_p}\Big|_{\mathbf{H}_{(1)p}\mathbf{H}_{(2)p}} = -\mathbf{K}_{\beta \mathbf{y}_{(1)}} \mathbf{H}_{(1)_\phi} \left(\langle \epsilon_{\mathbf{x}_\phi}^b \rangle - \mathbf{D}_{p\phi\psi} \langle \epsilon_{\mathbf{x}_\psi}^b \rangle \right), \tag{38}$$

which holds because $\mathbf{H}_{(1)_\psi} = 0$.

The right-hand side of Equation 38 will reduce if $\langle \epsilon_{\mathbf{x}_\phi}^b \rangle - \mathbf{D}_{p\phi\psi} \langle \epsilon_{\mathbf{x}_\psi}^b \rangle$ tends towards zero. $\mathbf{D}_{p\phi\psi}$, from Equation 32, is linearly dependent on $\mathbf{B}_{\mathbf{x}_{\phi\psi}}$. Therefore, the strength of the background-error covariances between $\mathbf{x}_\phi^b$ and $\mathbf{x}_\psi^b$ will influence how small the difference term in Equation 38 is. In the case in which the model biases $\langle \epsilon_{\mathbf{x}_\phi}^b \rangle$ and $\langle \epsilon_{\mathbf{x}_\psi}^b \rangle$ have similar sign and magnitude, if $\mathbf{D}_{p\phi\psi}$ tends to the identity, the difference term will reduce and the expected value of the bias-coefficient analysis error will tend to zero. $\mathbf{D}_{p\phi\psi}$ will tend to the identity if the background-error correlations between the states are large. In practice, the model biases may be similar for different parts of the state and the background-error correlations between these two states could be large when restricted to specific atmospheric layers higher in the atmosphere. However, this will not be the case when the bias-corrected and anchor observations observe states vertically far apart: for example, biases in stratospheric

temperatures tend to be much bigger than biases in tropospheric temperatures.

To understand the importance of $\mathbf{B}_{x_{\phi\psi}}$ when the anchor and bias-corrected observations observe different parts of the state, let us look at the case when $\mathbf{B}_{x_{\phi\psi}} = \mathbf{0}$. In this case we can simplify $\mathbf{K}_v$ from Equation 8, as the off-diagonal blocks in $(\mathbf{H}_v\mathbf{B}_v\mathbf{H}_v^T + \mathbf{R})$ are zero:

$$\mathbf{H}_{(1)p}\mathbf{B}_x\mathbf{H}_{(2)p}^T = \mathbf{H}_{(1)_\phi}\mathbf{B}_{x_{\phi\psi}}\mathbf{H}_{(2)_\psi}^T = \mathbf{0} \qquad (39)$$

and

$$\mathbf{H}_{(2)p}\mathbf{B}_x\mathbf{H}_{(1)p}^T = \mathbf{H}_{(2)_\psi}\mathbf{B}_{x_{\phi\psi}}^T\mathbf{H}_{(1)_\phi}^T = \mathbf{0}. \qquad (40)$$

Putting Equations 39 and 40 into $\mathbf{H}_v\mathbf{B}_v\mathbf{H}_v^T + \mathbf{R}$ gives a block diagonal matrix. Then $\mathbf{K}_v$ from Equation 8 can be simplified to

$$\mathbf{K}_v\Big|_{\mathbf{B}_{x_{\phi\psi}}=0} = \begin{pmatrix} \mathbf{K}_{xy_{(1)}}\big|_{\mathbf{B}_{x_{\phi\psi}}=0} & \mathbf{K}_{xy_{(2)}}\big|_{\mathbf{B}_{x_{\phi\psi}}=0} \\ \mathbf{K}_{\beta y_{(1)}}\big|_{\mathbf{B}_{x_{\phi\psi}}=0} & \mathbf{K}_{\beta y_{(2)}}\big|_{\mathbf{B}_{x_{\phi\psi}}=0} \end{pmatrix}, \qquad (41)$$

where $\mathbf{K}_{\beta y_{(1)}}$ and $\mathbf{K}_{\beta y_{(2)}}$ can be simplified from Equations A11 and A12 to give

$$\mathbf{K}_{\beta y_{(1)}}\Big|_{\mathbf{B}_{x_{\phi\psi}}=0} = \mathbf{B}_\beta\mathbf{C}_\beta^T(\mathbf{H}_{(1)_\phi}\mathbf{B}_{x_\phi}\mathbf{H}_{(1)_\phi}^T + \mathbf{C}_\beta\mathbf{B}_\beta\mathbf{C}_\beta^T + \mathbf{R}_{(1)})^{-1}, \qquad (42)$$

$$\mathbf{K}_{\beta y_{(2)}}\Big|_{\mathbf{B}_{x_{\phi\psi}}=0} = \mathbf{0}. \qquad (43)$$

The derivation of $\mathbf{K}_{\beta y_{(1)}}$ and $\mathbf{K}_{\beta y_{(2)}}$ for the general case case be found in Appendix A. As $\mathbf{K}_{\beta y_{(2)}}\big|_{\mathbf{B}_{x_{\phi\psi}}=0} = \mathbf{0}$, from Equation 13 the sensitivity of the bias-coefficient analysis is independent of the anchor observations when there are no background-error cross-correlations between the states observed by the anchor observations and the states observed by the bias-corrected observations. This means anchor observations do not play a role in determining the bias coefficients if anchor observations do not observe the same state variables as the bias-corrected observations and information is not passed between $\mathbf{x}_\phi$ and $\mathbf{x}_\psi$ via $\mathbf{B}_{x_{\phi\psi}}$. Therefore, if bias-corrected and anchor observations observe different state variables, nonzero background-error covariances allow the anchor observations to be used to determine the bias coefficients and therefore reduce the effect of the model bias in $\boldsymbol{\beta}^a$. We will use this result in Section 3.3.

So far, we have not considered that the model bias may change magnitude depending on location. In the next

section, we will investigate scenarios where the location of the model bias differs relative to the location of the anchor observations.

3.3 | Position of anchor observations relative to model bias

In reality, biases in the model will not be uniformly distributed throughout the domain. For example, a version of the ECMWF Integrated Forecast System model has a cold bias between 100 and 10 hPa and a warm bias above 10 hPa (Laloyaux *et al.*, 2020b). In this section, we explore the case in which the model bias varies across the domain. For simplicity, we assume that some parts of the domain are biased and others not. We continue to assume that the anchor observations observe only a subset of the state, but assume that the bias-corrected observations could observe the whole state.

Figure 1 is a schematic diagram that depicts the different possibilities of the locations of the model bias in relation to the observations. The presence of model bias is shown by the patterned background; the circles with $\mathbf{H}_{(1)}$ show the parts of the state observed by the bias-corrected observations; and the circles with $\mathbf{H}_{(2)}$ show the parts of the state observed by anchor observations. In Figure 1 a the model bias is in $\mathbf{x}_\psi$, which is observed by anchor observations and some bias-corrected observations, in Figure 1 b the model bias is in $\mathbf{x}_\phi$, which is only observed by bias-corrected observations, and in Figure 1 c the model bias is in both $\mathbf{x}_\phi$ and $\mathbf{x}_\psi$, such that the whole state that is observed has model bias. In the next sections we assume that the model biases in $\mathbf{x}_\phi$ and $\mathbf{x}_\psi$ (when present) have the same sign and magnitude. These scenarios are discussed below.

3.3.1 | Model bias in state variables observed by anchor observations, but not in all state variables observed by bias-corrected observations

If $\langle\epsilon_{x_\phi}^b\rangle = \mathbf{0}$ (i.e., there is no model bias in $\mathbf{x}_\phi^b$, the state variables are observed only by the bias-corrected observations) but $\langle\epsilon_{x_\psi}^b\rangle \neq \mathbf{0}$ (i.e., there is model bias in $\mathbf{x}_\psi^b$, the state variables are observed by both the bias-corrected and anchor observations), then the expected value of the bias-coefficient analysis, Equation 36, becomes

$$\langle\epsilon_\beta^a\rangle_{t=0,\mathbf{D}_p}\Big|_{\langle\epsilon_{x_\phi}^b\rangle=0} = \mathbf{K}_{\beta y_{(1)}}\left[\mathbf{H}_{(1)_\phi}\mathbf{D}_{p\phi\psi} - \mathbf{H}_{(1)_\psi}(\mathbf{I} - \mathbf{D}_{p\psi})\right]$$

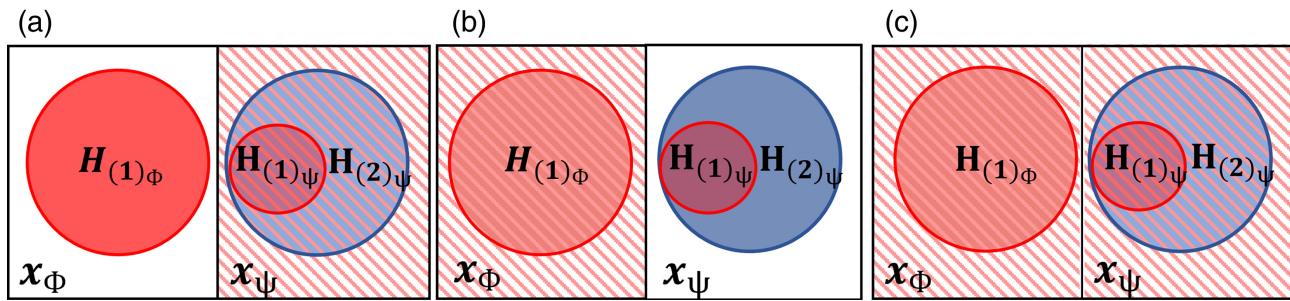$$\langle\epsilon_{x_\psi}^b\rangle. \qquad (44)$$

FIGURE 1 A schematic diagram depicting the location of the model bias in relation to the bias-corrected and anchor observations, where $\mathbf{H}_{(1)_\phi}$ is the observation operator relating the bias-corrected observations to $\mathbf{x}_\phi$; $\mathbf{H}_{(1)_\psi}$ is the observation operator relating the bias-corrected observations to $\mathbf{x}_\psi$; and $\mathbf{H}_{(2)_\psi}$ is the observation operator relating the anchor observations to $\mathbf{x}_\psi$. The model bias is shown by the patterned background. In (a) there is only model bias in $\mathbf{x}_\psi$, which is observed by the anchor observations and some bias-corrected observations, in (b) there is only model bias in $\mathbf{x}_\phi$, which is only observed by the bias-corrected observations, and in (c) there is model bias in both $\mathbf{x}_\phi$ and $\mathbf{x}_\psi$, which is observed by both bias-corrected and anchor observations. [Colour figure can be viewed at wileyonlinelibrary.com]

The effect of model bias from $\mathbf{x}_\psi^b$ in Equation 44 would be reduced if the matrix in the square brackets tends to zero. This could occur if both $\mathbf{H}_{(1)_\phi}\mathbf{D}_{p\phi\psi}$ and $\mathbf{H}_{(1)_\psi}(\mathbf{I} - \mathbf{D}_{p\psi})$ tend to zero, or if their difference tends to zero. If the bias-corrected observations observe state variables in $\mathbf{x}_\psi$ (so they observe the same part of the state as the anchor observations) such that $\mathbf{H}_{(1)_\psi}$ is nonzero, then the second term in the square brackets would tend to zero if $\mathbf{D}_{p\psi}^T$ tends to the identity. From Equation 34, this would occur if $\mathbf{R}_{(2)_\psi}$ is smaller than $\mathbf{B}_{\mathbf{x}_\psi}$, or, in other words, if the anchor observations were more precise than the state backgrounds that they observe, as we saw for a simplified case in Section 3.2.1. If the bias-corrected observations observe state variables in $\mathbf{x}_\phi$ (so they observe different parts of the state from the anchor observations) such that $\mathbf{H}_{(1)_\phi}$ is nonzero, then $\mathbf{H}_{(1)_\phi}\mathbf{D}_{p\phi\psi}$ would tend to zero if $\mathbf{D}_{p\phi\psi}$ tends to zero. Using Equation 32, $\mathbf{D}_{p\phi\psi}$ would tend to zero if the background-error covariances $\mathbf{B}_{\mathbf{x}_\phi\psi}$ are small, which would mean less information is passed between $\mathbf{x}_\phi^b$ and $\mathbf{x}_\psi^b$ than when the background-error covariances are larger.

