



**University of
Reading**

Bias Correction of Satellite Data in Data Assimilation for Numerical Weather Prediction

Devon Francis

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

PhD in Atmosphere, Oceans and Climate

Department of Meteorology

School of Mathematical, Physical and Computational Sciences

August 2023

University of Reading

Abstract

Data assimilation is a statistical technique that combines information from observations and a mathematical model in order to make the best estimate of a state at the current time, where the best estimate is known as the analysis. Basic data assimilation theory relies on the assumption that the background, model and observations are unbiased. However, this is often not the case and, if biases are left uncorrected, can cause significant systematic errors in the analysis. When bias is only present in the observations, VarBC (Variational Bias Correction) can correct for observation bias, and when bias is only present in the model, WC4DVar (Weak-Constraint 4D Variational Assimilation) can correct for model bias. However, when both observation and model biases are present, it is unknown how the different bias correction methods interact, and the role of anchor (unbiased) observations becomes crucial for providing a frame of reference from which the biases may be estimated.

We highlight the importance of correctly specifying the background error statistics in VarBC to ensure that the analysis is more precise than the prior estimate. We then demonstrate the characteristics needed in anchor observations to effectively reduce the contamination of biases in the analysis, when one or both types of bias are corrected for. We find that the location and timing of anchor observations is important in their ability to reduce the contamination of bias, as well as having precise anchor observations. In this thesis we study the mathematical theory underpinning VarBC and WC4DVar, and demonstrate our results in a toy numerical system.

Declaration

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

Devon Francis

Acknowledgements

I would like to give a massive thanks to my four supervisors: Dr Alison Fowler, Prof. Amos Lawless, Dr Stefano Migliorini and Dr John Eyre. Amos and Alison, thank you so much for your constant guidance, willingness to answer my stupid questions and all of the time you put into reading my paper, chapters and general writing; the thesis would not have existed without your support and guidance. John and Stefano, thank you for all of the insights into operational data assimilation at the Met Office and your enthusiasm for all of the research that has been haphazardly explained to you in our brief meetings! Thanks also for setting up my visit at the Met Office, and for Ruth Taylor for hosting me and guiding me through the operational VarBC system.

This work was funded by NERC as part of the Scenario DTP, and had CASE funding from the Met Office. A huge thank you to Wendy Neale and everyone else in Scenario for organising funding, travel to conferences and for running the numerous training, networking and social activities provided by the Scenario DTP over the past four years.

The Met department has been such a lovely place to work; to the whole of cohort 6, everyone in Lyle (Wednesday afternoons are much more suited to unicycling/frisbee-ing/tea-drinking than working...), the rest of my Harry Pitt office, DARCies, and all of the PhD students who were lucky enough to be in the Met building from the start - there are too many people to name, but I will miss you all! And thank you to everyone outside of Reading for giving me a break from thinking about research life - everyone at (canoe) polo, everyone who I've gone for tea and cake with and just everyone who's made me laugh. Finally thanks to Cam, for letting me whine about PhD life to you and then knowing how to cheer me up, for pretending to understand what data assimilation is, for reminding me that I am capable of finishing, you're the best.

Contents

1	Introduction	1
2	Background data assimilation theory and sources of Earth observations	4
2.1	Data assimilation in NWP	5
2.2	Observing network	14
2.3	Observation bias correction	23
2.4	Models used in NWP	34
2.5	Model bias correction	37
2.6	Summary	44
3	Motivation for research questions	45
3.1	Specifying background error covariances in VarBC	45
3.2	Correcting for either observation or model bias and the importance of un- biased observations	53
3.3	Research questions	61
3.4	Summary	63
4	Mis-specification of background error statistics on a VarBC system	64
4.1	Introduction	64

4.2	Sub-optimal analysis error covariance matrix equations for the state and observation bias coefficient	65
4.3	Scalar example	71
4.4	Conclusions and summary	76
5	Experimental design	79
5.1	Lorenz 96	80
5.2	Data assimilation set up	84
5.2.1	Simulating observations	84
5.2.2	Generating the background error covariance matrix	85
5.2.2.1	Generating a climatological \mathbf{B}	86
5.2.3	Correcting for observation bias (VarBC)	90
5.2.4	Correcting for model bias (WC4DVar)	91
5.3	Summary	92
6	The role of anchor observations in VarBC in the presence of model bias.	
	Part I: The importance of the location of anchor observations	94
6.1	Introduction	94
6.2	The importance of anchor observations in 3DVarBC: theoretical results	95
6.2.1	Position of anchor observations relative to bias-corrected observations	102
6.2.1.1	Anchor observations fully observe the domain	102
6.2.1.2	Anchor observations partially observe the domain	103
6.2.1.3	Special case: anchor observations and bias-corrected observations observe different parts of the state	105
6.2.2	Position of anchor observations relative to model bias	108
6.2.2.1	Background bias in state variables that are observed by anchor observations, but not in all state variables observed by bias-corrected observations (figure 6.1a):	108

6.2.2.2	Background bias not in state variables observed by anchor observations (figure 6.1b):	110
6.2.2.3	Background bias in state variables observed by both anchor and bias-corrected observations (figure 6.1c):	111
6.2.3	The effect of a biased bias coefficient analysis on the state analysis in further cycles	112
6.3	Numerical results	114
6.3.1	One Cycle Experiments (No model evolution)	114
6.3.1.1	The effect of varying the anchor observation error variance	115
6.3.1.2	The location of the anchor observations relative to the background bias	116
6.3.2	Cycled Experiment	120
6.3.2.1	How bias in the state analysis and bias coefficient analysis accumulates when the system is cycled	121
6.4	Conclusions and Discussion	123
7	The role of anchor observations in VarBC in the presence of model bias.	
	Part II: The importance of the timing of the anchor observations	127
7.1	Introduction	127
7.2	4DVarBC with two observation types	128
7.3	4DVarBC with model bias that is not accounted for	130
7.3.1	Bias-corrected observations at $t = 1$, anchor observations at $t = 2$. .	137
7.3.1.1	Anchor and bias-corrected observations observe different spatial states	140
7.3.2	Anchor observations at $t = 1$, bias-corrected observations at $t = 2$. .	145
7.3.2.1	Anchor and bias-corrected observations observe different spatial states	147

7.3.3	Anchor and bias-corrected observations at the same time	150
7.3.3.1	Anchor and bias-corrected observations observe different spatial states	151
7.3.4	Summary of theoretical results	152
7.4	Numerical Experiments	153
7.5	Conclusions and Discussion	159
7.6	Summary	163
8	The interaction between VarBC and WC4DVar and the role of anchor observations	164
8.1	Introduction	164
8.2	WC4DVar with uncorrected observation bias	165
8.2.1	Scalar system with anchor and biased observations at one time step	169
8.2.1.1	Biased and anchor observations at $t = 1$	172
8.2.1.2	Biased observations at $t = 1$, anchor observations at $t = 2$.	175
8.2.1.3	Biased observations at $t = 2$, anchor observations at $t = 1$.	177
8.2.2	Numerical experiments in a multi-variable system	181
8.2.3	Summary of WC4DVar when not correcting for observation bias . .	186
8.3	Comparison of the role of anchor observations when correcting for observa- tion bias, model bias, or both	188
8.3.1	Model and observation biases acting in different directions	196
8.4	Conclusions	197
8.5	Summary	199
9	Conclusions	201
	References	209

List of Figures

2.1	A schematic diagram of an LEO (low earth orbiting) satellite receiving a signal from the GNSS (global navigation system satellite) that has been refracted in the atmosphere. Note that the bending angle has been exaggerated for the illustration. Figure from Gleisner, Ringer, and Healy (2022) under the Creative Commons Attribution 4.0 International License: http://creativecommons.org/licenses/by/4.0/	18
2.2	Geographical coverage of GPS-RO data on 18th October 2023 at 00UTC. Figure from ECMWF website, accessed on 18th October 2023 (<i>ECMWF Geographical Coverage</i> , 2023)	19
2.3	Geographical coverage of used radiosondes that gave temperature data on 13th February 2023 at 12UTC. Figure from ECMWF website, accessed on 14th February 2023 (<i>ECMWF Geographical Coverage</i> , 2023).	20
2.4	Uncorrected scan bias by latitude band, NOAA 12, MSU 2, February 1997. Figure from Harris and Kelly (2001), copyright ©2001 John Wiley & Sons, Ltd	22
2.5	Air mass thickness bias predictors p_1 : 850-300hPa (a) and p_2 : 200-50hPa (b) on world maps for the ATMS (Advanced Microwave Technology Sounder) satellite on the 9th April 2013.	26
2.6	Bias predictor p_6 , a scan angle bias predictor.	27

3.1	Analysis error variance, A , as a function of the true background error variance, B , where the assumed background error variance is $B_A = 1$, and the assumed (and true) observation error variance is $R = 1$. A_{opt} is calculated from equation (3.6) and A is A_{sub} from equation (3.9). The diagonal line through the origin is the line $A = B$. The dashed vertical line shows the value of $B = B_d$, at which the analysis enters the danger zone for A_{sub} . ©British Crown Copyright, the Met Office, from Eyre and Hilton (2013). Reproduced by permission of John Wiley and Sons Inc.	52
3.2	Time series of prior RMSE for different experiments (colours) using (a) a perfect model with and without spatially constant observation bias = 0.3, and (b) an imperfect model with model bias = 2. Legends show specified observation biases and the augmented vector estimated in the assimilation. From Lorente-Plazas and Hacker (2017) ©American Meteorological Society. Used with permission.	59
4.1	σ_{ax}^2 varying both k_x^a and σ_{bx}^{a2} . $h = c_\beta = 1$. $\sigma_{\text{b}\beta}^{a2} = 1$, $\sigma_{\text{bx}}^2 = 0.75$. (a) $\sigma_{\text{b}\beta}^2 = 1.5$ (underestimating $\sigma_{\text{b}\beta}^2$) and (b) $\sigma_{\text{b}\beta}^2 = 0.5$ (overestimating $\sigma_{\text{b}\beta}^2$). White dashed line shows when $\sigma_{\text{ax}}^2 = \sigma_{\text{bx}}^2$. Black dotted line shows when $\sigma_{\text{bx}}^{a2} = \sigma_{\text{bx}}^2$	73
4.2	As in figure 4.1 but varying k_x^a and $\sigma_{\text{b}\beta}^{a2}$. $\sigma_{\text{b}\beta}^2 = 0.75$, $\sigma_{\text{bx}}^{a2} = 1$. (a) $\sigma_{\text{bx}}^2 = 1.5$ (underestimating σ_{bx}^2) and (b) $\sigma_{\text{bx}}^2 = 0.5$ (overestimating σ_{bx}^2). White dashed line shows when $\sigma_{\text{ax}}^2 = \sigma_{\text{bx}}^2$. Black dotted line shows when $\sigma_{\text{b}\beta}^{a2} = \sigma_{\text{b}\beta}^2$	75
5.1	Two model runs with the Lorenz 96 model from true initial conditions (black) and perturbed initial conditions (blue) for 100 time steps, where the time step is given by 0.0125.	82
5.2	Two model runs with the Lorenz 96 model with different forcing values, initialised from the same initial conditions, time step = 0.0125.	83

5.3	The climatological \mathbf{B}_x calculated from an ensemble, after the initial iteration.	89
5.4	The climatological \mathbf{B}_x calculated from an ensemble, after the second iteration.	89
6.1	A schematic diagram depicting the location of the background 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}_ϕ ; $\mathbf{H}_{(1)\psi}$ is the observation operator relating the bias-corrected observations to \mathbf{x}_ψ ; and $\mathbf{H}_{(2)\psi}$ is the observation operator relating the anchor observations to \mathbf{x}_ψ . The background bias is shown by the patterned background. In (a) there is only background bias in \mathbf{x}_ψ , which is observed by the anchor observations and some bias-corrected observations, (b) there is only background bias in \mathbf{x}_ϕ , which is only observed by the bias-corrected observations and (c) there is background bias in both \mathbf{x}_ϕ and \mathbf{x}_ψ , which is observed by both bias-corrected and anchor observations.	109
6.2	The bias ratio, equation (6.62), 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 section 5.2.2.1. A bias of 0.15 was added to all background state variables. The orange line is the analytic ratio, calculated from the linearised observation operators and error covariance matrices in the system and the blue crosses are the numerical ratios, calculated by averaging over 3000 realisations.	117

6.3	The bias ratio in equation (6.62) 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 (blue); in state variables only observed by anchor observations (orange); in state variables observed by both types of observations (green).	119
6.4	The ratios in equations (6.62) and (6.65) over 200 cycles for 100 realisations. $F_{\text{biased}} = 8.8$, error variances are 1, \mathbf{B}_v is calculated by an ensemble (section 5.2.2.1). Anchor and biased observations observe every state variable. . . .	122
7.1	VarBC to correct for observation bias in a 4DVar system, with anchor and bias-corrected observations at different times in the window. $\sigma_{o(2)}^2 = 1$. . .	155
7.2	As in figure 7.1 but with more precise anchor observations ($\sigma_{o(2)}^2 = 0.1$). . .	158
8.1	$K_{\eta y(1)}$ (blue) and $K_{xy(1)}$ (red) when biased observations at $t = 1$, anchor observations at $t = 1$ for different values of α	174
8.2	$K_{\eta y(1)}$ (blue) and $K_{xy(1)}$ (red) with biased observations at $t = 1$ and anchor observations at $t = 2$ for varied α	176
8.3	$K_{\eta y(1)}$ (blue) and $K_{xy(1)}$ (red) when biased observations at $t = 2$, anchor observations at $t = 1$, for varied α	179
8.4	$K_{\eta y(1)}$ (blue) and $K_{xy(1)}$ (red) when biased observations at $t = 2$, anchor observations at $t = 1$ for different values of α and 3 values of $\sigma_{o(2)}$	180
8.5	WC4DVar to correct for model bias only, with anchor and biased observations at different locations in time. $\sigma_{o(2)}^2 = 1$	183
8.6	As in figure 8.5, but with more precise anchor observations. $\sigma_{o(2)}^2 = 0.1$. . .	185

1 Chapter 1

2 Introduction

3 The mathematical models that describe the dynamics of the atmosphere are sensitive to
4 their initial conditions, which means that small errors in the initial conditions grow into
5 larger errors in the forecast. To minimise errors in the initial state, observations are in-
6 corporated with the model in a technique known as data assimilation, which provides a
7 better estimate of the initial state, known as the analysis, and thus provide more accurate
8 and more precise forecasts. Satellite observations are vital for improving weather fore-
9 casts (English et al., 2013; Eyre et al., 2022), but satellite radiance data often contain
10 significant biases (Dee & Uppala, 2009) which would reduce the accuracy of the numerical
11 weather predictions. Thankfully, most centres now correct for observation bias using a
12 technique known as VarBC (variational bias correction) (Dee, 2004), which allows more
13 of the satellite data available to be used effectively. VarBC has been developed under the
14 assumption that model biases are negligible. However, when both observation and model
15 biases are present, the observation bias is 'corrected' towards the model's (biased) clima-
16 tology, leading to an incorrect observation bias correction and therefore a biased analysis.
17 To overcome this, unbiased observations (known as anchor observations) are used as an
18 unbiased reference point to anchor the observation bias correction to the truth. Unfor-
19 tunately, typical anchor observations such as radiosondes can be sparse and infrequent

1 (*ECMWF Geographical Coverage*, 2023). Therefore it is important to know how to use
2 anchor observations effectively so that they can have the most impact, and it is important
3 to have knowledge of where newer developments of anchor observations, such as radio
4 occultation measurements, are needed the most.

5 Most previous studies have focused on the overall consequences of correcting for obser-
6 vation bias in the presence of model bias (e.g. Auligné et al., 2007; Dee & Uppala, 2009;
7 Han & Bormann, 2016) (or vice versa, the overall consequences of correcting for model
8 bias in the presence of observation bias (Laloyaux et al., 2020a)). They have touched on
9 the importance of unbiased reference information such as anchor observations, but have
10 not studied the characteristics of anchor observations needed to effectively reduce the con-
11 tamination of bias. Eyre (2016) discussed the importance of using anchor observations
12 when correcting for observation bias in the presence of model bias, presenting results from
13 a scalar system, and highlighted the consequences of giving too much weight to previous
14 estimates of the state, rather than to the anchor observations.

15 This thesis aims to understand the multivariate theory of why observation bias correc-
16 tion can give inaccurate or imprecise results, and to understand the role of anchor obser-
17 vations in improving these results. We will study the equations of VarBC and illustrate
18 our results in a simple numerical system. By understanding the underlying theory behind
19 VarBC, we will begin to be able to explain why more complicated systems can struggle,
20 and therefore where future research and developments should focus their improvements of
21 operational VarBC.

22 The structure of the thesis is as follows. In chapter 2 we will describe the background
23 theory needed for the rest of the thesis. We will begin by describing basic data assimilation
24 theory in section 2.1. We will then describe the operational observing network in section
25 2.2, and how VarBC corrects for observation biases in section 2.3. We will discuss the
26 mathematical models in use in numerical weather prediction in section 2.4, and how model
27 biases can be corrected for using a technique known as WC4DVar (weak-constraint 4-

1 dimensional variational assimilation) in section 2.5. In chapter 3 we will present the current
2 state of work on observation bias correction and identify the gaps that will be addressed by
3 our work, with the final section (section 3.3) highlighting our research questions which we
4 will answer throughout the thesis. Chapter 4 is our first results chapter, with results and
5 discussion on the negative impact that mis-specifying background error statistics has on
6 the analysis in VarBC. In chapter 5, we describe the set up of an idealised numerical system,
7 that is used to illustrate results of the subsequent chapters. In section 5.1 we discuss the
8 simple model chosen and in section 5.2 we describe the set up of the data assimilation
9 system, including how we generated the observations, background and constructed the
10 bias correction schemes. In chapters 6 and 7 we present results for the role of anchor
11 observations in VarBC, in the presence of both observation and model biases: in chapter 6
12 we study a VarBC system in the context of 3-dimensional variational assimilation, focusing
13 on the importance of the location of the anchor observations; in chapter 7 we study a
14 VarBC system in the context of 4-dimensional variational assimilation, focusing on the
15 importance of the timing of the anchor observations. Chapter 8 is our final research
16 chapter, which compares: correcting for observation bias in the presence of model bias,
17 correcting for model bias in the presence of observation bias, and correcting for both
18 sources of bias simultaneously. Finally, we present our overall conclusions and discussion
19 in chapter 9.

1 Chapter 2

2 Background data assimilation 3 theory and sources of Earth 4 observations

5 In this chapter we will discuss the background information needed to understand the ma-
6 jority of the thesis. In section 2.1 basic data assimilation theory will be introduced, both
7 for 3D and 4D variational assimilation. In section 2.2 we will introduce the current observ-
8 ing network and discuss the uncertainties associated with different observations. We will
9 then describe current bias correction techniques to account for some of these uncertainties
10 in section 2.3, with a particular interest in VarBC (variational bias correction). In sec-
11 tion 2.4 we will introduce the mathematical models used in numerical weather prediction
12 and discuss how errors can arise in them. Finally, in section 2.5 we will discuss a model
13 bias correction technique known as WC4DVar (weak-constraint 4-dimensional variational
14 assimilation).

2.1 Data assimilation in NWP

In numerical weather prediction (NWP), we have access to prior information from previous forecasts and current observations of surface and atmospheric geophysical fields. Bayes' theorem combines the probability that a prior event will occur with the conditional probability of a current event, so can be used to link data from forecasts and observations (Efron, 2013). Bayes theorem is given by,

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}, \quad (2.1)$$

where \mathbf{x} and \mathbf{y} are random variables and p indicates a probability distribution function.

In NWP, we can define $\mathbf{y} \in \mathbb{R}^m$ to be the data from observations and $\mathbf{x} \in \mathbb{R}^n$ to be the state we are interested in modelling. Therefore $p(\mathbf{x})$, the probability distribution function (pdf) describing the state, is the prior pdf that contains all of the previous knowledge about the system before assimilating the new observations; $p(\mathbf{y}|\mathbf{x})$ is the likelihood of the observations, given the state; and $p(\mathbf{y})$ is the prior probability of the observations (Lorenz, 1986).

If we assume that the errors in the model state and the observations are Gaussian and unbiased, then the pdfs for the prior and the observation likelihood are,

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{B}_x|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}_x^{-1} (\mathbf{x} - \mathbf{x}^b)\right), \quad (2.2)$$

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{(2\pi)^{\frac{m}{2}} |\mathbf{R}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{y} - h(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y} - h(\mathbf{x}))\right), \quad (2.3)$$

where $|\cdot|$ indicates the determinant; \mathbf{x}^b is known as the background state and is generated by a previous forecast; $h(\mathbf{x})$ is known as the observation operator, which is a function that transforms the state from the vector space of the model variables into the vector space of the observations; $\mathbf{B}_x \in \mathbb{R}^{n \times n}$ is known as the background error covariance matrix and

1 $\mathbf{R} \in \mathbb{R}^{m \times m}$ is known as the observation error covariance matrix. \mathbf{B}_x is the matrix that
2 describes the covariance of errors in the background and \mathbf{R} is the matrix that describes
3 the covariance of the errors in the observation process, for example instrument errors and
4 representativeness errors (Bouttier & Courtier, 2002). The diagonals of the background
5 and observation error covariance matrices are the error variances, which describe how
6 precise the background states and observations are expected to be. The off-diagonals
7 are the error cross covariances between the components of the background state (for the
8 background error covariance matrix) or between the observations (for the observation
9 error covariance matrix). It is assumed that the background and observation errors are
10 uncorrelated.

11 Therefore, by combining equation (2.1) with equations (2.2) and (2.3), the posterior
12 pdf of the state given the observations, is proportional to

$$p(\mathbf{x}|\mathbf{y}) \propto \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{B}_x|^{\frac{1}{2}} (2\pi)^{\frac{p}{2}} |\mathbf{R}|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} \left((\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}_x^{-1} (\mathbf{x} - \mathbf{x}^b) + (\mathbf{y} - h(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y} - h(\mathbf{x})) \right) \right\}. \quad (2.4)$$

13 The probability distribution function with the smallest variance will give the most precise
14 estimate. If the statistics are of Gaussian form, then the state with the minimum *a*
15 *posteriori* variance (the mean) is also the state with the maximum *a posteriori* probability
16 (the mode) (Van Leeuwen & Evensen, 1996). The maximum *a posteriori* probability is
17 known as the MAP estimate. Therefore, to find the minimum error variance estimate,
18 the MAP estimate can be found by minimising the negative of the expression inside the
19 exponential function of equation (2.4), i.e.

$$\max_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}) = \min_{\mathbf{x}} \frac{1}{2} \left((\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}_x^{-1} (\mathbf{x} - \mathbf{x}^b) + (\mathbf{y} - h(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y} - h(\mathbf{x})) \right). \quad (2.5)$$

20 The function to be minimised is known as the cost function. When observations are
21 only taken at one time step then the cost function for 3DVar (3-dimensional variational

1 assimilation) is given by (Lorenc, 1986)

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}_x^{-1}(\mathbf{x} - \mathbf{x}^b) + \frac{1}{2}(\mathbf{y} - h(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y} - h(\mathbf{x})). \quad (2.6)$$

2 Minimising the cost function simultaneously minimises the difference between the state
 3 and the background in the first term and minimises the difference between the state and
 4 the observations in the second term. The first product in equation (2.6) measures the
 5 difference between the state \mathbf{x} and the prior state \mathbf{x}^b . This is weighted by the inverse of
 6 the error covariance matrix \mathbf{B}_x , which describes the uncertainty in the background state.
 7 The second product in equation (2.6) measures the difference between the observations \mathbf{y}
 8 and the state in observation space $h(\mathbf{x})$. These are weighted by the inverse of the error
 9 covariance matrix \mathbf{R} , which describes the uncertainty in the observations. Both \mathbf{B}_x and
 10 \mathbf{R} are symmetric and positive definite.

11 In order to explicitly minimise equation (2.6), it is assumed that $h(\mathbf{x})$ is approximately
 12 linear. Therefore, $h(\mathbf{x})$ can be approximated by the first order Taylor expansion around
 13 \mathbf{x}^b to give,

$$h(\mathbf{x}) \approx h(\mathbf{x}^b) + \mathbf{H}|_{\mathbf{x}^b}(\mathbf{x} - \mathbf{x}^b), \quad (2.7)$$

14 where $\mathbf{H}|_{\mathbf{x}^b} \in \mathbb{R}^{m \times n}$ is the Jacobian of $h(\mathbf{x})$ with respect to \mathbf{x} at \mathbf{x}^b , which for simplicity,
 15 we will denote as \mathbf{H} . Differentiating equation (2.6) with respect to \mathbf{x} gives,

$$\nabla J(\mathbf{x}) = \mathbf{B}_x^{-1}(\mathbf{x} - \mathbf{x}^b) - \mathbf{H}^T \mathbf{R}^{-1}(\mathbf{y} - h(\mathbf{x})), \quad (2.8)$$

16 Substituting equation (2.7) into equation (2.8) linearises the gradient of the cost function,

$$\nabla J(\mathbf{x}) = \mathbf{B}_x^{-1}(\mathbf{x} - \mathbf{x}^b) - \mathbf{H}^T \mathbf{R}^{-1}(\mathbf{y} - (h(\mathbf{x}^b) + \mathbf{H}(\mathbf{x} - \mathbf{x}^b))). \quad (2.9)$$

17 The state that minimises equation (2.6) will be the best estimate of the state, given pre-

1 vious knowledge of the state from the background and observations. In data assimilation
 2 this is known as the analysis state and is denoted by \mathbf{x}^a (Lorenc, 1986). By substituting \mathbf{x}
 3 with \mathbf{x}^a in equation (2.9), equating equation (2.9) to 0 and rearranging, the state analysis
 4 is given by,

$$\mathbf{x}^a = \mathbf{x}^b + (\mathbf{B}_x^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} (\mathbf{y} - h(\mathbf{x}^b)), \quad (2.10)$$

5 where $\mathbf{y} - h(\mathbf{x}^b)$ is known as the innovation vector. The product of error covariance
 6 matrices can be rewritten using the Sherman-Morrison-Woodbury formula to give,

$$(\mathbf{B}_x^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} = \mathbf{B}_x \mathbf{H}^T (\mathbf{H} \mathbf{B}_x \mathbf{H}^T + \mathbf{R})^{-1} := \mathbf{K}, \quad (2.11)$$

7 which is denoted as \mathbf{K} and is known as the Kalman gain matrix. In this thesis we will use
 8 the form of \mathbf{K} as given on the right hand side of equation (2.11) as it does not include
 9 explicit inverses of \mathbf{B}_x and \mathbf{R} .

10 Therefore substituting \mathbf{K} back into equation (2.10) gives us (Lorenc, 1986),

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}(\mathbf{y} - h(\mathbf{x}^b)), \quad (2.12)$$

11 which is known as the best linear unbiased estimate (or BLUE) solution, when Gaussian-
 12 ity is assumed and the observation operator is approximately linear. The solution, given
 13 by the analysis \mathbf{x}^a , depends linearly on the difference between the observations and the
 14 background state. The Kalman gain matrix describes the sensitivity of the state to the ob-
 15 servations: if \mathbf{K} is small then the state analysis will be more dependent on the background
 16 and if \mathbf{K} is large then the state analysis will be more dependent on the observations. As
 17 the solution to equation (2.12) requires explicitly minimising the cost function, equation
 18 (2.6), it is in general too computationally expensive to calculate the BLUE for a realistic
 19 system, but it is useful in understanding the theory behind the bigger systems. Oper-
 20 ationally, equation (2.6) would be minimised by using numerical minimisation methods

1 such as the conjugate gradient method or quasi-Newton methods (e.g. Liu et al., 2018;
 2 Gratton & Tshimanga, 2009; Courtier et al., 1994).

3 The BLUE solution can be used to understand the errors in the analysis. The analysis,
 4 background and observation errors can respectively be denoted as

$$\boldsymbol{\epsilon}_x^a = \mathbf{x}^a - \mathbf{x}^t, \quad \boldsymbol{\epsilon}_x^b = \mathbf{x}^b - \mathbf{x}^t, \quad \boldsymbol{\epsilon}^o = \mathbf{y} - h(\mathbf{x}^t), \quad (2.13)$$

5 where \mathbf{x}^t is the true state and $\boldsymbol{\epsilon}$ denotes the error.

6 The analysis error can then be defined in terms of the background and observation
 7 errors. This can be calculated by first subtracting the true state from both sides of
 8 equation (2.12) and then adding and subtracting $h(\mathbf{x}^t)$ from within the innovation vector
 9 to give,

$$\mathbf{x}^a - \mathbf{x}^t = \mathbf{x}^b - \mathbf{x}^t + \mathbf{K}(\mathbf{y} - h(\mathbf{x}^t) + h(\mathbf{x}^t) - h(\mathbf{x}^b)). \quad (2.14)$$

10 The analysis, background and observation errors from equation (2.13) can be substituted
 11 into equation (2.14). Then writing \mathbf{x}^t as $\mathbf{x}^b - \boldsymbol{\epsilon}_x^b$, equation (2.14) becomes,

$$\boldsymbol{\epsilon}_x^a = \boldsymbol{\epsilon}_x^b + \mathbf{K}(\boldsymbol{\epsilon}^o - (h(\mathbf{x}^b) - h(\mathbf{x}^b - \boldsymbol{\epsilon}_x^b))). \quad (2.15)$$

12 Approximating $h(\mathbf{x}^b - \boldsymbol{\epsilon}_x^b)$ using the Taylor expansion about the background state cancels
 13 the $h(\mathbf{x}^b)$ term, so that equation (2.15) becomes,

$$\boldsymbol{\epsilon}_x^a = \boldsymbol{\epsilon}_x^b + \mathbf{K}(\boldsymbol{\epsilon}^o - \mathbf{H}\boldsymbol{\epsilon}_x^b), \quad (2.16)$$

$$\therefore \boldsymbol{\epsilon}_x^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\boldsymbol{\epsilon}_x^b + \mathbf{K}\boldsymbol{\epsilon}^o, \quad (2.17)$$

14 where we have again used that \mathbf{H} is the Jacobian of the observation operator h with respect
 15 to the state at the background state \mathbf{x}^b .

1 The expectation value of the error is formally defined as,

$$\langle \boldsymbol{\epsilon} \rangle = \int_{-\infty}^{\infty} \boldsymbol{\epsilon} p(\boldsymbol{\epsilon}) d\boldsymbol{\epsilon}, \quad (2.18)$$

2 where $\boldsymbol{\epsilon}$ is the error and $p(\boldsymbol{\epsilon})$ is the probability distribution function of the error (Baoding,
3 2007). Taking the expected value of equation (2.17) gives the expected value of the state
4 analysis errors,

$$\langle \boldsymbol{\epsilon}_x^a \rangle = (\mathbf{I} - \mathbf{KH}) \langle \boldsymbol{\epsilon}_x^b \rangle + \mathbf{K} \langle \boldsymbol{\epsilon}^o \rangle, \quad (2.19)$$

5 where the angled brackets denote the expected value. Taking the expected value of the
6 errors is a generalisation of the mean error. Therefore, if the expected value of the errors
7 is zero, this implies that the error is unbiased. Since so far we have assumed that the
8 observation and background errors are both unbiased, equation (2.19) implies that the
9 expected value of the analysis error is zero and therefore the state analysis is also unbiased.

10 Assuming that the background and observation errors are unbiased, the background
11 and observation error covariance matrices can be expressed as the expected value of
12 the background/observation errors multiplied by their respective transposes (Bouttier &
13 Courtier, 2002),

$$\mathbf{B}_x = \langle \boldsymbol{\epsilon}_x^b \boldsymbol{\epsilon}_x^{bT} \rangle, \quad \mathbf{R} = \langle \boldsymbol{\epsilon}^o \boldsymbol{\epsilon}^{oT} \rangle. \quad (2.20)$$

14 When \mathbf{B}_x and \mathbf{R} are expressed in this way, we will refer to them as the true background
15 and observation error covariance matrices for the system.

16 The analysis error covariance matrix, \mathbf{A} , can be defined similarly as

$$\mathbf{A} = \langle \boldsymbol{\epsilon}_x^a \boldsymbol{\epsilon}_x^{aT} \rangle. \quad (2.21)$$

17 By substituting equation (2.17) into equation (2.21), \mathbf{A} can be expressed in terms of the

1 background and observation errors as,

$$\mathbf{A} = \langle \langle (\mathbf{I} - \mathbf{KH})\boldsymbol{\epsilon}_x^b + \mathbf{K}\boldsymbol{\epsilon}^o \rangle \rangle \langle \langle (\mathbf{I} - \mathbf{KH})\boldsymbol{\epsilon}_x^b + \mathbf{K}\boldsymbol{\epsilon}^o \rangle \rangle^T \rangle. \quad (2.22)$$

2 It is assumed that the background and observation errors are uncorrelated such that
 3 $\langle \boldsymbol{\epsilon}^o \boldsymbol{\epsilon}_x^{bT} \rangle = 0$ and $\langle \boldsymbol{\epsilon}_x^b \boldsymbol{\epsilon}^{oT} \rangle = 0$. Then substituting equation (2.20) into equation (2.22) gives
 4 the most general expression for the analysis error covariance matrix,

$$\mathbf{A} = (\mathbf{I} - \mathbf{KH})\mathbf{B}_x(\mathbf{I} - \mathbf{KH})^T + \mathbf{K}\mathbf{R}\mathbf{K}^T. \quad (2.23)$$

5 When \mathbf{K} is dependent on the true background and observation error covariance matri-
 6 ces, as defined in equation (2.20), we can simplify \mathbf{A} further. We will show the derivation
 7 here, and in chapters 3 and 4 we will extend the derivation to the VarBC case analyse the
 8 analysis error covariance matrix in the case when the background error covariance matrix
 9 is ill-defined. First, we expand the second brackets in equation (2.23) and factorise the
 10 resulting last three terms by \mathbf{K}^T ,

$$\mathbf{A} = (\mathbf{I} - \mathbf{KH})\mathbf{B}_x + (-\mathbf{B}_x\mathbf{H}^T + \mathbf{K}\mathbf{H}\mathbf{B}_x\mathbf{H}^T + \mathbf{K}\mathbf{R})\mathbf{K}^T. \quad (2.24)$$

11 Substituting $\mathbf{K} = \mathbf{B}_x\mathbf{H}^T(\mathbf{H}\mathbf{B}_x\mathbf{H}^T + \mathbf{R})^{-1}$ from equation (2.11) into the second brackets
 12 in equation (2.24) gives,

$$\mathbf{A} = (\mathbf{I} - \mathbf{KH})\mathbf{B}_x + (-\mathbf{B}_x\mathbf{H}^T + \mathbf{B}_x\mathbf{H}^T(\mathbf{H}\mathbf{B}_x\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{B}_x\mathbf{H}^T + \mathbf{B}_x\mathbf{H}^T(\mathbf{H}\mathbf{B}_x\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{R})\mathbf{K}^T. \quad (2.25)$$

13 Factorising out the the last two terms in the second brackets leaves,

$$\mathbf{A} = (\mathbf{I} - \mathbf{KH})\mathbf{B}_x + (-\mathbf{B}_x\mathbf{H}^T + \mathbf{B}_x\mathbf{H}^T(\mathbf{H}\mathbf{B}_x\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{H}\mathbf{B}_x\mathbf{H}^T + \mathbf{R}))\mathbf{K}^T. \quad (2.26)$$

1 Therefore, cancelling out the terms multiplied by \mathbf{K}^T leaves the optimal analysis error
 2 covariance matrix, given by

$$\mathbf{A}_{\text{opt}} = (\mathbf{I} - \mathbf{KH})\mathbf{B}_x \quad (2.27)$$

3 Equation (2.27) shows that the analysis error covariance matrix in an optimal system will
 4 always have elements that are reduced from the equivalent elements in the background
 5 error covariance matrix by the term $-\mathbf{KH}$, as the analysis is a combination of the back-
 6 ground and the observations.

7 So far we have only studied the 3-dimensional case, which is where observations from
 8 one point in time are compared to a model at the same time. An equivalent state analysis
 9 can be formed for the 4-dimensional case, which compares observations with a model
 10 across a time window in order to get the best fit of the analysis in the whole window
 11 instead of at just one time step. The cost function for 4DVar (4-dimensional variational
 12 assimilation) is given by,

$$J(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^T \mathbf{B}_x^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \frac{1}{2} \sum_{i=0}^N (\mathbf{y}_i - h_i(\mathbf{x}_i))^T \mathbf{R}_i^{-1}(\mathbf{y}_i - h_i(\mathbf{x}_i)), \quad (2.28)$$

13 where \mathbf{x}_0 is the state at the initial time of the window; i is the observation time; $N + 1$
 14 is the total number of time steps where states are observed; and \mathbf{x}_i is the state at time i
 15 given by,

$$\mathbf{x}_i = m_{i-1}(\mathbf{x}_{i-1}) = m_{0 \rightarrow i}(\mathbf{x}_0), \quad (2.29)$$

16 where $m_{i-1}(\cdot)$ is the model that evolves the state from time $i - 1$ to time i and $m_{0 \rightarrow i}(\cdot)$ is
 17 the model that evolves the state from the initial time to time i , given by,

$$m_{0 \rightarrow i}(\mathbf{x}_0) = m_{i-1}(m_{i-2}(\dots m_0(\mathbf{x}_0)\dots)). \quad (2.30)$$

18 In equation (2.28), the summation has been taken over independent observation times.

1 This means we are assuming that the observations at different times are uncorrelated
2 (Thépaut et al., 1993). Many operational centres now use 4DVar (Gustafsson et al., 2018;
3 Kwon et al., 2018) to numerically estimate the state analysis, but here we will explicitly
4 minimise equation (2.28) to understand how the errors in the background and observations
5 can contaminate the analysis.

6 To minimise equation (2.28), we first take the gradient of the cost function with respect
7 to the initial state \mathbf{x}_0 ,

$$\nabla J = \mathbf{B}_x^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \sum_{i=0}^N \mathbf{M}_0^T \dots \mathbf{M}_{i-1}^T \mathbf{H}_i^T \mathbf{R}_i^{-1} (\mathbf{y}_i - h_i(m_{0 \rightarrow i}(\mathbf{x}_0))), \quad (2.31)$$

8 where \mathbf{M}_i^T for $i = 0, \dots, N$ is the transpose of the Jacobian of $m_i(\cdot)$, known as the adjoint
9 model at time i . Equation (2.31) can be rewritten by writing the summation in vector
10 form to give,

$$\nabla J = \mathbf{B}_x^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \hat{\mathbf{H}}^T \hat{\mathbf{R}}^{-1} \hat{\mathbf{d}}, \quad (2.32)$$

11 where the hat notation denotes a matrix in time and $\hat{\mathbf{H}}$, $\hat{\mathbf{R}}$ and $\hat{\mathbf{d}}$ are given by,

$$\hat{\mathbf{H}} = \begin{pmatrix} \mathbf{H}_0 \\ \mathbf{H}_1 \mathbf{M}_0 \\ \vdots \\ \mathbf{H}_N \mathbf{M}_{N-1} \dots \mathbf{M}_0 \end{pmatrix}, \quad \hat{\mathbf{R}} = \begin{pmatrix} \mathbf{R}_0 & \dots & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & \dots & \mathbf{R}_N \end{pmatrix}, \quad (2.33)$$

$$\hat{\mathbf{d}} = \begin{pmatrix} \mathbf{y}_0 - h_0(\mathbf{x}_0) \\ \mathbf{y}_1 - h_1(m_0(\mathbf{x}_0)) \\ \vdots \\ \mathbf{y}_N - h_N(m_{0 \rightarrow N}(\mathbf{x}_0)) \end{pmatrix}. \quad (2.34)$$

12 Approximating $h_i(m_{0 \rightarrow i}(\mathbf{x}_0))$ using the Taylor expansion around \mathbf{x}_0^b , as in equation

1 (2.7), linearises the innovation vector to give,

$$\hat{\mathbf{d}} = \begin{pmatrix} \mathbf{y}_0 - (h_0(\mathbf{x}_0^b) + \mathbf{H}_0(\mathbf{x}_0 - \mathbf{x}_0^b)) \\ \mathbf{y}_1 - (h_1(m_0(\mathbf{x}_0^b)) + \mathbf{H}_1\mathbf{M}_0(\mathbf{x}_0 - \mathbf{x}_0^b)) \\ \vdots \\ \mathbf{y}_N - (h_N(m_{0 \rightarrow N}(\mathbf{x}_0^b)) + \mathbf{H}_N\mathbf{M}_{N-1}\dots\mathbf{M}_0(\mathbf{x}_0 - \mathbf{x}_0^b)) \end{pmatrix} \quad (2.35)$$

2 By substituting equation (2.35) into equation (2.32) and setting equation (2.32) to zero
3 gives the state analysis for the 4-dimensional case,

$$\mathbf{x}_0^a = \mathbf{x}_0^b + (\mathbf{B}_x^{-1} + \hat{\mathbf{H}}^T \hat{\mathbf{R}}^{-1} \hat{\mathbf{H}})^{-1} \hat{\mathbf{H}}^T \hat{\mathbf{R}}^{-1} \hat{\mathbf{d}}^b, \quad (2.36)$$

4 where $\hat{\mathbf{d}}^b$ is equation (2.34) when $\mathbf{x}_0 = \mathbf{x}_0^b$. Using the Sherman-Morrison-Woodbury
5 formula to rewrite $(\mathbf{B}_x^{-1} + \hat{\mathbf{H}}^T \hat{\mathbf{R}}^{-1} \hat{\mathbf{H}})^{-1} \hat{\mathbf{H}}^T \hat{\mathbf{R}}^{-1}$ as $\mathbf{B}_x \hat{\mathbf{H}}^T (\hat{\mathbf{H}} \mathbf{B}_x \hat{\mathbf{H}}^T + \hat{\mathbf{R}})^{-1}$ as in equation
6 (2.11) gives,

$$\mathbf{x}_0^a = \mathbf{x}_0^b + \mathbf{B}_x \hat{\mathbf{H}}^T (\hat{\mathbf{H}} \mathbf{B}_x \hat{\mathbf{H}}^T + \hat{\mathbf{R}})^{-1} \hat{\mathbf{d}}^b, \quad (2.37)$$

7 which is the form of the 4-dimensional state analysis equation that we will use in this
8 thesis.

9 Note that in 4DVar, \mathbf{B}_x is defined as a static matrix, but can be considered to be
10 quasi-static, as the implied covariances are given by $\mathbf{M} \mathbf{B}_x \mathbf{M}^T$. This means that although
11 at the beginning of the window, \mathbf{B}_x is given by the static background error covariance
12 matrix, it will evolve throughout the window according to the model dynamics. This is
13 an important advantage of 4DVar to 3DVar.

14 2.2 Observing network

15 Data assimilation requires a range of observations to give global coverage and to give
16 information on many variables in a dynamical system. Observations can come from, for

1 example, satellites, land surface measurements, marine measurements, upper air mea-
2 surements and ground-based radar, respectively giving observations of radiances, air/sea
3 temperatures, humidity, wind speed/direction, pressure and rainfall, to name a few.

4 A large part of the observing network is made up of satellite observations, which have
5 made a large contribution to the improvement of NWP forecasts (English et al., 2013;
6 Eyre et al., 2022). There are two types of Earth observation orbits: polar orbiting/LEO
7 (low Earth orbiting) satellites and geostationary satellites (Eyre, 2000). Polar orbiting
8 satellites follow an orbit that crosses both polar regions. They are relatively close to the
9 Earth, normally about 850 km above the surface, so have high horizontal resolution, but
10 are only able to get one or two observations of a particular point per day. Geostationary
11 satellites rotate at the same speed as the Earth, which means they always observe the
12 same location, so they have great temporal resolution. However, to rotate at the same
13 speed, geostationary satellites must be approximately 35800km away from the surface
14 which means that they have a lower horizontal resolution than polar orbiting instruments
15 with the same field of view.

16 Satellites provide various information, which can either be heavily pre-processed to
17 extract model variables such as temperature and humidity, or can be taken as raw mea-
18 surements. Some examples of the preferred data types for assimilation options, as reviewed
19 in Eyre (1997) and updated in Eyre et al. (2022) are:

- 20 • passive temperature/humidity soundings (as radiances);
- 21 • wind information, in the form of atmospheric motion vectors (AMVs) where features
22 are tracked using geostationary satellites;
- 23 • passive microwave imagery which give information on water vapour, cloud water and
24 ice, precipitation and wind speed, but in recent years there has been a move towards
25 assimilating microwave imagery as radiances;
- 26 • passive visible/infra-red imagery, primarily giving information on water vapour;

- radio occultation (RO), known as GPS-RO, which retrieves the refractivity or bending angle, which can be related to meteorological variables, as described below in equation (2.39).

In this thesis we will focus on satellite radiances and radio occultation.

In order to compare radiance observations with model states such as temperature and pressure, the observation operator, $h(\mathbf{x})$, is a form of the radiative transfer equation, which models the radiative processes along an optical path, thus predicting the model's version of the radiance observations. An approximate radiative transfer equation that is used to model the atmosphere upwelling radiance $L(\nu, \theta)$ at a frequency ν and viewing angle θ , neglecting scattering effects, is given by,

$$L(\nu, \theta) = (1 - N)L^{\text{Clr}}(\nu, \theta) + NL^{\text{Cld}}(\nu, \theta) \quad (2.38)$$

where N is the fractional cloud cover (assumes unit cloud top emissivity), and $L^{\text{Clr}}(\nu, \theta)$ and $L^{\text{Cld}}(\nu, \theta)$ are the clear sky and fully cloudy top of atmosphere upwelling radiances. The clear sky top of atmosphere upwelling radiance data is a function made up of radiances from the surface and radiances emitted by the atmosphere, which are dependent on frequency, viewing angle, mixing ratios, surface emissivity and temperature. The fully cloudy top of atmosphere upwelling radiance is a function of frequency, viewing angle, temperature and cloud top temperature (Eyre, 1991; Eyre & Woolf, 1988; Saunders et al., 1999, 2018). The combined radiance observations from IASI (Infrared Atmospheric Sounding Interferometer), AIRS (Atmospheric Infrared Sounder) and AMSU-A (Advanced Microwave Sounding Unit-A) instruments have the largest observation impact on numerical weather prediction systems compared to any other observation type (Lorenc & Marriott, 2014). Radiance observations give global coverage, with observations of the same location generally available twice daily (*Met Office NRT Quality Monitoring*, 2023).

RO uses both GPS (global positioning system) and LEO satellites. Data is received

1 from the LEO satellite whenever a GPS satellite rises or sets such that its ray path can
 2 traverse the Earth’s atmospheric limb. The ray is refracted through the atmosphere and
 3 can be characterised by the bending angle, as shown in figure 2.1, which can be transformed
 4 to a vertical profile of refractive indexes (Kursinski et al., 1997). The refractive index and
 5 refractivity can then be related to meteorological variables by the following approximation
 6 (Eyre et al., 2022),

$$N = \kappa_1 \frac{p}{T} + \kappa_2 \frac{e}{T^2} + \kappa_3 \frac{n_e}{f^2} + \kappa_4 W \quad (2.39)$$

7 where $N = (n - 1) \times 10^6$ is the refractivity with n the refractive index; p is pressure; T is
 8 temperature; e is the water vapour pressure; n_e is the electron density; f is the frequency;
 9 W is the liquid water density; and κ_i for $i = 1, \dots, 4$ are empirically derived coefficients.
 10 RO provides information on temperature for the stratosphere and upper troposphere and
 11 humidity for the lower troposphere, due to different terms in equation (2.39) becoming
 12 more/less significant at different atmospheric levels.

13 Figure 2.2 shows the horizontal distribution of all assimilated GPS-RO data for a given
 14 6 hour 4DVar assimilation window, which has been taken from *ECMWF Geographical*
 15 *Coverage* (2023). The observations are distributed fairly evenly across the horizontal
 16 domain, with good coverage over both the land and ocean. As the data comes from a 6
 17 hour window, this shows that the majority of the Earth is observed at least twice every 12
 18 hours, but the observation time for a particular location depends on when the satellites
 19 are in the correct position for that location.

20 Another important source of observation data used in NWP is radiosondes, which have
 21 been used in NWP for over seven decades and provide upper air temperature, pressure,
 22 humidity and wind measurements (Sun et al., 2013). There are approximately 800 land
 23 stations that report temperature measurements regularly, which usually give data either
 24 once or twice a day (Ingleby, 2017). There is also some coverage over the oceans, when
 25 radiosondes are released from ships, but this is very limited. The density of radiosondes

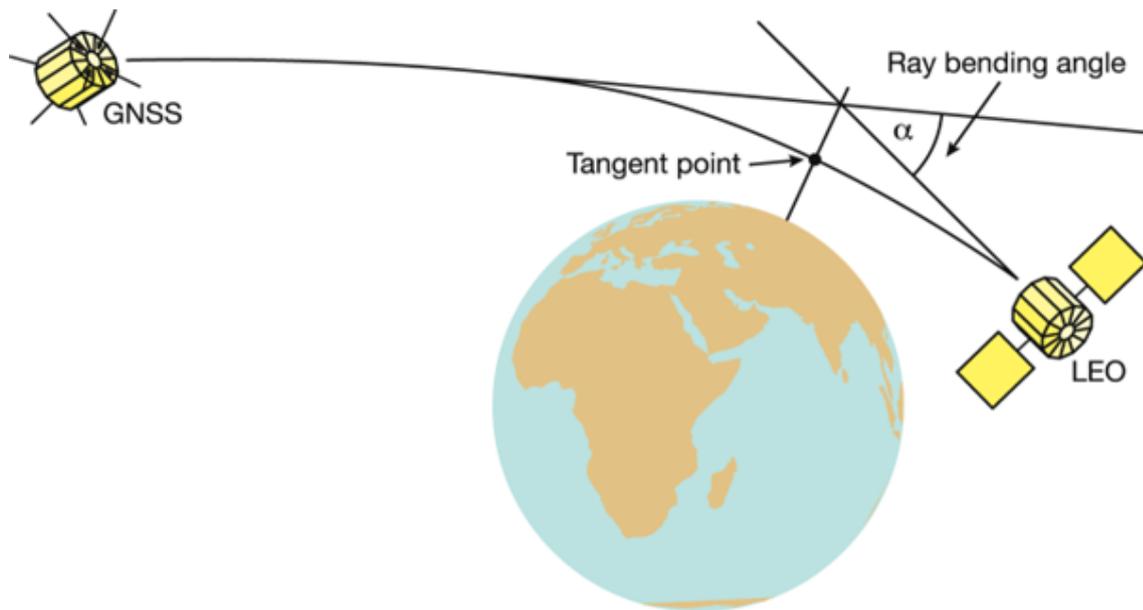
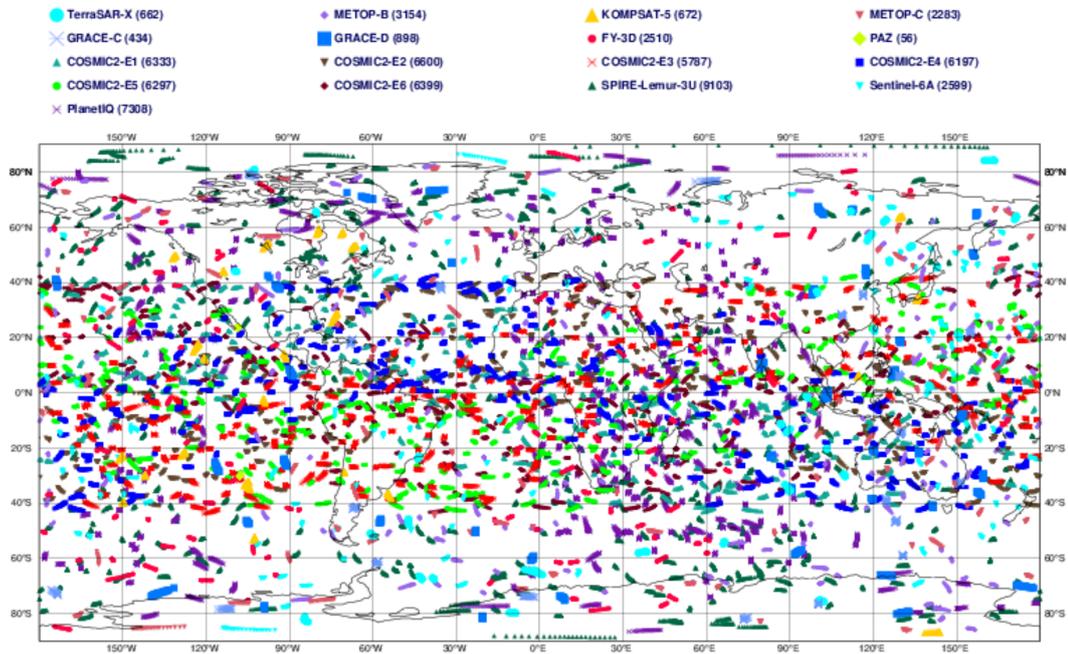


Figure 2.1: A schematic diagram of an LEO (low earth orbiting) satellite receiving a signal from the GNSS (global navigation system satellite) that has been refracted in the atmosphere. Note that the bending angle has been exaggerated for the illustration. Figure from Gleisner et al. (2022) under the Creative Commons Attribution 4.0 International License: <http://creativecommons.org/licenses/by/4.0/>

Geographical coverage

ECMWF data coverage (all observations) - GPSRO
2023101721 to 2023101803
Total number of obs = 67292



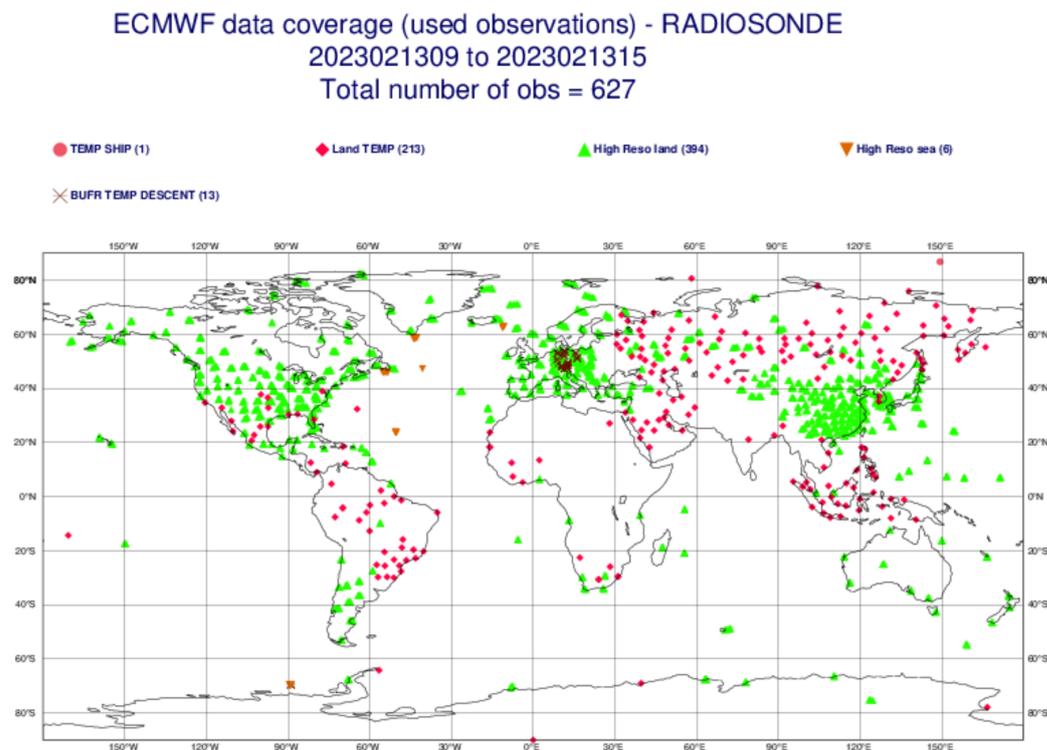
© 2023 European Centre for Medium-Range Weather Forecasts (ECMWF)
Source: www.ecmwf.int
Created at 2023-10-18T15:17:35.167Z



Figure 2.2: Geographical coverage of GPS-RO data on 18th October 2023 at 00UTC. Figure from ECMWF website, accessed on 18th October 2023 (ECMWF Geographical Coverage, 2023)

1 increases in the northern extratropics, as can be seen in figure 2.3 which shows the tem-
 2 perature radiosonde data coverage from 5 radiosonde types from ECMWF on the 13th
 3 February 2023 at 12UTC. This shows the disparity between the radiosonde data coverage
 4 between the land and the ocean and thus between the northern and southern hemispheres.
 5 There are also large gaps in the radiosonde coverage across Africa, south-western Asia and
 6 Australia, which can be due to, for example, less funding for meteorological observations,
 7 less populated areas, or difficult to access land coverage.

Geographical coverage



© 2020 European Centre for Medium-Range Weather Forecasts (ECMWF)
 Source: www.ecmwf.int
 Licence: CC-BY-4.0 and ECMWF Terms of Use(<https://apps.ecmwf.int/datasets/licences/general/>)



Figure 2.3: Geographical coverage of used radiosondes that gave temperature data on 13th February 2023 at 12UTC. Figure from ECMWF website, accessed on 14th February 2023 (ECMWF Geographical Coverage, 2023).

1 Errors arise across all observation types and can be categorised into random and sys-
2 tematic error; a systematic error (bias) is an error that does not average out to zero,
3 whereas a random error averaged across many realisations would give zero. Errors can
4 occur due to damaged, old or limited observations, approximations in the observation
5 operator and representativity between the state and observation. Random error can be
6 accounted for within a data assimilation system through the use of \mathbf{R} , the observation er-
7 ror covariance matrix, as discussed in section 2.1, but biases cause an additional problem
8 to the DA system, as DA theory is based on the assumption that observation errors are
9 unbiased.

10 Radiance observations have a large impact on the NWP DA system, but also contain
11 significant biases (Dee & Uppala, 2009). Radiance biases can be due to problems with an
12 instrument’s calibration, deficiencies in the instrument, and due to the assumptions made
13 in the radiative transfer equation, to estimate variations of equation (2.38). The magni-
14 tudes of the biases vary across instruments and channels, but can be up to approximately
15 3.5K, when transferred to the model variable temperature (Saunders et al., 2013). By
16 comparing the observations with the background state, it has been found that the bias
17 can be parameterised into three structures: the scan bias, the air mass thickness bias,
18 and the orbital bias. The edges of a scan tend to have a larger bias compared to the
19 centre, as can be seen in figure 2.4, which is from Harris and Kelly (2001) and shows the
20 uncorrected scan bias across the satellite’s swath in different latitudinal bands. This bias
21 occurs as the radiation measurements across a satellite’s swath are different, even if view-
22 ing a horizontally homogeneous atmosphere, due to different solar radiation effects such
23 as, reflection off the clouds and surface, and scattering in the upper atmosphere (Wark,
24 1993). Biases that occur due to mis-representing the physics of the atmosphere have been
25 shown to be correlated with the air mass thicknesses, as different air mass thicknesses are
26 a good representation of different parts of the state, such as temperature or humidity (He
27 et al., 2016). Finally, the SSMIS (Special Sensor Microwave Imager/Sounder) and MWRI

1 (Microwave Radiation Imager) instruments suffer from specific instrument biases, which
 2 are caused by the open design of the conical scanners. Although this leads to errors due
 3 to poor calibration and due to solar intrusion (solar radiation hitting the inside of the
 4 instrument and therefore affecting the observations), the periodical nature of the errors
 5 means that the general structure of the biases remains the same across each orbit, hence
 these biases are known as orbital biases (Booton et al., 2014).

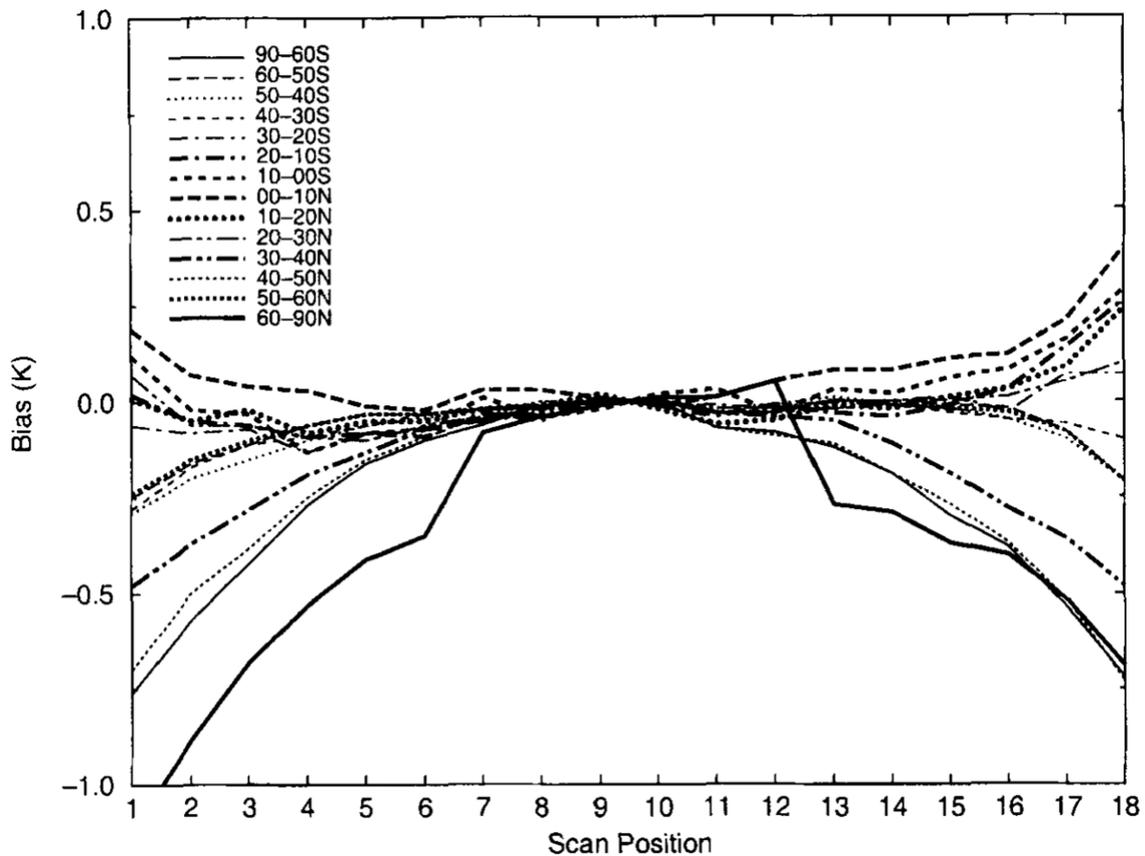


Figure 2.4: Uncorrected scan bias by latitude band, NOAA 12, MSU 2, February 1997. Figure from Harris and Kelly (2001), copyright ©2001 John Wiley & Sons, Ltd

6

7 RO measurements in the lower atmosphere contain cold biases due to the large refrac-
 8 tivity gradients in the lower troposphere and to the spatial distribution of water vapour
 9 (Ao et al., 2003). RO measurements above 5km do not need to be bias-corrected as any

1 biases occurring in the bending angles are small in comparison to other biases within the
2 DA system (Healy & Thépaut, 2006; Cucurull et al., 2014).

3 Radiosondes also contain biases in their raw data, which are mainly caused by radiative
4 effects; a warm bias is present during the day due to sunlight heating the sensor and a
5 cold bias is present at night as the sensor emits longwave radiation. There are also some
6 smaller biases caused by lags in sensor response to changing temperatures as the radiosonde
7 rises. An attempt to correct these biases at each radiosonde site is made, based on simple
8 algorithms given by the radiosonde manufacturers and meteorological agencies (Sun et
9 al., 2013). Therefore, radiosondes are considered to be unbiased when used in the data
10 assimilation system.

11 **2.3 Observation bias correction**

12 Radiance observations have the largest impact in reducing forecast error in a data assimi-
13 lation compared to other observations, but only if their biases can be corrected (Cardinali,
14 2009). There are broadly three types of observation bias correction technique. The first is
15 a static bias correction, which is currently used to correct some scan biases and has been
16 used to correct air mass biases. It works by estimating the bias by comparing observa-
17 tions and the background states, and correcting the observations using an assumed bias
18 correction function (Harris & Kelly, 2001). The estimated bias is not usually updated at
19 the beginning of each cycle, so if the observation biases vary over time, the bias estimates
20 need to be manually updated.

21 The next observation bias correction technique is defined as the “offline” bias correc-
22 tion in Auligné et al. (2007). The bias coefficient is calculated by minimising a similar
23 cost function to equation (2.6), but which is dependent on on a new parameter: β , the
24 bias coefficient, where there would be a unique bias coefficient for each channel of each
25 instrument. This minimisation is performed prior to the estimation of the state analysis.

1 Therefore, like the static technique, the offline correction is dependent on the state back-
 2 ground, but unlike the static technique, is updated at each cycle. This allows the bias
 3 coefficient to react to changes in the observation bias, due to, for example, an instrumen-
 4 tal failure or contamination, but is quite susceptible to contamination from biases in the
 5 background (Eyre, 2016).

6 The final observation bias correction technique is known as VarBC (variational bias
 7 correction), which adaptively calculates the bias coefficient, β , by defining the cost func-
 8 tion, equation (2.6), to be dependent on both the state and the bias coefficient, such
 9 that the bias coefficient is updated at the same time as the state (Dee, 2004). VarBC is
 10 the most common operational observation bias correction technique for satellite radiances
 11 (Gustafsson et al., 2018; Kwon et al., 2018), as it is updated at each cycle and is the least
 12 affected by model bias (Eyre, 2016), so this thesis will only study VarBC further.

13 In VarBC, a bias correction function is added to the observations to be corrected, such
 14 that the k^{th} observation type (for example, an individual channel of a particular satellite
 15 instrument) is estimated to be,

$$\mathbf{y}_k = h_k(\mathbf{x}) + \mathbf{e}^o + c_k(\mathbf{x}^b) \quad (2.40)$$

16 where \mathbf{y}_k is the k^{th} observation type; h_k is the corresponding observation operator; \mathbf{x} is
 17 the state; \mathbf{e}^o is the random error in the observation; c_k is the bias correction function.
 18 Operationally, the bias correction function for the k^{th} observation type is given by

$$c_k(\mathbf{x}^b) = \mathbf{s}_k + \sum_{i=1}^{r_k} \beta_{k,i} p_{k,i}(\mathbf{x}^b) \quad (2.41)$$

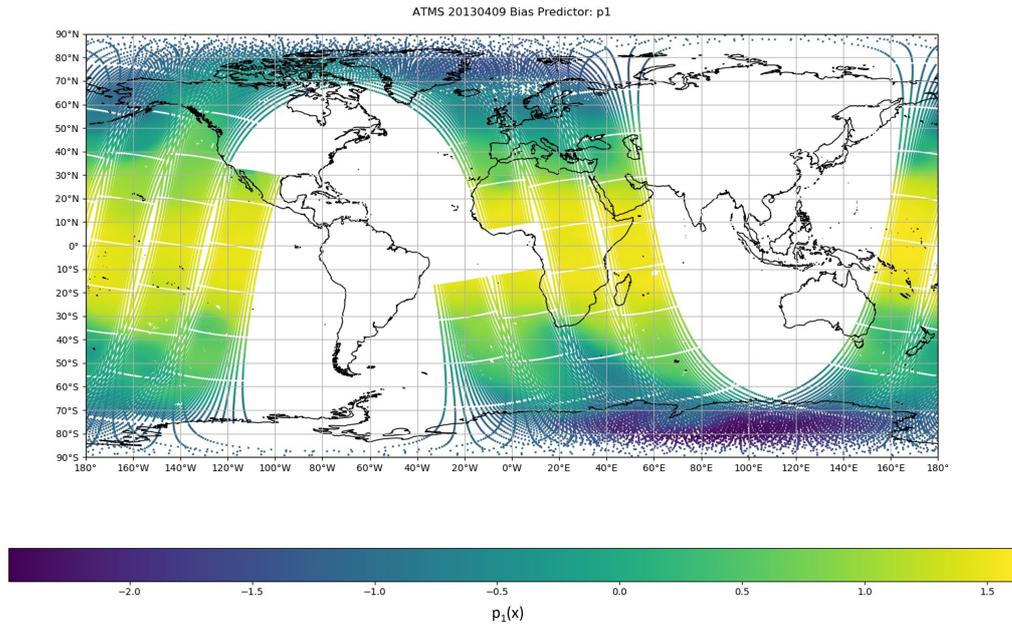
19 where s_k is constant per scan angle and is used to partially describe the scan angle bias;
 20 $p_{k,i}$ are the r_k predictors used for the k^{th} observation; and $\beta_{k,i}$ are the corresponding
 21 observation bias correction coefficients that are updated at each cycle in VarBC for each
 22 observation type (Cameron & Bell, 2018). The vector $\beta \in \mathbb{R}^r$ is defined to hold all

1 values of $\beta_{k,i}$ and the function $c(\mathbf{x}, \boldsymbol{\beta}) \in \mathbb{R}^{m_1}$ is defined to hold the bias correction for all
2 observations that are bias-corrected, where m_1 is the number of biased observations. For
3 simplicity, we will refer to $\boldsymbol{\beta}$ as the observation bias coefficient.