Therefore, if anchor and bias-corrected observations observe the same state, the anchor observations can reduce the effect of the model bias in the state when the anchor observations are more precise than the state backgrounds they observe. If there are bias-corrected observations that observe a different part of the state from the anchor observations, which do not have model bias, then smaller background-error covariances between the two parts of the state will limit the amount of model bias able to contaminate the estimate of the bias coefficient, as the anchor observations cannot share information about the model bias with the bias-corrected observations. This is in contrast to Section 3.2.3, in which we showed that if anchor and bias-corrected observations observe different

states, but the model bias observed was the same, then they would need nonzero background-error cross-correlations for the anchor observations to pass information about the model bias to the bias-corrected observations.

3.3.2 | Model bias not in state variables observed by anchor observations

If there is no model bias in $\mathbf{x}_\psi^b$ (i.e., there is no model bias in state variables observed by anchor observations), such that $\langle \epsilon_{\mathbf{x}_\psi}^b \rangle = \mathbf{0}$, but there is model bias in $\mathbf{x}_\phi^b$ (i.e., there is model bias in state variables only observed by bias-corrected observations), then the expected value of the bias-coefficient analysis, Equation 36, reduces to

$$\langle \epsilon_\beta^a \rangle_{t=0,\mathbf{D}_p}\Big|_{\langle \epsilon_{\mathbf{x}_2}^b \rangle=\mathbf{0}} = -\mathbf{K}_{\beta\mathbf{y}_{(1)}}\mathbf{H}_{(1)_\phi}\langle \epsilon_{\mathbf{x}_\phi}^b \rangle. \tag{45}$$

In this case, the anchor observations cannot reduce the effect of the model bias in $\mathbf{x}_\phi^b$ via $\mathbf{D}_p$, as $\langle \epsilon_\beta^a \rangle_{t=0,\mathbf{D}_p}$ is no longer dependent on $\mathbf{D}_p$ and no other variables in Equation 45 are explicitly dependent on the anchor observations. Therefore, if anchor observations do not observe the parts of the state that have model bias, they cannot explicitly reduce the effect of model bias on $\boldsymbol{\beta}^a$. Anchor observations will only have an effect on model bias implicitly through $\mathbf{K}_{\beta\mathbf{y}_{(1)}}$, as $\mathbf{K}_{\beta\mathbf{y}_{(1)}}$ is implicitly dependent on $\mathbf{H}_{(2)}$ and $\mathbf{R}_{(2)}$ (see Equation A12). However, as was discussed at the end of Section 3.1, a small $\mathbf{K}_{\beta\mathbf{y}_{(1)}}$ would mean the bias-coefficient analysis is independent of the bias-corrected observations, so we are not interested in this case.

3.3.3 | Model bias in state variables observed by both anchor and bias-corrected observations

If we have model bias in both $\mathbf{x}_\phi^b$ and $\mathbf{x}_\psi^b$, then we come back to Equation 36, which gives the same results as in Section 3.2 : that is, giving a higher weighting to the anchor observations will reduce the contamination of model bias in the observation bias coefficient. If there are similar model biases in state variables observed by either bias-corrected or anchor observations, background-error covariances between state variables become more important in sharing information about the model bias.

In Sections 3.2 and 3.3, we have shown that, in order for the anchor observations to have the biggest impact on reducing the effect of model bias on $\boldsymbol{\beta}^a$, both anchor and bias-corrected observations need to observe model bias. If both anchor and bias-corrected observations observe the same parts of the state, the effect of the model bias is smallest when $\mathbf{D}_{p\psi}$ tends towards the identity, as was shown in Section 3.2.1. If both anchor and bias-corrected observations observe model bias, but do not observe the same state variables, background-error correlations with large magnitudes will have a bigger impact in reducing the effect of model bias on $\boldsymbol{\beta}^a$, as was shown in Section 3.2.3, as $\mathbf{D}_{p\phi\psi}$ is linearly dependent on $\mathbf{B}_{x_{\phi\psi}}$ (Equation 32).

3.4 | The effect of a biased bias-coefficient analysis on the state analysis in further cycles

So far we have only looked at the contamination of model bias in $\boldsymbol{\beta}^a$, not in $\mathbf{x}^a$. However, any bias in the bias-coefficient analysis will filter into the bias correction and therefore the state analysis in the next cycle. Within this section we extend the theory developed so far to understand the impact of model bias on the state analysis via the implementation of VarBC.

Cycle 1 (theory so far): At the initial time, we assume that $\boldsymbol{\beta}^b$ is unbiased. We assume the mean anchor observation errors are zero by definition. When the bias correction function is dependent on the true state and bias coefficient, $c(\mathbf{x}^t, \beta^t)$, we assume it is perfect, such that the expected value of the bias-corrected observation errors are zero. Then the mean values of the errors in the observations and bias-coefficient background at the first cycle are denoted by

$$\langle \epsilon_1^o \rangle_{t=0} = \langle \mathbf{y}_{(1)t=0} - h_{(1)}(\mathbf{x}_{t=0}^t) - c(\mathbf{x}_{t=0}^t, \beta_{t=0}^t) \rangle = \mathbf{0},$$
$$\langle \epsilon_2^o \rangle_{t=0} = \langle \mathbf{y}_{(2)t=0} - h_{(2)}(\mathbf{x}_{t=0}^t) \rangle = \mathbf{0}, \quad (46)$$

$$\langle \epsilon_\beta^b \rangle_{t=0} = \mathbf{0}. \quad (47)$$

If we assume that there is a bias in the state background, which arises from a model bias, then from Equations 22 and 23 we have that the expected value of the analysis errors at the first cycle for the state and bias coefficients respectively are

$$\langle \epsilon_x^a \rangle_{t=0} = (\mathbf{I} - \mathbf{K}_{xy_{(1)}} \mathbf{H}_{(1)} - \mathbf{K}_{xy_{(2)}} \mathbf{H}_{(2)}) \langle \epsilon_x^b \rangle_{t=0}, \quad (48)$$

$$\langle \epsilon_\beta^a \rangle_{t=0} = (-\mathbf{K}_{\beta y_{(1)}} \mathbf{H}_{(1)} - \mathbf{K}_{\beta y_{(2)}} \mathbf{H}_{(2)}) \langle \epsilon_x^b \rangle_{t=0}. \quad (49)$$

Note that Equation 49 is equivalent to Equation 24.

Cycle 2: We assume the bias coefficients are approximately constant between cycles, such that $\langle \epsilon_\beta^b \rangle_{t=1} = \langle \epsilon_\beta^a \rangle_{t=0}$. We assume that the expected values of the errors in both bias-corrected and unbiased observations are still zero, as the observation errors in Equation 47 are not dependent on the background state and background bias coefficient. The state background is the previous state analysis evolved forward via the linearised model $\mathbf{M}$, plus a bias increment $\boldsymbol{\eta}\Delta t$. The expected values of the background and observation errors at the second cycle are therefore given by

$$\langle \epsilon_\beta^b \rangle_{t=1} = \langle \epsilon_\beta^a \rangle_{t=0} = (-\mathbf{K}_{\beta y_{(1)}} \mathbf{H}_{(1)} - \mathbf{K}_{\beta y_{(2)}} \mathbf{H}_{(2)}) \langle \epsilon_x^b \rangle_{t=0}, \quad (50)$$

$$\langle \epsilon_x^b \rangle_{t=1} = \mathbf{M}_{0 \to 1}(\langle \epsilon_x^a \rangle_{t=0}) + \boldsymbol{\eta}\Delta t, \quad (51)$$

$$\langle \epsilon_1^o \rangle_{t=1} = \langle \mathbf{y}_{(1)t=1} - h_{(1)}(\mathbf{x}_{t=1}^t) - c(\mathbf{x}_{t=1}^t, \beta_{t=1}^t) \rangle = \mathbf{0},$$
$$\langle \epsilon_2^o \rangle_{t=1} = \langle \mathbf{y}_{(2)t=1} - h_{(2)}(\mathbf{x}_{t=1}^t) \rangle = \mathbf{0}. \quad (52)$$

Then, substituting Equations 50 – 52 into Equations 22 and 23, we have that the expected values of the analysis errors for the state and bias coefficients at the second cycle are

$$\langle \epsilon_x^a \rangle_{t=1} = (\mathbf{I} - \mathbf{K}_{xy_{(1)}} \mathbf{H}_{(1)} - \mathbf{K}_{xy_{(2)}} \mathbf{H}_{(2)}) \langle \epsilon_x^b \rangle_{t=1} - \mathbf{K}_{xy_{(1)}} \mathbf{C}_\beta \langle \epsilon_\beta^a \rangle_{t=0}, \quad (53)$$

$$\langle \epsilon_\beta^a \rangle_{t=1} = -(\mathbf{K}_{\beta y_{(1)}} \mathbf{H}_{(1)} + \mathbf{K}_{\beta y_{(2)}} \mathbf{H}_{(2)}) \langle \epsilon_x^b \rangle_{t=1} + (\mathbf{I} - \mathbf{K}_{\beta y_{(1)}} \mathbf{C}_\beta) \langle \epsilon_\beta^a \rangle_{t=0}. \quad (54)$$

The expected value of the state analysis is now dependent on both $\langle \epsilon_x^b \rangle_{t=1}$ and $\langle \epsilon_\beta^a \rangle_{t=0}$. This shows that, if the bias-coefficient analysis is biased at time n, then this will contaminate the estimate of both state and bias-coefficient analysis at time $n+1$. As $\langle \epsilon_\beta^b \rangle_{t=1} = \langle \epsilon_\beta^a \rangle_{t=0}$, we see that in future cycles we cannot assume $\langle \epsilon_\beta^b \rangle = 0$, as any bias in the bias-coefficient analysis will become the bias in the bias-coefficient background. Therefore, this shows that it is important to reduce the contamination of model bias in

the observation bias coefficient, as $\langle \epsilon_\beta^a \rangle$ feeds into future cycles and contaminates the state analysis.

4 | EXPERIMENTAL DESIGN

In this section, we demonstrate our general theory for a linear system by considering a simplified experiment that uses the nonlinear Lorenz 96 model (Lorenz, 1996), which we will denote as L96, as the basis for our system. L96 has been used as a way to study predictability, particularly in weather and other atmospheric systems, and is often used today as a test problem for numerical weather prediction and in data assimilation (e.g. Ott *et al.*, 2004; Fertig *et al.*, 2007; Brajard *et al.*, 2020). We use L96 as it allows us to choose certain parameters that control how quickly the system displays chaotic behaviour (Lorenz, 1996; Kerin and Engler, 2020). Model bias is added to the equations over more than one cycle by changing the forcing term in the model that propagates the state analysis forward, but not in the model that calculates the true state, which is used to generate the observations.

The L96 model is a system of coupled ordinary differential equations which describes the evolution of a quantity via advection, dissipation, and external forcing. The system contains N variables $x_1, ..., x_N$, and is determined by N equations, where the kth equation is given by

$$\frac{dx_k}{dt} = -x_{k-2}x_{k-1} + x_{k-1}x_{k+1} - x_k + F, \quad (55)$$

where F is independent of k (Lorenz, 1996). This is solved numerically using the fourth-order Runge–Kutta scheme. We assume that we are on a periodic circular domain such that $x_0 = x_N$ and $x_1 = x_{N+1}$. The first two terms of Equation 55 are the advection terms, which conserve the total energy: they simulate flow out of and into the grid point k, respectively. The third term is the internal dissipation, where a fraction of the quantity present is destroyed or dissipated, and the fourth term, F, shows how much of the current quantity x_k is added to each state (Kerin and Engler, 2020).

As in many toy data assimilation experiments utilising the L96 model, we choose the number of spatial variables N to be 40 and the forcing term in the true model, F, to be 8 (e.g. Ott *et al.*, 2004; Fertig *et al.*, 2007; Brajard *et al.*, 2020). We set the distance between grid points $\Delta L_k = 1$ for all k, such that the length of the domain is 40 (equal to the number of spatial variables). We initialise the L96 model from a sine curve and run it for 10^5 time steps of $\Delta t = 0.0125$, in order to allow the system to relax

into its dynamics. The time step $\Delta t = 0.0125$ is equivalent to approximately 45 min in the real atmosphere (Ott *et al.*, 2004; Brajard *et al.*, 2020), based on the assumption that the error-doubling time of the atmosphere is approximately 1 day (Simmons and Hollingsworth, 2002). We then use the final time step as the initial condition for the true trajectory in our numerical experiments.

In our data assimilation setup, we set the true model to be the L96 model with forcing $F = 8$. We define the biased observations to be the truth plus an error consisting of random and systematic error,

$$\mathbf{y}_{(1)} = \mathbf{x}^t + \mathbf{e}_{(1)}^o + \mathbf{b}^o \in \mathbb{R}^{m_{(1)}}, \quad (56)$$

where $\mathbf{x}^t$ is the true state; $\mathbf{e}_{(1)}^o$ is the uncorrelated random error in the biased observations, calculated from a Gaussian distribution with zero mean and error variance $\sigma_{o(1)}^2$; and $\mathbf{b}^o$ is the observation bias. We define the bias to be constant over all observed variables, and set it to 2 ($:= \beta^t$). The unbiased observations are given by the truth plus a random error,

$$\mathbf{y}_{(2)} = \mathbf{x}^t + \mathbf{e}_{(2)}^o \in \mathbb{R}^{m_{(2)}}, \quad (57)$$

where $\mathbf{e}_{(2)}^o$ is the random error in the unbiased observations, calculated from a Gaussian distribution with zero mean and error variance $\sigma_{o(2)}^2$. These will be our anchor observations. The data assimilation method used for our experiments with the L96 model is 1DVar, which is analogous to the theory presented on 3DVar in Section 3, as we do not use a time dimension. The number of observations in space varies depending on the experiment. For each experiment there is at least one observation (anchor and/or bias-corrected observations) observing every state variable. Although such a high frequency of observations is unrealistic operationally, this is necessary when using the L96 model, as variables that are spatially close in L96 are not highly correlated (Lorenz, 2005), which implies that their errors are not highly correlated. We attempted similar experiments with sparser observations, but it produced background-error covariance matrices with unstable values.

We simulate direct observations, hence $\mathbf{H}_{(1)} \in \mathbb{R}^{m_1 \times 40}$ and $\mathbf{H}_{(2)} \in \mathbb{R}^{m_2 \times 40}$ have values of one at the locations of observations and zero at all other locations. As the biases added to the observations are constant, we set $c(\mathbf{x}, \beta) = \beta$ such that $\mathbf{C}_\mathbf{x} = \mathbf{0}$ and $\mathbf{C}_\beta = 1$. This means that we only use one predictor, which is equal to 1, that is, $p_{k,0} = 1$ in Equation 1. $\mathbf{R}_{(1)}$ and $\mathbf{R}_{(2)}$ are set to be the identity, unless stated otherwise as in experiment 1 (Section 5.1.1). The matrix $\mathbf{B}_v$ (the matrix that combines $\mathbf{B}_\mathbf{x}$ and $\mathbf{B}_\beta$: see Equation 10) is given by two different matrices in the

following experiments. In experiments 1 and 3, $\mathbf{B}_v$ is calculated from an initial ensemble run, so that it is the optimal background-error covariance matrix for the system. The calculation of $\mathbf{B}_v$ for these experiments is explained in more detail in Appendix E. In experiment 2 (Section 5.1.2), in which we do not cycle the system, the correlations in $\mathbf{B}_x$ are given by the second-order autoregressive (SOAR) correlation function, which has previously been used to model the background-error correlations of the atmosphere (Ingleby, 2001; Simonin *et al.*, 2014), allowing us to control the length-scales of the error correlations. The SOAR correlation function is given by

$$s_k = \left(1 + \frac{k}{L_b}\right) e^{-k/L_b}, \tag{58}$$

where k is the index from 1 to $N/2$; L_b is the length-scale; and e is the exponential function. For values greater than $N/2$, s_k is repeated in opposite order such that the array $\mathbf{s}$ is palindromic. To calculate $\mathbf{B}_x$, $\mathbf{s}$ is multiplied by σ_{bx}^2 to create a circulant matrix. The background-error covariances are varied by varying the length-scale of the SOAR function. $\mathbf{B}_\beta$ is given by the scalar $\sigma_{b\beta}^2$.

5 | NUMERICAL RESULTS

In this section, we demonstrate the theoretical results from Section 3 using the L96 model and the metrics we have introduced in Section 4.

We want to demonstrate the ability of anchor observations to reduce the contamination of model bias acting on the analysed observation bias coefficient. In each experiment we compute the bias-coefficient analysis over a given number of realisations, with random error in the observations and initial background values. The number of realisations varies in each experiment, due to the computational cost and the error variances chosen. From these realisations we obtain the mean and standard deviation of the bias-coefficient analysis. We can illustrate the bias in the analysed bias coefficient with the ratio

$$\frac{|\overline{\beta^a} - \beta^t|}{\sigma_{a\beta}}, \tag{59}$$

where $|\cdot|$ is the absolute value; $\overline{\beta^a}$ is the mean β^a over all realisations; β^t is the true bias; and $\sigma_{a\beta}$ is the standard deviation of β^a from all realisations. This shows the bias associated with the mean value of β^a in relation to the error variance. We will refer to this ratio as the bias ratio. If the bias ratio is large, then the bias can be considered significant in comparison with the random noise, but if the bias ratio is small, then the bias in the bias-coefficient analysis
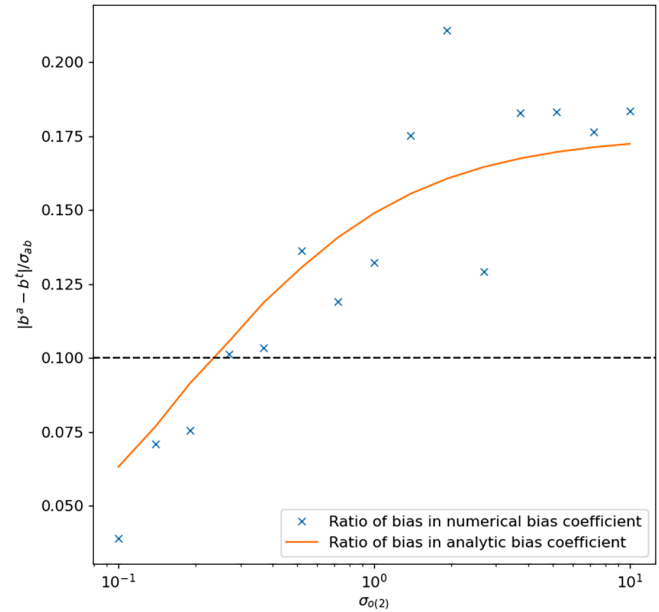

FIGURE 2 The ratio (Equation 59) from running the system with different values of anchor observation-error standard deviation ($\sigma_{o(2)}$). Both biased and anchor observations are at every location in the domain. The bias-corrected observation-error standard deviations are set to 1. The background-error covariance matrix is calculated from an ensemble, see Appendix E. A bias of 0.15 was added to all background state variables. The solid line is the analytic ratio, calculated from the linearised observation operators and error covariance matrices in the system, and the crosses are the numerical ratios, calculated by averaging over 3000 realisations. [Colour figure can be viewed at wileyonlinelibrary.com]

will be lost within the random error and so will be insignificant. The ratio at 0.1 is plotted for reference as a dotted line in Figures 2, 3, and 4 and is referred to as the reference ratio.