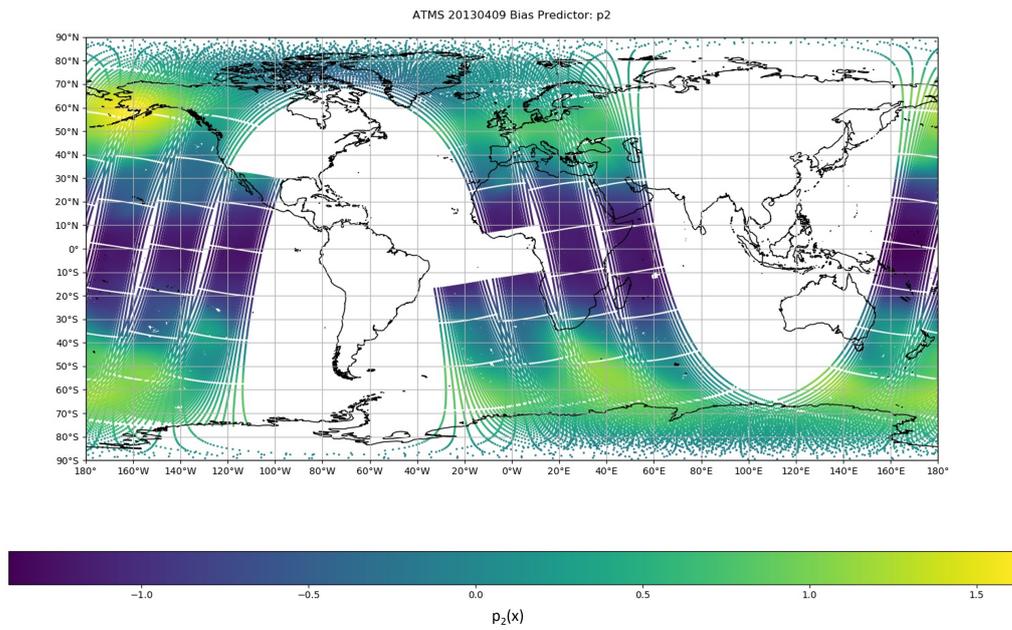
4 Generally, the following predictors are used operationally: air mass thicknesses, which
5 are the differences in height between different pressure levels; Legendre polynomials to
6 describe the spatial variation of bias across a scan; Fourier series to describe orbital bias
7 (only used for SSMIS and MWRI); and a constant predictor which is not dependent on
8 location, which is given by $p_{k,0} = 1$ (Cameron & Bell, 2018; Di Tomaso & Bormann, 2011;
9 Dee & Uppala, 2009).

10 In figure 2.5 we have reproduced a global map from the Met Office from the ATMS
11 instrument on the 9th April 2013, of the air mass thickness predictors p_1 and p_2 , which are
12 the thicknesses between 850-300hPa (figure 2.5a) and 200-50hPa (figure 2.5b) respectively.
13 Air mass thicknesses are a function of temperature, so as the temperature increases, the
14 thicknesses will increase. It has been shown that there are high correlations between
15 radiance biases and some air mass thicknesses, reflecting the channel weighting functions
16 and the background error correlations (Harris & Kelly, 2001). Figures 2.5a and 2.5b show
17 that, in general, as one thickness increases, the other decreases, for example around the
18 equator, the lower altitude thickness bias is positive, whereas the higher altitude thickness
19 bias is negative. However, this relationship is not linear: there are locations where one
20 thickness picks up a bias that the other does not, for example at 60°N, 170°W there is a
21 location of positive bias in figure 2.5b that has zero bias in figure 2.5a.

22 In figure 2.6 we have produced a similar global map, but of p_6 , one of the scan angle
23 bias predictors. In general, satellite swaths will have a larger bias on the edges, compared
24 to the centre, but as the relationship is not as simple as a quadratic curve, the scan
25 angle biases are depicted by a Legendre polynomial, with each predictor representing the
26 different polynomial terms. Figure 2.6 shows the positive values of the predictor on the
27 edges of the swath, which dip to negative in the middle.



(a)



(b)

Figure 2.5: Air mass thickness bias predictors p_1 : 850-300hPa (a) and p_2 : 200-50hPa (b) on world maps for the ATMS (Advanced Microwave Technology Sounder) satellite on the 9th April 2013.

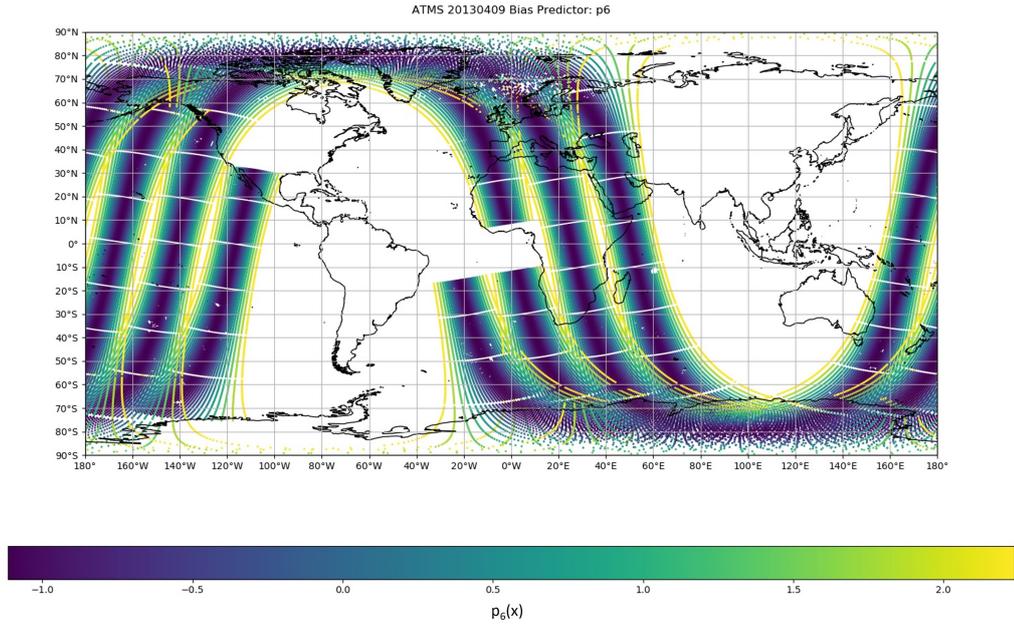


Figure 2.6: Bias predictor p_6 , a scan angle bias predictor.

1 In VarBC a new control vector is defined as the state and the observation bias coeffi-
 2 cient, given by,

$$\mathbf{v} = \begin{pmatrix} \mathbf{x} \\ \boldsymbol{\beta} \end{pmatrix} \in \mathbb{R}^{n+r}, \quad (2.42)$$

3 where n is the size of the state and r is the size of the observation bias coefficient. A new
 4 background error covariance matrix, dependent on both the state and the observation bias
 5 coefficient, is defined as,

$$\mathbf{B}_v = \begin{pmatrix} \mathbf{B}_x & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_\beta \end{pmatrix} \in \mathbb{R}^{(n+r) \times (n+r)}, \quad (2.43)$$

6 where \mathbf{B}_x is the state background error covariance matrix as defined in equation (2.20);
 7 \mathbf{B}_β is the observation bias coefficient background error covariance matrix and is a measure
 8 of how large the observation bias coefficient background errors are. It is assumed that the

1 background error covariances between the state and the observation bias coefficient are
2 zero (Dee, 2004). The observation operator acting on the control vector, which contains
3 the bias correction and the observation operator that transforms the state to observation
4 space, is defined as,

$$h_v(\mathbf{v}) = h(\mathbf{x}) + c(\mathbf{x}^b, \boldsymbol{\beta}), \quad (2.44)$$

5 where the bias correction function is dependent on a previously derived background state,
6 but the observation bias coefficient $\boldsymbol{\beta}$ is variable. The bias correction function $c(\mathbf{x}^b, \boldsymbol{\beta})$
7 will be zero for observations that are not bias-corrected within VarBC.

8 Therefore, the cost function for VarBC is given by,

$$J(\mathbf{v}) = \frac{1}{2}(\mathbf{v} - \mathbf{v}^b)^T \mathbf{B}_v^{-1}(\mathbf{v} - \mathbf{v}^b) + \frac{1}{2}(\mathbf{y} - h_v(\mathbf{v}))^T \mathbf{R}^{-1}(\mathbf{y} - h_v(\mathbf{v})), \quad (2.45)$$

9 where \mathbf{v}^b is the background of the state and bias coefficients; \mathbf{y} is the vector that contains
10 all observations; and \mathbf{R} is the observation error covariance matrix, as defined in equation
11 (2.20) (Auligné et al., 2007).

12 In order to explicitly minimise equation (2.45), it is assumed that $h_v(\mathbf{v})$ is approxi-
13 mately linear. The term $h_v(\mathbf{v})$ can be approximated using the Taylor expansion about \mathbf{v}^b
14 by approximating $h(\mathbf{x})$ and $c(\mathbf{x}^b, \boldsymbol{\beta})$ separately,

$$h(\mathbf{x}) \approx h(\mathbf{x}^b) + \mathbf{H}(\mathbf{x} - \mathbf{x}^b) \quad (2.46)$$

$$c(\mathbf{x}^b, \boldsymbol{\beta}) = c(\mathbf{x}^b, \boldsymbol{\beta}^b) + \mathbf{C}_\beta(\boldsymbol{\beta} - \boldsymbol{\beta}^b). \quad (2.47)$$

15 where $\mathbf{C}_\beta \in \mathbb{R}^{m_1 \times r}$ is the Jacobian of $c(\mathbf{x}^b, \boldsymbol{\beta})$ with respect to the observation bias coeffi-
16 cient $\boldsymbol{\beta}$ at \mathbf{x}^b and $\boldsymbol{\beta}^b$ and m_1 is the number of bias-corrected observations. The linearised
17 observation operator and bias correction can be stored in one matrix as,

$$\mathbf{H}_v = \begin{pmatrix} \mathbf{H} & \mathbf{C}_\beta \end{pmatrix} \in \mathbb{R}^{m_1 \times (n+r)}. \quad (2.48)$$

1 Differentiating equation (2.45) and equating it to zero, as in equations (2.8) to (2.12),
 2 gives the minimum of the cost function for VarBC, which we will refer to as the analysis
 3 control vector, given by,

$$\mathbf{v}^a = \mathbf{v}^b + \mathbf{K}_v(\mathbf{y} - h_v(\mathbf{v}^b)), \quad (2.49)$$

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

$$\mathbf{K}_v = \mathbf{B}_v \mathbf{H}_v^T (\mathbf{H}_v \mathbf{B}_v \mathbf{H}_v^T + \mathbf{R})^{-1} \in \mathbb{R}^{(n+r) \times m_1}. \quad (2.50)$$

6 The Kalman Gain matrix, \mathbf{K}_v , can be split into its \mathbf{x} and β parts as follows,

$$\mathbf{K}_v = \begin{pmatrix} \mathbf{K}_x \\ \mathbf{K}_\beta \end{pmatrix}, \quad (2.51)$$

7 where

$$\mathbf{K}_x = \mathbf{B}_x \mathbf{H}^T (\mathbf{H}_v \mathbf{B}_v \mathbf{H}_v^T + \mathbf{R})^{-1} \in \mathbb{R}^{n \times m} \quad (2.52)$$

$$\mathbf{K}_\beta = \mathbf{B}_\beta \mathbf{C}_\beta^T (\mathbf{H}_v \mathbf{B}_v \mathbf{H}_v^T + \mathbf{R})^{-1} \in \mathbb{R}^{r \times m}. \quad (2.53)$$

8 Since \mathbf{B}_v is block diagonal, equation (2.49) can be split into its state and observation
 9 bias coefficient parts:

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}_x(\mathbf{y} - h_v(\mathbf{v}^b)), \quad (2.54)$$

$$\beta^a = \beta^b + \mathbf{K}_\beta(\mathbf{y} - h_v(\mathbf{v}^b)). \quad (2.55)$$

10 Equations (2.49), (2.54) and (2.55) can be used to derive expressions for the errors in
 11 both the state and the observation bias coefficient in an equivalent method to calculating
 12 the state analysis error equation for the general 3D DA case in equations (2.14)-(2.17).

1 The errors in the control vector, state and observation bias coefficient analysis, background
 2 and observations are given by,

$$\boldsymbol{\epsilon}_v^a = \mathbf{v}^a - \mathbf{v}^t = \begin{pmatrix} \boldsymbol{\epsilon}_x^a \\ \boldsymbol{\epsilon}_\beta^a \end{pmatrix} = \begin{pmatrix} \mathbf{x}^a - \mathbf{x}^t \\ \boldsymbol{\beta}^a - \boldsymbol{\beta}^t \end{pmatrix}, \quad (2.56)$$

$$\boldsymbol{\epsilon}_v^b = \mathbf{v}^b - \mathbf{v}^t = \begin{pmatrix} \boldsymbol{\epsilon}_x^b \\ \boldsymbol{\epsilon}_\beta^b \end{pmatrix} = \begin{pmatrix} \mathbf{x}^b - \mathbf{x}^t \\ \boldsymbol{\beta}^b - \boldsymbol{\beta}^t \end{pmatrix}, \quad (2.57)$$

$$\boldsymbol{\epsilon}_v^o = \mathbf{y} - h_v(\mathbf{v}^t) = \mathbf{y} - h(\mathbf{x}^t) - c(\mathbf{x}^t, \boldsymbol{\beta}^t), \quad (2.58)$$

3 where the t superscript denotes the true value of the state, bias coefficient or control
 4 vector respectively. The true bias coefficient is the theoretical vector that would perfectly
 5 correct the observation bias in the bias correction. The control vector analysis equation,
 6 equation (2.56), can be written in terms of the background and observation error equations.
 7 Subtracting the true control vector from both sides of equation (2.49) and adding and
 8 subtracting $h(\mathbf{v}^t)$ in the innovation vector gives,

$$\mathbf{v}^a - \mathbf{v}^t = \mathbf{v}^b - \mathbf{v}^t + \mathbf{K}_v(\mathbf{y} - h_v(\mathbf{v}^t) + h_v(\mathbf{v}^t) - h_v(\mathbf{v}^b)). \quad (2.59)$$

9 Substituting the analysis, background and observation control vector error equations from
 10 equations (2.56)-(2.58) and substituting $\mathbf{v}^t = \mathbf{v}^b - \boldsymbol{\epsilon}_v^b$ into the expanded $h_v(\mathbf{v}^t)$ gives,

$$\boldsymbol{\epsilon}_v^a = \boldsymbol{\epsilon}_v^b + \mathbf{K}_v(\boldsymbol{\epsilon}_v^o + h(\mathbf{x}^b - \boldsymbol{\epsilon}_x^b) + c(\mathbf{x}^b - \boldsymbol{\epsilon}_x^b, \boldsymbol{\beta}^b - \boldsymbol{\epsilon}_\beta^b) - h_v(\mathbf{v}^b)). \quad (2.60)$$

11 Approximating $h(\mathbf{x}^b - \boldsymbol{\epsilon}_x^b)$ and $c(\mathbf{x}^b - \boldsymbol{\epsilon}_x^b, \boldsymbol{\beta}^b - \boldsymbol{\epsilon}_\beta^b)$ using the first order Taylor expansion
 12 around \mathbf{x}^b and $\boldsymbol{\beta}^b$ respectively gives,

$$\boldsymbol{\epsilon}_v^a = \boldsymbol{\epsilon}_v^b + \mathbf{K}_v(\boldsymbol{\epsilon}_v^o - (\mathbf{H} + \mathbf{C}_x)\boldsymbol{\epsilon}_x^b - \mathbf{C}_\beta\boldsymbol{\epsilon}_\beta^b), \quad (2.61)$$

1 where \mathbf{C}_x is the Jacobian of the bias correction with respect to the state \mathbf{x} . In this thesis we
 2 will assume that the errors that come from mis-specifying the state in the bias correction
 3 function are very small, so that we can ignore the \mathbf{C}_x term. Operationally, this is realistic
 4 as the bias correction function (equation (2.41)) only changes over large regions of the
 5 state, so inaccuracies in the state will be negligible. We can therefore assume $\mathbf{C}_x \approx 0$.

6 Rearranging equation (2.61) to factorise the background errors together gives,

$$\boldsymbol{\epsilon}_v^a = (\mathbf{I} - \mathbf{K}_v \mathbf{H}_v) \boldsymbol{\epsilon}_v^b + \mathbf{K}_v \boldsymbol{\epsilon}_v^o. \quad (2.62)$$

7 Equation (2.62) can be separated into its \mathbf{x} and $\boldsymbol{\beta}$ parts using equations (2.56)-(2.58) to
 8 give,

$$\boldsymbol{\epsilon}_x^a = (\mathbf{I} - \mathbf{K}_x \mathbf{H}) \boldsymbol{\epsilon}_x^b - \mathbf{K}_x \mathbf{C}_\beta \boldsymbol{\epsilon}_\beta^b + \mathbf{K}_x \boldsymbol{\epsilon}_v^o, \quad (2.63)$$

$$\boldsymbol{\epsilon}_\beta^a = (\mathbf{I} - \mathbf{K}_\beta \mathbf{C}_\beta) \boldsymbol{\epsilon}_\beta^b - \mathbf{K}_\beta \mathbf{H} \boldsymbol{\epsilon}_x^b + \mathbf{K}_\beta \boldsymbol{\epsilon}_v^o. \quad (2.64)$$

9 Equation (2.63) differs from the state analysis error equation when not using VarBC,
 10 equation (2.17), as equation (2.63) is dependent on both the state background error and
 11 the bias coefficient background error. Taking the expected value of equations (2.63) and
 12 (2.64) gives,

$$\langle \boldsymbol{\epsilon}_x^a \rangle = (\mathbf{I} - \mathbf{K}_x \mathbf{H}) \langle \boldsymbol{\epsilon}_x^b \rangle - \mathbf{K}_x \mathbf{C}_\beta \langle \boldsymbol{\epsilon}_\beta^b \rangle + \mathbf{K}_x \langle \boldsymbol{\epsilon}_v^o \rangle, \quad (2.65)$$

$$\langle \boldsymbol{\epsilon}_\beta^a \rangle = (\mathbf{I} - \mathbf{K}_\beta \mathbf{C}_\beta) \langle \boldsymbol{\epsilon}_\beta^b \rangle - \mathbf{K}_\beta \mathbf{H} \langle \boldsymbol{\epsilon}_x^b \rangle + \mathbf{K}_\beta \langle \boldsymbol{\epsilon}_v^o \rangle. \quad (2.66)$$

13 Equations (2.65) and (2.66) show that the biases in the state and observation bias co-
 14 efficient analyses are dependent on both the biases from the state and observation bias
 15 coefficient backgrounds.

16 As in equation (2.20), assuming that the background errors are unbiased, the control
 17 vector background error covariance matrix can be expressed as the expected value of the

1 background error multiplied by its transpose,

$$\mathbf{B}_v = \langle \boldsymbol{\epsilon}_v^b \boldsymbol{\epsilon}_v^{bT} \rangle. \quad (2.67)$$

2 Assuming that the analysis errors are unbiased, which means from equations (2.65) and
 3 (2.66) that assuming the observations have been correctly bias-corrected and that the
 4 background errors are unbiased, the analysis error covariance matrix for the control vector
 5 can also be expressed as,

$$\mathbf{A}_v = \langle \boldsymbol{\epsilon}_v^a \boldsymbol{\epsilon}_v^{aT} \rangle. \quad (2.68)$$

6 By substituting equation (2.62) into equation (2.68) and following the same method as in
 7 equations (2.22) to (2.23), \mathbf{A}_v can be expressed in terms of the background and observation
 8 errors as,

$$\mathbf{A}_v = (\mathbf{I} - \mathbf{K}_v \mathbf{H}_v) \mathbf{B}_v (\mathbf{I} - \mathbf{K}_v \mathbf{H}_v)^T + \mathbf{K}_v \mathbf{R} \mathbf{K}_v^T. \quad (2.69)$$

9 The analysis error covariance matrix \mathbf{A}_v is a matrix which can be split into its state and
 10 bias coefficient parts as,

$$\mathbf{A}_v = \begin{pmatrix} \mathbf{A}_x & \mathbf{A}_{x\beta} \\ \mathbf{A}_{\beta x} & \mathbf{A}_\beta \end{pmatrix} \in \mathbb{R}^{(n+r) \times (n+r)}, \quad (2.70)$$

11 where \mathbf{A}_x describes the covariances in the state analysis errors; \mathbf{A}_β describes the covari-
 12 ances in the observation bias coefficient analysis errors; and $\mathbf{A}_{x\beta}$ and $\mathbf{A}_{\beta x}$ describe the
 13 covariances between the state analysis errors and the observation bias coefficient analysis
 14 errors. Therefore, in the case where the true background and observation error covari-
 15 ance matrices are used within the system such that \mathbf{K}_v is optimal and following the same
 16 method as for the general DA case (equations (2.24) - (2.27)), the optimal control vector

1 analysis error covariance matrix is given by,

$$\mathbf{A}_{\mathbf{v},\text{opt}} = (\mathbf{I} - \mathbf{K}_{\mathbf{v}}\mathbf{H}_{\mathbf{v}})\mathbf{B}_{\mathbf{v}}. \quad (2.71)$$

2 Equation (2.71) has the same form as equation (2.27), but includes both the state and the
3 observation bias coefficient.

4 The VarBC analysis equations (2.54) and (2.55) can also be calculated in a 4D system,
5 by extending general 3D DA theory to 4D DA theory (equations (2.12) to (2.37)). By
6 defining the observation operator on \mathbf{v} at the i^{th} time step as,

$$h_{v_i}(\mathbf{v}) = h_i(m_{0 \rightarrow i}(\mathbf{x}_0)) + c_i(m_{0 \rightarrow i}(\mathbf{x}_0))^{\text{b}}, \boldsymbol{\beta}, \quad (2.72)$$

7 where $m_{0 \rightarrow i}$ is the model that takes the state from the initial time step to the i^{th} time
8 step as defined in equation (2.29) and both the observation operator transforming \mathbf{x}_0 to
9 observation space and the bias correction function are dependent on time. Note that even
10 if h and c do not explicitly change in time, their evaluation will change in time as \mathbf{x}_i
11 changes in time. Therefore the linearised observation operator and bias correction are
12 given by,

$$\hat{\mathbf{H}} = \begin{pmatrix} \mathbf{H}_0 \mathbf{M}_{0 \rightarrow 0} \\ \vdots \\ \mathbf{H}_N \mathbf{M}_{0 \rightarrow N} \end{pmatrix} \in \mathbb{R}^{(N+1)m_1 \times n}, \quad \hat{\mathbf{C}}_{\beta} = \begin{pmatrix} \mathbf{C}_{\beta_0} \\ \vdots \\ \mathbf{C}_{\beta_N} \end{pmatrix} \in \mathbb{R}^{(N+1)m_1 \times r}, \quad (2.73)$$

13 which can be defined in one matrix as

$$\hat{\mathbf{H}}_{\mathbf{v}} = \left(\hat{\mathbf{H}}, \hat{\mathbf{C}}_{\beta} \right) \in \mathbb{R}^{(N+1)m_1 \times (n+r)}. \quad (2.74)$$

1 Therefore the analysis equations for 4DVarBC (4-dimensional VarBC) are given by

$$\mathbf{x}_0^a = \mathbf{x}_0^b + \hat{\mathbf{K}}_x \hat{\mathbf{d}}_v^b, \quad (2.75)$$

$$\boldsymbol{\beta}^a = \boldsymbol{\beta}^b + \hat{\mathbf{K}}_\beta \hat{\mathbf{d}}_v^b, \quad (2.76)$$

2 where $\hat{\mathbf{K}}_x$ and $\hat{\mathbf{K}}_\beta$ are the time dependent Kalman gain matrices for the state and bias
3 coefficient given by

$$\hat{\mathbf{K}}_x = \mathbf{B}_x \hat{\mathbf{H}} \left(\hat{\mathbf{H}} \mathbf{B}_x \hat{\mathbf{H}}^T + \hat{\mathbf{C}}_\beta \mathbf{B}_\beta \hat{\mathbf{C}}_\beta^T + \hat{\mathbf{R}} \right)^{-1}, \quad (2.77)$$

$$\hat{\mathbf{K}}_\beta = \mathbf{B}_\beta \hat{\mathbf{C}}_\beta^T \left(\hat{\mathbf{H}} \mathbf{B}_x \hat{\mathbf{H}}^T + \hat{\mathbf{C}}_\beta \mathbf{B}_\beta \hat{\mathbf{C}}_\beta^T + \hat{\mathbf{R}} \right)^{-1} \quad (2.78)$$

4 and $\hat{\mathbf{d}}_v^b$ is the time dependent innovation vector given by

$$\hat{\mathbf{d}}_v^b = \begin{pmatrix} \mathbf{y}_0 - h_0(\mathbf{x}_0^b) - c_0(\mathbf{x}_0^b, \boldsymbol{\beta}^b) \\ \mathbf{y}_1 - h_1(m_0(\mathbf{x}_0^b)) - c_1(m_0(\mathbf{x}_0^b), \boldsymbol{\beta}^b) \\ \vdots \\ \mathbf{y}_N - h_N(m_{0 \rightarrow N}(\mathbf{x}_0^b)) - c_N(m_{0 \rightarrow N}(\mathbf{x}_0^b), \boldsymbol{\beta}^b) \end{pmatrix}. \quad (2.79)$$

5 We will use the 4DVarBC analysis equations in chapter 7.

6 In the next section we will discuss the different models used in data assimilation for
7 numerical weather prediction, how uncertainties can occur, and how we can correct for
8 model biases.

9 **2.4 Models used in NWP**

10 Mathematical models in NWP are equations that are used to represent the evolution of the
11 atmosphere. They describe the approximate physics and dynamics of the atmosphere to
12 give information on, for example, temperature, pressure and zonal/meridional velocities,

1 which are known as the model states. Models are used in NWP both to give temporally
2 continuous global information and to create forecasts of the weather for the next days
3 and weeks. The atmospheric dynamics can be approximated numerically by discretising
4 the equations into grid boxes to cover the given region of interest. Their scales can vary
5 from the global, where the Earth's atmosphere is projected onto a grid with grid boxes
6 representing around 10km in length, to regional scales, where smaller regions, such as the
7 UK, are projected onto grids with grid boxes representing 1-3km each (Gustafsson et al.,
8 2018).

9 Most global models solve the compressible non-hydrostatic equations of motion (Walters
10 et al., 2017; Prill et al., 2022; Gustafsson et al., 2018). Using the non-hydrostatic model
11 allows smaller scale dynamics to play a role in the evolution of the model when using a
12 high resolution grid scale (Staniforth & Wood, 2008). Some centres use nested models,
13 which use the same model for both global and regional scales (Gustafsson et al., 2018).
14 There are nested short to medium range forecasts which have high resolution grid boxes:
15 used to give a more detailed representation of smaller scale atmospheric processes as well
16 as to represent surface features such as coastlines and orography. The global configura-
17 tions provide short to medium-range weather forecasts and are used in the nested models
18 as the boundary conditions for the regional models.

19 As is the case with observational error, models also contain sources of uncertainty
20 which can cause both systematic and random errors. In a model simulation, three types
21 of uncertainty can occur: initial condition uncertainty; boundary condition uncertainty;
22 and model uncertainty. When the model is very sensitive to the initial conditions, a small
23 perturbation from the true value changes the model drastically. Data assimilation helps
24 to reduce initial condition uncertainty, as it compares the most recent model forecast with
25 the latest observations to provide more accurate initial conditions. Boundary condition
26 uncertainty occurs both at the boundaries of regional models, as only a limited domain is
27 modelled, as well as on the boundaries between atmosphere and the land/ocean. Model

1 uncertainty stems from both structural and parametric uncertainty. As parameters vary,
2 sometimes the solution can bifurcate to an entirely different behaviour. Therefore, as-
3 sumptions made in the parameters of models, or in the models themselves can lead to
4 both systematic and random model error as the model variables and their dynamics are
5 misrepresented.

6 Some amount of random model error can be accounted for within the system by us-
7 ing perturbations to atmospheric state variables to give an ensemble of forecasts. The
8 perturbations are amplified by chaotic processes, resulting in forecasts that diverge from
9 each other and thus give a better understanding of the model error present in forecasts
10 (Slingo & Palmer, 2011). The use of ensemble systems will not be discussed in this thesis.
11 However, systematic model error is more difficult to account for and can arise due to a
12 variety of reasons, with some examples given below.

13 There are known temperature and humidity biases in the upper troposphere/lower
14 stratosphere. Temperature biases have been observed at the ECMWF (European Centre
15 for Medium-Range Weather Forecasts) and Met Office in the extratropical lowermost
16 stratosphere to be within 0.3K in short forecasts of less than 24 hours (Dyroff et al.,
17 2015; Carminati et al., 2019), but if left to run, this bias will grow. Moist biases have been
18 found in both the IFS (Integrated Forecasting System - ECMWF) and MetUM (Met Office
19 Unified Model) in the lowermost stratosphere, with a maximum of approximately 170%
20 of the observed values (Bland et al., 2021). Humidity biases will also cause collocated
21 cold biases due to the additional radiative cooling (Forster & Shine, 2002; Maycock et al.,
22 2011).

23 At the extratropical tropopause there are sharp vertical gradients of water vapour, po-
24 tential vorticity and temperature that the NWP models struggle to resolve, which creates
25 a moist bias in the lower stratosphere, leading to a cold bias (Krüger et al., 2022; Stenke
26 et al., 2008). There is some uncertainty in the cause of the moist bias, but explanations in-
27 clude: the misrepresentation of dynamical transport and mixing processes; and numerical

1 diffusion and insufficient model resolution in the semi-Lagrangian advection scheme (al-
2 though a moist bias of similar order is also found in Eulerian formulated models) (Krüger
3 et al., 2022).

4 Temperature biases have also been shown to be caused by the under-representation of
5 aerosols in NWP models. There tends to be a cold bias in regions with a high concentration
6 of aerosols, for example over India and China, due to anthropogenic activity. There also
7 exists aerosol-caused warming effects, which tend to be over remote areas or oceans, where
8 aerosol-radiation interaction or aerosol-cloud interaction seem to have large effects on the
9 temperature (Huang & Ding, 2021).

10 The air that flows over and around hills and mountains causes a drag on the atmo-
11 sphere, and is another large source of uncertainty in NWP due to the governing role the
12 orographic drag has on the atmosphere’s general circulation. Global models now have
13 sufficient resolution to resolve large-scale mountain waves, which have an impact on the
14 location of midlatitude jets (Brayshaw et al., 2009). However, the drag exerted on the at-
15 mosphere due to subgrid-scale orography, which generates gravity waves that create drag
16 forces on flows up to the stratosphere and mesosphere (Bacmeister, 1993), still needs to
17 be parametrised and is parametrised differently depending on the horizontal scales of the
18 orographic features and across different models. The representation of this orographic
19 drag is still a cause of major uncertainty in NWP models (Elvidge et al., 2019).

20 **2.5 Model bias correction**

21 Strong-constraint data assimilation assumes that the model is perfect (ie. has no error)
22 so that the state at the $i + 1^{\text{th}}$ time step is given by,

$$\mathbf{x}_{i+1}^t = m_i(\mathbf{x}_i^t) \tag{2.80}$$

1 where m_i is the model that evolves the true state \mathbf{x}^t at time i to time $i + 1$, as in equation
 2 (2.29). However, as we have seen that model biases do exist, there are several ways that
 3 they can be corrected. Longer timescale biases within the model climatologies can be
 4 handled by adding constant artificial sources and sinks to the equations of motion (Saha,
 5 1992), whereas methods which allow corrections of shorter timescale model errors use
 6 the difference between analyses and 6-hour forecasts to calculate an explicit correction
 7 (Jean ThiéBaux & Morone, 1990). This thesis will focus on another method known as
 8 WC4DVar (weak-constraint 4-dimensional variational assimilation), in which model error
 9 is accounted for by relaxing the assumption that the model is perfect and calculating the
 10 model error simultaneously with the estimate of the state (Sasaki, 1970; Derber, 1989). We
 11 have chosen to focus on WC4DVar as it has been used operationally to correct for model
 12 bias (Laloyaux et al., 2020a) and is a comparable method to VarBC, as both methods
 13 correct for the biases adaptively within the data assimilation system.

14 WC4DVar is used to estimate model error between time steps in an assimilation win-
 15 dow or forecast. In this formulation of 4DVar, an extra parameter $\boldsymbol{\eta}$ is included within
 16 the variational assimilation to represent model error. Therefore, in the weak-constraint
 17 formulation, equation (2.80) is extended to,

$$\mathbf{x}_{i+1}^t = m_i(\mathbf{x}_i^t) + \boldsymbol{\eta}_{i+1}^t, \quad (2.81)$$

18 where $\boldsymbol{\eta}_{i+1}^t$ is the true model error that comes from evolving the state between time i and
 19 $i + 1$. If the model error is approximately equal at each time step within the assimilation
 20 window such that $\boldsymbol{\eta}_i^t = \boldsymbol{\eta}_{i+1}^t$, then the model error $\boldsymbol{\eta}^t$ describes a bias, such that WC4DVar
 21 can be used to solve model bias. In this thesis, we assume the model error is constant,
 22 such that $\boldsymbol{\eta}^t$ will be referred to as the model bias. As each \mathbf{x}_i can be written in terms
 23 of the state at the previous time step, \mathbf{x}_{i+1} can be written in terms of \mathbf{x}_0 and $\boldsymbol{\eta}^t$ in the

1 following way:

$$\mathbf{x}_{i+1}^t = m_i(m_{i-1}(\dots(m_0(\mathbf{x}_0^t) + \boldsymbol{\eta}^t)\dots) + \boldsymbol{\eta}^t) + \boldsymbol{\eta}^t := \tilde{m}_{0 \rightarrow i+1}(\mathbf{x}_0^t, \boldsymbol{\eta}^t), \quad (2.82)$$

2 where the model taking the state from the initial time step to the $(i+1)^{\text{th}}$ time step when
 3 model bias is present, is a function of both the state and $\boldsymbol{\eta}^t$ and is denoted as $\tilde{m}_{0 \rightarrow i+1}$.

4 In reality, the true model bias is unknown so the model bias is estimated, which is
 5 defined as $\boldsymbol{\eta}$, known as the model bias parameter. The error in $\boldsymbol{\eta}$ is assumed to be
 6 Gaussian with random error, with mean $\mathbf{0}$ and error covariance matrix \mathbf{Q} . In order to
 7 make the best estimate of the state and the best estimate of the model bias parameter, \mathbf{x}_0
 8 and $\boldsymbol{\eta}$ are simultaneously calculated in each assimilation cycle (Wergen, 1992; Zupanski,
 9 1993; Bennett et al., 1996; Vidard et al., 2004). Defining the control vector as a vector
 10 that holds both the state and the model bias parameter, gives,

$$\mathbf{p} = \begin{pmatrix} \mathbf{x}_0 \\ \boldsymbol{\eta} \end{pmatrix} \in \mathbb{R}^{2n}, \quad (2.83)$$

11 where $\mathbf{x}_0 \in \mathbb{R}^n$ is the state at the initial time in the assimilation window; and $\boldsymbol{\eta} \in \mathbb{R}^n$ is
 12 the model bias parameter. Defining the background error covariance matrix as the state
 13 and model bias parameter parts gives,

$$\mathbf{B}_p = \begin{pmatrix} \mathbf{B}_x & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{pmatrix} \quad (2.84)$$

14 where $\mathbf{B}_x \in \mathbb{R}^{n \times n}$ is the state background error covariance matrix as defined in equation
 15 (2.20) and $\mathbf{Q} \in \mathbb{R}^{n \times n}$ is the model error covariance matrix which describes the random
 16 error in the estimate of the model bias parameter. It is assumed that the background
 17 error covariances between the state and the model bias parameter are zero.

18 When the state at a particular time step is given by the model evolution of the initial

1 state plus the model bias parameter, the observation operator is therefore a function of
 2 both the state and the model bias parameter,

$$h_{\mathbf{p}_i}(\mathbf{x}_i) = h_i(\tilde{m}_{0 \rightarrow i}(\mathbf{x}_0, \boldsymbol{\eta})), \quad i = 1, \dots, N, \quad (2.85)$$

$$h_{\mathbf{p}_0}(\mathbf{x}_0), \quad i = 0. \quad (2.86)$$

3 Then the cost function in terms of \mathbf{p} (Laloyaux et al., 2020a) is given by,

$$J(\mathbf{p}) = \frac{1}{2}(\mathbf{p} - \mathbf{p}^b)^T \mathbf{B}_p^{-1}(\mathbf{p} - \mathbf{p}^b) + \frac{1}{2} \sum_{i=0}^N (\mathbf{y}_i - h_i(\mathbf{x}_i))^T \mathbf{R}_i^{-1}(\mathbf{y}_i - h_i(\mathbf{x}_i)). \quad (2.87)$$

Considering \mathbf{x}_i as a function of \mathbf{x}_0 and $\boldsymbol{\eta}$ ($\mathbf{x}_i = \tilde{m}_{\mathbf{p},0 \rightarrow i}(\mathbf{x}_0, \boldsymbol{\eta})$) and approximating $h_i(\mathbf{x}_i)$
 by the first order Taylor expansion around \mathbf{x}_0^b and $\boldsymbol{\eta}^b$ gives,

$$\begin{aligned} h_i(\mathbf{x}_i) \approx & h_i(\tilde{m}_{\mathbf{p},0 \rightarrow i}(\mathbf{x}_0^b, \boldsymbol{\eta}^b)) + \mathbf{H}_{\mathbf{x},i} \mathbf{M}_{i-1} \times \dots \times \mathbf{M}_0 (\mathbf{x}_0 - \mathbf{x}_0^b) \\ & + \mathbf{H}_{\mathbf{x},i} \left(\sum_{j=1}^{i-1} \mathbf{M}_{i-1} \times \dots \times \mathbf{M}_j + \mathbf{I} \right) (\boldsymbol{\eta} - \boldsymbol{\eta}^b), \end{aligned} \quad (2.88)$$

4 where $\mathbf{H}_{\mathbf{x},k}$ and \mathbf{M}_k are the linearised observation operator and model with respect to \mathbf{x}_0
 5 at the k^{th} time step and \mathbf{I} is the identity matrix.

6 Therefore the gradient of the cost function, equation (2.87), is given by,

$$\nabla J = \mathbf{B}_p^{-1}(\mathbf{p} - \mathbf{p}^b) - \hat{\mathbf{H}}_p^T \hat{\mathbf{R}}^{-1} \hat{\mathbf{d}}_p. \quad (2.89)$$

7 The hat symbols denote matrices that hold values from more than one time step. $\hat{\mathbf{d}}_p$
 8 contains the innovation vectors $\mathbf{y}_i - h_i(\mathbf{x}_i)$ for all $i \in [0, N]$ and $\hat{\mathbf{H}}_p$ is the linearised
 9 observation operator with respect to both \mathbf{x} and $\boldsymbol{\eta}$, given by,

$$\hat{\mathbf{H}}_p = \begin{pmatrix} \hat{\mathbf{H}}_{\mathbf{p}_x}, & \hat{\mathbf{H}}_{\mathbf{p}_\eta} \end{pmatrix} \quad (2.90)$$

1 where we have defined

$$\hat{\mathbf{H}}_{\mathbf{p}_x} = \begin{pmatrix} \mathbf{H}_{x,0} \\ \vdots \\ \mathbf{H}_{x,i}\mathbf{M}_{i-1}\dots\mathbf{M}_0 \\ \vdots \\ \mathbf{H}_{x,N}\mathbf{M}_{N-1}\dots\mathbf{M}_0 \end{pmatrix}, \quad (2.91)$$

$$\hat{\mathbf{H}}_{\mathbf{p}_\eta} = \begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{H}_{x,i}(\sum_{j=1}^{i-1}\mathbf{M}_{i-1}\dots\mathbf{M}_j + \mathbf{I}) \\ \vdots \\ \mathbf{H}_{x,N}(\sum_{j=1}^{N-1}\mathbf{M}_{N-1}\dots\mathbf{M}_j + \mathbf{I}) \end{pmatrix}. \quad (2.92)$$

2 Note that $\hat{\mathbf{H}}_{\mathbf{p}_\eta}$ is the linearised observation operator with respect to $\boldsymbol{\eta}$, but, as h_i is a
 3 function of \mathbf{x}_i (which is a function of \mathbf{x}_0 and $\boldsymbol{\eta}$) the linearised observation operator is
 4 calculated via the chain rule, such that $\hat{\mathbf{H}}_{\mathbf{p}_\eta}$ is dependent on \mathbf{H}_x .

5 Using the same method as in equations (2.8) to (2.12), the gradient of the cost function
 6 for WC4DVar can be rearranged to give the analysis vector for both the state and the
 7 model bias estimate,

$$\mathbf{p}^a = \mathbf{p}^b + \mathbf{B}_p \hat{\mathbf{H}}_p^T (\hat{\mathbf{H}}_p \mathbf{B}_p \hat{\mathbf{H}}_p^T + \hat{\mathbf{R}}_p)^{-1} \hat{\mathbf{d}}_p^b. \quad (2.93)$$

8 We denote $\mathbf{B}_p \hat{\mathbf{H}}_p^T (\hat{\mathbf{H}}_p \mathbf{B}_p \hat{\mathbf{H}}_p^T + \hat{\mathbf{R}}_p)^{-1} = \mathbf{K}_p$ as the Kalman gain matrix for WC4DVar.
 9 Equation (2.93) can be split into the state and model bias parameter by first separating
 10 \mathbf{K}_p ,

$$\mathbf{K}_p = \begin{pmatrix} \mathbf{B}_x & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{H}}_{\mathbf{p}_x}^T \\ \hat{\mathbf{H}}_{\mathbf{p}_\eta}^T \end{pmatrix} \left[\begin{pmatrix} \hat{\mathbf{H}}_{\mathbf{p}_x} & \hat{\mathbf{H}}_{\mathbf{p}_\eta} \end{pmatrix} \begin{pmatrix} \mathbf{B}_x & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{H}}_{\mathbf{p}_x}^T \\ \hat{\mathbf{H}}_{\mathbf{p}_\eta}^T \end{pmatrix} + \hat{\mathbf{R}} \right]^{-1}. \quad (2.94)$$

1 Therefore, substituting equation (2.94) into equation (2.93) and splitting \mathbf{p} into its \mathbf{x}_0 and
 2 $\boldsymbol{\eta}$ parts gives the state and model bias parameter analyses,

$$\begin{pmatrix} \mathbf{x}_0^a \\ \boldsymbol{\eta}^a \end{pmatrix} = \begin{pmatrix} \mathbf{x}_0^b \\ \boldsymbol{\eta}^b \end{pmatrix} + \begin{pmatrix} \mathbf{B}_x \hat{\mathbf{H}}_{p_x}^T [\hat{\mathbf{H}}_{p_x} \mathbf{B}_x \hat{\mathbf{H}}_{p_x}^T + \hat{\mathbf{H}}_{p_\eta} \mathbf{Q} \hat{\mathbf{H}}_{p_\eta}^T + \hat{\mathbf{R}}]^{-1} \hat{\mathbf{d}}_p^b \\ \mathbf{Q} \hat{\mathbf{H}}_{p_\eta}^T [\hat{\mathbf{H}}_{p_x} \mathbf{B}_x \hat{\mathbf{H}}_{p_x}^T + \hat{\mathbf{H}}_{p_\eta} \mathbf{Q} \hat{\mathbf{H}}_{p_\eta}^T + \hat{\mathbf{R}}]^{-1} \hat{\mathbf{d}}_p^b \end{pmatrix}. \quad (2.95)$$

3 Laloyaux et al. (2020a) explored the use of these coupled equations in correcting for
 4 systematic errors arising in a simple two-layer quasi-geostrophic channel model. They
 5 suggested that WC4DVar could be used to correct for model biases if the background and
 6 model error covariances have different spatial structures. This is achievable operationally,
 7 as model biases, which have large scales, are more prominent in the stratosphere, and
 8 background (initial condition) errors are more prominent in the troposphere, which have
 9 smaller scales across shorter timescales. As a result of this, and the large amount of radio
 10 occultation data available in these areas, WC4DVar is now implemented at ECMWF in
 11 the stratosphere, reducing temperature biases in the analysis by up to 50% (Laloyaux et
 12 al., 2020b).

13 The error equations for the analyses, backgrounds and observations in WC4DVar are
 14 given by,

$$\boldsymbol{\epsilon}_x^a = \mathbf{x}_0^a - \mathbf{x}_0^t, \quad \boldsymbol{\epsilon}_x^b = \mathbf{x}_0^b - \mathbf{x}_0^t, \quad (2.96)$$

$$\boldsymbol{\epsilon}_i^o = \mathbf{y}_i - h_i(\mathbf{x}_i^t) = \mathbf{y}_i - h_i(\tilde{m}_{0 \rightarrow i}(\mathbf{x}_0^t, \boldsymbol{\eta}^t)), \quad (2.97)$$

$$\boldsymbol{\epsilon}_\eta^a = \boldsymbol{\eta}^a - \boldsymbol{\eta}^t, \quad \boldsymbol{\epsilon}_\eta^b = \boldsymbol{\eta}^b - \boldsymbol{\eta}^t. \quad (2.98)$$

15 In order to write the state and model bias parameter analysis error equations in terms
 16 of the background and observation error equations, we initially need to expand $\hat{\mathbf{d}}_p^b$. For
 17 simplicity we expand $\hat{\mathbf{d}}_p^b$ at the i^{th} time step, but this can easily be extended to the full

1 vector. Initially, we write out the full definition of $\mathbf{d}_{p_i}^b$,

$$\mathbf{d}_{p_i}^b = \mathbf{y}_i - h_i(\tilde{m}_{0 \rightarrow i}(\mathbf{x}_0^b, \boldsymbol{\eta}^b)). \quad (2.99)$$

2 Next we add and subtract the true observation operator at the i^{th} time step,

$$\mathbf{d}_{p_i}^b = \mathbf{y}_i - h_i(\tilde{m}_{0 \rightarrow i}(\mathbf{x}_0^t, \boldsymbol{\eta}^t)) + h_i(\tilde{m}_{0 \rightarrow i}(\mathbf{x}_0^t, \boldsymbol{\eta}^t)) - h_i(\tilde{m}_{0 \rightarrow i}(\mathbf{x}_0^b, \boldsymbol{\eta}^b)). \quad (2.100)$$

3 Substituting equation (2.97) into equation (2.100) and substituting the background state
4 and background model bias parameter for their respective background errors plus their
5 truths (as in equations (2.96) and (2.98) respectively) gives,

$$\mathbf{d}_{p_i}^b = \boldsymbol{\epsilon}_i^o + h_i(\tilde{m}_{0 \rightarrow i}(\mathbf{x}_0^b - \boldsymbol{\epsilon}_x^b, \boldsymbol{\eta}^b - \boldsymbol{\epsilon}_\eta^b)) - h_i(\tilde{m}_{0 \rightarrow i}(\mathbf{x}_0^b, \boldsymbol{\eta}^b)). \quad (2.101)$$

6 Approximating the observation operator using a first order Taylor expansion around the
7 true state and true model bias parameter, and cancelling out the observation operator
8 terms gives $\mathbf{d}_{p_i}^b$ in terms of the background and observation errors,

$$\mathbf{d}_{p_i}^b = \boldsymbol{\epsilon}_i^o - \mathbf{H}_{x,i} \mathbf{M}_{i-1} \times \dots \times \mathbf{M}_0 \boldsymbol{\epsilon}_x^b - \mathbf{H}_{x,i} \left(\sum_{j=1}^{i-1} \mathbf{M}_{i-1} \times \dots \times \mathbf{M}_j + \mathbf{I} \right) \boldsymbol{\epsilon}_\eta^b. \quad (2.102)$$

9 Therefore, by using the definitions of $\hat{\mathbf{H}}_{p_x}$ and $\hat{\mathbf{H}}_{p_\eta}$ from equations (2.91) and (2.92)
10 respectively, the full innovation vector can be written in terms of the observation and
11 background errors as,

$$\hat{\mathbf{d}}_p^b = \hat{\boldsymbol{\epsilon}}^o - \hat{\mathbf{H}}_{p_x} \boldsymbol{\epsilon}_x^b - \hat{\mathbf{H}}_{p_\eta} \boldsymbol{\epsilon}_\eta^b, \quad (2.103)$$

12 where $\hat{\boldsymbol{\epsilon}}^o$ is the vector that holds the observation errors at all times.

13 In order to calculate the state analysis error equation in terms of the background and
14 observation errors, we subtract \mathbf{x}_0^t from both sides of the state part of equation (2.95),

1 and substitute the analysis and background error equations, equation (2.96). Finally, we
 2 substitute equation (2.103) for $\hat{\mathbf{d}}_p^b$. Therefore the state analysis error equation is given by,

$$\boldsymbol{\epsilon}_x^a = \boldsymbol{\epsilon}_x^b + \mathbf{B}_x \hat{\mathbf{H}}_{p_x}^T [\hat{\mathbf{H}}_{p_x} \mathbf{B}_x \hat{\mathbf{H}}_{p_x}^T + \hat{\mathbf{H}}_{p_\eta} \mathbf{Q} \hat{\mathbf{H}}_{p_\eta}^T + \hat{\mathbf{R}}]^{-1} (\hat{\boldsymbol{\epsilon}}^o - \hat{\mathbf{H}}_{p_x} \boldsymbol{\epsilon}_x^b - \hat{\mathbf{H}}_{p_\eta} \boldsymbol{\epsilon}_\eta^b) \quad (2.104)$$

3 Similarly, by subtracting $\boldsymbol{\eta}^t$ from both sides of the model bias parameter part of
 4 equation (2.95), substituting the error equations from equation (2.98) and substituting
 5 equation (2.103) for $\hat{\mathbf{d}}_p^b$, gives the model bias parameter analysis error equation as,

$$\boldsymbol{\epsilon}_\eta^a = \boldsymbol{\epsilon}_\eta^b + \mathbf{Q} \hat{\mathbf{H}}_{p_\eta}^T [\hat{\mathbf{H}}_{p_x} \mathbf{B}_x \hat{\mathbf{H}}_{p_x}^T + \hat{\mathbf{H}}_{p_\eta} \mathbf{Q} \hat{\mathbf{H}}_{p_\eta}^T + \hat{\mathbf{R}}]^{-1} (\hat{\boldsymbol{\epsilon}}^o - \hat{\mathbf{H}}_{p_x} \boldsymbol{\epsilon}_x^b - \hat{\mathbf{H}}_{p_\eta} \boldsymbol{\epsilon}_\eta^b). \quad (2.105)$$

6 Both the state and model bias parameter analysis errors are dependent on the state and
 7 model bias parameter background errors. This shows that, when the system is cycled,
 8 errors in the state are coupled with errors in the model bias parameter.

9 2.6 Summary

10 In chapter 2 we have introduced basic variational data assimilation theory and have demon-
 11 strated how data assimilation combines data from observations and mathematical models.
 12 We have shown that both observations and models are sources for uncertainty, which can
 13 have a significant impact on an accurate estimate of the state if the error is systematic,
 14 rather than random. To combat these biases, we have shown two adaptive bias correc-
 15 tion techniques that are designed to correct for observation and model biases respectively,
 16 by including an additional parameter in the control vector. In chapter 3 we will present
 17 several studies that explore bias correction in data assimilation, which will motivate our
 18 research questions.

1 Chapter 3

2 Motivation for research questions

3 In this thesis we wish to understand how best to correct for observation and model bi-
4 ases, using VarBC and WC4DVar respectively, and how to reduce the contamination of
5 model/observation biases when correcting for observation/model biases by utilising unbi-
6 ased observations. In this chapter we will discuss some important studies in the field of
7 data assimilation to understand where the gaps in knowledge lie, in order to form our re-
8 search questions that will create the basis of this thesis. In section 3.1 we will discuss how
9 the background error covariance for the state and observation bias coefficient are chosen
10 operationally within VarBC, and present a previous study on the impact of mis-specifying
11 the state background error covariance matrix in a general DA system. In section 3.2 we
12 will discuss previous studies on observation/model bias correction and their reliance on
13 unbiased reference data. In section 3.3 we pose our research questions.

14 3.1 Specifying background error covariances in VarBC

15 We have shown in chapter 2 that, in data assimilation, the background error covariance
16 matrix describes the relationships between the errors in the background states. In VarBC
17 the background error covariance matrix is made up of the state and observation bias

1 coefficient background error covariance matrices, as was shown in equation (2.43). Op-
 2 erationally, both the state and observation bias coefficient background error covariance
 3 matrices are only approximations of the true background error covariance matrices of the
 4 system, which means that the optimal background error covariance matrices are not used.
 5 Eyre and Hilton (2013) demonstrated the impact of mis-specifying the state background
 6 error covariance matrix in a data assimilation system without bias correction, but there
 7 have not been any studies on the mis-specification of the background error covariance
 8 matrices in VarBC.

9 In this section we will initially describe how the state and observation bias coefficient
 10 background error covariance matrices are estimated operationally, by following two papers:
 11 Bannister (2008a) and Cameron and Bell (2018). We will then describe the method of
 12 quantifying the impact of mis-specifying the state background error covariance matrix in
 13 Eyre and Hilton (2013), so that we can extend this method to a VarBC system in chapter
 14 4. The discussion of these papers will lead to developing research question 1, which will
 15 be explained in section 3.3.

16 Bannister (2008a) describes the role of \mathbf{B}_x to spread information (both about the
 17 background and observations) horizontally and vertically so that information is shared
 18 across the domain and between variables. In a cycled system, the true \mathbf{B}_x can be calculated
 19 by evolving the analysis error covariance matrix from the previous cycle forward to give
 20 the forecast error covariances using the Kalman filter equations (Kalman, 1960). These
 21 evolve the previous analysis error covariance matrix according to the model and account
 22 for some model error. The forecast error covariances \mathbf{P}^f are given by,

$$\mathbf{P}^f = \mathbf{M}_{0 \rightarrow N} \mathbf{P}^a \mathbf{M}_{0 \rightarrow N}^T + \mathbf{Q}, \quad (3.1)$$

23 where \mathbf{P}^a is the analysis error covariance matrix from the previous cycle; $\mathbf{M}_{0 \rightarrow N}$ is the
 24 Jacobian of the forecast model from the beginning to the end of the assimilation window,

1 as described in equation (2.31); and \mathbf{Q} describes the error variances in the model. If \mathbf{P}^f
2 were known, then it would be the optimal choice for \mathbf{B}_x , given the linearity, Gaussianity
3 and unbiased model property assumptions hold. However, operationally it would be too
4 computationally expensive to calculate \mathbf{P}^f explicitly, as \mathbf{P}^f would require matrix multipli-
5 cations of matrix dimension $n \times n$, where n is the size of the state. In a global system, n
6 could be all state variables (temperature, pressure, humidity etc) across all latitude and
7 longitude values, as well as all vertical values, so n could have a size of approximately 10^9 .
8 Therefore, \mathbf{B}_x is taken to be an approximation of the background error statistics in the
9 system.

10 In Bannister's 2008 review paper (Bannister, 2008a), three methods are described to
11 estimate the background error covariance matrix. The first method relates the innovation
12 vector to the background and observation errors, which means that the background errors
13 of model variables that are directly observed can be approximated (Rutherford, 1972).
14 However, this method is limited to the number of *in-situ* observations (such as radiosondes)
15 available and can only be calculated for background variables that have direct observations.
16 The NMC (National Meteorological Centre) method (Parrish & Derber, 1992) uses the
17 statistics of the differences between pairs of forecasts valid at the same time but taken
18 from different initial times. As the forecast differences are usually averaged over a period
19 of a few months, the NMC method is useful for estimating climatological background
20 error covariances (Bouttier, 1996). However, this method struggles in poorly observed
21 locations as there will be smaller updates to the forecasts in those regions. The ensemble
22 method uses an ensemble of forecasts, to calculate the statistics of the difference between
23 ensemble members (based on Houtekamer et al., 1996). As the ensemble method uses
24 forecasts from the current time it is able to produce a more flow-dependent \mathbf{B}_x than
25 the NMC method. However, it can only provide as many modes of errors as there are
26 ensemble members. Therefore operationally, a mixture or combination of these methods
27 is used to estimate the state background error covariance matrix, with some centres using a

1 hybrid approach that combines a climatological state background error covariance matrix
 2 with a flow-dependent background error covariance matrix that includes 'errors of the
 3 day' (Gustafsson et al., 2018). However, the calculation of the state background error
 4 covariance matrix is, unfortunately, still only an estimate of the true background error
 5 covariance matrix.

6 Unfortunately, there have been far fewer studies into how the observation bias coef-
 7 ficient background error covariance matrix \mathbf{B}_β in VarBC should be calculated. The bias
 8 coefficient background error covariance matrix, \mathbf{B}_β , determines how large the update on
 9 the estimate of β will be, given the uncertainty in the previous estimate. The uncertainty
 10 in the observation bias coefficient can originate from the assumed structure of the obser-
 11 vation biases, as well as the model error impacting the previous estimate. The methods
 12 used to estimate the state background error covariance matrix cannot be used for the
 13 observation bias coefficient background error covariance matrix, as it is assumed that the
 14 observation bias coefficients are roughly constant between assimilation windows, such that
 15 the previous analysis observation bias coefficient is taken to be the background observation
 16 bias coefficient of the subsequent cycle (Dee, 2005). At the Met Office, the method for
 17 approximating \mathbf{B}_β is given in Cameron and Bell (2018). \mathbf{B}_β is approximated as diagonal,
 18 with error variances V_{β_i} defined by,

$$V_{\beta_i}^{-1} = \frac{N_{\text{bgerr}}}{m_1} \sum_{k=1}^{m_1} p_{k,i}^2 R_k^{-1}, \quad (3.2)$$

19 where $p_{k,i}$ are the predictors corresponding to the i^{th} observation bias coefficient for the
 20 k^{th} bias-corrected observation type, as defined in equation (2.41); R_k are the k observa-
 21 tion error variances; m_1 is the number of bias-corrected observations; and N_{bgerr} is the
 22 weighting (divided by m_1) given to the observations when determining the observation
 23 bias coefficient. N_{bgerr} is defined at the Met Office so that the difference between the
 24 initial estimate of β and the best estimate of β halves in h_t cycles, such that N_{bgerr} is

1 given by,

$$N_{\text{bgerr}} = \max(m_{\text{avg}}, m_{\text{min}}) \frac{1}{2^{\frac{1}{h_t}} - 1}, \quad (3.3)$$

2 with m_{avg} the expected number of observations per cycle for the given channel; m_{min}
3 is the minimum number of observations; and h_t is the chosen difference halving time
4 in units of data assimilation cycles. m_{min} is included in case there is a period of low
5 observation numbers, to stop the observation bias coefficient being based on a very small
6 number of observations. In the usual case when $m_{\text{avg}} > m_{\text{min}}$, the difference halving time
7 is determined by h_t (which at the Met Office is 8 DA cycles, corresponding to 2 days).
8 This results in a minimum difference halving time across all observations, which means
9 the observation bias coefficient can be calculated from information from all bias-corrected
10 observations. The calculation of the bias coefficient background error covariance matrix
11 is also clearly an approximation and is defined to limit how much the bias coefficient can
12 vary between each cycle, rather than the true bias coefficient background error covariance
13 matrix.

14 We have shown that it is too computationally expensive to calculate the true back-
15 ground error covariance matrices for the state and observation bias coefficient and that
16 operationally, only assumptions can be used. Eyre and Hilton (2013) showed the impact
17 of mis-specifying the state background error covariance matrix on the analysis error co-
18 variance matrix in a system without bias correction, but as the bias coefficient background
19 error covariance matrix may also be mis-specified, this is clearly a problem that can be
20 extended to the VarBC case. We will describe Eyre and Hilton's method here as we will
21 extend it to include VarBC in chapter 4. Eyre and Hilton (2013) began with the general
22 equation for the analysis error covariance matrix, given by equation (2.23) and repeated
23 here for simplicity:

$$\mathbf{A} = (\mathbf{I} - \mathbf{KH})\mathbf{B}_x(\mathbf{I} - \mathbf{KH})^T + \mathbf{KRK}^T. \quad (3.4)$$

24 If the background error covariance matrix used in the system is the true background error

1 covariance matrix, then the optimal \mathbf{K} is given by,

$$\mathbf{K}(\mathbf{B}_x) = \mathbf{B}_x \mathbf{H}^T (\mathbf{H} \mathbf{B}_x \mathbf{H}^T + \mathbf{R})^{-1}, \quad (3.5)$$

2 where although \mathbf{K} is dependent on \mathbf{B}_x , \mathbf{H} and \mathbf{R} , it is denoted to be a function of \mathbf{B}_x
 3 as this dependency is what is being investigated. Therefore the optimal analysis error
 4 covariance matrix is given by,

$$\mathbf{A}_{\text{opt}} = (\mathbf{I} - \mathbf{K} \mathbf{H}) \mathbf{B}_x, \quad (3.6)$$

5 as was derived in equation (2.27). As the state background error covariance matrix is
 6 usually only an approximation, Eyre and Hilton (2013) consider a sub-optimal case of
 7 equation (3.4), whereby \mathbf{B}_x is an approximation, denoted by \mathbf{B}_x^a . Therefore the sub-
 8 optimal \mathbf{K} , denoted by \mathbf{K}^a would be given by,

$$\mathbf{K}^a(\mathbf{B}_x^a) = \mathbf{B}_x^a \mathbf{H}^T (\mathbf{H} \mathbf{B}_x^a \mathbf{H}^T + \mathbf{R})^{-1} =: \mathbf{K}^a \quad (3.7)$$

9 and thus the sub-optimal analysis error covariance matrix would be given by,

$$\mathbf{A}_{\text{sub}} = (\mathbf{I} - \mathbf{K}^a \mathbf{H}) \mathbf{B}_x (\mathbf{I} - \mathbf{K}^a \mathbf{H})^T + \mathbf{K}^a \mathbf{R} \mathbf{K}^{aT}. \quad (3.8)$$

10 Note that equation (3.8) is now linear in \mathbf{B}_x , as the true \mathbf{B}_x comes from multiplying the
 11 background error in the system with its transpose, whereas \mathbf{B}_x^a in \mathbf{K}^a originates from using
 12 a mis-specified background error covariance matrix in the DA system. By multiplying out
 13 the brackets from the first term in equation (3.8), \mathbf{A}_{sub} can be rewritten in terms of the
 14 optimal analysis error covariance matrix ($\mathbf{A}_{\text{opt}}^a$), which is equal to \mathbf{A}_{opt} in equation (3.6),

1 but using \mathbf{B}_x^a instead of \mathbf{B}_x , plus an additional term,

$$\mathbf{A}_{\text{sub}} = \mathbf{A}_{\text{opt}}^a + (\mathbf{I} - \mathbf{K}^a \mathbf{H})(\mathbf{B}_x - \mathbf{B}_x^a)(\mathbf{I} - \mathbf{K}^a \mathbf{H})^T, \quad (3.9)$$

2 (see also section 4.2) (Watts & McNally, 1988). In order to demonstrate the impact of the
 3 difference term on the sub-optimal state analysis error covariance matrix, Eyre and Hilton
 4 set up a simple numerical system. Their simple experiment consisted of one state variable
 5 which was directly observed by one observation, such that the Jacobian of the observation
 6 operator was $H = 1$, with observation error variance $R = 1$ (H and R are no longer in
 7 bold to show that they are scalar). Their assumed background error variance is constant,
 8 given by $B_A = 1$ and they vary the true background error variance. For each value of
 9 the true background error variance, the optimal and sub-optimal analysis error variances
 10 are plotted, shown in figure 3.1 as blue diamonds and red crosses respectively. Figure 3.1
 11 comes from Eyre and Hilton (2013). When B_A is smaller than the true background error
 12 variance (ie. the right hand side of the figure, $B > 1$), then the background error variance
 13 has been underestimated, so the system risks giving too much weight to the background
 14 state. When B_A is greater than the true background error variance (ie. the left hand side
 15 of the figure, $B < 1$), then the background error variance has been overestimated, so the
 16 system risks giving too much weight to the observations. In figure 3.1, the 1-1 line has
 17 been plotted as the solid black line which shows the value of the analysis error variance if
 18 it were equal to the true background error variance.

19 The process of data assimilation is used in order to make a better estimate of the
 20 state, by combining the observations and background. Therefore, it is assumed that the
 21 analysis error variance will be 'better' (ie. smaller) than the background error variance.
 22 From the equation for \mathbf{A}_{opt} , equation (3.4), this is known to be the case for the optimal
 23 analysis error variance, and is shown in figure 3.1, as A_{opt} is always smaller than the
 24 true background error variance (the black line). However, figure 3.1 shows that there is a

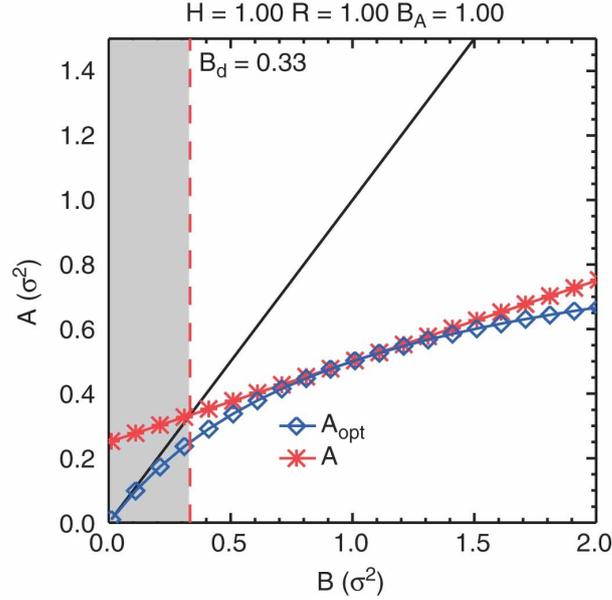


Figure 3.1: Analysis error variance, A , as a function of the true background error variance, B , where the assumed background error variance is $B_A = 1$, and the assumed (and true) observation error variance is $R = 1$. A_{opt} is calculated from equation (3.6) and A is A_{sub} from equation (3.9). The diagonal line through the origin is the line $A = B$. The dashed vertical line shows the value of $B = B_d$, at which the analysis enters the danger zone for A_{sub} . ©British Crown Copyright, the Met Office, from Eyre and Hilton (2013). Reproduced by permission of John Wiley and Sons Inc.

1 range, defined by Eyre and Hilton as the “danger zone”, where A_{sub} is greater than the
 2 true background error variance, shown by the shaded region. In this region, using data
 3 assimilation would actually be detrimental to the estimate of the state and highlights the
 4 danger of overestimating the value of the background error variance.

5 The existence of the danger zone is a cause for concern when estimating the state
 6 background error covariance matrix in a system without bias correction. In VarBC, the
 7 observation bias coefficient background error covariance matrix is also only an estimate of
 8 the true background error covariance matrix, which therefore poses the question of what is
 9 the effect of mis-specifying both the state and the observation bias coefficient background
 10 error covariance matrix in VarBC? This leads to research question 1, given in section 3.3,
 11 and will be studied and discussed in chapter 4, by extending the work by Eyre and Hilton
 12 (2013) which has just been described.

3.2 Correcting for either observation or model bias and the importance of unbiased observations

Standard data assimilation methods assume that there are no biases present in the system. However, as shown in sections 2.2 and 2.4, biases can occur in both observations and in the model. Dee (2005) discusses the complications of having both observation and model biases present. If the source of the bias (ie. observation or model) is unknown, Dee suggests that it could be better to use 'bias-blind' data assimilation, whereby no bias correction technique is used, otherwise the source of the bias may be wrongly attributed, which can lead to a biased analysis. If the source of the bias is known and can be characterised by known parameters, then both observation and model biases can be corrected by including additional parameters into the data assimilation system, as described in chapters 2.3 and 2.5 respectively. The separation of biases requires additional reference information, for example, knowledge of the causes of biases, or independent observations.