5.1 | One-cycle experiments (no model evolution)

In the first two experiments (Sections 5.1.1 and 5.1.2) we use 1DVar to calculate the analysis of the state and bias coefficients at the initial time step. To simulate a model bias, we set the state background to have a bias, to represent any bias that has accumulated from running a previous cycle forward using the model. The background state is given by the truth plus an error,

$$\mathbf{x}^b = \mathbf{x}^t + \mathbf{e}_x^b + \mathbf{b}^b \in \mathbb{R}^{40}, \tag{60}$$

$$\beta^b = \beta^t + e_\beta^b \in \mathbb{R}, \tag{61}$$

FIGURE 3 The ratio in Equation 59 at one cycle when the length-scale of $\mathbf{B}_x$ has been varied. Biased observations observe all even state variables (x_0, x_2 etc), anchor observations observe all odd state variables. The dotted line is the reference ratio. A bias of 0.3 was added to the background: in state variables only observed by biased observations (solid line); in state variables only observed by anchor observations (dotted line); and in state variables observed by both types of observations (dash-dotted line). [Colour figure can be viewed at wileyonlinelibrary.com]
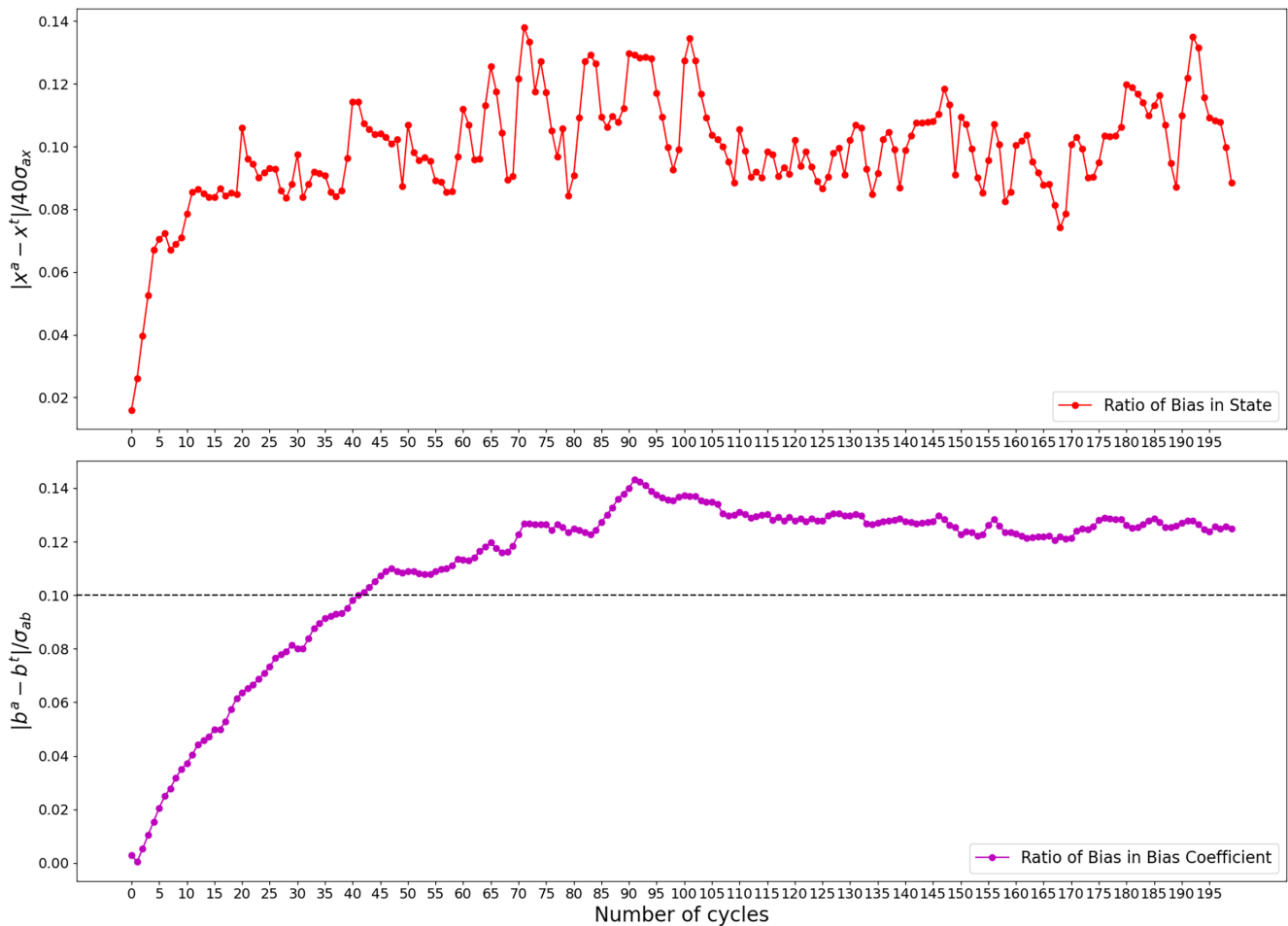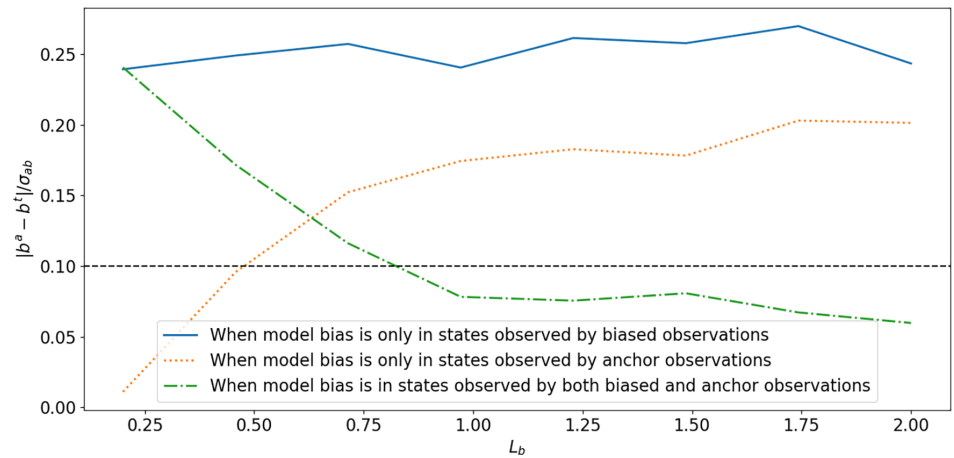

FIGURE 4 The ratios in Equations 59 and 62 over 200 cycles for 100 realisations. $F_{\text{biased}} = 8.8$, error variances are 1, $\mathbf{B}_v$ is calculated by an ensemble (Appendix E). Anchor and biased observations observe every state variable. [Colour figure can be viewed at wileyonlinelibrary.com]

where $\mathbf{e}_x^b$ is the random error in the state background calculated from a Gaussian distribution with zero mean and from the error covariance matrix $\mathbf{B}_x$; $\mathbf{b}^b$ is the bias

in the initial state background; and e_β^b is the random error in the bias-coefficient background calculated from a Gaussian distribution with zero mean and error variance

$\sigma_{\mathrm{b}\beta}^2$. In experiment 1 this is set from the climatological $\mathbf{B}_\mathrm{v}$ (as described in Appendix E) and in experiment 2 this is set to be 1, so that it is equal to the other error variances.

5.1.1 | The effect of varying the anchor observation-error variance

First we present an experiment to illustrate the results found in Section 3.2.1. The theoretical results showed that, in a system observed fully by anchor observations, small anchor observation-error variances reduce the contamination of model bias in β^a. Therefore, in this experiment, we have observed all spatial variables with both bias-corrected and anchor observations, such that every spatial variable is observed twice, once by the bias-corrected observations and once by the anchor observations.

In Figure 2, we plot the bias ratio as in Equation 59 from 3000 realisations after one cycle and vary the anchor observation-error standard deviation, to test the importance of the weighting given to anchor observations in reducing the effect of model bias in the analysed bias coefficient. The bias-corrected observation-error standard deviations ($\sigma_{\mathrm{o}(1)}$) are set to 1, so that they are approximately 10% of the variability of the state. $\mathbf{B}_\mathrm{v}$ is calculated via an ensemble, as described in Appendix E, and is kept constant for the different values of $\sigma_{\mathrm{o}(2)}$. Although this gives a suboptimal $\mathbf{B}_\mathrm{v}$ for the system, it isolates the effect of varying the anchor observation-error variance. A bias of 0.15 has been added to the initial conditions, as in Equation 60, which represents our model bias. This is approximately 1.5% of the variability of the state and was chosen so that the system could still control the model bias. $\sigma_{\mathrm{o}(2)}$ has been varied from one-tenth of $\sigma_{\mathrm{o}(1)}$ to ten times $\sigma_{\mathrm{o}(1)}$ (i.e., between 0.1 and 10). The crosses are the results from the 3000 realisations and the solid line is the analytic result given by Equation 24, to give the "true" bias ratio for the system.

In Figure 2, the bias in the observation bias coefficient becomes more significant as the anchor observation-error standard deviation is increased. There is some variation in the bias ratio when calculated from the realisations, which is due to the random error in the observations and background, but they follow the shape of the analytic solution. The bias ratio increasing with larger anchor observation-error variance is in line with the results found from Equation 27, as we showed that if the anchor observation-error variance is small, then the effect of the model bias on β^a would be small. This is because the ratio $\mathbf{D}$ in Equation 18 will tend to the identity as $\sigma_{\mathrm{o}(2)}$ reduces. The ratio increases as $\sigma_{\mathrm{o}(2)}$ increases, which shows that, as less trust is put into the anchor observations, the bias in the analysed bias coefficient increases. This occurs as model bias dilutes the analysis of the bias coefficient and pulls the VarBC system away from the truth. In the limit where the anchor observation-error variance is large, the anchor observations receive insignificant weight in the analysis and so the biased observations are bias-corrected towards the model bias, instead of to the truth.

5.1.2 | The location of the anchor observations relative to the model bias

Next we present experiments to demonstrate the results found in Sections 3.2.3 and 3.3, such that we have anchor and biased observations observing different state variables that do and do not have model biases. In Section 3.3, we showed how the role of the anchor observations in reducing the effect of model bias changes depending on whether or not anchor observations observe state variables that have model bias. When anchor and biased observations observe different parts of the state, as in Section 3.2.3, then the influence of the anchor observations is dependent on the background-error covariances between state variables.

In Figure 3, we have plotted the ratio in Equation 59 from 1000 realisations after one cycle. The observations are spaced evenly at every other spatial variable, such that the biased observations observe the even state variables (x_0, x_2, x_4, etc.) and anchor observations observe all odd state variables (x_1, x_3, x_5, etc.). The observation-error standard deviations are equal to 1. $\mathbf{B}_\mathrm{x}$ is given by the SOAR correlation function, with $\sigma_{\mathrm{bx}}^2 = 1$. $\mathbf{B}_\beta$ is scalar, also equal to 1. To see the effect of different background-error covariances, we have varied $\mathbf{B}_\mathrm{x}$ by varying the length-scale, L_b, which is varied in Figure 3 along the x-axis. When L_b is small, the background-error covariances between state variables are small, and when L_b is increased, so are the background-error covariances between state variables. Larger length-scales will mean more background information is shared between parts of the state that are spatially further away from each other. The dotted line is 0.1, below which any biases are considered to be insignificant. We have considered three cases where a bias of 0.3 has been added to the background state in three different locations: in state variables that are observed by biased observations; in state variables that are observed by anchor observations; and in state variables that are observed by both types of observations.

Model bias only in state variables observed by anchor observations (dotted line): In Figure 3, when the model bias is only in state variables observed by anchor observations (dotted line), the bias in the observation bias coefficient is insignificant when the error covariances of $\mathbf{B}_\mathrm{x}$ are small. As the length-scale of $\mathbf{B}_\mathrm{x}$ is increased, the ratio of the bias also increases, which means β^a has a more

significant bias in comparison with the random error. This reflects the results found from Equation 44, as we found that, in order to reduce the contamination of model bias in state variables observed by the anchor observations in $\langle \epsilon_\beta^a \rangle_{t=0}$, less information had to be passed between the state variables observed by the anchor and bias-corrected observations: that is, smaller background-error correlations. However, in Section 3.2.3, we showed that, if the anchor and bias-corrected observations observe different parts of the state and $\mathbf{B}_{x_{\phi\psi}}$ tends to zero, then the anchor observations would not be used in the VarBC system. Hence this scenario would not be ideal, as it would mean that either β^a is affected by model bias or the anchor observations are not used within the VarBC system.

Model bias only in state variables not observed by anchor observations (solid line): In Figure 3, when model bias is only in state variables that are observed by the biased observations (solid line), the bias in $\langle \epsilon_\beta^a \rangle_{t=0}$ remains significant regardless of the magnitude of the covariances in $\mathbf{B}_x$. This reflects the theoretical results, as Equation 45 is independent of $\mathbf{D}$, which means that the anchor observations cannot reduce the effect of model bias in state variables directly if they do not observe the parts of the state that have model bias.

Model bias in state variables observed by both anchor and bias-corrected observations (dash-dotted line): In Figure 3, if model bias is in both state variables observed by biased observations and state variables observed by anchor observations (dash-dotted line), then the bias in $\langle \epsilon_\beta^a \rangle_{t=0}$ is initially significant compared with the random error when the length-scales of $\mathbf{B}_x$ are small and becomes less significant as more information between background state variables is shared. As was shown in Equation 38, Figure 3 shows that varying $\mathbf{B}_{x_{\phi\psi}}$ has a large effect in reducing the effect of the model bias from parts of the state that both are and are not observed by the anchor observations. If model bias is in the system, then having strong error correlations between bias-corrected and anchor observations that observe state variables with similar model biases is the best possibility, as the anchor observations have the biggest effect in reducing the contamination of model bias acting on the estimate of the observation bias coefficient, whilst still being used within VarBC.

5.2 | Cycled experiment

In the third experiment (Section 5.2.1) the system is run over 50 cycles to demonstrate how the bias in the bias coefficient will accumulate and be passed into the state analysis. On the first cycle, the background state and bias coefficient are set up as in Equations 60 and 61, but with zero added background bias, so that $\mathbf{b}^b = 0$. This means

the first few cycles will act as a spin-up period to allow the background bias to settle into an equilibrium. In future cycles, the background values at cycle T are given by the analysis values at cycle $T-1$ after they are evolved forwards in time. The bias-coefficient background is evolved by the identity, that is, the previous bias-coefficient analysis is taken to be the bias-coefficient background. The state background is evolved forward via the model from Equation 55, but replacing F with $F_{\text{biased}} = 8.8$ to add a bias to the model by changing the forcing term.

We set the assimilation window length to be 10 time steps (which represents approximately 7.5 hr in the real atmosphere, as discussed at the beginning of Section 4) to allow the background errors to grow sufficiently within each window.

5.2.1 | How bias in the state analysis and bias-coefficient analysis accumulates when the system is cycled

Finally, we have an experiment to demonstrate the results from Section 3.4, which show the effect of having a bias in β^a on the state analysis in the next cycle. These show how the contamination of the model bias in β^a leads to a biased β^b, which in turn contributes to the bias in the state analysis.

We run an experiment that has 100 realisations over 200 cycles. In Figure 4, we plot the ratio from Equation 59 for the bias-coefficient analysis (bottom panel), but have included a similar ratio for the norm of the state analysis bias, given by

$$\frac{\sqrt{\sum_{i=1}^{40}(\overline{x}_i^a - x_i^t)^2}}{40\,\overline{\sigma}_{ax}}, \tag{62}$$

where $\overline{x}_i^a$ is the average state analysis across all realisations for the spatial variable x_i; x_i^t is the true value of the spatial variable x_i; 40 is the number of state variables; and $\overline{\sigma}_{ax}$ is the mean value of the standard deviations of the state analysis from 100 realisations, where the mean is calculated over all state variables. We have both anchor and biased observations at every location; the observation-error standard deviations are 1; and $\mathbf{B}_v$ is calculated via an ensemble, see Appendix E. We have added a model bias by multiplying the forcing term in the model used to evolve the analysis by 1.1, such that F_{biased} is 8.8, but have not added a model bias via the initial conditions, as explained in Section 5.2. Multiplying the forcing term by 1.1 only changes the state values by 0.01 over one cycle, allowing the model bias still to be constrained. We plot the ratios over 200 cycles to show the evolution of the bias in the state and the bias coefficient.

In Figure [4], the ratios for the state and the bias coefficient start at approximately zero and then the state ratio increases towards 0.1 and the bias coefficient ratio increases towards 0.12. When the ratios are above 0.1, we consider the biases to be significant in comparison with the random error. Note that the magnitude of the ratios would be larger if a larger model bias were chosen, fewer state variables were observed, or less trust were put into the observations, which we do not show here. Therefore the 0.1 line has been included in the bias-coefficient bias ratio to show general trends, not the exact value of parameters that causes the bias to be significant. The bias in the first cycle is zero for both the state and the bias coefficient, because no model bias has been added to the initial conditions. There is only a model bias from the second cycle, as model bias is added via the forcing term F_{biased}, which evolves the state analysis forward, so that the state analysis of the first cycle becomes the state background of the second cycle and so on. We would expect the bias in both the state and the bias coefficient to grow with each cycle, as we showed in Equations [53] and [54]; the biases in $\mathbf{x}^{\text{a}}$ and β^{a} are the accumulation of the previous state and bias coefficient biases. In Figure [4], the error in the state appears to reach saturation after approximately 20 cycles and the error in the bias coefficient reaches saturation after approximately 90 cycles. As the model bias, η_i, is constant in time, the state and bias coefficient analyses reach an equilibrium between the background bias and the truth from the anchor observations. We would expect an equilibrium to exist, as VarBC relies on the background and the unbiased observations as sources of the truth, so if they are different then the analysis will be pulled between the two until it reaches an equilibrium. However, it is not straightforward to evaluate analytically what the equilibrium should be.

6 | CONCLUSIONS AND DISCUSSION

In order to study how model bias can contaminate observation-bias correction, we have looked at the role of unbiased (anchor) observations in variational bias correction (VarBC). The conclusions are based on general optimal estimation theory and we have demonstrated the results with idealised experiments.

In this study, we have focused on the importance of the anchor observations in reducing the contamination of model bias in bias-coefficient analysis, as opposed to reducing the effect on the state analysis. This is because if the bias-coefficient analysis has systematic error, then observations will be wrongly bias-corrected and the systematic error will filter into the state analysis, as was shown when the system was cycled in Section [3.4] and Figure [4].

In a theoretical world where we could have full coverage of anchor observations, we showed in Equation [27] that the model bias can still contaminate the bias-coefficient analysis. The only way that model bias could not affect the bias-coefficient analysis at all would be to have zero random error in the anchor observations. However, as this is only possible theoretically, we can only look at reducing the contamination of model bias in the bias-coefficient analysis, not removing it completely. We showed in Equation [26] and Figure [2] that the effect of model bias is reduced when the anchor observation-error variance is smaller than the state backgrounds that the anchor observations observe. Operationally, the anchor observation-error variance needs to reflect the uncertainty associated with the observations. However, it is possible to change the weighting given to the anchor observations by using more anchor observations within the system, if they are available. Therefore, although we have shown that more precise anchor observations will reduce the contamination of model bias in the observation-bias correction, this result also extends to the importance of a higher spatial frequency of anchor observations, to increase the weighting given to the anchor observations within the system and thus have the biggest impact in reducing the contamination of model bias.

In Equations [36], [44], and [45] and Figure [3], we showed that the contamination of model bias in the bias-coefficient analysis can only be reduced by the anchor observations if the anchor observations also observe state variables that have similar model bias. Anchor observations have typically been from radiosondes, which mostly have coverage over land, but with the increased use of radio occultation (RO) the spatial and temporal distribution of anchor observations in the upper troposphere and stratosphere has increased. Radiosondes and RO give a good coverage of anchor observations for temperature in the troposphere and stratosphere, but variables such as humidity and wind speed are less well observed by unbiased observations. This means that, although the spatial coverage of anchor observations is increasing, there are still significant parts of the model domain where model bias could still contaminate the bias-coefficient analysis.

We showed in Equation [44] and Figure [3] that, if model bias only exists in state variables observed by the anchor observations, such that bias-corrected observations do not observe state variables with model bias, then the anchor observations will pass the model bias into the bias-coefficient analysis if there are strong error correlations between the background state variables. This could occur, for example, in the lower troposphere, as radiosonde measurements are mostly taken over land, with little coverage over sea. If a temperature bias exists over the land but not the sea and there are strong horizontal error

correlations between the land and the sea, then the model bias could be passed into the bias coefficients via radiosondes.

In Equation 45 and Figure 3, we showed that, if there is no model bias in the state variables that are observed by the anchor observations but there is model bias in state variables observed by the bias-corrected observations, then the anchor observations cannot explicitly reduce the effect of model bias in other state variables. Therefore, areas that are most at risk of contamination of model bias within the observation-bias correction will be locations with fewer unbiased observations, such as humidity in the upper troposphere, which is only sparsely observed by radiosondes but is known to have model biases.

We showed in Equation 36 and at the end of Section 3.3 that, if both biased and anchor observations observe state variables that have model bias, then anchor observations can reduce the contamination of model bias in the bias-coefficient analysis. We saw in Figure 3 that, if the anchor and biased observations observe different parts of the state that have similar model bias characteristics, the background-error covariances become more important: larger background-error correlations will transfer more information about the model bias between state variables and so will reduce the contamination of the model bias in the bias-coefficient analysis. In an operational system, the length-scales of $\mathbf{B}_x$ will be locationally dependent: for example, the length-scales of $\mathbf{B}_x$ will be larger at higher altitudes (Ingleby, 2001). There will also be an element of flow dependence, such that the background-error covariance matrix will deform with the flow (Bannister, 2008). This article shows that anchor observations that observe different variables/locations from the bias-corrected observations, when both variables/locations have similar model biases, will have the largest impact in reducing the contamination of model bias when they are in systems with larger background-error covariances, such as in the upper atmosphere. When similar model biases are present in systems where background-error covariances are small between variables observed by bias-corrected and anchor observations, such as between locations of observations across a front, then anchor observations will only have a small impact in reducing the contamination of model bias in the observation-bias correction.

This article has aimed to derive new insight into the role of anchor observations for mitigating the impact of model bias in VarBC. Our theoretical findings have been tested in a toy system. The next steps for this work should be to extend the theory to 4D variational data assimilation and to look at how the assumptions made hold in an operational system: for example, how the bias predictors used operationally allow the state and bias coefficient domains to be separated more clearly.