Some observations are used to anchor biased observations and/or biased models to the truth, which are known as anchor observations (Eyre, 2016). Anchor observations will not be completely free of biases, but we assume that the biases they contain are significantly smaller than other biases in the system, or have been bias-corrected prior to use in the data assimilation cycle. Currently, several types of observations are used as anchor observations within NWP systems. Radiosondes are used to anchor temperature, pressure and humidity in the troposphere, but as was shown in figure 2.3, these measurements are mostly over the land in the northern hemisphere. RO measurements provide anchor information about temperature in the stratosphere and upper troposphere with good spatial coverage across the Earth. Some radiance observations are 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, or they are bias-corrected prior to use in the DA system. For example, AMSU-A channel 14 is used to anchor the upper stratosphere (Di Tomaso & Bormann, 2011);

1 selected hyperspectral infra-red window channels are used to anchor skin temperature at
2 the Met Office; and selected ozone channels from hyperspectral infrared instruments are
3 used to anchor the upper tropospheric/lower stratospheric ozone analysis at ECMWF
4 (Han & McNally, 2010). However, the assumption that observation biases are small for
5 these observations can be wrong, for example biases in the observation operator may be
6 large (Han & Bormann, 2016).

7 In this section we will explore three studies that discuss correcting for observation
8 and/or model biases: Eyre (2016), Laloyaux et al. (2020a) and Lorente-Plazas and Hacker
9 (2017). In this thesis we are interested in how one bias can be corrected for, when the
10 other bias is present. The first study (Eyre, 2016) demonstrates the importance of unbiased
11 observations when correcting for observation bias in the presence of model bias. We will
12 discuss the results from their scalar theoretical example in this section and, in sections 6, 7
13 and 8, we will further this discussion by extending the theoretical study from a scalar to a
14 vector system. The second study (Laloyaux et al., 2020b) demonstrates when WC4DVar
15 is able to correct for model bias and the effect of correcting for it in the presence of
16 observation bias. In this section we will discuss the results, with particular interest on
17 when both biases are present. In chapter 8 we will extend this by exploring the role of
18 unbiased observations when correcting for model bias in the presence of observation bias.
19 The final study (Lorente-Plazas & Hacker, 2017) demonstrates the ability of a simple
20 numerical system in the presence of both observation and model biases to correct for one
21 or both of the biases. We will discuss their experiments and results, as in chapter 8 we
22 will set up a similar numerical system, but that also includes unbiased observations. We
23 will extend the work of Lorente-Plazas and Hacker (2017) to understand the importance
24 of the unbiased observations when correcting for either type of bias.

25 Eyre (2016) explores the use of anchor observations, as additional reference informa-
26 tion when correcting for observation biases in the presence of model bias in a simple
27 scalar system. By defining the analysis as the weighting between the background, anchor

1 observations and bias-corrected observations, Eyre analytically studied how the anchor
2 observations could limit the contamination of model bias on the analysis in a scalar sys-
3 tem. Eyre (2016) described three simple experiments to test the sensitivity of the fraction
4 of analysis bias to model bias when different weightings and model relaxation rate were
5 used.

6 In the first experiment, the weighting given to the anchor observations was varied,
7 whilst keeping the total weight of both the biased and anchor observations constant.
8 Keeping the total observation weight constant implicitly kept the background weighting
9 constant as the background and observations weights were constrained to sum up to 1.
10 When no weight was given to the anchor observations, the analysis bias was given by
11 the model bias. When no weight was given to the biased observations then the fraction
12 of analysis bias to model bias dramatically reduced. Therefore, this experiment showed
13 that, when less weight is given to the anchor observations, the analysis will have a greater
14 contamination from model bias.

15 In the second experiment, the importance of the weighting of both observations com-
16 pared to the background was tested, by varying the total observation weight and thus
17 reducing the background weight. The anchor and biased observation weights were taken
18 to be equal. When the observation weight was zero, the analysis bias was given by the
19 model bias. When the background weight was zero, the model bias did not influence the
20 analysis bias at all. Therefore, as more weight went towards the background, there was a
21 greater contamination of model bias on the analysis.

22 In the third experiment, the rate that the model relaxed to its climatology was varied.
23 When the model relaxation rate was zero, the analysis bias was independent of the model
24 bias, but as the relaxation rate was increased, the analysis bias became a larger fraction
25 of the model bias.

26 The second and third experiments showed that the contamination of model bias on
27 the analysis would be reduced, if the calculation of observation bias coefficients came from

1 data where model biases were small and slowly varying, or calculated in regions where
2 there are many anchor observations, such as the troposphere in the northern hemisphere.

3 Overall, Eyre (2016) demonstrated that model bias will contaminate the observation
4 bias correction, and that anchor observations play an important role in mitigating the
5 damaging effects of model bias. However, what is not discussed is the role of anchor
6 observations within a non-scalar system.

7 Some research has been carried out to understand how model bias can be corrected
8 using WC4DVar. Laloyaux et al. (2020a) used WC4DVar in a simple numerical system
9 (both with and without observation biases) to correct for model bias and then applied
10 this to the ECMWF system in Laloyaux et al. (2020b). Laloyaux et al. (2020a) tested
11 how well the simple system was able to differentiate between background and model error
12 by using one of two different model error covariance matrices in their experiments. One
13 model error covariance matrix \mathbf{Q}_s was defined to have short length scales, such that it
14 was equal to the given background error covariance matrix ($\mathbf{Q}_s = \mathbf{B}_x$). The second
15 model error covariance matrix \mathbf{Q}_l was defined to have large length scales, so that it was
16 different to the background error covariance matrix and to align with the hypothesis that
17 model error has longer length scales to background error. Experiments were undertaken
18 to demonstrate how WC4DVar could distinguish between background and model error,
19 and how it performs in the presence of observation bias.

20 In experiments without observation biases, Laloyaux et al. (2020a) found that the
21 experiment that used \mathbf{Q}_s struggled to estimate model bias, as WC4DVar was attributing
22 the model error to background error. However, when \mathbf{B}_x and \mathbf{Q} had different length scales
23 (ie. \mathbf{Q}_l was used), WC4DVar was able to correctly attribute the model error, and thus
24 performed well in correcting for the model bias. They repeated the experiment using \mathbf{Q}_l
25 when the observations were limited to one area (instead of being homogeneous across the
26 domain), to simulate, for example, many observations taken around an airport. In this
27 case, they found that WC4DVar was unable to estimate the model error in regions without

1 observations, despite having a scale separation between \mathbf{B}_x and \mathbf{Q}_1 .

2 In their final experiment, Laloyaux et al. added an uncorrected bias to all observations
3 within the system. As was found in Dee (2005), they found that there was less overall bias
4 in the analysis if model bias correction was not used at all, because the model bias was
5 wrongly corrected towards the observation bias when model bias was explicitly accounted
6 for.

7 Overall, Laloyaux et al. (2020a) demonstrated that model bias could be corrected
8 for using WC4DVar, if the length scales of the background and model error covariance
9 matrices were significantly different and if observations were available across the domain.
10 The separation of scales can be achieved if the model error covariance matrix is restricted
11 to represent errors on length-scales where the spectral energy density of background errors
12 is small. This is a well-verified approximation in the IFS, as model biases are prevalent
13 in the stratosphere and initial condition errors on smaller spatial and temporal scales are
14 prevalent in the troposphere (Laloyaux et al., 2020b). However, Laloyaux et al. (2020a)
15 also showed that WC4DVar failed in the presence of biased observations, so concluded
16 that unbiased observations were crucial in correcting for model bias in the presence of
17 observation bias. In the stratosphere, the use of unbiased observations is very achievable,
18 due to the large amount of radio occultation data available, but in the lower or upper
19 atmosphere, unbiased observations can be more sparse.

20 Both Eyre (2016) and Laloyaux et al. (2020a) have shown the need for unbiased obser-
21 vations when correcting for one bias in the presence of the other, but more work needs to
22 be done to understand what characteristics are needed in anchor observations to properly
23 mitigate the contamination of bias. How does the location of the anchor observations in
24 space and time influence their ability to reduce the contamination of model bias? How
25 does the state background error covariance matrix transfer information between observed
26 and unobserved states? This gap in knowledge in the use of unbiased observations in bias
27 correction schemes leads to research question 2, discussed in section 3.3.

1 In Lorente-Plazas and Hacker's (2017) paper, the interaction between observation and
2 model biases without unbiased observations in data assimilation is tested. Lorente-Plazas
3 and Hacker (2017) performed several numerical experiments on a simple system to test how
4 much bias filtered into the state analysis when each bias was corrected for, by including
5 additional parameters into the control vector, as in VarBC and WC4DVar (as discussed
6 in chapter 2, sections 2.3 and 2.5). The experiments used Model III developed by Lorenz
7 (2005), because the model had large-scale correlations between neighbouring grid points
8 and it combined small and large scales, which can be compared to the interaction between
9 mesoscales and synoptic scales in the atmosphere. Model bias was added to the system
10 by changing the forcing term in the model that evolved the system forwards in time.
11 Observation bias was added to the observations by either adding a constant homogeneously
12 to all observations; or by spatially varying the observation bias. The latter case will not
13 be discussed here. Two types of experiments were described in order to compare how well
14 the bias corrections performed in comparison to not bias-correcting: bias-aware, where
15 the biases were explicitly corrected for; and bias-blind, where biases were ignored, both
16 of which were described in Dee (2005). In this way, the contamination of the model and
17 observation biases were isolated and they were able to demonstrate the contamination of
18 one bias when the other was being corrected.

19 In total, nine experiments were presented with different combinations of biases present,
20 and corrected for: one control experiment which had no observation or model biases, with
21 no bias correction; three 'bias-blind' experiments that had observation and/or model
22 biases that were not corrected for; five bias correction experiments that had observation
23 and/or model biases present and one/both were corrected. Each experiment used 100
24 ensemble members which had random error in the observations with error variance = 0.5
25 and random error in the background state, where the error variance was calculated using
26 the Kalman filter and inflated by a mean factor of 1.1 and standard deviation of 0.6 at
27 the initial time.

1 In the control experiment, the state RMSE (root mean square error) was equal to the
 2 standard deviation, with a negligible bias. In the completely bias-blind experiments, the
 3 state RMSEs and the magnitudes of the biases were comparable when only one source
 4 of bias was present, although the signs of the biases were opposites. As the state RMSE
 5 and the standard deviations of the bias-blind experiments were similar, these biases were
 6 chosen for the later experiments.

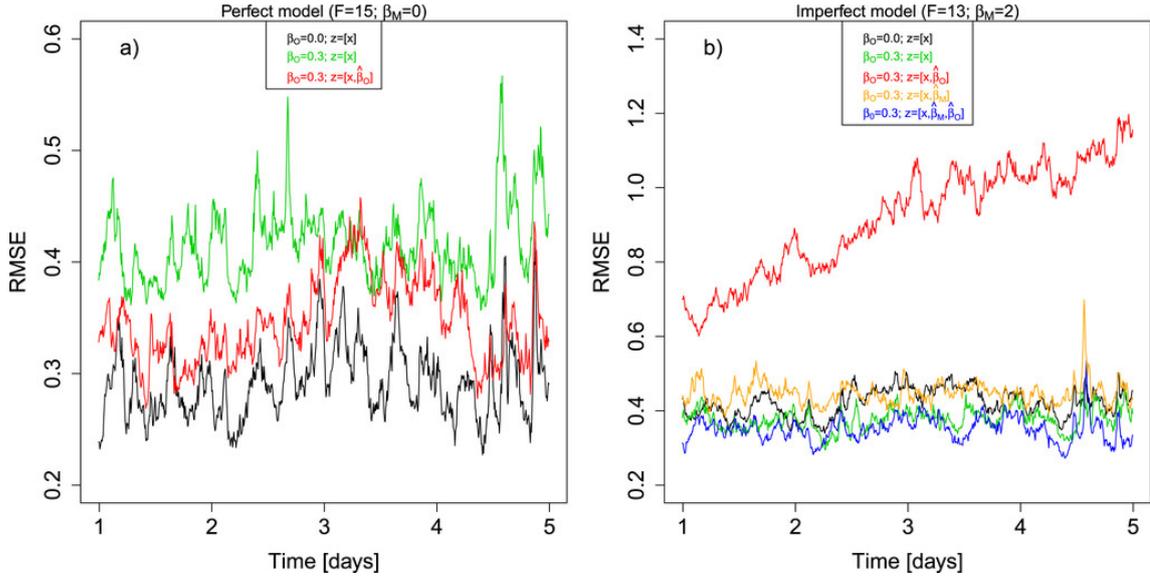


Figure 3.2: Time series of prior RMSE for different experiments (colours) using (a) a perfect model with and without spatially constant observation bias = 0.3, and (b) an imperfect model with model bias = 2. Legends show specified observation biases and the augmented vector estimated in the assimilation. From Lorente-Plazas and Hacker (2017) ©American Meteorological Society. Used with permission.

7 Figure 3.2 comes from Lorente-Plazas and Hacker (2017) and shows the time series
 8 of the state RMSE for eight of the experiments described above (other than when both
 9 model and observation biases were added, but only model bias was corrected for), when
 10 a perfect model was used (figure 3.2a) and when a biased (imperfect) model was used
 11 (figure 3.2b). We will focus on figure 3.2b, where both observation and model biases were
 12 present.

13 When the experiment was bias-blind to both observation and model biases (green line,

1 figure 3.2b), the state RMSE was surprisingly small, despite not correcting for either bias.
2 This was due to the experiment design, as the observation and model biases had opposite
3 signs, so combining them naturally cancelled them both out.

4 When observation bias was corrected, but the system was bias-blind to the model bias
5 (red line, figure 3.2b), the state RMSE was much larger compared to the other experiments
6 and diverged as time evolved. However, when model bias was corrected, but the system
7 was bias-blind to observation bias (gold line), the state RMSE was much lower. Later
8 experiments suggested that this was because the state was more sensitive to error variances
9 in the background observation bias estimate than error variances in the background model
10 bias estimate. If only one bias was corrected, then the non-corrected bias would filter
11 into the bias-corrected estimate. Therefore, uncorrected model bias would have a larger
12 impact on observation bias correction as it was absorbed into the observation bias estimate
13 and therefore affected the state estimate, whereas when uncorrected observation bias was
14 absorbed into the model bias estimate, it did not have as large an effect on the state
15 estimate.

16 When both model and observation biases were corrected for (blue line), the state
17 RMSE was the lowest of all the experiments when both observation and model biases
18 were present. Lorente-Plazas and Hacker suggested that knowledge of the *a priori* model
19 error appeared to be unnecessary, and the system only needed to know that model bias
20 existed. Although, they did acknowledge that in a complex geophysical model, the form
21 of a parametric model to represent model errors is not always known.

22 Lorente-Plazas and Hacker (2017) demonstrated that in the presence of both obser-
23 vation and model biases, correcting for model bias and not observation bias reduced the
24 state RMSE more than only correcting for observation bias, but correcting for both biases
25 simultaneously reduced the state RMSE the most. However, their experiments do not
26 include any unbiased observations when both sources of bias were present, which both
27 Eyre (2016) and Laloyaux et al. (2020a) showed to be crucial in anchoring the bias correc-

1 tion techniques. This raises the question of how well both VarBC and WC4DVar perform
2 comparatively to each other in the presence of both sources of bias, when unbiased ob-
3 servations can be used. This therefore leads to research question 3, which is discussed in
4 section 3.3.

5 **3.3 Research questions**

6 In chapter 3 we have presented important studies in the field of bias correction in data
7 assimilation to understand where the gaps in knowledge lie. We will consider the following
8 research questions which will be the basis for the rest of this thesis:

9

10 **RQ 1: What are the consequences of mis-specifying the background error**
11 **statistics in VarBC?**

12 The background error covariance matrices for both the state and the observation bias
13 coefficient are not exactly known and, operationally, are an estimate of the true back-
14 ground error covariance matrix (Bannister, 2008a; Cameron & Bell, 2018). Eyre and
15 Hilton (2013) showed the damage that a mis-specified state background error covariance
16 matrix can make in a 3DVar data assimilation system without bias correction, but did
17 not investigate the impact of mis-specifying the background error covariance matrices in a
18 VarBC system. We will demonstrate the impact that mis-specifying both \mathbf{B}_x and \mathbf{B}_β have
19 on the state and bias coefficient analysis error covariance matrices, when using VarBC.
20 RQ 1 will be answered in chapter 4.

21

22 **RQ 2: What criteria are needed in the anchor observations in order to suc-**
23 **cessfully reduce bias in the analysis when both model and observation biases**
24 **are present but only observation bias is accounted for?**

25 Eyre (2016) and Laloyaux et al. (2020a) both demonstrated the need for anchor obser-

1 vations when correcting for observation/model bias in the presence of both sources of bias,
2 so that the bias correction was not corrected towards the wrong source of bias. However,
3 neither studied which properties of anchor observations were more important in reducing
4 the contamination of bias within a bias correction scheme. Therefore, we will study the
5 following:

- 6 • **RQ 2.1: Where are anchor observations most effective in reducing the
7 contamination of bias?** In chapter 2.2, we have discussed that, geographically,
8 anchor observations do not cover all atmospheric variables across the Earth. There-
9 fore, where do anchor observations need to lie in comparison to model biases and
10 biased observations in order to successfully mitigate the contamination of model bias
11 on the observation bias correction? RQ 2.1 will be answered in chapter 6.
 - 12 • **RQ 2.2: When are anchor observations most effective in reducing the
13 contamination of bias?** In 4DVar, observations can be spread throughout the
14 assimilation window. Does the ability of anchor observations to reduce the contam-
15 ination of bias in the bias correction differ, depending on whether they are closer to
16 the beginning or end of the window? RQ 2.2 will be answered in chapters 7 and 8.
 - 17 • **RQ 2.3: Does the quality of the anchor observations matter in reducing
18 the contamination of bias?** Eyre (2016) demonstrated how the weighting given
19 to the anchor observations impacted how sensitive the bias in the analysis was to
20 the model bias. In a non-scalar system, how do the error covariances between the
21 background errors and the observation errors interact, and would it be better to
22 have a full coverage of anchor observations with low precision, or a smaller coverage
23 with higher precision? RQ 2.3 will be answered in chapters 6, 7 and 8.
- 24 **RQ3: How important are anchor observations when correcting for model
25 and/or observation bias?**

1 Lorente-Plazas and Hacker (2017) compared correcting for observation and/or model
2 bias in the presence of both biases and found that correcting for only model bias reduced
3 the error in the state analysis more than only correcting for observation bias, although
4 they found correcting for both simultaneously reduced the error even more. Operationally,
5 bias correction methods are mostly used when anchor observations are available. There-
6 fore, what difference does using anchor observations make when trying to correct one or
7 both biases in the presence of both? How much better do the bias correction methods
8 perform when they are able to use anchor observations to mitigate the contamination of
9 bias? RQ 3 will be answered in chapter 8.

10

11 These research questions will be tackled by analytically studying the relevant analysis
12 equations in VarBC or WC4DVar and will be illustrated using simple numerical systems.
13 These techniques will allow us to understand the basic theory and to visualise the equa-
14 tions, which allows us to interpret the equations in both scalar and vector systems under
15 specific assumptions.

16 **3.4 Summary**

17 In chapter 3 we have presented several papers that have demonstrated bias correction
18 techniques when both model and observation biases were present. We showed the need
19 for more understanding of when unbiased observations are able to disentangle model and
20 observation biases and developed research questions which will be used to shape the re-
21 maining ideas discussed in this thesis. In the next chapter we will answer research question
22 1, by studying the effect of mis-specifying the observation bias coefficient background error
23 covariance matrix in VarBC.

1 Chapter 4

2 Mis-specification of background 3 error statistics on a VarBC system

4 4.1 Introduction

5 As already discussed in section 2.3, VarBC is very important operationally as it allows the
6 use of biased radiance observations in NWP. However, we showed in section 3.1 that the
7 background error covariance matrices of both the state and the observation bias coefficient
8 are only approximations of the true background error covariance matrices. In this chapter
9 we will study the implication of mis-specifying both \mathbf{B}_x and \mathbf{B}_β in VarBC.

10 In section 3.1 we gave an overview of Eyre and Hilton (2013), which demonstrated how
11 mis-specifying the state background error covariance matrix can cause the state analysis er-
12 ror covariance matrix to be larger than the background error covariance matrix in a 3DVar
13 system. In section 3.1 we also discussed how, when using VarBC, the observation bias co-
14 efficient background error covariance matrix is only an estimate of the true background
15 error covariance matrix, designed to limit the growth in the observation bias coefficient,
16 rather than reflect the 'true' bias coefficient background error covariance matrix (Cameron
17 & Bell, 2018). Therefore, in this chapter we extend the work of Eyre and Hilton (2013)

1 to include VarBC, in order to answer research question 1: what are the consequences of
 2 mis-specifying the background error statistics in VarBC? We will demonstrate the effect of
 3 under- and overestimating both the state and observation bias coefficient background error
 4 covariance matrices on the state and observation bias coefficient analysis error covariance
 5 matrices, in order to highlight the consequences of mis-specifying them.

6 We will start by calculating the VarBC analysis error covariance matrix equation when
 7 the background error covariance matrix is mis-specified, by following the method in Eyre
 8 and Hilton (2013). The results are then presented in a simple scalar system to give further
 9 understanding.

10 **4.2 Sub-optimal analysis error covariance matrix equations** 11 **for the state and observation bias coefficient**

12 In section 2.3 the analysis error covariance matrix in a VarBC system that contains both
 13 the state and observation bias coefficient analysis error covariance matrices, was given in
 14 its most general form in equation (2.69) as,

$$\mathbf{A}_v = (\mathbf{I} - \mathbf{K}_v \mathbf{H}_v) \mathbf{B}_v (\mathbf{I} - \mathbf{K}_v \mathbf{H}_v)^T + \mathbf{K}_v \mathbf{R} \mathbf{K}_v^T \in \mathbb{R}^{(n+r) \times (n+r)}, \quad (4.1)$$

15 where \mathbf{K}_v is the sensitivity of the control vector to the observations; \mathbf{H}_v is the linearised
 16 observation operator which is a Jacobian of the control vector with respect to the state
 17 \mathbf{x} and the observation bias coefficient β ; \mathbf{B}_v is the background error covariance matrix
 18 for both \mathbf{x} and β ; and \mathbf{R} is the observation error covariance matrix. Note that equation
 19 (4.1) is equivalent to equation (2.69) and has just been repeated here for convenience
 20 for the reader. If the background and observation error covariance matrices are the true
 21 background and observation error covariance matrices, then \mathbf{K}_v is optimal. In this case,
 22 we showed in equation (2.71) that the optimal analysis error covariance matrix can be

1 written as,

$$\mathbf{A}_{v,\text{opt}} = (\mathbf{I} - \mathbf{K}_v \mathbf{H}_v) \mathbf{B}_v, \quad (4.2)$$

2 which is equivalent to equation (2.71), but has been shown here again for convenience.

3 However, when the background error covariance matrix is mis-specified, i.e. the as-
 4 sumed background error covariance matrix \mathbf{B}_v^a inputted into the system is not equal to the
 5 true background error covariance matrix \mathbf{B}_v , then the Kalman gain matrix will be given
 6 by,

$$\mathbf{K}_v^a = \mathbf{B}_v^a \mathbf{H}_v^T (\mathbf{H}_v \mathbf{B}_v^a \mathbf{H}_v^T + \mathbf{R})^{-1}, \quad (4.3)$$

7 where we have denoted \mathbf{K}_v^a to be the suboptimal Kalman gain matrix when the assumed
 8 background error covariance matrix is used. Note that, operationally, the observation
 9 error covariance matrix would also only be an assumption of the truth, but, as we are
 10 only interested in the effect of mis-specifying the background error covariance matrix in
 11 this work, we assume the true \mathbf{R} has been used. Then, following the same method as
 12 in Eyre and Hilton (2013) but for the control vector that contains both the state and
 13 the observation bias coefficient, we can calculate the sub-optimal analysis error covariance
 14 matrix. Substituting \mathbf{K}_v^a into equation (4.1) gives,

$$\mathbf{A}_v^a = (\mathbf{I} - \mathbf{K}_v^a \mathbf{H}_v) \mathbf{B}_v (\mathbf{I} - \mathbf{K}_v^a \mathbf{H}_v)^T + \mathbf{K}_v^a \mathbf{R} \mathbf{K}_v^{aT}. \quad (4.4)$$

15 The analysis error covariance matrix, \mathbf{A}_v^a , is still dependent on the true \mathbf{B}_v as \mathbf{B}_v comes
 16 from the background errors in the data assimilation system itself, whereas the \mathbf{B}_v^a in \mathbf{K}_v^a
 17 comes from the assumed error covariance matrices inputted into the system. As \mathbf{K}_v^a is
 18 no longer dependent on \mathbf{B}_v (but instead is dependent on \mathbf{B}_v^a), \mathbf{A}_v^a can not be reduced
 19 in the same way as in equations (4.1) and (4.2). Adding and subtracting the assumed
 20 background error covariance matrix to the true background error covariance matrix in

1 equation (4.4) gives,

$$\mathbf{A}_v^a = (\mathbf{I} - \mathbf{K}_v^a \mathbf{H}_v)(\mathbf{B}_v + \mathbf{B}_v^a - \mathbf{B}_v^a)(\mathbf{I} - \mathbf{K}_v^a \mathbf{H}_v)^T + \mathbf{K}_v^a \mathbf{R} \mathbf{K}_v^{aT}, \quad (4.5)$$

2 and expanding gives,

$$\mathbf{A}_v^a = (\mathbf{I} - \mathbf{K}_v^a \mathbf{H}_v) \mathbf{B}_v^a (\mathbf{I} - \mathbf{K}_v^a \mathbf{H}_v)^T + \mathbf{K}_v^a \mathbf{R} \mathbf{K}_v^{aT} + (\mathbf{I} - \mathbf{K}_v^a \mathbf{H}_v)(\mathbf{B}_v - \mathbf{B}_v^a)(\mathbf{I} - \mathbf{K}_v^a \mathbf{H}_v)^T. \quad (4.6)$$

3 The first two terms are equal to equation (4.1) for $\mathbf{B}_v = \mathbf{B}_v^a$, which we showed in equation
 4 (4.2) can be simplified to $\mathbf{A}_{v,\text{opt}}$. Therefore, equation (4.6) can be simplified to

$$\mathbf{A}_v^a = \mathbf{A}_{v,\text{opt}}^a + (\mathbf{I} - \mathbf{K}_v^a \mathbf{H}_v)(\mathbf{B}_v - \mathbf{B}_v^a)(\mathbf{I} - \mathbf{K}_v^a \mathbf{H}_v)^T, \quad (4.7)$$

5 where $\mathbf{A}_{v,\text{opt}}^a$ is equation (4.2) when $\mathbf{B}_v = \mathbf{B}_v^a$. Equation (4.7) is an extension of equation
 6 (10) in Eyre and Hilton (2013) to include the observation bias coefficient in the control
 7 vector in the analysis error variance when both the state and background error covariance
 8 matrices are mis-specified. It is the sum of the optimal analysis error covariance matrix
 9 if the assumed background error covariance matrix is equal to the true background error
 10 covariance matrix, and a product of matrices which is linearly dependent on the difference
 11 between the true background error covariance matrix and the assumed background error
 12 covariance matrix.

13 In order to determine how the state analysis error covariance matrix and the obser-
 14 vation bias coefficient analysis error covariance matrix are dependent on the errors in the
 15 assumed state and bias coefficient background error covariance matrices, we can separate
 16 \mathbf{A}_v^a from equation (4.7) into its \mathbf{x} and β parts, as

$$\mathbf{A}_v = \begin{pmatrix} \mathbf{A}_x & \mathbf{A}_{x\beta} \\ \mathbf{A}_{\beta x} & \mathbf{A}_\beta \end{pmatrix} \quad (4.8)$$

1 as in equation (2.70) in section 2.3. To write equation (4.7) in terms of its state and
 2 observation bias coefficient parts, we first expand $\mathbf{A}_{v,\text{opt}}^a$ by writing: \mathbf{K}_v^a in terms of \mathbf{K}_x^a
 3 and \mathbf{K}_β^a ; \mathbf{H}_v in terms of \mathbf{H} and \mathbf{C}_β ; and \mathbf{B}_v^a in terms of \mathbf{B}_x^a and \mathbf{B}_β^a :

$$\mathbf{A}_{v,\text{opt}}^a = \left(\mathbf{I}^{n+r} - \begin{pmatrix} \mathbf{K}_x^a \\ \mathbf{K}_\beta^a \end{pmatrix} \begin{pmatrix} \mathbf{H} & \mathbf{C}_\beta \end{pmatrix} \right) \begin{pmatrix} \mathbf{B}_x^a & 0 \\ 0 & \mathbf{B}_\beta^a \end{pmatrix}, \quad (4.9)$$

4 where $\mathbf{I}^{n+r} \in \mathbb{R}^{(n+r) \times (n+r)}$ is the identity matrix of dimension $n+r$ (ie. the dimension of
 5 the state and bias coefficient combined); $\mathbf{H} \in \mathbb{R}^{m \times n}$ is the linearised observation operator
 6 as defined in section 2.1; $\mathbf{C}_\beta \in \mathbb{R}^{m \times r}$ is the linearised bias correction with respect to β as
 7 defined in section 2.3; $\mathbf{B}_x^a \in \mathbb{R}^{n \times n}$ and $\mathbf{B}_\beta^a \in \mathbb{R}^{r \times r}$ are the assumed state and bias coefficient
 8 background error covariance matrices respectively; $\mathbf{K}_x^a \in \mathbb{R}^{n \times m}$ is the sensitivity of the
 9 state to the observations and is given by,

$$\mathbf{K}_x^a = \mathbf{B}_x^a \mathbf{H}^T (\mathbf{H} \mathbf{B}_x^a \mathbf{H}^T + \mathbf{C}_\beta \mathbf{B}_\beta^a \mathbf{C}_\beta^T + \mathbf{R})^{-1} \quad (4.10)$$

10 which is equivalent to the \mathbf{K}_x as defined in equation (2.52), but for $\mathbf{B}_x = \mathbf{B}_x^a$ and $\mathbf{B}_\beta = \mathbf{B}_\beta^a$;
 11 and $\mathbf{K}_\beta^a \in \mathbb{R}^{r \times m}$ is the sensitivity of the bias coefficient to the observations and is given
 12 by,

$$\mathbf{K}_\beta^a = \mathbf{B}_\beta^a \mathbf{C}_\beta^T (\mathbf{H} \mathbf{B}_x^a \mathbf{H}^T + \mathbf{C}_\beta \mathbf{B}_\beta^a \mathbf{C}_\beta^T + \mathbf{R})^{-1} \quad (4.11)$$

13 which is equivalent to the \mathbf{K}_β as defined in equation (2.53) but for $\mathbf{B}_x = \mathbf{B}_x^a$ and $\mathbf{B}_\beta = \mathbf{B}_\beta^a$.
 14 Simplifying equation (4.9) gives the form of the optimal analysis error covariance matrix
 15 for the state and observation bias coefficient when $\mathbf{B}_v = \mathbf{B}_v^a$ as,

$$\mathbf{A}_{v,\text{opt}}^a = \begin{pmatrix} (\mathbf{I}^n - \mathbf{K}_x^a \mathbf{H}) \mathbf{B}_x^a & -\mathbf{K}_x^a \mathbf{C}_\beta \mathbf{B}_\beta^a \\ -\mathbf{K}_\beta^a \mathbf{H} \mathbf{B}_x^a & (\mathbf{I}^r - \mathbf{K}_\beta^a \mathbf{C}_\beta) \mathbf{B}_\beta^a \end{pmatrix}, \quad (4.12)$$

16 where $\mathbf{A}_{v,\text{opt}}^a$ is symmetric, and can be shown by expanding $\mathbf{K}_x^a \mathbf{C}_\beta \mathbf{B}_\beta^a$ and $\mathbf{K}_\beta^a \mathbf{H} \mathbf{B}_x^a$ respec-

1 tively to show they are equal.