ORCID
Devon J. Francis ⓘ https://orcid.org/0000-0001-5351-5065
Alison M. Fowler ⓘ https://orcid.org/0000-0003-3650-3948
Amos S. Lawless ⓘ https://orcid.org/0000-0002-3016-6568
John Eyre ⓘ https://orcid.org/0000-0002-9877-392X
Stefano Migliorini ⓘ https://orcid.org/0000-0002-9489-9867

REFERENCES
Auligné, T., McNally, A.P. and Dee, D.P. (2007) Adaptive bias correction for satellite data in a numerical weather prediction system. *Quarterly Journal of the Royal Meteorological Society: A Journal of the Atmospheric Sciences, Applied Meteorology and Physical Oceanography*, 133, 631–642.

Bannister, R.N. (2008) 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, 1951–1970.

Brajard, J., Carrassi, A., Bocquet, M. and Bertino, L. (2020) Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations: A case study with the Lorenz 96 model. *Journal of Computational Science*, 44, 101171.

Cameron, J. and Bell, W. (2018) The testing and implementation of variational bias correction (VarBC) in the Met Office global NWP system.

Candy, B., Cotton, J. and Eyre, J.R. (2021) *Recent results of observation data denial experiments. Tech. rep.* Exeter UK: Met Office.

Cucurull, L., Anthes, R. and Tsao, L.-L. (2014) Radio occultation observations as anchor observations in numerical weather prediction models and associated reduction of bias corrections in microwave and infrared satellite observations. *Journal of Atmospheric and Oceanic Technology*, 31, 20–32.

Dee, D.P. (2004) Variational bias correction of radiance data in the ECMWF system. Proceedings of the ECMWF workshop on assimilation of high spectral resolution sounders in NWP, Reading, UK, vol. 28, 97–112.

Dee, D.P. and Uppala, S. (2009) Variational bias correction of satellite radiance data in the ERA-Interim reanalysis. *Quarterly Journal of the Royal Meteorological Society: A Journal of the Atmospheric*

Sciences, Applied Meteorology and Physical Oceanography, 135, 1830–1841.

Di Tomaso, E. and Bormann, N. (2011) *Assimilation of ATOVS radiances at ECMWF: First year EUMETSAT fellowship report. Tech. rep*. Reading UK: European Centre for Medium-Range Weather Forecasts.

English, S., McNally, T., Bormann, N., Salonen, K., Matricardi, M., Moranyi, A., Rennie, M., Janisková, M., Di Michele, S. and Geer, A. (2013) *Impact of satellite data. Tech. rep*. Reading UK: European Centre for Medium-Range Weather Forecasts.

Eyre, J.R. (2016) Observation bias correction schemes in data assimilation systems: A theoretical study of some of their properties. *Quarterly Journal of the Royal Meteorological Society*, 142, 2284–2291.

Eyre, J.R., Bell, W., Cotton, J., English, S.J., Forsythe, M., Healy, S.B. and Pavelin, E.G. (2022) Assimilation of satellite data in numerical weather prediction. Part II: Recent years. *Quarterly Journal of the Royal Meteorological Society*, 148, 521–556.

Eyre, J.R., English, S.J. and Forsythe, M. (2020) Assimilation of satellite data in numerical weather prediction. Part I: The early years. *Quarterly Journal of the Royal Meteorological Society*, 146, 49–68.

Fertig, E.J., Harlim, J. and Hunt, B.R. (2007) A comparative study of 4D-VAR and a 4D ensemble Kalman filter: Perfect model simulations with Lorenz-96. *Tellus A: Dynamic Meteorology and Oceanography*, 59, 96–100.

Han, W. and Bormann, N. (2016) *Constrained adaptive bias correction for satellite radiance assimilation in the ECMWF 4D-Var system. Tech. rep*. Reading, UK: European Centre for Medium Range Weather Forecasts.

Han, W. and McNally, A.P. (2010) The 4D-Var assimilation of ozone-sensitive infrared radiances measured by IASI. *Quarterly Journal of the Royal Meteorological Society*, 136, 2025–2037.

Healy, S.B. and Thépaut, J.-N. (2006) Assimilation experiments with CHAMP GPS radio occultation measurements. *Quarterly Journal of the Royal Meteorological Society*, 132, 605–623.

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, 209–231.

Kerin, J. and Engler, H. (2020) On the Lorenz'96 model and some generalizations. arXiv preprint arXiv:2005.07767.

Laloyaux, P., Bonavita, M., Chrust, M. and Gürol, S. (2020a) Exploring the potential and limitations of weak-constraint 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, 146, 4067–4082.

Laloyaux, P., Bonavita, M., Dahoui, M., Farnan, J., Healy, S., Hólm, E. and Lang, S.T.K. (2020b) Towards an unbiased stratospheric analysis. *Quarterly Journal of the Royal Meteorological Society*, 146, 2392–2409.

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

Lorenz, E.N. (2005) Designing chaotic models. *Journal of the Atmospheric Sciences*, 62, 1574–1587.

Lu, T.-T. and Shiou, S.-H. (2002) Inverses of 2× 2 block matrices. *Computers & Mathematics with Applications*, 43, 119–129.

Lu, T.-T. and Shiou, S.-H. (2005) Designing chaotic models. *Journal of the Atmospheric Sciences*, 62, 1574–1587.

Ott, E., Hunt, B.R., Szunyogh, I., Zimin, A.V., Kostelich, E.J., Corazza, M., Kalnay, E., Patil, D.J. and Yorke, J.A. (2004) A local ensemble Kalman filter for atmospheric data assimilation. *Tellus A: Dynamic Meteorology and Oceanography*, 56, 415–428.

Saunders, R. (2021) The use of satellite data in numerical weather prediction. *Weather*, 76, 95–97.

Simmons, A.J. and Hollingsworth, A. (2002) Some aspects of the improvement in skill of numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 128, 647–677.

Simonin, D., Ballard, S.P. 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, 2298–2314.

Sun, B., Reale, A., Schroeder, S., Seidel, D.J. and Ballish, B. (2013) Toward improved corrections for radiation-induced biases in radiosonde temperature observations. *Journal of Geophysical Research: Atmospheres*, 118, 4231–4243.

Talagrand, O. (1998) A posteriori evaluation and verification of analysis and assimilation algorithms. Proc. of ECMWF Workshop on Diagnosis of Data Assimilation System, 17–28.

APPENDIX A. CALCULATION OF KV FOR TWO OBSERVATION TYPES

If we have two types of observation: $\mathbf{y}_{(1)}$ which are bias-corrected and $\mathbf{y}_{(2)}$ which are unbiased, then the linearised observation operator is given by

$$\mathbf{H}_v(\mathbf{v}) = \begin{pmatrix} \mathbf{H}_{(1)} + \mathbf{C}_x & \mathbf{C}_\beta \\ \mathbf{H}_{(2)} & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{(m_1+m_2)\times(n+r)}, \quad (A1)$$

and the observation-error variance for both types of observations is given by

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_{(1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{(2)} \end{pmatrix} \in \mathbb{R}^{(m_1+m_2)\times(m_1+m_2)}. \quad (A2)$$

Hence $\mathbf{K}_v$ becomes

$$\mathbf{K}_v = \mathbf{B}_v \mathbf{H}_v^T (\mathbf{H}_v \mathbf{B}_v \mathbf{H}_v^T + \mathbf{R})^{-1}, \quad (A3)$$

$$= \begin{pmatrix} \mathbf{B}_x(\mathbf{H}_{(1)} + \mathbf{C}_x)^T & \mathbf{B}_x\mathbf{H}_{(2)}^T \\ \mathbf{B}_\beta \mathbf{C}_\beta^T & \mathbf{0} \end{pmatrix}$$

$$\times \begin{pmatrix} (\mathbf{H}_{(1)} + \mathbf{C}_x)\mathbf{B}_x(\mathbf{H}_{(1)} + \mathbf{C}_x)^T & (\mathbf{H}_{(1)} + \mathbf{C}_x)\mathbf{B}_x\mathbf{H}_{(2)}^T \\ + \mathbf{C}_\beta \mathbf{B}_\beta \mathbf{C}_\beta^T + \mathbf{R}_{(1)} & \\ \mathbf{H}_{(2)}\mathbf{B}_x(\mathbf{H}_{(1)} + \mathbf{C}_x)^T & \mathbf{H}_{(2)}\mathbf{B}_x\mathbf{H}_{(2)}^T + \mathbf{R}_{(2)} \end{pmatrix}^{-1}.$$

$$(A4)$$

As $(\mathbf{H}_v \mathbf{B}_v \mathbf{H}_v^T + \mathbf{R})$ is symmetric, from equation 4.2 in Lu and Shiou (2002) we can calculate its inverse by writing

$$(\mathbf{H}_v \mathbf{B}_v \mathbf{H}_v^T + \mathbf{R})^{-1}$$

$$= \begin{pmatrix} \mathbf{W} & \mathbf{X} \\ \mathbf{X}^T & \mathbf{z} \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} (\mathbf{W} - \mathbf{X}\mathbf{Z}^{-1}\mathbf{X}^T)^{-1} & -(\mathbf{W} - \mathbf{X}\mathbf{Z}^{-1}\mathbf{X}^T)^{-1}\mathbf{X}\mathbf{Z}^{-1} \\ -\mathbf{Z}^{-1}\mathbf{X}^T(\mathbf{W} - \mathbf{X}\mathbf{Z}^{-1}\mathbf{X}^T)^{-1} & \mathbf{Z}^{-1} + \mathbf{Z}^{-1}\mathbf{X}^T(\mathbf{W} - \mathbf{X}\mathbf{Z}^{-1}\mathbf{X}^T)^{-1}\mathbf{X}\mathbf{Z}^{-1} \end{pmatrix},$$

where $\qquad$ (A5)

$$\mathbf{W} = (\mathbf{H}_{(1)} + \mathbf{C}_x)\mathbf{B}_x(\mathbf{H}_{(1)} + \mathbf{C}_x)^T + \mathbf{C}_\beta \mathbf{B}_\beta \mathbf{C}_\beta^T + \mathbf{R}_{(1)}, \quad (A6)$$

$$\mathbf{X} = (\mathbf{H}_{(1)} + \mathbf{C}_x)\mathbf{B}_x\mathbf{H}_{(2)}^T, \qquad (A7)$$

$$\mathbf{Z} = \mathbf{H}_{(2)}\mathbf{B}_x\mathbf{H}_{(2)}^T + \mathbf{R}_{(2)}. \qquad (A8)$$