2 Next, to continue expanding \mathbf{A}_v^a into its state and observation bias coefficient parts,
 3 we expand the second term of equation (4.7) into its \mathbf{x} and $\boldsymbol{\beta}$ parts,

$$\begin{aligned} & (\mathbf{I} - \mathbf{K}_v^a \mathbf{H}_v)(\mathbf{B}_v - \mathbf{B}_v^a)(\mathbf{I} - \mathbf{K}_v^a \mathbf{H}_v)^T \\ &= \left(\mathbf{I}^{n+r} - \begin{pmatrix} \mathbf{K}_x^a \\ \mathbf{K}_\beta^a \end{pmatrix} \begin{pmatrix} \mathbf{H} & \mathbf{C}_\beta \end{pmatrix} \right) \begin{pmatrix} \mathbf{B}_x - \mathbf{B}_x^a & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_\beta - \mathbf{B}_\beta^a \end{pmatrix} \left(\mathbf{I}^{n+r} - \begin{pmatrix} \mathbf{K}_x^a \\ \mathbf{K}_\beta^a \end{pmatrix} \begin{pmatrix} \mathbf{H} & \mathbf{C}_\beta \end{pmatrix} \right)^T, \end{aligned} \quad (4.13)$$

4 which gives,

$$= \begin{pmatrix} (\mathbf{I}^n - \mathbf{K}_x^a \mathbf{H})(\mathbf{B}_x - \mathbf{B}_x^a) & -\mathbf{K}_x^a \mathbf{C}_\beta (\mathbf{B}_\beta - \mathbf{B}_\beta^a) \\ -\mathbf{K}_\beta^a \mathbf{H} (\mathbf{B}_x - \mathbf{B}_x^a) & (\mathbf{I}^r - \mathbf{K}_\beta^a \mathbf{C}_\beta) (\mathbf{B}_\beta - \mathbf{B}_\beta^a) \end{pmatrix} \times \begin{pmatrix} (\mathbf{I}^n - \mathbf{K}_x^a \mathbf{H})^T & -(\mathbf{K}_\beta^a \mathbf{H})^T \\ -(\mathbf{K}_x^a \mathbf{C}_\beta)^T & (\mathbf{I}^r - \mathbf{K}_\beta^a \mathbf{C}_\beta)^T \end{pmatrix}. \quad (4.14)$$

5 Therefore expanding the final matrix multiplication from equation (4.14) and adding to
 6 equation (4.12) gives us the analysis error covariance matrix in terms of its \mathbf{x} and $\boldsymbol{\beta}$
 7 elements. We can separate the four elements of the matrix into \mathbf{A}_x , $\mathbf{A}_{x\beta}$, $\mathbf{A}_{\beta x}$ and \mathbf{A}_β as
 8 in equation (4.8) to give,

$$\mathbf{A}_x = (\mathbf{I}^n - \mathbf{K}_x^a \mathbf{H}) \mathbf{B}_x^a + (\mathbf{I}^n - \mathbf{K}_x^a \mathbf{H})(\mathbf{B}_x - \mathbf{B}_x^a)(\mathbf{I}^n - \mathbf{K}_x^a \mathbf{H})^T + \mathbf{K}_x^a \mathbf{C}_\beta (\mathbf{B}_\beta - \mathbf{B}_\beta^a)(\mathbf{K}_x^a \mathbf{C}_\beta)^T, \quad (4.15)$$

9

$$\mathbf{A}_{x\beta} = -\mathbf{K}_x^a \mathbf{C}_\beta \mathbf{B}_\beta^a + (\mathbf{I}^n - \mathbf{K}_x^a \mathbf{H})(\mathbf{B}_x - \mathbf{B}_x^a)(-\mathbf{K}_\beta^a \mathbf{H})^T - \mathbf{K}_x^a \mathbf{C}_\beta (\mathbf{B}_\beta - \mathbf{B}_\beta^a)(\mathbf{I}^r - \mathbf{K}_\beta^a \mathbf{C}_\beta)^T, \quad (4.16)$$

10

$$\mathbf{A}_{\beta x} = -\mathbf{K}_\beta^a \mathbf{H} \mathbf{B}_x^a - \mathbf{K}_\beta^a \mathbf{H} (\mathbf{B}_x - \mathbf{B}_x^a)(\mathbf{I}^n - \mathbf{K}_x^a \mathbf{H})^T + (\mathbf{I}^r - \mathbf{K}_\beta^a \mathbf{C}_\beta) (\mathbf{B}_\beta - \mathbf{B}_\beta^a)(-\mathbf{K}_x^a \mathbf{C}_\beta)^T, \quad (4.17)$$

11

$$\mathbf{A}_\beta = (\mathbf{I}^r - \mathbf{K}_\beta^a \mathbf{C}_\beta) \mathbf{B}_\beta^a + \mathbf{K}_\beta^a \mathbf{H} (\mathbf{B}_x - \mathbf{B}_x^a)(\mathbf{K}_\beta^a \mathbf{H})^T + (\mathbf{I}^r - \mathbf{K}_\beta^a \mathbf{C}_\beta) (\mathbf{B}_\beta - \mathbf{B}_\beta^a)(\mathbf{I}^r - \mathbf{K}_\beta^a \mathbf{C}_\beta)^T. \quad (4.18)$$

12 It is simple to check that $\mathbf{A}_{x\beta}$ and $\mathbf{A}_{\beta x}$ are transposes of each other and can be calculated
 13 by expanding \mathbf{K}_x^a and \mathbf{K}_β^a respectively as in equations (4.10) and (4.11). $\mathbf{A}_{x\beta}$ and $\mathbf{A}_{\beta x}$
 14 represent the analysis error covariances between the state and the bias coefficient. In

1 this chapter we are only concerned with how the mis-specification of \mathbf{B}_x^a and \mathbf{B}_β^a impact
2 \mathbf{A}_x and \mathbf{A}_β , not $\mathbf{A}_{x\beta}$ and $\mathbf{A}_{\beta x}$, as the cross-covariance matrices are less important when
3 estimating the state and the observation bias coefficient. Therefore, we will not study $\mathbf{A}_{x\beta}$
4 and $\mathbf{A}_{\beta x}$ any further.

5 Equations (4.15) and (4.18) show that the analysis error covariance matrices for both
6 the state and observation bias coefficient are sensitive to the mis-specification of both
7 the state and the observation bias coefficient background error covariance matrices, as
8 both are dependent on $\mathbf{B}_x - \mathbf{B}_x^a$ and $\mathbf{B}_\beta - \mathbf{B}_\beta^a$. The sensitivity of the state analysis error
9 covariance matrix to the mis-specification of \mathbf{B}_x^a and \mathbf{B}_β^a in equation (4.15) is controlled
10 by \mathbf{K}_x^a and the sensitivity of the bias coefficient analysis error covariance matrix to the
11 mis-specification of \mathbf{B}_x^a and \mathbf{B}_β^a in equation (4.18) is controlled by \mathbf{K}_β^a .

12 In equation (4.15), if $\mathbf{K}_x^a \mathbf{C}_\beta$ tends to the zero matrix, which would occur when: the
13 bias correction is independent of the observation bias coefficient ($\mathbf{C}_\beta = 0$), which would be
14 more similar to a static correction scheme as described in section 2.3; or when the assumed
15 state background error variance is very small ($\mathbf{B}_x^a \rightarrow \mathbf{0}$, therefore $\mathbf{K}_x^a \rightarrow \mathbf{0}$), then \mathbf{A}_x will
16 be independent of the mis-specification of \mathbf{B}_β . However, these are uninteresting cases as
17 they would mean the data assimilation is entirely reliant on the state background, rather
18 than observations, to make an update to the system. Equivalently, in equation (4.18),
19 if $\mathbf{K}_\beta^a \mathbf{H}$ tends to the zero vector, which would occur when: there are no observations
20 ($\mathbf{H} = 0$); or when the assumed bias coefficient background error variance is very small
21 ($\mathbf{B}_\beta^a \rightarrow \mathbf{0}$, therefore $\mathbf{K}_\beta^a \rightarrow \mathbf{0}$), then \mathbf{A}_β would be independent of the mis-specification of
22 \mathbf{B}_x^a . However, these are also both uninteresting cases as the observations would not be
23 used in the VarBC system. Therefore it is clear that, in general, both \mathbf{A}_x and \mathbf{A}_β are
24 sensitive to the mis-specification of both \mathbf{B}_x^a and \mathbf{B}_β^a .

25 In order to understand equations (4.15) and (4.18) further, we will study the theoretical
26 scalar case when there is one state that is observed by one biased observation, which is
27 corrected using VarBC.

1 4.3 Scalar example

2 The danger zone of the analysis error variance in the scalar case was defined in Eyre
 3 and Hilton (2013) to be when the analysis error variance is larger than the corresponding
 4 background error variance. We can extend this to VarBC, such that the danger zone of
 5 the state analysis is when $\sigma_{ax}^2 > \sigma_{bx}^2$, where σ_{ax}^2 is the scalar state analysis error variance
 6 and σ_{bx}^2 is the scalar state background error variance. Furthermore, we can define the
 7 danger zone of the bias coefficient analysis to be when $\sigma_{a\beta}^2 > \sigma_{b\beta}^2$, where $\sigma_{a\beta}^2$ is the scalar
 8 bias coefficient analysis error variance and $\sigma_{b\beta}^2$ is the scalar bias coefficient background
 9 error variance.

10 In order to investigate when the state and bias coefficient analysis error variances fall
 11 into the danger zones based on the choice of state and bias-coefficient background error
 12 variances, we look at the scalar cases of equations (4.15) and (4.18), which are given by,

$$\sigma_{ax}^2 = (1 - k_x^a h) \sigma_{bx}^{a2} + (1 - k_x^a h)^2 (\sigma_{bx}^2 - \sigma_{bx}^{a2}) + (k_x^a c_\beta)^2 (\sigma_{b\beta}^2 - \sigma_{b\beta}^{a2}), \quad (4.19)$$

$$\sigma_{a\beta}^2 = (1 - k_\beta^a c_\beta) \sigma_{b\beta}^{a2} + (k_\beta^a h)^2 (\sigma_{bx}^2 - \sigma_{bx}^{a2}) + (1 - k_\beta^a c_\beta)^2 (\sigma_{b\beta}^2 - \sigma_{b\beta}^{a2}), \quad (4.20)$$

13 where k_x^a , k_β^a , h and c_β are given in italic lower case to show that they are scalar and k_x^a
 14 and k_β^a are given by,

$$k_x^a = \frac{\sigma_{bx}^{a2} h}{\sigma_{bx}^{a2} h^2 + \sigma_{b\beta}^{a2} c_\beta^2 + \sigma_o^2}, \quad (4.21)$$

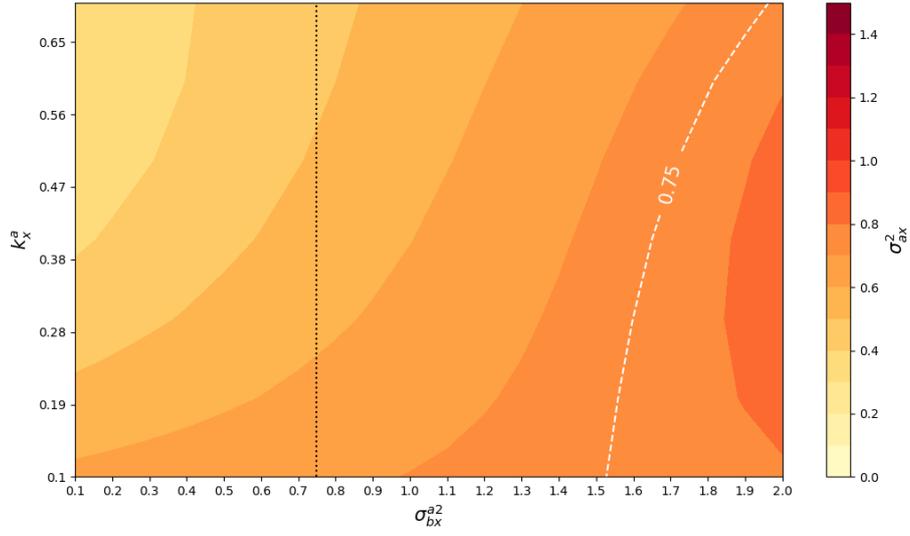
$$k_\beta^a = \frac{\sigma_{b\beta}^{a2} c_\beta}{\sigma_{bx}^{a2} h^2 + \sigma_{b\beta}^{a2} c_\beta^2 + \sigma_o^2}. \quad (4.22)$$

15 Equations (4.19) and (4.20) are symmetric to each other. The first term for both the state
 16 and bias coefficient analysis error variance are the optimal analysis error variances if the
 17 assumed background error variances are correct. The self-sensitivity background terms are
 18 the difference terms for their own background error variances, multiplied by $(1 - k_x^a h)^2$ or

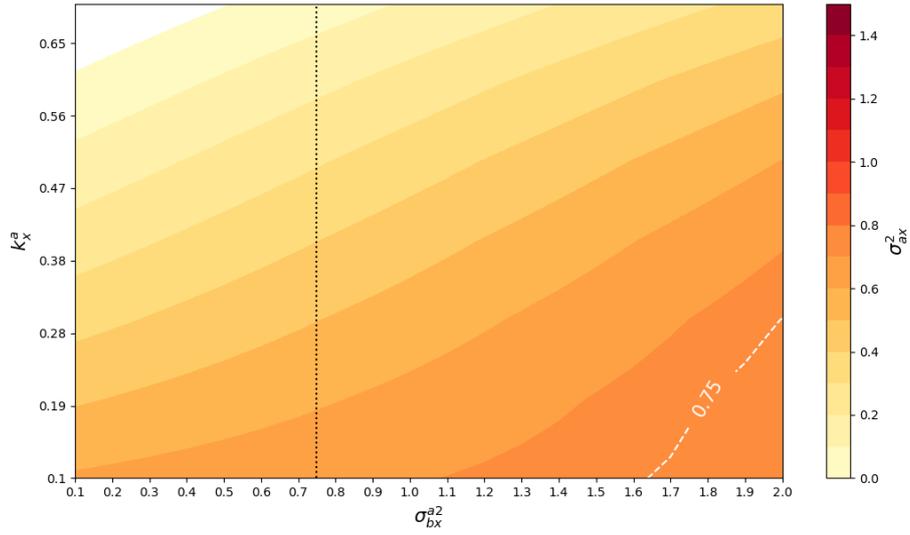
1 $(1 - k_\beta^a c_\beta)^2$ respectively. The cross-sensitivity background terms are the difference terms
 2 for the other background error variances (ie. the bias coefficient background error variance
 3 for the state analysis error variance and the state background error variance for the bias
 4 coefficient analysis error variance), multiplied by $(k_x^a c_\beta)^2$ or $(k_\beta^a h)^2$ respectively.

5 In order to test when equations (4.19) and (4.20) fall into their respective danger
 6 zones, we study σ_{ax}^2 and $\sigma_{a\beta}^2$ when the parameters have given values. The observations are
 7 assumed to be direct and the biases linear with respect to β , such that $h = c_\beta = 1$. For
 8 each experiment we vary either k_x^a or k_β^a and either σ_{bx}^2 or $\sigma_{b\beta}^2$ and keep the other values
 9 constant. Note that k_x^a and k_β^a would be dependent on σ_{bx}^2 and $\sigma_{b\beta}^2$, but for simplicity, we
 10 have varied them mutually exclusively so that our results are less reliant on the specific
 11 parameters chosen. In practice, varying k_x^a would be equivalent to varying the observation
 12 error variance.

13 In figure 4.1 we have plotted σ_{ax}^2 from equation (4.19) whilst varying the sensitivity of
 14 the state to the observations (k_x^a) and the assumed state background error variance (σ_{bx}^2).
 15 The true state background error variance (σ_{bx}^2) is 0.75, which is denoted by the dotted line:
 16 when σ_{bx}^2 is to the left of the line it has been underestimated; and when it is to the right of
 17 the line it has been overestimated. The assumed bias coefficient background error variance
 18 is set to $\sigma_{b\beta}^2 = 1$. The state analysis error variance has been plotted for two cases of the
 19 true bias coefficient background error variance: when $\sigma_{b\beta}^2 = 1.5$ (ie. the bias coefficient
 20 background error variance has been underestimated) in figure 4.1a; and when $\sigma_{b\beta}^2 = 0.5$
 21 (ie. the bias coefficient background error variance has been overestimated) in figure 4.1b.
 22 The boundary of the danger zone is marked by the white dashed contour line at the value
 23 of the true state background error variance. State analysis error variance values larger
 24 than the contour fall into the danger zone, as $\sigma_{ax}^2 > \sigma_{bx}^2$ in these areas. The white spaces
 25 are where σ_{ax}^2 are negative, so have been removed. These negative error variance values
 26 come from impossible values of k_x^a as it would require the observation error variance to
 27 also be negative, so are unrealistic.



(a) Underestimating $\sigma_{b\beta}^2$



(b) Overestimating $\sigma_{b\beta}^2$

Figure 4.1: σ_{ax}^2 varying both k_x^a and σ_{bx}^2 . $h = c_\beta = 1$. $\sigma_{b\beta}^{a2} = 1$, $\sigma_{bx}^2 = 0.75$. (a) $\sigma_{b\beta}^2 = 1.5$ (underestimating $\sigma_{b\beta}^2$) and (b) $\sigma_{b\beta}^2 = 0.5$ (overestimating $\sigma_{b\beta}^2$). White dashed line shows when $\sigma_{ax}^2 = \sigma_{bx}^2$. Black dotted line shows when $\sigma_{b\beta}^{a2} = \sigma_{bx}^2$.

1 In figure 4.1a, the state analysis error variance only falls into the danger zone when
 2 $\sigma_{\text{bx}}^{\text{a}2}$ is greater than twice the true value, as the true state background error variance was
 3 defined as 0.75. This suggests that underestimating the state background error variance
 4 keeps the analysis error variance away from the danger zone.

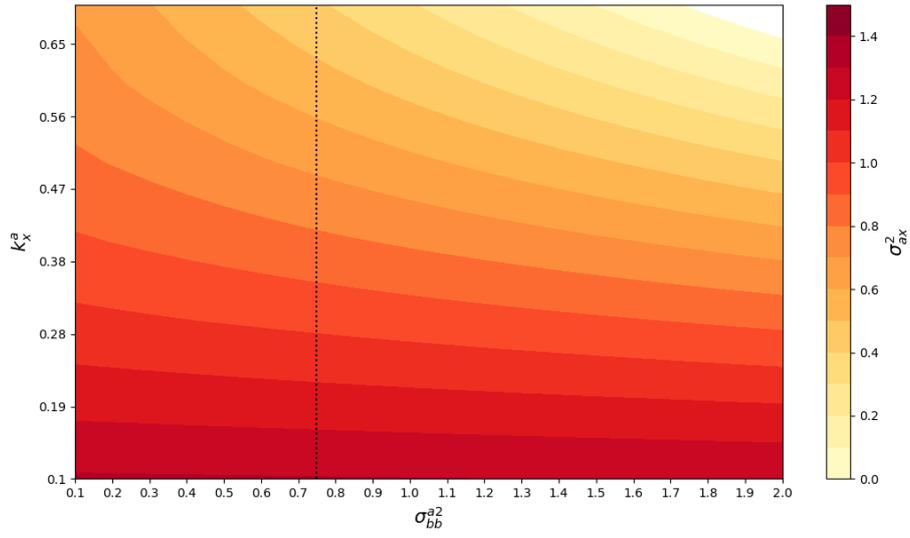
5 In figure 4.1b, the danger zone has been reduced compared to figure 4.1a, which implies
 6 that overestimating $\sigma_{\text{b}\beta}^{\text{a}2}$ reduces the likelihood that σ_{ax}^2 will fall into the danger zone. As
 7 in figure 4.1a, underestimating the background error variance, or also putting more trust
 8 into the observations (such that $k_{\text{x}}^{\text{a}} > 0.3$) keeps the state analysis error variance away
 9 from the danger zone.

10 In figure 4.2 we have again plotted σ_{ax}^2 from equation (4.19), but have now varied k_{x}^{a} and
 11 $\sigma_{\text{b}\beta}^{\text{a}2}$. The true observation bias coefficient background error variance is $\sigma_{\text{b}\beta}^2 = 0.75$, which
 12 is shown by the dotted black line. When $\sigma_{\text{b}\beta}^{\text{a}2}$ is to the left of the dotted line it has been
 13 underestimated and when it is to the right of the dotted line it has been overestimated.
 14 The assumed state background error variance is set to $\sigma_{\text{bx}}^{\text{a}2} = 1$. The state analysis error
 15 variance has been plotted for two cases of the true state background error variance: when
 16 $\sigma_{\text{bx}}^2 = 1.5$ (ie. the state background error variance is underestimated) in figure 4.2a; and
 17 when $\sigma_{\text{bx}}^2 = 0.5$ (ie. the state background error variance is overestimated) in figure 4.2b.

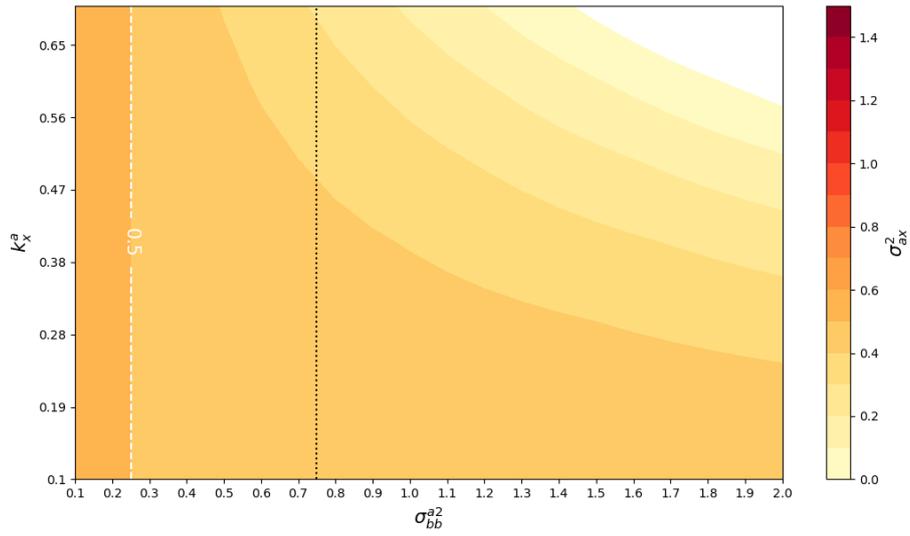
18 In figure 4.2a, we have underestimated the state background error variance, which has
 19 meant that σ_{ax}^2 is always less than the true state background error variance, 1.5, and thus
 20 the danger zone has disappeared, which is consistent with figure 4.1. However, the values
 21 of the state analysis error variance are more dependent on k_{x}^{a} than $\sigma_{\text{b}\beta}^{\text{a}2}$, with larger state
 22 analysis error variance when k_{x}^{a} is smaller.

23 In figure 4.2b, we have overestimated the state background error variance. The danger
 24 zone occurs when the bias coefficient background error variance has been underestimated,
 25 but in this case, the sensitivity of the observations (ie. the value of k_{x}^{a}) does not impact
 26 whether the state analysis error variance is in the danger zone.

27 Overall, figures 4.1 and 4.2 show that, for the scalar case, the danger zone of σ_{ax}^2 is



(a) Underestimating σ_{bx}^2



(b) Overestimating σ_{bx}^2

Figure 4.2: As in figure 4.1 but varying k_x^a and σ_{bb}^{a2} . $\sigma_{b\beta}^2 = 0.75$, $\sigma_{bx}^{a2} = 1$. (a) $\sigma_{bx}^2 = 1.5$ (underestimating σ_{bx}^2) and (b) $\sigma_{bx}^2 = 0.5$ (overestimating σ_{bx}^2). White dashed line shows when $\sigma_{ax}^2 = \sigma_{bx}^2$. Black dotted line shows when $\sigma_{b\beta}^{a2} = \sigma_{b\beta}^2$.

1 avoided when the state background error variance is underestimated, as was the case in
2 a data assimilation system without bias correction (Eyre & Hilton, 2013). Furthermore,
3 the likelihood of falling into the danger zone is further reduced when the bias coefficient
4 background error variance is overestimated. In general, the likelihood of falling into the
5 danger zone is also reduced when k_x^a is larger, which can happen when more weight is
6 given to the observations, as can be seen by decreasing σ_o^2 in equation (4.21). This makes
7 sense as the higher the weighting of the observations, the less the weighting of the back-
8 ground error variances, so mis-specifying the background error variances will have less
9 of an impact. Therefore, mis-specifying the background error covariance matrices will
10 be more detrimental in regions with sparser observations, for example, over the southern
11 hemisphere in the lower troposphere, as we discussed in chapter 2.2.

12 The scalar equations for σ_{ax}^2 and $\sigma_{a\beta}^2$ mirror each other, so the results for σ_{ax}^2 and $\sigma_{a\beta}^2$
13 are symmetric between the self-sensitivity and the cross sensitivity parts. We found that
14 in order to avoid the state analysis danger zone, it is better to underestimate the state
15 background error variance and overestimate the bias coefficient background error variance.
16 Therefore, in order to avoid the bias coefficient analysis danger zone the opposite is true: it
17 is better to underestimate the bias coefficient background error variance and overestimate
18 the state background error variance. Therefore, in order to safely avoid one danger zone,
19 we risk falling into the other.

20 4.4 Conclusions and summary

21 To understand why VarBC may not produce the optimal estimate when correcting for
22 observation bias, we have studied the consequences of mis-specifying the state and obser-
23 vation bias coefficient background error covariance matrices in VarBC, in order to answer
24 research question 1.

25 Eyre and Hilton (2013) found that mis-specifying the background error covariance

1 matrix in a system without bias correction could lead to the occurrence of a danger zone,
2 in which analysis error variances are larger than the background error variances. In this
3 chapter we have extended their work to a VarBC system, to account for mis-specifying both
4 the state and bias coefficient background error covariance matrices. In VarBC, both the
5 state and bias coefficient analysis error covariance matrices can be calculated, therefore we
6 found that both a state and bias coefficient danger zone can exist. In equations (4.15) and
7 (4.18) we have shown that both of these analysis error covariance matrices will be different
8 from the optimal analysis error covariance matrices if the state and/or bias coefficient
9 background error covariance matrices have been mis-specified.

10 In order to more clearly understand the impact of mis-specifying the background error
11 covariance matrices, we have studied the analysis error variance in a scalar example (equa-
12 tions (4.19) and (4.20)), where we were easily able to vary some of the parameters. As in
13 Eyre and Hilton (2013), we defined the state danger zone to be when the state analysis
14 error variance was larger than the true state background error variance. We also extended
15 Eyre and Hilton (2013) to include a bias coefficient danger zone, which we defined as
16 when the bias coefficient analysis error variance was larger than the true bias coefficient
17 background error variance. In these danger zones, the data assimilation system was detri-
18 mental in estimating the analysis statistics, as the analysis had larger error variances than
19 the background.

20 Across four different experiments (figures 4.1 - 4.2) we varied the values given for the
21 state sensitivities to the observations and the assumed state/bias coefficient background
22 error variances to test when the danger zone was avoided. Our results showed that un-
23 derestimating the state background error variance and overestimating the bias coefficient
24 background error variance avoided the state analysis danger zone, but that the opposite
25 was true to avoid the bias coefficient analysis danger zone: we needed to overestimate the
26 state background error variance and underestimate the bias coefficient background error
27 variance. As avoiding one danger zone made the other danger zone more likely, it poses

1 the question: which danger zone is more detrimental to the overall analysis estimate? If
2 the state analysis error variance is larger than the state background error variance, then
3 the data assimilation system has not improved the state analysis estimate from the back-
4 ground estimate and hence there has been no improvement in using data assimilation to
5 estimate the state. However, if the bias coefficient analysis error variance is larger than
6 the bias coefficient background error variance, then the observation bias correction will
7 not adapt to changing observations and would perhaps be more similar to a static bias cor-
8 rection scheme, which can be more sensitive to model biases, as was discussed in chapter
9 2.3.

10 In this chapter we have mostly studied the impact of mis-specifying the background
11 error variances in a scalar system, to begin to answer research question 1. However,
12 this work should be extended to study the impact of mis-specifying the background error
13 covariance matrices in a vector system, particularly understanding the impact of mis-
14 specifying the background error covariances compared to mis-specifying the background
15 error variances.

16 Chapter 4 has demonstrated the impact of mis-specifying the state and bias coefficient
17 background error variances in VarBC, in order to understand why VarBC may give sub-
18 optimal results when correcting for observation bias. In the next chapter we will discuss
19 the design of the numerical experiments that will later be used to support our theoret-
20 ical findings in chapters that study the role of unbiased observations in the presence of
21 observation and model biases.

1 Chapter 5

2 Experimental design

3 In this chapter we will discuss the design of a simple numerical DA system that uses the
4 Lorenz 96 model (Lorenz, 1996). This system will be used in chapters 6 - 8 to demonstrate
5 our theoretical results in a multi-variable system. The Lorenz 96 model has been used
6 in many DA test systems to study predictability, particularly in weather and climate
7 prediction (e.g Brajard et al., 2020; Fertig et al., 2007; Ott et al., 2004). We have chosen
8 the Lorenz 96 model as on average the errors from a perturbation grow, and because the
9 model parameters can be chosen to control how quickly it displays chaotic behaviour, so we
10 can control how quickly a perturbed model deviates from the true model and therefore can
11 perform controlled experiments with model error. It is also useful to use a low-dimensional
12 model compared to an operational model, as it allows us to visualise our results, without
13 having to control or assign too many parameters. However, the Lorenz 96 model is a
14 simplified model compared to operational models, which means that the experiments will
15 miss out some complexities, for example, all variables have the same length scale and
16 therefore will have similar dynamics to each other.

1 5.1 Lorenz 96

2 The Lorenz 96 model (Lorenz, 1996) (which we will refer to just as Lorenz 96) is a system
3 of coupled ordinary differential equations which describe the transfer of a quantity via
4 advection, dissipation and external forcing. The system contains n variables: X_1, \dots, X_n
5 on a periodic circular domain, and is governed by n equations:

$$\frac{dX_k}{dt} = -X_{k-2}X_{k-1} + X_{k-1}X_{k+1} - X_k + F, \quad (5.1)$$

6 where $k = 0, \dots, n-1$ is the spatial index and F is independent of k . Note that as the
7 spatial domain is circular, $X_0 = X_n$ and $X_1 = X_{n+1}$. The first two terms, $-X_{k-2}X_{k-1} +$
8 $X_{k-1}X_{k+1}$, are the advection terms: they simulate the flow out of and into the k^{th} variable.
9 The third term, $-X_k$, is the internal dissipation, where a fraction of the quantity present
10 is destroyed or dissipated. The fourth term, F , is the simulated forcing which is added to
11 each variable X_k .

12 For very small values of F , the solutions of equation (5.1) decays to the steady solution,
13 such that $X_0 = \dots = X_{n-1} = F$. If F is larger, then the solutions have periodic behaviour
14 and if F is larger still, the solutions have chaotic behaviour. Karimi and Paul (2010)
15 tested how Lorenz 96 changes behaviour with different sized spatial domains (ie. varying
16 n) and with different forcing parameters (ie. varying F). They found that the solutions to
17 equation (5.1) switched between chaotic and periodic behaviour when the spatial domain
18 was varied. For example, when $n = 38$ and $F = 5$, the solution had periodic behaviour,
19 but when $n = 22$ or $n = 47$ (and $F = 5$) the solution had chaotic behaviour. Increasing
20 the forcing term to 10, regardless of the size of n , also caused the solution to have chaotic
21 behaviour. Several data assimilation experiments that have used Lorenz 96 have chosen
22 $n = 40$ and $F = 8$, which gives a solution with chaotic behaviour (e.g. Brajard et al.,
23 2020; Fertig et al., 2007; Ott et al., 2004).

1 Lorenz (1996) argues that equation (5.1) can be reasonably compared to an atmo-
2 spheric variable such as atmospheric temperature on a latitudinal circle. Each state so-
3 lution tends to lie between about -5 and 10 with no true periodicity. Also, due to the
4 symmetry in the model, all the variables have statistically similar behaviour. When the
5 number of variables is around 36 (or higher) the error doubling time for the model can
6 be comparable to the error doubling time of global circulation models, which is about 1.5
7 days, assuming 1 time unit of the model is 5 days (Lorenz, 2005). However, it should be
8 noted that this is a simplification, partly because the global circulation models have errors
9 that grow differently in different scales. For example, planetary and synoptic scales may
10 grow more slowly, whereas smaller scales are more likely to have errors that would grow
11 (and dissipate) much more quickly. Equation (5.1) does not account for these different
12 scales, but is useful for a basic comparison to some atmospheric variables.

13 In our experiments, we have numerically solved equation (5.1) using the fourth order
14 Runge-Kutta scheme, with time step $dt = 0.0125$, which is approximately equal to 1.5
15 hours (given that the time unit of equation (5.1) is 5 days). As in other previous data
16 assimilation studies that used Lorenz 96 (e.g Brajard et al., 2020; Fertig et al., 2007; Ott
17 et al., 2004), we chose $n = 40$ and $F = 8$ to give a solution with chaotic behaviour. We
18 initially ran the model for 10^5 time steps from a sine curve to allow the model to settle into
19 its climatology and took the final time step as the initial conditions for our experiments.

20 In figure 5.1 we have plotted Lorenz 96 for 100 time steps for four evenly distributed
21 state variables (\mathbf{x}_0 , \mathbf{x}_{10} , \mathbf{x}_{20} and \mathbf{x}_{30}) from our initial conditions (black lines) and from
22 perturbed initial conditions (blue lines), where the perturbation has come from adding
23 a random number from a Gaussian distribution with mean zero and variance one to the
24 initial conditions. The values of the perturbed and true trajectories roughly stay within the
25 same maxima and minima, but the perturbed trajectory deviates from the true trajectory
26 across the time steps, especially when the initial condition is very different to the true
27 initial conditions, as for state x_{10} . Increasing the time step would exacerbate the difference

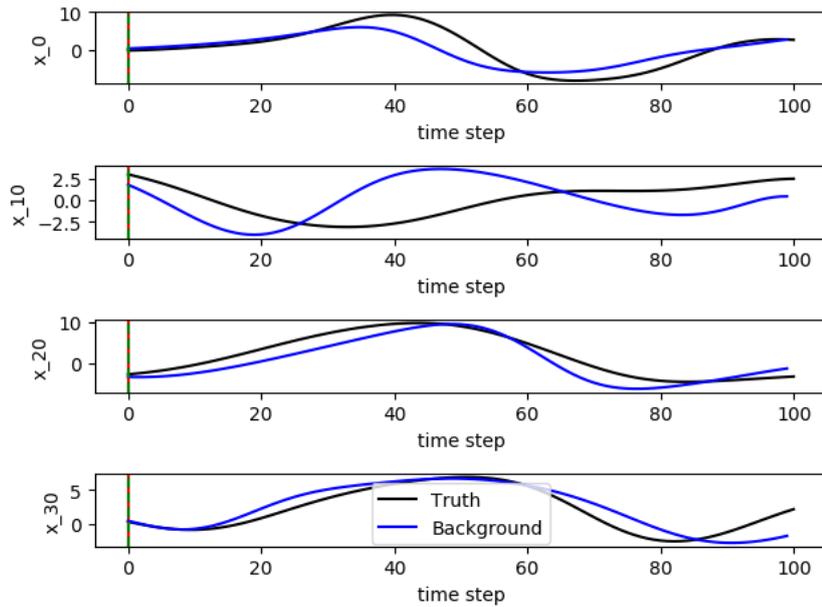


Figure 5.1: Two model runs with the Lorenz 96 model from true initial conditions (black) and perturbed initial conditions (blue) for 100 time steps, where the time step is given by 0.0125.

1 between a perturbation and the true run. In our numerical experiments in chapters 6 -
 2 8 we have chosen the time step of 0.0125 so that the system does not exhibit chaotic
 3 behaviour quickly and random perturbations do not immediately deviate substantially
 4 from the true trajectory. This allows us to have more control over the system.

5 In our numerical experiments in chapters 6 - 8 we are interested in how model bias
 6 contaminates the data assimilation system. We can add model bias to the system by
 7 changing the forcing term in equation (5.1). In figure 5.2 we have run an experiment,
 8 whereby we have changed the forcing parameter for each run. The black line has the
 9 'true' forcing of 8, and the other trajectories have forcings of 8.8 and 12 respectively.
 10 When the forcing has only been changed by 10% to 8.8, the trajectory is very similar
 11 to the true model, but when the forcing has been increased by 50%, the trajectory has
 12 changed more drastically. As we saw in chapter 2, operationally, the model biases tend
 13 to be smaller than the observation biases. Therefore, in chapter 6 we have set the biased

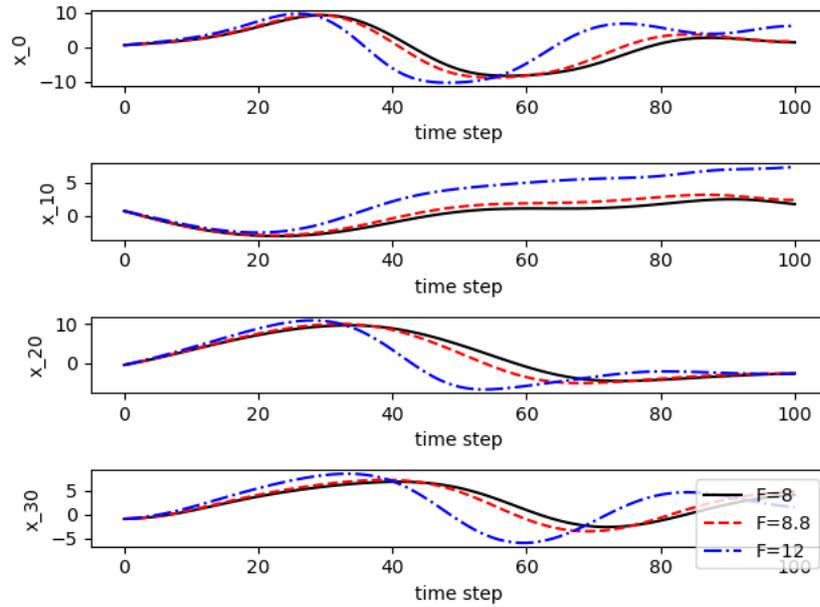


Figure 5.2: Two model runs with the Lorenz 96 model with different forcing values, initialised from the same initial conditions, time step = 0.0125.

1 model to have a forcing of 8.8, so that the model trajectory has not significantly changed,
 2 and the bias only has a small effect on the analysis. In chapters 7 and 8 we want to directly
 3 compare the observation and model bias corrections, so want the observation and model
 4 biases to have similar impacts on the analysis. Therefore, we have chosen the larger forcing
 5 of 12 in those chapters so that the model causes a similar overall bias in the analysis as
 6 the biased observations cause in the analysis. Obviously, in an operational system model
 7 biases will often be more complicated and could vary across the domain, for example land
 8 surfaces can be difficult to model and sharp vertical gradients can be difficult to capture
 9 in the equations, as we discussed in section 2.4. However, this simple introduction of
 10 model biases is useful as an initial step in understanding how well the data assimilation
 11 can estimate a model bias with known structure.

1 5.2 Data assimilation set up

2 In this section we will discuss the setup of the DA system we will use in our numerical
3 experiments.

4 Although usually in data assimilation the truth is unknown, in our numerical experi-
5 ments we are able to use the numerical solution of equation (5.1) as our true state (denoted
6 by \mathbf{x}^t), which forms the basis of our identical twin experiments. It is useful to know the
7 truth to demonstrate our theoretical findings as we can explicitly know the errors in the
8 system, so can easily compare different scenarios. We have set the data assimilation cycles
9 to occur every 10 time steps to allow perturbations in the model to slightly deviate away
10 from the truth between analyses. The cost function was minimised using the conjugate-
11 gradient method.

12 5.2.1 Simulating observations

13 We set the observations to directly observe the state, such that

$$\mathbf{y} = \mathbf{H}\mathbf{x}^t + \mathbf{e}^o \in \mathbb{R}^m, \quad (5.2)$$

14 where $\mathbf{y} \in \mathbb{R}^m$ is a vector that holds all observations, with m the number of observations;
15 $\mathbf{x}^t \in \mathbb{R}^{40}$ is a vector that holds the true state; $\mathbf{H} \in \mathbb{R}^m$ is the linearised observation operator
16 that maps the state from state space to observation space; and \mathbf{e}^o is the observation
17 error which could be random and/or systematic. We have defined the observation error
18 covariance matrix \mathbf{R} to be σ_o^2 multiplied by the identity matrix. This means that we have
19 assumed that the random part of the observation errors are uncorrelated, an assumption
20 that has been used in the past in operational numerical weather prediction in order to
21 simplify the inverse matrix, although modern numerical weather prediction systems tend
22 to now use satellite inter-channel correlations (Waller et al., 2016). \mathbf{e}^o is a vector of

1 random numbers from a Gaussian distribution with mean b^o and variance σ_o^2 . We will
 2 set some observations to have a known bias and some to be unbiased (which will be the
 3 anchor observations). For the biased observations, b^o is nonzero so that the observation
 4 error is both random and biased. For the unbiased observations, b^o is zero such that the
 5 observation error is random and unbiased.

6 Operationally, the majority of observations come from indirect observations such as
 7 radiance observations (Lorenc & Marriott, 2014). In our numerical experiments we have
 8 chosen to include only direct observations, as this simplification allows us to understand
 9 where the errors in the observations come from, as well as the structure of the errors.
 10 Therefore the linearised observation operator, \mathbf{H} , has values of 1 where the state is observed
 11 and 0 if the state is not observed.

12 5.2.2 Generating the background error covariance matrix

13 At the initial time the background states are defined as the truth plus an error, given by,

$$\mathbf{x}^b = \mathbf{x}^t + \mathbf{e}_x^b \in \mathbb{R}^{40}, \quad (5.3)$$

14 where \mathbf{x}^b is a vector containing the background states and \mathbf{e}_x^b is the background error
 15 which could be random and/or systematic. \mathbf{e}_x^b is a vector of random numbers from a
 16 Gaussian distribution with mean \mathbf{b}^b and error covariance matrix \mathbf{B}_x . If elements from \mathbf{b}^b
 17 are zero, then the background states associated with those elements will have unbiased
 18 random error and if elements of \mathbf{b}^b are nonzero then the background states associated
 19 with those elements will have random and biased error. Two forms of \mathbf{B}_x are used within
 20 this thesis. One is created using the climatological background error covariance matrix of
 21 the system, which is explained below in section 5.2.2.1. This is used so that there are no
 22 additional errors from mis-specifying the background error covariance matrix, as explored
 23 in chapter 4. The other background error covariance matrix used has error variances

1 given by a prescribed σ_{bx}^2 and error correlations given by the SOAR (second order auto
 2 regressive) correlation function, which has previously been used to model the background
 3 error correlations of the atmosphere (e.g. Ingleby, 2001; Simonin et al., 2014). The SOAR
 4 correlation function is given by,

$$s_k = \left(1 + \frac{k}{L_b}\right) e^{-\frac{k}{L_b}} \quad (5.4)$$

5 where k is the index from 1 to $\frac{n}{2}$; L_b is the length scale; and e is the exponential function.
 6 For values greater than $\frac{n}{2}$, s_k is repeated in opposite order such that the vector \mathbf{s} (which
 7 contains all s_k) is palindromic. To calculate \mathbf{B}_x , \mathbf{s} is transformed into a circular matrix
 8 and multiplied by σ_{bx}^2 . We have used the SOAR correlation function to calculate \mathbf{B}_x , to
 9 control the length scales of the error correlations.

10 At subsequent cycles, the background states are taken to be the analysis states of the
 11 previous cycle after they have been evolved forward via the numerical solution of equation
 12 (5.1) to the correct time step. However, the background error covariance matrix is taken
 13 to be static, such that it is constant for different assimilation cycles.

14 **5.2.2.1 Generating a climatological \mathbf{B}**

15 Note that this explanation of calculating the climatological background error covariance
 16 matrix has been adapted from our description in Francis et al. (2023).

17 A sample estimate of the climatological \mathbf{B}_x matrix is derived from an ensemble of DA
 18 experiments cycled in time. The estimate of \mathbf{B}_x is sensitive to the assumed value of \mathbf{B}_x used
 19 within the data assimilation system for the experiments. We therefore repeat the method
 20 a second time with the new estimate of \mathbf{B}_x , to get a second estimate of \mathbf{B}_x . We found
 21 that two iterations were sufficient for the estimate of \mathbf{B}_x to converge to an appropriate
 22 climatological estimate for Lorenz 96, assimilating observations at every spatial variable.

23 On the first iteration, the background error covariance matrix for the state is calculated

1 using the SOAR error correlation function as in equation (5.4), with error variance $\sigma_{\text{bx}}^2 = 1$
2 and length scale $L_b = 1$. At the initial time, an ensemble of 15 background states are
3 calculated using equation (5.3), each with a different random error and $\mathbf{b}^b = \mathbf{0}$. From the
4 observations and background states, an ensemble of analyses was generated. The analysis
5 ensemble was then evolved forward via the model by 10 time steps to give the background
6 ensemble at the start of the next assimilation window. The background ensemble at any
7 given time step greater than 0 is given by,

$$\mathbf{x}_{t+1,i}^b = m_t(\mathbf{x}_{t,i}^a), \quad \beta_{t+1,i}^b = \beta_{t,i}^a, \quad (5.5)$$

8 where t is the time index; i is the ensemble member index; and $m_t(\mathbf{x}_{t,i})$ is the assimilation
9 model that takes the state from time step t to $t + 1$. Note that if the forcing has been
10 changed to add a model bias, then m_t could be different to the true model. The errors
11 between the background of each ensemble member and the true state at the same time
12 are given by,

$$\boldsymbol{\epsilon}_{\mathbf{x}_{t,i}}^b = \mathbf{x}_{t,i}^b - \mathbf{x}_{t,i}^t. \quad (5.6)$$

13 The analysis ensemble was cycled for 700 assimilation windows to provide in total 10500
14 samples (700 cycles \times 15 realisations) of the background error, equation (5.6). 700 cycles
15 were chosen to ensure that \mathbf{B}_x was not dependent on a particular time step, so that it
16 could be used statically across multiple windows. We chose to only use 15 realisations as
17 we needed some realisations to account for the background error at the initial cycle, but
18 as there was already a lot of data from the 700 cycles, to save on computing power, a
19 small number of realisations was adequate. From these 10500 samples of $\boldsymbol{\epsilon}_x^b$ we were then
20 able to update the climatological estimate of the \mathbf{B}_x matrix, by calculating the sample

1 covariances of all state background errors $\epsilon_{x_{t,i}}^b$,

$$\text{Cov}(\epsilon_{x_{j,t,i}}^b, \epsilon_{x_{k,t,i}}^b) = \frac{\sum_{t=1}^{700} \sum_{i=1}^{15} (\epsilon_{x_{j,t,i}}^b - \bar{\epsilon}_{x_j}^b)(\epsilon_{x_{k,t,i}}^b - \bar{\epsilon}_{x_k}^b)}{10500 - 1}, \quad (5.7)$$

2 where $\epsilon_{x_{j,t,i}}^b$ and $\epsilon_{x_{k,t,i}}^b$ are the background errors for the j^{th} and k^{th} variables of the state
3 at time t , ensemble member i ; and $\bar{\epsilon}_{x_j}^b$ and $\bar{\epsilon}_{x_k}^b$ are the mean background errors from the
4 10500 samples at state variables j and k respectively. If the model has a bias, then the
5 background errors will also be biased. However, as we are comparing the background
6 errors to each other (and they would all have the same bias), the bias would not affect the
7 error covariances, as the values are just how much they vary around the mean.

8 To remove noise caused by the limited sample estimate of \mathbf{B}_x , we set the covariances
9 between state variables further than 5 grid points apart to 0. An example of the back-
10 ground error covariance matrix from the initial iteration is shown in figure 5.3, when the
11 observation and background error variances are initially equal and set to 1 and the true
12 forcing is used. Figure 5.3a shows the random noise on the error covariances and figure
13 5.3b shows \mathbf{B}_x when the noise has been removed. As the state variables are on a circular
14 domain, x_0 and x_{39} have a correlation to each other as they would be next to each other
15 in the domain.

16 Using the new \mathbf{B}_x , a new background error covariance matrix was calculated from
17 equations (5.5) to (5.7), again using 700 cycles and 15 realisations. After removing the
18 random noise, this background error covariance matrix was taken to be the climatological
19 background error covariance matrix in the numerical experiments, as further iterations
20 were not found to make significant changes. Again, an example of the state background
21 error covariance matrix is shown in figure 5.4, when the observation and background error
22 variances are initialised to equal 1 and the true F is used. Figure 5.4a shows the back-
23 ground error covariance matrix after the cycling, and figure 5.4b shows the background
24 error covariance matrix after the random noise has been removed. When the observa-

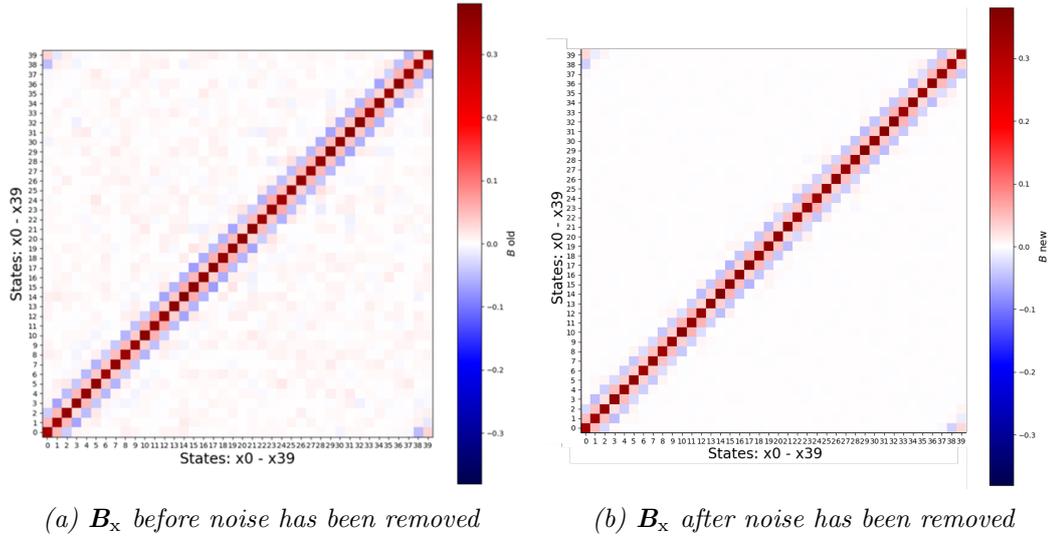


Figure 5.3: The climatological B_x calculated from an ensemble, after the initial iteration.

1 tion error variances are larger, then the state analysis error variances are larger, so the
 2 background error variances in the next cycle are also larger and when the observation
 3 error variances are smaller, the opposite is true, ie. the background error variances of the
 4 next cycles are smaller. However, the general structure of B_x remains the same for all
 experiments that use the climatological B_x .

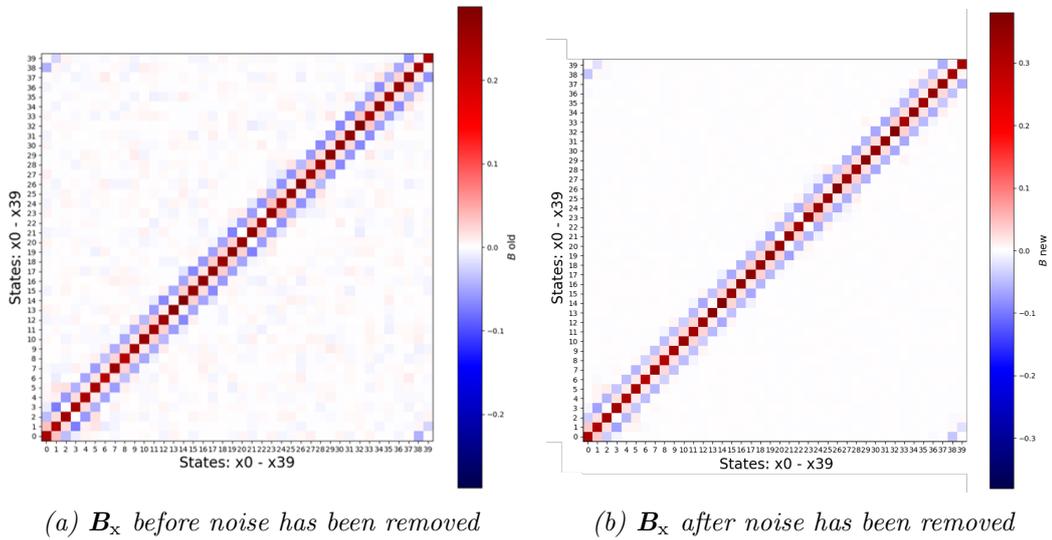


Figure 5.4: The climatological B_x calculated from an ensemble, after the second iteration.

1 5.2.3 Correcting for observation bias (VarBC)

2 We will now describe the observation bias correction setup that uses VarBC, as described
3 in section 2.3. The observation bias is added to the observations linearly, as shown in
4 equation (5.2), with a constant bias across all observations. This means that, in this
5 case, the observation bias correction is not a function of \mathbf{x}^b , only a function of the scalar
6 observation bias coefficient such that it is given by,

$$c(\mathbf{x}^b, \boldsymbol{\beta}) = \beta, \quad (5.8)$$

7 where the β that is to be predicted is a direct estimate of the true bias ($\beta^t = b^o$). This
8 can be compared to equation (2.41), where we have chosen one predictor, with $p(\mathbf{x}) = 1$.
9 The linearised bias correction about the state (\mathbf{C}_x) and observation bias coefficient (\mathbf{C}_β)
10 respectively as in equation (2.48) are therefore,

$$C_x = 0, \quad C_\beta = 1. \quad (5.9)$$

11 As the bias is constant across all observations, β is a scalar. Therefore the control vector
12 is given by,

$$\mathbf{v} = \begin{pmatrix} \mathbf{x} \\ \beta \end{pmatrix}. \quad (5.10)$$

13 At the initial cycle, the background observation bias coefficient is defined as the true
14 observation bias coefficient plus an error,

$$\beta^b = \beta^t + e_\beta^b, \quad (5.11)$$

1 where β^t is the true observation bias coefficient; and e_β^b is the error in the observation bias
2 coefficient background, generated by a random number from a Gaussian distribution with
3 mean zero and variance $\sigma_{b\beta}^2$. As for the state background error covariance matrix, $\sigma_{b\beta}^2$
4 is defined in two different ways, depending on the experiment. Either the climatological
5 $\sigma_{b\beta}^2$ is calculated simultaneously with the state background error covariance matrix, in the
6 same method as described in section 5.2.2.1, or it is predefined by a given value, which
7 will be defined in each experiment. When the climatological $\sigma_{b\beta}^2$ is calculated, it is an
8 order of magnitude smaller than the state background error variances. This is because the
9 observation bias is set to be constant across cycles, so there is less room for error growth.
10 The background error covariances between the state and the observation bias coefficient
11 are manually set to zero, to align with the separation of the state and observation bias
12 coefficient in the cost function, equation (2.45).

13 In subsequent cycles, β^b is defined as the β^a from the previous cycle. Note that the
14 observation bias coefficient is considered to be roughly constant between cycles, so the
15 model evolving the observation bias coefficient analysis between cycles is just the identity.

16 **5.2.4 Correcting for model bias (WC4DVar)**

17 In this section we will describe the model bias correction set up, that uses WC4DVar, as
18 described in chapter 2.5. A model or background bias can be added to the system in two
19 ways: by adding a biased error to the background (ie. \mathbf{b}^b is nonzero) to give a background
20 bias; or by changing the forcing term in equation (5.1) to give a model bias, such that the
21 model that evolves the state forward in time is different to the model that generates the
22 observations. Note that artificially adding a background bias is only useful if the system is
23 not cycled, as a biased analysis will lead to a biased background in the next cycle anyway.

24 If the forcing term in equation (5.1) is changed from F to F_{biased} , then a model bias
25 will occur in the form of

$$\mathbf{x}_{i+1}^t = m_i(\mathbf{x}_i^t) + \boldsymbol{\eta}^t, \quad (5.12)$$

1 where \mathbf{x}_i is the state at time i ; m_i is the assimilation model which now uses F_{biased} instead
 2 of F ; and $\boldsymbol{\eta}^t$ is the true model bias, calculated by

$$\boldsymbol{\eta}^t = (F_{\text{biased}} - F) \times dt. \quad (5.13)$$

3 As F_{biased} is constant across all states, the model bias can be estimated as a scalar η
 4 multiplied by a vector of ones, such that the control vector for WC4DVar becomes,

$$\mathbf{p} = \begin{pmatrix} \mathbf{x}_0 \\ \eta \end{pmatrix}. \quad (5.14)$$

5 The background model bias estimate is defined as the true model bias plus a random error,

$$\eta^b = \eta^t + e_\eta^b, \quad (5.15)$$

6 where e_η^b is the error in the model bias estimate background, given by a random number
 7 from a Gaussian distribution with mean zero and variance $\sigma_{b\eta}^2$. The background model
 8 bias estimate error variance is defined as ten percent of the average state background
 9 error variances, as the model errors will be much smaller than the background errors,
 10 because the model errors are the errors in the estimate of η at one time step, whereas the
 11 background errors are the accumulation of errors across the whole window. The model
 12 bias is taken to be approximately constant between cycles so that η^b in future cycles is
 13 given by the η^a of the previous cycle (so the model between cycles for the model bias
 14 estimate is the identity).

15 **5.3 Summary**

16 In chapter 5 we have discussed the set up for the numerical experiments that will be used
 17 in chapters 6-8. The Lorenz 96 model was described in more detail, with some basic exper-

1 iments to demonstrate the dynamical behaviour. An explanation of the data assimilation
2 set up was given, as well as an explanation of the two bias correction techniques used. In
3 the next chapter, we study the characteristics of anchor observations needed to be able to
4 reduce the contamination of model bias in observation bias correction in 3DVar to begin to
5 answer research question 2. We will use numerical experiments with the set up described
6 in chapter 5 to demonstrate our theoretical results.

1 Chapter 6

2 The role of anchor observations in 3 VarBC in the presence of model 4 bias. Part I: The importance of 5 the location of anchor observations

6 6.1 Introduction

7 In section 3.2 we discussed the results of Eyre (2016) who demonstrated, in a scalar system,
8 the need for anchor observations when correcting for observation bias in the presence of
9 model bias. Eyre (2016) showed that, if model bias is present in a VarBC system, then, as
10 the number of anchor observations reduces, the observation bias correction is more affected
11 by model bias, so the analysis will be pulled towards the model bias. In this study, we
12 extend the work of Eyre (2016) by looking at a multivariate system with explicit random
13 error to understand the importance of the locations of anchor observations and the anchor
14 observation uncertainty characteristics in reducing the effect of the model bias on the

1 observation bias correction in order to answer research questions 2.1 and 2.3. Note that
2 this chapter has been adapted for this thesis from our paper: Francis et al. (2023).

3 In section 6.2 we will extend current VarBC theory to include both bias-corrected
4 and anchor observations to demonstrate the role of anchor observations in reducing the
5 contamination of model bias in 3DVarBC: in section 6.2.1 we study the importance of the
6 location of the anchor observations relative to the bias-corrected observations; in section
7 6.2.2 we study the importance of the location of the anchor observations relative to the
8 locations of model bias; and in section 6.2.3 we demonstrate how bias in the analysed bias
9 correction coefficients filters into the state analysis in subsequent cycles. In section 6.3
10 we test the theory using the idealised 40 variable model of the atmosphere as described
11 in chapter 5 to show: how the observation bias coefficients are affected by model bias,
12 depending on the precision of the anchor observations (section 6.3.1.1); the effect of anchor
13 and bias-corrected observations observing different parts of the state (section 6.3.1.2);
14 and how the model bias contaminates both the observation bias correction and the state
15 estimation when VarBC is cycled (section 6.3.2.1). Finally we present our conclusions and
16 further discussion in section 6.4.

17 **6.2 The importance of anchor observations in 3DVarBC:** 18 **theoretical results**

19 We can extend current VarBC theory, as described in section 2.3, to include two obser-
20 vation types: one that is bias-corrected, which will be labelled $\mathbf{y}_{(1)}$, and one that is not
21 bias-corrected, which will act as our anchor observations, labelled $\mathbf{y}_{(2)}$. Note that we
22 assume that the anchor observations are unbiased.

23 In order to separate the roles of the bias-corrected and anchor observations, we write

1 equations (2.54) and (2.55) in terms of $\mathbf{y}_{(1)} \in \mathbb{R}^{m_1}$ and $\mathbf{y}_{(2)} \in \mathbb{R}^{m_2}$, as

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}_{xy(1)} \mathbf{d}_{(1)} + \mathbf{K}_{xy(2)} \mathbf{d}_{(2)}, \quad (6.1)$$

$$\boldsymbol{\beta}^a = \boldsymbol{\beta}^b + \mathbf{K}_{\beta y(1)} \mathbf{d}_{(1)} + \mathbf{K}_{\beta y(2)} \mathbf{d}_{(2)}, \quad (6.2)$$

2 where we have assumed that there are no correlations between the errors in the bias-
3 corrected and anchor observations and where the innovation vectors for the bias-corrected
4 and anchor observations are given by,

$$\mathbf{d}_{(1)} = \mathbf{y}_{(1)} - h_{(1)}(\mathbf{x}^b) - c(\mathbf{x}^b, \boldsymbol{\beta}^b); \quad \mathbf{d}_{(2)} = \mathbf{y}_{(2)} - h_{(2)}(\mathbf{x}^b). \quad (6.3)$$

5 We have denoted $\mathbf{K}_{xy(1)}$ and $\mathbf{K}_{xy(2)}$ to be the sensitivities of the state analysis to the
6 bias-corrected and anchor observations respectively; and $\mathbf{K}_{\beta y(1)}$ and $\mathbf{K}_{\beta y(2)}$ to be the sen-
7 sitivities of the bias coefficient analysis to the bias-corrected and anchor observations
8 respectively. These four Kalman gain matrices are stored in the matrix \mathbf{K}_v :

$$\mathbf{K}_v = \mathbf{B}_v \mathbf{H}_v^T (\mathbf{H}_v \mathbf{B}_v \mathbf{H}_v^T + \mathbf{R})^{-1} = \begin{pmatrix} \mathbf{K}_{xy(1)} & \mathbf{K}_{xy(2)} \\ \mathbf{K}_{\beta y(1)} & \mathbf{K}_{\beta y(2)} \end{pmatrix} \in \mathbb{R}^{(n+r) \times (m_1+m_2)}, \quad (6.4)$$

9 where \mathbf{H}_v now includes the linearised observation operators from both types of observa-
10 tions and is given by,

$$\mathbf{H}_v(\mathbf{v}) = \begin{pmatrix} \mathbf{H}_{(1)} & \mathbf{C}_\beta \\ \mathbf{H}_{(2)} & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{(m_1+m_2) \times (n+r)}. \quad (6.5)$$

11 $\mathbf{H}_{(1)}$ is the Jacobian of the bias-corrected observation operator $h_{(1)}(\mathbf{x})$; $\mathbf{H}_{(2)}$ is the Jacobian
12 of the anchor observation operator $h_{(2)}(\mathbf{x})$; and \mathbf{C}_β is the Jacobian of $c(\mathbf{x}^b, \boldsymbol{\beta})$ with respect
13 to $\boldsymbol{\beta}$. \mathbf{B}_v is the background error covariance matrix for the state (\mathbf{B}_x) and the bias

1 coefficient (\mathbf{B}_β), assuming that there are no error correlations between the state and
 2 observation bias coefficient, as given by equation (2.43) in section 2.3. The observation
 3 error covariance matrix (\mathbf{R}) for the bias-corrected and anchor observations, is given by,

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_{(1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{(2)} \end{pmatrix} \in \mathbb{R}^{(m_1+m_2) \times (m_1+m_2)}. \quad (6.6)$$

4 From equation (6.2) the sensitivities of the bias coefficient analysis to the bias-corrected
 5 observations, anchor observations, state background and bias coefficient background re-
 6 spectively are given by:

$$\frac{\partial \beta^a}{\partial \mathbf{y}_{(1)}} = \mathbf{K}_{\beta y_{(1)}}, \quad (6.7)$$

$$\frac{\partial \beta^a}{\partial \mathbf{y}_{(2)}} = \mathbf{K}_{\beta y_{(2)}}, \quad (6.8)$$

$$\frac{\partial \beta^a}{\partial \mathbf{x}^b} = -\mathbf{K}_{\beta y_{(1)}} \mathbf{H}_{(1)} - \mathbf{K}_{\beta y_{(2)}} \mathbf{H}_{(2)}, \quad (6.9)$$

$$\frac{\partial \beta^a}{\partial \beta^b} = \mathbf{I} - \mathbf{K}_{\beta y_{(1)}} \mathbf{C}_\beta, \quad (6.10)$$

7 where we have used the convention that a partial derivative is a row vector. We will use
 8 equation (6.9) to study the dependency of β^a on the state background in 3DVarBC, which
 9 is the mechanism for model bias to be passed into the bias coefficient analysis.

10 In order to understand the sensitivity of β^a to \mathbf{x}^b via equation (6.9), we need to
 11 know what $\mathbf{K}_{\beta y_{(1)}}$ and $\mathbf{K}_{\beta y_{(2)}}$ are dependent on. Expanding equation (6.4) into its \mathbf{x} , β ,
 12 bias-corrected and anchor observation parts gives,

$$\mathbf{K}_v = \begin{pmatrix} \mathbf{B}_x \mathbf{H}_{(1)}^T & \mathbf{B}_x \mathbf{H}_{(2)}^T \\ \mathbf{B}_\beta \mathbf{C}_\beta^T & \mathbf{0} \end{pmatrix} \times \begin{pmatrix} \mathbf{H}_{(1)} \mathbf{B}_x \mathbf{H}_{(1)}^T + \mathbf{C}_\beta \mathbf{B}_\beta \mathbf{C}_\beta^T + \mathbf{R}_{(1)} & \mathbf{H}_{(1)} \mathbf{B}_x \mathbf{H}_{(2)}^T \\ \mathbf{H}_{(2)} \mathbf{B}_x \mathbf{H}_{(1)}^T & \mathbf{H}_{(2)} \mathbf{B}_x \mathbf{H}_{(2)}^T + \mathbf{R}_{(2)} \end{pmatrix}^{-1}. \quad (6.11)$$

1 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
 2 calculate its inverse by denoting,

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

3 such that the inverse of a block symmetric matrix is given by,

$$= \begin{pmatrix} (\mathbf{W} - \mathbf{XZ}^{-1}\mathbf{X}^T)^{-1} & -(\mathbf{W} - \mathbf{XZ}^{-1}\mathbf{X}^T)^{-1}\mathbf{XZ}^{-1} \\ -\mathbf{Z}^{-1}\mathbf{X}^T(\mathbf{W} - \mathbf{XZ}^{-1}\mathbf{X}^T)^{-1} & \mathbf{Z}^{-1} + \mathbf{Z}^{-1}\mathbf{X}^T(\mathbf{W} - \mathbf{XZ}^{-1}\mathbf{X}^T)^{-1}\mathbf{XZ}^{-1} \end{pmatrix}, \quad (6.12)$$

4 where

$$\mathbf{W} = \mathbf{H}_{(1)} \mathbf{B}_x \mathbf{H}_{(1)}^T + \mathbf{C}_\beta \mathbf{B}_\beta \mathbf{C}_\beta^T + \mathbf{R}_{(1)}, \quad (6.13)$$

$$\mathbf{X} = \mathbf{H}_{(1)} \mathbf{B}_x \mathbf{H}_{(2)}^T, \quad (6.14)$$

$$\mathbf{Z} = \mathbf{H}_{(2)} \mathbf{B}_x \mathbf{H}_{(2)}^T + \mathbf{R}_{(2)}. \quad (6.15)$$

5 Hence combining equations (6.11) and (6.12), the expressions for $\mathbf{K}_{xy(1)}$, $\mathbf{K}_{xy(2)}$, $\mathbf{K}_{\beta y(1)}$
 6 and $\mathbf{K}_{\beta y(2)}$ are given by,

$$\begin{aligned} \mathbf{K}_{xy(1)} &= \mathbf{B}_x \mathbf{H}_{(1)}^T (\mathbf{W} - \mathbf{XZ}^{-1}\mathbf{X}^T)^{-1} \\ &\quad - \mathbf{B}_x \mathbf{H}_{(2)}^T \mathbf{Z}^{-1} \mathbf{X}^T (\mathbf{W} - \mathbf{XZ}^{-1}\mathbf{X}^T)^{-1}, \end{aligned} \quad (6.16)$$

$$\begin{aligned} \mathbf{K}_{xy(2)} &= -\mathbf{B}_x \mathbf{H}_{(1)}^T (\mathbf{W} - \mathbf{XZ}^{-1}\mathbf{X}^T)^{-1} \mathbf{XZ}^{-1} \\ &\quad + \mathbf{B}_x \mathbf{H}_{(2)}^T (\mathbf{Z}^{-1} + \mathbf{Z}^{-1} \mathbf{X}^T (\mathbf{W} - \mathbf{XZ}^{-1}\mathbf{X}^T)^{-1} \mathbf{XZ}^{-1}), \end{aligned} \quad (6.17)$$

$$\mathbf{K}_{\beta y(1)} = \mathbf{B}_\beta \mathbf{C}_\beta^T (\mathbf{W} - \mathbf{XZ}^{-1}\mathbf{X}^T)^{-1}, \quad (6.18)$$

$$\mathbf{K}_{\beta y(2)} = -\mathbf{B}_\beta \mathbf{C}_\beta^T (\mathbf{W} - \mathbf{XZ}^{-1}\mathbf{X}^T)^{-1} \mathbf{XZ}^{-1}. \quad (6.19)$$

1 Therefore we can rewrite $\mathbf{K}_{\beta_{y(2)}}$ in terms of $\mathbf{K}_{\beta_{y(1)}}$:

$$\mathbf{K}_{\beta_{y(2)}} = -\mathbf{K}_{\beta_{y(1)}} \mathbf{H}_{(1)} \mathbf{B}_x \mathbf{H}_{(2)}^T (\mathbf{H}_{(2)} \mathbf{B}_x \mathbf{H}_{(2)}^T + \mathbf{R}_{(2)})^{-1}. \quad (6.20)$$

2 This shows that the sensitivity of the bias coefficient analysis to the anchor observations
 3 is dependent on the sensitivity of the bias coefficient analysis to the bias-corrected obser-
 4 vations and vice versa ($\mathbf{K}_{\beta_{y(1)}}$ is also dependent on $\mathbf{K}_{\beta_{y(2)}}$).