Hence, combining Equations A4 and A5, the expressions for $\mathbf{K}_{xy_{(1)}}$, $\mathbf{K}_{xy_{(2)}}$, $\mathbf{K}_{\beta y_{(1)}}$, and $\mathbf{K}_{\beta y_{(2)}}$ are given by

$$\mathbf{K}_{xy_{(1)}} = \mathbf{B}_x(\mathbf{H}_{(1)} + \mathbf{C}_x)^T(\mathbf{W} - \mathbf{X}\mathbf{Z}^{-1}\mathbf{X}^T)^{-1} \\ - \mathbf{B}_x\mathbf{H}_{(2)}^T\mathbf{Z}^{-1}\mathbf{X}^T(\mathbf{W} - \mathbf{X}\mathbf{Z}^{-1}\mathbf{X}^T)^{-1}, \qquad (A9)$$

$$\mathbf{K}_{xy_{(2)}} = -\mathbf{B}_x(\mathbf{H}_{(1)} + \mathbf{C}_x)^T(\mathbf{W} - \mathbf{X}\mathbf{Z}^{-1}\mathbf{X}^T)^{-1}\mathbf{X}\mathbf{Z}^{-1} \\ + \mathbf{B}_x\mathbf{H}_{(2)}^T(\mathbf{Z}^{-1} + \mathbf{Z}^{-1}\mathbf{X}^T(\mathbf{W} - \mathbf{X}\mathbf{Z}^{-1}\mathbf{X}^T)^{-1}\mathbf{X}\mathbf{Z}^{-1}), \qquad (A10)$$

$$\mathbf{K}_{\beta y_{(1)}} = \mathbf{B}_\beta \mathbf{C}_\beta^T(\mathbf{W} - \mathbf{X}\mathbf{Z}^{-1}\mathbf{X}^T)^{-1}, \qquad (A11)$$

$$\mathbf{K}_{\beta y_{(2)}} = -\mathbf{B}_\beta \mathbf{C}_\beta^T(\mathbf{W} - \mathbf{X}\mathbf{Z}^{-1}\mathbf{X}^T)^{-1}\mathbf{X}\mathbf{Z}^{-1}. \qquad (A12)$$

Therefore, if we set $\mathbf{C}_x$ to be zero, then we can rewrite $\mathbf{K}_{\beta y_{(2)}}$ in terms of $\mathbf{K}_{\beta y_{(1)}}$:

$$\mathbf{K}_{\beta y_{(2)}} = -\mathbf{K}_{\beta y_{(1)}}\mathbf{X}\mathbf{Z}^{-1} \\ = -\mathbf{K}_{\beta y_{(1)}}\mathbf{H}_{(1)}\mathbf{B}_x\mathbf{H}_{(2)}^T(\mathbf{H}_{(2)}\mathbf{B}_x\mathbf{H}_{(2)}^T + \mathbf{R}_{(2)})^{-1}. \quad (A13)$$

APPENDIX B. BIAS-COEFFICIENT ANALYSIS ERROR

To calculate the error in the bias-coefficient analysis, we subtract the true bias from both sides of the bias-coefficient analysis Equation 6 and add and subtract $h_{v_{(1)}}(\mathbf{v}^t)$ and $h_{v_{(2)}}(\mathbf{v}^t)$ in the respective innovation vectors, to find

$$\boldsymbol{\beta}^a - \boldsymbol{\beta}^t \\ = \boldsymbol{\beta}^b - \boldsymbol{\beta}^t + \mathbf{K}_{\beta y_{(1)}}(\mathbf{y}_{(1)} - h_{v_{(1)}}(\mathbf{v}^t) + h_{v_{(1)}}(\mathbf{v}^t) - h_{v_{(1)}}(\mathbf{v}^b)) \\ + \mathbf{K}_{\beta y_{(2)}}(\mathbf{y}_{(2)} - h_{v_{(2)}}(\mathbf{v}^t) + h_{v_{(2)}}(\mathbf{v}^t) - h_{v_{(2)}}(\mathbf{v}^b)). \qquad (B1)$$

Using Equations 19, 20, and 21, we can simplify this using the error equations, plus rewriting $\mathbf{v}^t$ as $\mathbf{v}^b - \epsilon_v^b$ and

expanding h_v around $\mathbf{v}^b$; we find that the bias-coefficient analysis error is given by

$$\epsilon_\beta^a = \epsilon_\beta^b + \mathbf{K}_{\beta y_{(1)}}(\epsilon_1^o - \mathbf{H}_{v_{(1)}}\epsilon_v^b) + \mathbf{K}_{\beta y_{(2)}}(\epsilon_2^o - \mathbf{H}_{v_{(2)}}\epsilon_v^b), \quad (B2)$$

where $\mathbf{H}_{v_{(1)}} = (\mathbf{H}_{(1)}, \mathbf{C}_\beta)$ and $\mathbf{H}_{v_{(2)}} = (\mathbf{H}_{(2)}, \mathbf{0})$ (see Equation 9). To find the error in the state, we would go through the same procedure, but subtracting the true state from both sides of the state analysis Equation 5 initially, which gives

$$\epsilon_x^a = \epsilon_x^b + \mathbf{K}_{xy_{(1)}}(\epsilon_1^o - \mathbf{H}_{v_{(1)}}\epsilon_v^b) + \mathbf{K}_{xy_{(2)}}(\epsilon_2^o - \mathbf{H}_{v_{(2)}}\epsilon_v^b) \quad (B3)$$

(full derivation not shown).

APPENDIX C. SIMPLIFYING THE EXPECTED VALUE OF βA AFTER THE ASSUMPTIONS THAT THERE IS ONLY BIAS IN THE MODEL

If we assume that $\langle \epsilon_\beta^b \rangle$, $\langle \epsilon_1^o \rangle$, and $\langle \epsilon_2^o \rangle$ are equal to zero, then Equation 23 becomes

$$\langle \epsilon_\beta^a \rangle \Big|_{\langle \epsilon_1^o \rangle = \mathbf{0}, \langle \epsilon_2^o \rangle = \mathbf{0}, \langle \epsilon_\beta^b \rangle = \mathbf{0}}$$

$$= -\mathbf{K}_{\beta y_{(1)}} \left(\mathbf{H}_{(1)}, \quad \mathbf{C}_\beta \right) \begin{pmatrix} \langle \epsilon_x^b \rangle \\ \langle \epsilon_\beta^b \rangle \end{pmatrix}$$

$$- \mathbf{K}_{\beta y_{(2)}} \left(\mathbf{H}_{(2)} \quad \mathbf{0} \right) \begin{pmatrix} \langle \epsilon_x^b \rangle \\ \langle \epsilon_\beta^b \rangle \end{pmatrix}$$

$$= -\mathbf{K}_{\beta y_{(1)}}\mathbf{H}_{(1)}\langle \epsilon_x^b \rangle - \mathbf{K}_{\beta y_{(2)}}\mathbf{H}_{(2)}\langle \epsilon_x^b \rangle.$$

As we can rewrite $\mathbf{K}_{\beta y_{(2)}}$ in terms of $\mathbf{K}_{\beta y_{(1)}}$, this simplifies to

$$\langle \epsilon_\beta^a \rangle \Big|_{\langle \epsilon_1^o \rangle = \mathbf{0}, \langle \epsilon_2^o \rangle = \mathbf{0}, \langle \epsilon_\beta^b \rangle = \mathbf{0}} = -\mathbf{K}_{\beta y_{(1)}}\mathbf{H}_{(1)}(\mathbf{I} - \mathbf{D})\langle \epsilon_x^b \rangle := \langle \epsilon_\beta^a \rangle_{t=0}, \qquad (C1)$$

where $\mathbf{D}$ is as defined in Equation 18 and we denote $\langle \epsilon_\beta^a \rangle$ at the first cycle to be $\langle \epsilon_\beta^a \rangle_{t=0}$.

APPENDIX D. SIMPLIFYING D WHEN THE ANCHOR OBSERVATIONS PARTIALLY OBSERVE THE SYSTEM

When the anchor observations only observe a subset of the state, then $\mathbf{D}$ is given by

$$\mathbf{D}|_{H_{(2)}=H_{(2)p}} = \begin{pmatrix} \mathbf{B}_{x_\phi} & \mathbf{B}_{x_{\phi\psi}} \\ \mathbf{B}_{x_{\phi\psi}} & \mathbf{B}_{x_\psi} \end{pmatrix} \begin{pmatrix} \mathbf{0} \\ \mathbf{H}_{(2)_\psi}^T \end{pmatrix}$$

$$\left(\left(\mathbf{0}, \mathbf{H}_{(2)_\psi} \right) \begin{pmatrix} \mathbf{B}_{x_\phi} & \mathbf{B}_{x_{\phi\psi}} \\ \mathbf{B}_{x_{\phi\psi}} & \mathbf{B}_{x_\psi} \end{pmatrix} \begin{pmatrix} \mathbf{0} \\ \mathbf{H}_{(2)_\psi}^T \end{pmatrix} + \mathbf{R}_{(2)} \right)^{-1}$$

$$\times \left(\mathbf{0}, \mathbf{H}_{(2)_\psi} \right),$$

$$= \begin{pmatrix} \mathbf{B}_{\mathrm{x}_{\phi\psi}} \mathbf{H}_{(2)_\psi}^{\mathrm{T}} \\ \mathbf{B}_{\mathrm{x}_\psi} \mathbf{H}_{(2)_\psi}^{\mathrm{T}} \end{pmatrix} (\mathbf{H}_{(2)_\psi} \mathbf{B}_{\mathrm{x}_\psi} \mathbf{H}_{(2)_\psi}^{\mathrm{T}} + \mathbf{R}_{(2)})^{-1}$$

$$\begin{pmatrix} \mathbf{0}, & \mathbf{H}_{(2)_\psi} \end{pmatrix},$$

$$= \begin{pmatrix} \mathbf{B}_{\mathrm{x}_{\phi\psi}} \mathbf{H}_{(2)_\psi}^{\mathrm{T}} (\mathbf{H}_{(2)_{\phi\psi}} \mathbf{B}_{\mathrm{x}_\psi} \mathbf{H}_{(2)_\psi}^{\mathrm{T}} + \mathbf{R}_{(2)})^{-1} \\ \mathbf{B}_{\mathrm{x}_\psi} \mathbf{H}_{(2)_\psi}^{\mathrm{T}} (\mathbf{H}_{(2)_\psi} \mathbf{B}_{\mathrm{x}_\psi} \mathbf{H}_{(2)_\psi}^{\mathrm{T}} + \mathbf{R}_{(2)})^{-1} \end{pmatrix}$$

$$\begin{pmatrix} \mathbf{0}, & \mathbf{H}_{(2)_\psi} \end{pmatrix},$$

$$= \begin{pmatrix} \mathbf{0} & \mathbf{B}_{\mathrm{x}_{\phi\psi}} \mathbf{H}_{(2)_\psi}^{\mathrm{T}} (\mathbf{H}_{(2)_\psi} \mathbf{B}_{\mathrm{x}_\psi} \mathbf{H}_{(2)_\psi}^{\mathrm{T}} + \mathbf{R}_{(2)})^{-1} \mathbf{H}_{(2)_\psi} \\ \mathbf{0} & \mathbf{B}_{\mathrm{x}_\psi} \mathbf{H}_{(2)_\psi}^{\mathrm{T}} (\mathbf{H}_{(2)_\psi} \mathbf{B}_{\mathrm{x}_\psi} \mathbf{H}_{(2)_\psi}^{\mathrm{T}} + \mathbf{R}_{(2)})^{-1} \mathbf{H}_{(2)_\psi} \end{pmatrix}$$

$$:= \mathbf{D}_{\mathrm{p}} = \begin{pmatrix} \mathbf{D}_{\mathrm{p}\phi} & \mathbf{D}_{\mathrm{p}\phi\psi} \\ \mathbf{D}_{\mathrm{p}\psi\phi} & \mathbf{D}_{\mathrm{p}\psi} \end{pmatrix}.$$

Then the sensitivity of $\boldsymbol{\beta}^{\mathrm{a}}$ to $\mathbf{x}^{\mathrm{b}}$ (Equation 17) when $\mathbf{H}_{(2)} = \mathbf{H}_{(2)\mathrm{p}}$ is given by

$$\left. \frac{\partial \boldsymbol{\beta}^{\mathrm{a}}}{\partial \mathbf{x}^{\mathrm{b}}} \right|_{\mathbf{H}_{(2)}=\mathbf{H}_{(2)\mathrm{p}}}$$
$$= -\mathbf{K}_{\beta \mathrm{y}_{(1)}} \Big|_{\mathbf{H}_{(2)}=\mathbf{H}_{(2)\mathrm{p}}} \mathbf{H}_{(1)} (\mathbf{I} - \mathbf{D}_{\mathrm{p}}) \quad (\mathrm{D1})$$

$$= -\mathbf{K}_{\beta \mathrm{y}_{(1)}} \Big|_{\mathbf{H}_{(2)}=\mathbf{H}_{(2)\mathrm{p}}} \begin{pmatrix} \mathbf{H}_{(1)_\phi}, & \mathbf{H}_{(1)_\psi} \end{pmatrix} \begin{pmatrix} \mathbf{I} & -\mathbf{D}_{\mathrm{p}\phi\psi} \\ \mathbf{0} & \mathbf{I} - \mathbf{D}_{\mathrm{p}\psi} \end{pmatrix} \quad (\mathrm{D2})$$

$$= -\mathbf{K}_{\beta \mathrm{y}_{(1)}} \Big|_{\mathbf{H}_{(2)}=\mathbf{H}_{(2)\mathrm{p}}} \begin{pmatrix} \mathbf{H}_{(1)_\phi}, & -\mathbf{H}_{(1)_\phi} \mathbf{D}_{\mathrm{p}\phi\psi} + \mathbf{H}_{(1)_\psi} (\mathbf{I} - \mathbf{D}_{\mathrm{p}\psi}) \end{pmatrix}. \quad (\mathrm{D3})$$

Substituting this into the equation for $\langle \epsilon_\beta^{\mathrm{a}} \rangle$ (Equation 24), we find

$$\langle \epsilon_\beta^{\mathrm{a}} \rangle \Big|_{\mathbf{D}=\mathbf{D}_{\mathrm{p}}}$$
$$= -\mathbf{K}_{\beta \mathrm{y}_{(1)}} \Big(\mathbf{H}_{(1)_\phi}, -\mathbf{H}_{(1)_\phi} \mathbf{D}_{\mathrm{p}\phi\psi} + \mathbf{H}_{(1)_\psi} (\mathbf{I} - \mathbf{D}_{\mathrm{p}\psi}) \Big) \begin{pmatrix} \langle \epsilon_{\mathrm{x}_\phi}^{\mathrm{b}} \rangle \\ \langle \epsilon_{\mathrm{x}_\psi}^{\mathrm{b}} \rangle \end{pmatrix} \quad (\mathrm{D4})$$

$$= -\mathbf{K}_{\beta \mathrm{y}_{(1)}} \Big(\mathbf{H}_{(1)_\phi} \langle \epsilon_{\mathrm{x}_\phi}^{\mathrm{b}} \rangle - \big[\mathbf{H}_{(1)_\phi} \mathbf{D}_{\mathrm{p}\phi\psi} - \mathbf{H}_{(1)_\psi} (\mathbf{I} - \mathbf{D}_{\mathrm{p}\psi}) \big] \langle \epsilon_{\mathrm{x}_\psi}^{\mathrm{b}} \rangle \Big). \quad (\mathrm{D5})$$

APPENDIX E. CALCULATING THE CLIMATOLOGICAL ESTIMATE OF THE BACKGROUND-ERROR COVARIANCE MATRIX FOR NUMERICAL EXPERIMENTS

A sample estimate of the climatological $\mathbf{B}_{\mathrm{v}}$ matrix is derived from an ensemble of 1DVar experiments cycled in time. The estimate of $\mathbf{B}_{\mathrm{v}}$ is sensitive to the $\mathbf{B}_{\mathrm{v}}$ used within the 1DVar experiments. We therefore repeat the method a second time with the new estimate of $\mathbf{B}_{\mathrm{v}}$, to get a second estimate of $\mathbf{B}_{\mathrm{v}}$. We found that two iterations was sufficient for the estimate of $\mathbf{B}_{\mathrm{v}}$ to converge to an appropriate climatological estimate for the Lorenz96 model assimilating both anchor and bias-corrected observations at every spatial variable.

The true trajectory is generated by solving Equation 55 numerically using the fourth-order Runge–Kutta scheme. The bias-corrected and anchor observations are at every location and are calculated as described in Equations 56 and 57.

On the first iteration, the background-error covariance matrix for the state is calculated using the SOAR error correlation function as in Equation 58 with standard deviation $\sigma_{\mathrm{bx}} = 1$ and length-scale $L_{\mathrm{b}} = 1$. The bias-coefficient background-error standard deviation, $\sigma_{\mathrm{b}\beta}$, is set to 0.5, and all error covariances between the state and bias coefficient were set to 0, so that $\mathbf{B}_{\mathrm{v}}$ was given by

$$\mathbf{B}_{\mathrm{v}} = \begin{pmatrix} \mathbf{B}_{\mathrm{x}} & \mathbf{0} \\ \mathbf{0} & \sigma_{\mathrm{b}\beta}^2 \end{pmatrix}. \quad (\mathrm{E1})$$

At the initial time, an ensemble of 15 background states and bias coefficients is calculated as in Section 5.1 using $\mathbf{B}_{\mathrm{v}}$ as the background-error variance. An ensemble of analyses was generated by updating each background member using Equations 5 and 6. The analysis ensemble was then evolved forward via the model by ten time steps to give the background ensemble at the next observation time. This method for calculating the background ensemble is repeated 700 times, such that the background ensemble at any given time step greater than 1 is given by

$$\mathbf{x}_{t+1,i}^{\mathrm{b}} = M_{t \rightarrow t+1}(\mathbf{x}_{t,i}^{\mathrm{a}}), \quad \beta_{t+1,i}^{\mathrm{b}} = \beta_{t,i}^{\mathrm{a}}, \quad (\mathrm{E2})$$

where t is the time index; i is the ensemble member index; and $M_{t \rightarrow t+1}(\mathbf{x}_{t,i})$ is the model that takes the state from time step t to $t+1$. The errors between the background of each ensemble member and the true state at the same time are given by

$$\epsilon_{\mathrm{x}_{t,i}}^{\mathrm{b}} = \mathbf{x}_{t,i}^{\mathrm{b}} - \mathbf{x}_{t,i}^{\mathrm{t}}, \quad \epsilon_{\beta_{t,i}}^{\mathrm{b}} = \beta_{t,i}^{\mathrm{b}} - \beta^{\mathrm{t}}. \quad (\mathrm{E3})$$

The analysis ensemble was cycled for 700 assimilation times to provide a total of 10,500 samples (700×15) of the background error (Equation E3). From these 10,500 samples of $\epsilon_{\mathrm{x}}^{\mathrm{b}}$ and $\epsilon_\beta^{\mathrm{b}}$, we were then able to update the climatological estimate of the $\mathbf{B}_{\mathrm{v}}$ matrix by calculating the

sample covariances of all state background errors $\epsilon_{x_{t,i}}^b$ and bias coefficient errors $\epsilon_{\beta_{t,i}}^b$:

$$\text{Cov}\left(\epsilon_{xj_{t,i}}^b, \epsilon_{xk_{t,i}}^b\right) = \frac{\sum_{t=1}^{700}\sum_{i=1}^{15}(\epsilon_{xj_{t,i}} - \overline{\epsilon}_{xj})(\epsilon_{xk_{t,i}} - \overline{\epsilon}_{xk})}{10,500 - 1},$$
(E4)

$$\sigma_{b\beta} = \frac{\sum_{t=1}^{700}\sum_{i=1}^{15}(\epsilon_{\beta_{t,i}} - \overline{\epsilon}_\beta)(\epsilon_{\beta_{t,i}} - \overline{\epsilon}_\beta)}{10,500 - 1},$$
(E5)

where $\epsilon_{xj_{t,i}}^b$ and $\epsilon_{xk_{t,i}}^b$ are the background errors for the jth and kth variables of the state at time t, ensemble i; and $\overline{\epsilon}_{xj}$ and $\overline{\epsilon}_{xk}$ are the mean background errors from the 10,500 samples at state variables j and k, respectively. The error covariances between the background state and bias coefficient were set to 0 as before.

The process described above was repeated in another iteration, but $\mathbf{B}_x$ used within the assimilation was given by the state background-error covariance matrix that had just been calculated. To remove noise caused by the limited sample estimate of $\mathbf{B}_x$, we set the covariances between state variables further than two grid points apart to 0. $\sigma_{b\beta}^2$ was given by the estimate of the bias-coefficient background-error variance, which had also just been calculated. Using the new $\mathbf{B}_v$, a new background-error covariance matrix was calculated from Equations E2 – E5, again
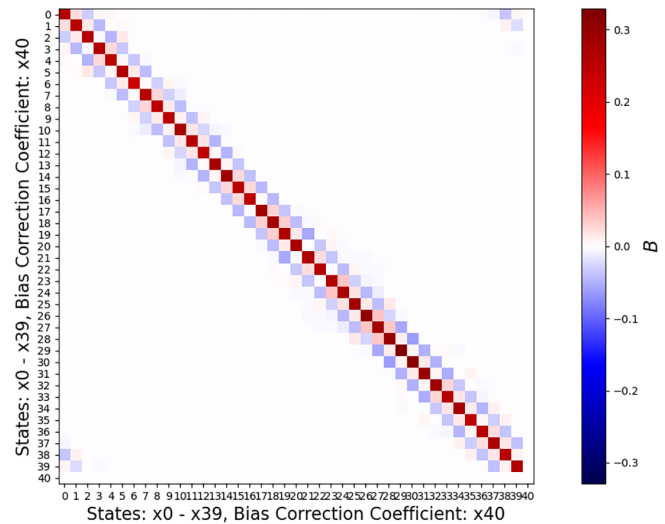

FIGURE E1 The climatological $\mathbf{B}_v$ calculated from an ensemble. [Colour figure can be viewed at wileyonlinelibrary.com]

using 700 cycles and 15 realisations. This was taken to be the optimal background-error covariance matrix in the numerical experiments in Sections 5.1.1 and 5.2.1.

Figure E1 is the error covariance matrix from the final iteration, which was used in our numerical experiments. As the state variables are on a circular domain, x_0 and x_{39} have a correlation to each other, as they would be next to each other in the domain.