5 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
 6 coefficient analysis to the state background, equation (6.9), so that it is written in terms
 7 of $\mathbf{K}_{\beta_{y(1)}}$:

$$\frac{\partial \beta^a}{\partial \mathbf{x}^b} = -\mathbf{K}_{\beta_{y(1)}} \mathbf{H}_{(1)} (\mathbf{I} - \mathbf{D}), \quad (6.21)$$

8 where we have defined \mathbf{D} as,

$$\mathbf{D} = \mathbf{B}_x \mathbf{H}_{(2)}^T (\mathbf{H}_{(2)} \mathbf{B}_x \mathbf{H}_{(2)}^T + \mathbf{R}_{(2)})^{-1} \mathbf{H}_{(2)}. \quad (6.22)$$

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

$$\boldsymbol{\epsilon}_x^a = \mathbf{x}^a - \mathbf{x}^t, \quad \boldsymbol{\epsilon}_\beta^a = \boldsymbol{\beta}^a - \boldsymbol{\beta}^t, \quad (6.23)$$

$$\boldsymbol{\epsilon}_x^b = \mathbf{x}^b - \mathbf{x}^t, \quad \boldsymbol{\epsilon}_\beta^b = \boldsymbol{\beta}^b - \boldsymbol{\beta}^t, \quad \boldsymbol{\epsilon}_v^b = \mathbf{v}^b - \mathbf{v}^t, \quad (6.24)$$

$$\boldsymbol{\epsilon}_1^o = \mathbf{y}_{(1)} - h_{(1)}(\mathbf{x}^t) - c(\mathbf{x}^t, \boldsymbol{\beta}^t), \quad \boldsymbol{\epsilon}_2^o = \mathbf{y}_{(2)} - h_{(2)}(\mathbf{x}^t), \quad (6.25)$$

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

15 The expected value of the analysis errors in the state and bias coefficient can be

1 calculated as in equations (2.65) and (2.66), but including both observation types and is
 2 given by,

$$\langle \boldsymbol{\epsilon}_x^a \rangle = \langle \boldsymbol{\epsilon}_x^b \rangle + \mathbf{K}_{xy(1)} \langle \boldsymbol{\epsilon}_1^o \rangle - \mathbf{K}_{xy(1)} \mathbf{H}_{v(1)} \langle \boldsymbol{\epsilon}_v^b \rangle + \mathbf{K}_{xy(2)} \langle \boldsymbol{\epsilon}_2^o \rangle - \mathbf{K}_{xy(2)} \mathbf{H}_{v(2)} \langle \boldsymbol{\epsilon}_v^b \rangle, \quad (6.26)$$

$$\langle \boldsymbol{\epsilon}_\beta^a \rangle = \langle \boldsymbol{\epsilon}_\beta^b \rangle + \mathbf{K}_{\beta y(1)} \langle \boldsymbol{\epsilon}_1^o \rangle - \mathbf{K}_{\beta y(1)} \mathbf{H}_{v(1)} \langle \boldsymbol{\epsilon}_v^b \rangle + \mathbf{K}_{\beta y(2)} \langle \boldsymbol{\epsilon}_2^o \rangle - \mathbf{K}_{\beta y(2)} \mathbf{H}_{v(2)} \langle \boldsymbol{\epsilon}_v^b \rangle, \quad (6.27)$$

3 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 (6.5). If the right
 4 hand side of equation (6.26) is nonzero then the state analysis is biased. If the right hand
 5 side of equation (6.27) is nonzero then the bias coefficient analysis has a bias. Biases in β^a
 6 will filter into \mathbf{x}^a when β^a becomes β^b in the next cycle, as will be shown in more depth
 7 in section 6.2.3.

8 In this chapter we are interested in the effect of model bias in the form of background
 9 bias, as any model bias will accumulate across the window and be seen as background
 10 bias in the next cycle. We therefore assume $\langle \boldsymbol{\epsilon}_x^b \rangle$ is nonzero. Assuming that $c(\mathbf{x}^t, \boldsymbol{\beta}^t)$ is
 11 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.
 12 We assume the anchor observations are unbiased ($\langle \boldsymbol{\epsilon}_2^o \rangle = \mathbf{0}$). To isolate the effect of
 13 background bias on the bias coefficient analysis, we make the theoretical assumption that
 14 the initial bias coefficient background has no bias ($\langle \boldsymbol{\epsilon}_\beta^b \rangle = \mathbf{0}$). In subsequent cycles, $\langle \boldsymbol{\epsilon}_\beta^b \rangle$
 15 will be propagated from the previous cycle, so this assumption can only be possible for the
 16 first cycle. We will demonstrate the impact that a biased background bias coefficient has
 17 on the state and bias coefficient analyses in sections 6.2.3 and 6.3.2.1. Therefore, when
 18 there is only background bias in the system, equation (6.27) becomes:

$$\begin{aligned} \langle \boldsymbol{\epsilon}_\beta^a \rangle \Big|_{\langle \boldsymbol{\epsilon}_1^o \rangle = \mathbf{0}, \langle \boldsymbol{\epsilon}_2^o \rangle = \mathbf{0}, \langle \boldsymbol{\epsilon}_\beta^b \rangle = \mathbf{0}} &= -\mathbf{K}_{\beta y(1)} \begin{pmatrix} \mathbf{H}_{(1)}, & \mathbf{C}_\beta \end{pmatrix} \begin{pmatrix} \langle \boldsymbol{\epsilon}_x^b \rangle \\ \mathbf{0} \end{pmatrix} - \mathbf{K}_{\beta y(2)} \begin{pmatrix} \mathbf{H}_{(2)} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \langle \boldsymbol{\epsilon}_x^b \rangle \\ \mathbf{0} \end{pmatrix}, \\ &= -\mathbf{K}_{\beta y(1)} \mathbf{H}_{(1)} \langle \boldsymbol{\epsilon}_x^b \rangle - \mathbf{K}_{\beta y(2)} \mathbf{H}_{(2)} \langle \boldsymbol{\epsilon}_x^b \rangle. \end{aligned} \quad (6.28)$$

1 As we can rewrite $\mathbf{K}_{\beta_{Y(2)}}$ in terms of $\mathbf{K}_{\beta_{Y(1)}}$ as in equation (6.20), this simplifies to,

$$\left. \langle \boldsymbol{\epsilon}_{\beta}^a \rangle \right|_{\langle \boldsymbol{\epsilon}_1^o \rangle = \mathbf{0}, \langle \boldsymbol{\epsilon}_2^o \rangle = \mathbf{0}, \langle \boldsymbol{\epsilon}_{\beta}^b \rangle = \mathbf{0}} = -\mathbf{K}_{\beta_{Y(1)}} \mathbf{H}_{(1)} (\mathbf{I} - \mathbf{D}) \langle \boldsymbol{\epsilon}_x^b \rangle := \langle \boldsymbol{\epsilon}_{\beta}^a \rangle_{t=0}, \quad (6.29)$$

2 where \mathbf{D} is as defined in equation (6.22) and we denote $\langle \boldsymbol{\epsilon}_{\beta}^a \rangle$ with the above assumptions to
 3 be $\langle \boldsymbol{\epsilon}_{\beta}^a \rangle_{t=0}$ to highlight that this is only valid for the initial cycle. In the next few sections
 4 we will study equation (6.29) to see how $\langle \boldsymbol{\epsilon}_{\beta}^a \rangle_{t=0}$ is affected by background bias when we
 5 vary the anchor observation parameters.

6 The effect of the background bias on β^a will be small if at least one term in the
 7 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
 8 coefficient analysis to the bias-corrected observations, equation (6.7), would be small. If
 9 $\mathbf{H}_{(1)}$ is small, then the bias-corrected observations would not be used to determine the
 10 state analysis. Therefore, as these are both uninteresting cases for determining the bias
 11 coefficient analysis, in this study we will focus instead on when the magnitude of $\mathbf{I} - \mathbf{D}$ is
 12 small depending on the parameters given in the system.

13 Note $\langle \boldsymbol{\epsilon}_{\beta}^a \rangle_{t=0}$ can also be rewritten in terms of the sensitivity of the bias coefficient
 14 analysis to the state background, equation (6.21), such that equation (6.29) becomes

$$\langle \boldsymbol{\epsilon}_{\beta}^a \rangle_{t=0} = \frac{\partial \beta^a}{\partial \mathbf{x}^b} \langle \boldsymbol{\epsilon}_x^b \rangle. \quad (6.30)$$

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

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

1 **6.2.1 Position of anchor observations relative to bias-corrected observa-**
2 **tions**

3 In this section we want to understand how anchor observations can reduce the contamina-
4 tion of model bias on the bias coefficient analysis depending on whether the bias-corrected
5 and anchor observations observe the same state variables.

6 **6.2.1.1 Anchor observations fully observe the domain**

7 In order to understand the role of anchor observations in reducing the contamination
8 of model bias in VarBC, we first consider an almost perfect case where we have anchor
9 observations everywhere in the domain. If $\mathbf{H}_{(1)}$ and $\mathbf{H}_{(2)}$ are both equal to the identity,
10 that is, the state is fully observed directly by both bias-corrected and anchor observations,
11 then \mathbf{D} , equation (6.22) is given by,

$$\mathbf{D}|_{\mathbf{H}_{(2)}=\mathbf{I}} = \mathbf{B}_x(\mathbf{B}_x + \mathbf{R}_{(2)})^{-1} := \mathbf{D}_I. \quad (6.31)$$

12 Equation (6.29) is then given by,

$$\langle \epsilon_\beta^a \rangle_{t=0} |_{\mathbf{D}=\mathbf{D}_I} = -\mathbf{K}_{\beta y_{(1)}} |_{\mathbf{H}_{(1)}=\mathbf{H}_{(2)}=\mathbf{I}} (\mathbf{I} - \mathbf{D}_I) \langle \epsilon_x^b \rangle := \langle \epsilon_\beta^a \rangle_{t=0, \mathbf{D}_I}. \quad (6.32)$$

13 Equation (6.32) shows that, even with complete observation coverage, $\langle \epsilon_\beta^a \rangle_{t=0, \mathbf{D}_I}$ is still
14 nonzero, as it is still a function of the state background bias. However, if \mathbf{D}_I tends to
15 the identity, then the right hand side of equation (6.32) would tend to zero, such that
16 the observation bias correction would no longer be contaminated by the background bias.
17 The term \mathbf{D}_I would tend to the identity when the elements of $\mathbf{R}_{(2)}$ are much smaller than
18 the equivalent elements of \mathbf{B}_x . This would occur when the anchor observations are much
19 more precise than the state backgrounds that they observe.

20 Overall, equation (6.31) shows that, even with anchor observations at every model

1 grid point, we still have model bias contaminating the bias coefficient analysis. This can
 2 only be reduced by having anchor observations that are much more precise than the state
 3 backgrounds that they observe.

4 **6.2.1.2 Anchor observations partially observe the domain**

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

9 Let the state \mathbf{x} be separated into two parts: \mathbf{x}_ϕ and \mathbf{x}_ψ such that $\mathbf{x} = \begin{pmatrix} \mathbf{x}_\phi^\top & \mathbf{x}_\psi^\top \end{pmatrix}^\top$.
 10 Let the anchor observations only observe a subset of the state, such that they only observe
 11 variables in \mathbf{x}_ψ . Bias-corrected observations could observe variables in \mathbf{x}_ϕ and \mathbf{x}_ψ . Then
 12 the linearised bias-corrected and anchor observation operators will be given by,

$$\mathbf{H}_{(1)} = \begin{pmatrix} \mathbf{H}_{(1)\phi} & \mathbf{H}_{(1)\psi} \end{pmatrix}, \quad (6.33)$$

$$\mathbf{H}_{(2)}|_{\mathbf{H}_{(2)\phi}=0} = \begin{pmatrix} \mathbf{0} & \mathbf{H}_{(2)\psi} \end{pmatrix} := \mathbf{H}_{(2)p}, \quad (6.34)$$

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

17 The background error covariance matrix which describes the relationships between \mathbf{x}_ϕ^b
 18 and \mathbf{x}_ψ^b is,

$$\mathbf{B}_x = \begin{pmatrix} \mathbf{B}_{x_\phi} & \mathbf{B}_{x_\phi\psi} \\ \mathbf{B}_{x_\phi\psi}^\top & \mathbf{B}_{x_\psi} \end{pmatrix}. \quad (6.35)$$

19 The magnitude of $\mathbf{B}_{x_\phi\psi}$ determines how much information about the observations will be
 20 shared between the state variables (Bannister, 2008a). For example, if the elements in

1 $\mathbf{B}_{\mathbf{x}_{\phi\psi}}$ are small then there are weak correlations between the errors in \mathbf{x}_{ϕ}^b and \mathbf{x}_{ψ}^b .

We are interested in how the value of \mathbf{D} , equation (6.22), 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,

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

and the block elements can be calculated by expanding equation (6.22) for \mathbf{x}_{ϕ} and \mathbf{x}_{ψ} ,

$$\begin{aligned} \mathbf{D}|_{H_{(2)}=H_{(2)\text{p}}} &= \begin{pmatrix} \mathbf{B}_{\mathbf{x}_{\phi}} & \mathbf{B}_{\mathbf{x}_{\phi\psi}} \\ \mathbf{B}_{\mathbf{x}_{\phi\psi}} & \mathbf{B}_{\mathbf{x}_{\psi}} \end{pmatrix} \begin{pmatrix} \mathbf{0} \\ \mathbf{H}_{(2)\psi}^{\text{T}} \end{pmatrix} \left(\begin{pmatrix} \mathbf{0}, & \mathbf{H}_{(2)\psi} \end{pmatrix} \begin{pmatrix} \mathbf{B}_{\mathbf{x}_{\phi}} & \mathbf{B}_{\mathbf{x}_{\phi\psi}} \\ \mathbf{B}_{\mathbf{x}_{\phi\psi}} & \mathbf{B}_{\mathbf{x}_{\psi}} \end{pmatrix} \begin{pmatrix} \mathbf{0} \\ \mathbf{H}_{(2)\psi}^{\text{T}} \end{pmatrix} + \mathbf{R}_{(2)} \right)^{-1} \\ &\quad \times \begin{pmatrix} \mathbf{0}, & \mathbf{H}_{(2)\psi} \end{pmatrix}, \\ &:= \mathbf{D}_{\text{p}} = \begin{pmatrix} \mathbf{D}_{\text{p}\phi} & \mathbf{D}_{\text{p}\phi\psi} \\ \mathbf{D}_{\text{p}\psi\phi} & \mathbf{D}_{\text{p}\psi} \end{pmatrix}. \end{aligned}$$

2 Such that the block elements of \mathbf{D}_{p} are given by,

$$\mathbf{D}_{\text{p}\phi} = \mathbf{0}, \tag{6.36}$$

$$\mathbf{D}_{\text{p}\phi\psi} = \mathbf{B}_{\mathbf{x}_{\phi\psi}} \mathbf{H}_{(2)\psi}^{\text{T}} (\mathbf{H}_{(2)\psi} \mathbf{B}_{\mathbf{x}_{\psi}} \mathbf{H}_{(2)\psi}^{\text{T}} + \mathbf{R}_{(2)})^{-1} \mathbf{H}_{(2)\psi}, \tag{6.37}$$

$$\mathbf{D}_{\text{p}\psi\phi} = \mathbf{0}, \tag{6.38}$$

$$\mathbf{D}_{\text{p}\psi} = \mathbf{B}_{\mathbf{x}_{\psi}} \mathbf{H}_{(2)\psi}^{\text{T}} (\mathbf{H}_{(2)\psi} \mathbf{B}_{\mathbf{x}_{\psi}} \mathbf{H}_{(2)\psi}^{\text{T}} + \mathbf{R}_{(2)})^{-1} \mathbf{H}_{(2)\psi}. \tag{6.39}$$

3 Then the sensitivity of β^a to \mathbf{x}^b , equation (6.21), when $\mathbf{H}_{(2)} = \mathbf{H}_{(2)\text{p}}$ is given by,

$$\left. \frac{\partial \beta^a}{\partial \mathbf{x}^b} \right|_{\mathbf{H}_{(2)}=\mathbf{H}_{(2)\text{p}}} = -\mathbf{K}_{\beta y_{(1)}}|_{\mathbf{H}_{(2)}=\mathbf{H}_{(2)\text{p}}} \begin{pmatrix} \mathbf{H}_{(1)\phi}, & -\mathbf{H}_{(1)\phi} \mathbf{D}_{\text{p}\phi\psi} + \mathbf{H}_{(1)\psi} (\mathbf{I} - \mathbf{D}_{\text{p}\psi}) \end{pmatrix}. \tag{6.40}$$

4 The first block element of the matrix in equation (6.40) (ie. $-\mathbf{K}_{\beta y_{(1)}}|_{\mathbf{H}_{(2)}=\mathbf{H}_{(2)\text{p}}} \mathbf{H}_{(1)\phi}$) gives

1 the sensitivity of β^a to \mathbf{x}_ϕ^b and the second block element of the matrix in equation (6.40)
 2 (ie. $-\mathbf{K}_{\beta Y(1)}|_{\mathbf{H}(2)=\mathbf{H}(2)_p}[-\mathbf{H}_{(1)\phi}\mathbf{D}_{p\phi\psi} + \mathbf{H}_{(1)\psi}(\mathbf{I} - \mathbf{D}_{p\psi})]$) gives the sensitivity of β^a to \mathbf{x}_ψ^b .
 3 The terms $\mathbf{D}_{p\phi}$ and $\mathbf{D}_{p\psi\phi}$ are zero (from equations (6.36) and (6.38)) so do not appear
 4 in this equation, but their role would be to vary the sensitivity of β^a on \mathbf{x}_ϕ^b . Therefore as
 5 $\mathbf{D}_{p\phi}$ and $\mathbf{D}_{p\psi\phi}$ are zero, the sensitivity of β^a to \mathbf{x}_ϕ^b is not explicitly dependent on anchor
 6 observations, when anchor observations do not observe state variables in \mathbf{x}_ϕ . However,
 7 there is some implicit dependence of anchor observations in $\mathbf{K}_{\beta Y(1)}$. If $\mathbf{D}_{p\phi\psi}$ and $\mathbf{D}_{p\psi}$
 8 change magnitude, then this will vary the sensitivity of β^a to \mathbf{x}_ψ^b . Therefore, as $\mathbf{D}_{p\phi\psi}$ and
 9 $\mathbf{D}_{p\psi}$ are explicitly dependent on the anchor observations, the sensitivity of β^a to \mathbf{x}_ψ is also
 10 explicitly dependent on the anchor observations, as, in this case, the anchor observations
 11 observe state variables in \mathbf{x}_ψ .

12 Substituting \mathbf{D}_p into equation (6.29) gives $\langle \epsilon_\beta^a \rangle_{t=0}$ when the anchor observations ob-
 13 serve part of the state:

$$\begin{aligned}
 \langle \epsilon_\beta^a \rangle &= -\mathbf{K}_{\beta Y(1)} \left(\mathbf{H}_{(1)\phi} \langle \epsilon_{\mathbf{x}_\phi}^b \rangle - [\mathbf{H}_{(1)\phi} \mathbf{D}_{p\phi\psi} - \mathbf{H}_{(1)\psi} (\mathbf{I} - \mathbf{D}_{p\psi})] \langle \epsilon_{\mathbf{x}_\psi}^b \rangle \right) \\
 &:= \langle \epsilon_\beta^a \rangle_{t=0, \mathbf{D}_p}
 \end{aligned} \tag{6.41}$$

14 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}_ϕ^b and \mathbf{x}_ψ^b respec-
 15 tively. We will use this equation further in sections 6.2.1.3 and 6.2.2 to show how the
 16 sensitivity of the bias in β^a varies depending on the location of the model bias in relation
 17 to the anchor observations.

18 **6.2.1.3 Special case: anchor observations and bias-corrected observations ob-** 19 **serve different parts of the state**

20 Next we look at the case where anchor and bias-corrected observations observe different
 21 parts of the state to understand the importance of the background error covariance matrix.
 22 Let the bias-corrected observations only observe \mathbf{x}_ϕ and anchor observations only observe

1 \mathbf{x}_ψ . Then the observation operators of both observation types will be given by:

$$\begin{pmatrix} \mathbf{H}_{(1)\phi} & \mathbf{0} \end{pmatrix} := \mathbf{H}_{(1)\text{P}}, \quad \begin{pmatrix} \mathbf{0} & \mathbf{H}_{(2)\psi} \end{pmatrix} := \mathbf{H}_{(2)\text{P}}, \quad (6.42)$$

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

4 Substituting equation (6.42) into equation (6.41) gives the expected value of the bias
5 coefficient analysis when anchor and bias-corrected observations observe different parts of
6 the state,

$$\langle \epsilon_\beta^a \rangle_{t=0, \mathbf{D}_\text{P}} \Big|_{\mathbf{H}_{(1)\text{P}} \mathbf{H}_{(2)\text{P}}} = -\mathbf{K}_{\beta \mathcal{Y}(1)} \mathbf{H}_{(1)\phi} \left(\langle \epsilon_{\mathbf{x}_\phi}^b \rangle - \mathbf{D}_{\text{P}\phi\psi} \langle \epsilon_{\mathbf{x}_\psi}^b \rangle \right), \quad (6.43)$$

7 which is equivalent to equation (6.41) but with $\mathbf{H}_{(1)\psi} = \mathbf{0}$.

8 The right hand side of equation (6.43) will reduce if $\langle \epsilon_{\mathbf{x}_\phi}^b \rangle - \mathbf{D}_{\text{P}\phi\psi} \langle \epsilon_{\mathbf{x}_\psi}^b \rangle$ tends towards
9 zero. The term $\mathbf{D}_{\text{P}\phi\psi}$, from equation (6.37), is linearly dependent on $\mathbf{B}_{\mathbf{x}_\phi\psi}$. Therefore, the
10 strength of the background error covariances between \mathbf{x}_ϕ^b and \mathbf{x}_ψ^b will influence how small
11 the difference term in equation (6.43) is. In the case when the model biases $\langle \epsilon_{\mathbf{x}_\phi}^b \rangle$ and
12 $\langle \epsilon_{\mathbf{x}_\psi}^b \rangle$ have similar sign and magnitude, if $\mathbf{D}_{\text{P}\phi\psi}$ tends to the identity, then the difference
13 term will reduce and the expected value of the bias coefficient analysis error will tend to
14 zero. In practice, the model biases may be similar for different parts of the state and
15 the background error correlations between these two states could be large when restricted
16 to specific atmospheric layers higher in the atmosphere. However, this will not be the
17 case when the bias-corrected and anchor observations observe states vertically far apart,
18 for example, biases in stratospheric temperatures tend to be much bigger than biases in
19 tropospheric temperatures.

20 To understand the importance of $\mathbf{B}_{\mathbf{x}_\phi\psi}$ when the anchor and bias-corrected observa-
21 tions observe different parts of the state, let us look at the case when $\mathbf{B}_{\mathbf{x}_\phi\psi} = \mathbf{0}$. In this
22 case we can simplify \mathbf{K}_v from equation (6.4), as the off-diagonal blocks in $(\mathbf{H}_\text{v} \mathbf{B}_\text{v} \mathbf{H}_\text{v}^\text{T} + \mathbf{R})$

1 are zero:

$$\mathbf{H}_{(1)p} \mathbf{B}_x \mathbf{H}_{(2)p}^T = \mathbf{H}_{(1)\phi} \mathbf{B}_{x_{\phi\psi}} \mathbf{H}_{(2)\psi}^T = \mathbf{0} \quad \text{and} \quad (6.44)$$

$$\mathbf{H}_{(2)p} \mathbf{B}_x \mathbf{H}_{(1)p}^T = \mathbf{H}_{(2)\psi} \mathbf{B}_{x_{\phi\psi}}^T \mathbf{H}_{(1)\phi}^T = \mathbf{0}. \quad (6.45)$$

2 Substituting equations (6.44) and (6.45) into $\mathbf{H}_v \mathbf{B}_v \mathbf{H}_v^T + \mathbf{R}$ gives a block diagonal matrix.

3 Then \mathbf{K}_v from equation (6.4), can be simplified to:

$$\mathbf{K}_v \Big|_{\mathbf{B}_{x_{\phi\psi}}=0} = \begin{pmatrix} \mathbf{K}_{xy(1)} \Big|_{\mathbf{B}_{x_{\phi\psi}}=0} & \mathbf{K}_{xy(2)} \Big|_{\mathbf{B}_{x_{\phi\psi}}=0} \\ \mathbf{K}_{\beta y(1)} \Big|_{\mathbf{B}_{x_{\phi\psi}}=0} & \mathbf{K}_{\beta y(2)} \Big|_{\mathbf{B}_{x_{\phi\psi}}=0} \end{pmatrix}, \quad (6.46)$$

4 where $\mathbf{K}_{\beta y(1)}$ and $\mathbf{K}_{\beta y(2)}$ can be simplified from equations (6.18) and (6.19) to give

$$\mathbf{K}_{\beta y(1)} \Big|_{\mathbf{B}_{x_{\phi\psi}}=0} = \mathbf{B}_\beta \mathbf{C}_\beta^T (\mathbf{H}_{(1)\phi} \mathbf{B}_{x_\phi} \mathbf{H}_{(1)\phi}^T + \mathbf{C}_\beta \mathbf{B}_\beta \mathbf{C}_\beta^T + \mathbf{R}_{(1)})^{-1}, \quad (6.47)$$

$$\mathbf{K}_{\beta y(2)} \Big|_{\mathbf{B}_{x_{\phi\psi}}=0} = \mathbf{0}. \quad (6.48)$$

5 As $\mathbf{K}_{\beta y(2)} \Big|_{\mathbf{B}_{x_{\phi\psi}}=0} = \mathbf{0}$, then, from equation (6.8), the sensitivity of the bias coefficient
6 analysis to the anchor observations is zero when there are no background error cross cor-
7 relations between the states observed by the anchor observations and the states observed
8 by the bias-corrected observations. This means anchor observations do not play a role
9 in determining the bias coefficient if anchor observations do not observe the same state
10 variables as the bias-corrected observations and no information is passed between \mathbf{x}_ϕ and
11 \mathbf{x}_ψ via $\mathbf{B}_{x_{\phi\psi}}$. Therefore, if bias-corrected and anchor observations observe different state
12 variables, nonzero background error covariances allow the anchor observations to be used
13 to determine the bias coefficient and therefore reduce the effect of the background bias in
14 β^a . We will use this result in section 6.2.2.

15 So far we have not considered that the background bias may change magnitude de-
16 pending on location. In the next section we will investigate scenarios where the location

1 of the model bias differs relative to the location of the anchor observations.

2 **6.2.2 Position of anchor observations relative to model bias**

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

10 Figure 6.1 is a schematic diagram which depicts the different possibilities of the loca-
11 tions of the background bias in relation to the observations. The presence of background
12 bias is shown by the patterned red background; the red circles show the parts of the state
13 observed by the bias-corrected observations; and the blue circles show the parts of the state
14 observed by anchor observations. In (a) the background bias is in \mathbf{x}_ψ , which is observed
15 by anchor observations and some bias-corrected observations, in (b) the background bias
16 is in \mathbf{x}_ϕ which is only observed by bias-corrected observations and in (c) the background
17 bias is in both \mathbf{x}_ϕ and \mathbf{x}_ψ such that the whole state that is observed has background bias.
18 In the next sections we assume that the background biases in \mathbf{x}_ϕ and \mathbf{x}_ψ (when present)
19 have the same sign and magnitude. These scenarios are discussed below.

20 **6.2.2.1 Background bias in state variables that are observed by anchor ob-** 21 **servations, but not in all state variables observed by bias-corrected** 22 **observations (figure 6.1a):**

23 If $\langle \epsilon_{\mathbf{x}_\phi}^b \rangle = \mathbf{0}$ (ie. there is no background bias in \mathbf{x}_ϕ^b , the state variables observed only by the
24 bias-corrected observations) but $\langle \epsilon_{\mathbf{x}_\psi}^b \rangle \neq \mathbf{0}$ (ie. there is background bias in \mathbf{x}_ψ^b , the state
25 variables observed by both the bias-corrected and anchor observations) then the expected

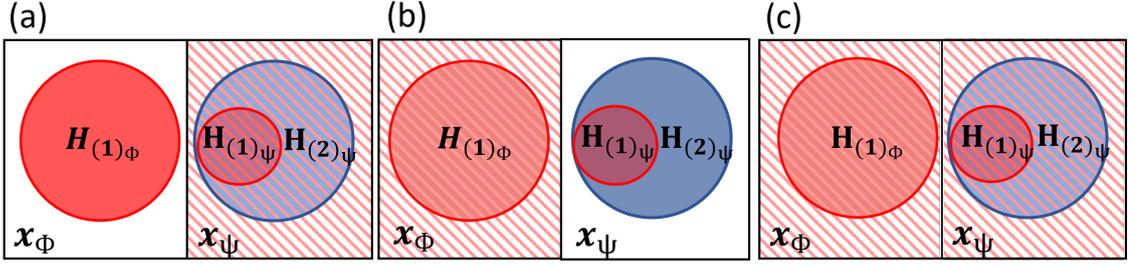


Figure 6.1: A schematic diagram depicting the location of the background bias in relation to the bias-corrected and anchor observations, where $\mathbf{H}_{(1)\phi}$ is the observation operator relating the bias-corrected observations to \mathbf{x}_ϕ ; $\mathbf{H}_{(1)\psi}$ is the observation operator relating the bias-corrected observations to \mathbf{x}_ψ ; and $\mathbf{H}_{(2)\psi}$ is the observation operator relating the anchor observations to \mathbf{x}_ψ . The background bias is shown by the patterned background. In (a) there is only background bias in \mathbf{x}_ψ , which is observed by the anchor observations and some bias-corrected observations, (b) there is only background bias in \mathbf{x}_ϕ , which is only observed by the bias-corrected observations and (c) there is background bias in both \mathbf{x}_ϕ and \mathbf{x}_ψ , which is observed by both bias-corrected and anchor observations.

1 value of the bias coefficient analysis, equation (6.41), becomes,

$$\langle \epsilon_\beta^a \rangle_{t=0, \mathbf{D}_p} \Big|_{\langle \epsilon_{\mathbf{x}_\phi}^b \rangle = 0} = \mathbf{K}_{\beta y(1)} [\mathbf{H}_{(1)\phi} \mathbf{D}_{p\phi\psi} - \mathbf{H}_{(1)\psi} (\mathbf{I} - \mathbf{D}_{p\psi})] \langle \epsilon_{\mathbf{x}_\psi}^b \rangle. \quad (6.49)$$

2 The effect of background bias from \mathbf{x}_ψ^b in equation (6.49) would be reduced if the matrix in
3 the square brackets tends to zero. This could occur if both $\mathbf{H}_{(1)\phi} \mathbf{D}_{p\phi\psi}$ and $\mathbf{H}_{(1)\psi} (\mathbf{I} - \mathbf{D}_{p\psi})$
4 tend to zero, or if their difference tends to zero. If the bias-corrected observations observe
5 state variables in \mathbf{x}_ψ (so they observe the same part of the state as the anchor observations)
6 such that $\mathbf{H}_{(1)\psi}$ is nonzero, then the second term in the square brackets would tend to zero
7 if $\mathbf{D}_{p\psi}$ tends to the identity. From equation (6.39), this would occur if $\mathbf{R}_{(2)\psi}$ is smaller
8 than $\mathbf{B}_{\mathbf{x}_\psi}$, or in other words, the anchor observations were more precise than the state
9 backgrounds that they observe, as we saw for a simplified case in section 6.2.1.1. If the
10 bias-corrected observations observe state variables in \mathbf{x}_ϕ (so they observe different parts
11 of the state to the anchor observations) such that $\mathbf{H}_{(1)\phi}$ is nonzero, then $\mathbf{H}_{(1)\phi} \mathbf{D}_{p\phi\psi}$ would
12 tend to zero if $\mathbf{D}_{p\phi\psi}$ tends to zero. Using equation (6.37), $\mathbf{D}_{p\phi\psi}$ would tend to zero if
13 the background error covariances $\mathbf{B}_{\mathbf{x}_\phi\psi}$ are small, which would mean less information is

1 passed between \mathbf{x}_ϕ^b and \mathbf{x}_ψ^b , than when the background error covariances are larger.

2 Therefore, if anchor and bias-corrected observations observe the same state, the anchor
 3 observations can reduce the effect of the background bias in the state when the anchor
 4 observations are more precise than the state backgrounds they observe. If there are bias-
 5 corrected observations that observe a different part of the state to the anchor observations,
 6 which does not have background bias, then smaller background error covariances between
 7 the two parts of the state will limit the amount of background bias able to contaminate
 8 the estimate of the bias coefficient, as the anchor observations cannot share information
 9 about the background bias to the bias-corrected observations. This is in contrast to section
 10 6.2.1.3, in which we showed that if anchor and bias-corrected observations observe different
 11 states, but the background bias observed was the same, then they would need nonzero
 12 background error cross correlations for the anchor observations to pass information about
 13 the background bias to the bias-corrected observations.

14 **6.2.2.2 Background bias not in state variables observed by anchor observa-**
 15 **tions (figure 6.1b):**

16 If there is no background bias in \mathbf{x}_ψ^b (ie. there is no background bias in state variables
 17 observed by anchor observations), such that $\langle \boldsymbol{\epsilon}_{\mathbf{x}_\psi}^b \rangle = \mathbf{0}$, but there is background bias
 18 in \mathbf{x}_ϕ^b (ie. there is background bias in state variables only observed by bias-corrected
 19 observations), then the expected value of the bias coefficient analysis, equation (6.41),
 20 reduces to,

$$\langle \boldsymbol{\epsilon}_\beta^a \rangle_{t=0, \mathbf{D}_p} \Big|_{\langle \boldsymbol{\epsilon}_{\mathbf{x}_2}^b \rangle = \mathbf{0}} = -\mathbf{K}_{\beta y(1)} \mathbf{H}_{(1)\phi} \langle \boldsymbol{\epsilon}_{\mathbf{x}_\phi}^b \rangle. \quad (6.50)$$

21 In this case, the anchor observations cannot reduce the effect of the background bias in \mathbf{x}_ϕ^b
 22 via \mathbf{D}_p as $\langle \boldsymbol{\epsilon}_\beta^a \rangle_{t=0, \mathbf{D}_p}$ is no longer dependent on \mathbf{D}_p and no other variables in equation (6.50)
 23 are explicitly dependent on the anchor observations. Therefore, if anchor observations
 24 do not observe the parts of the state that have background bias, they cannot explicitly

1 reduce the effect of background bias on β^a . Anchor observations will only have an effect
 2 on background bias implicitly through $\mathbf{K}_{\beta_{y(1)}}$, as $\mathbf{K}_{\beta_{y(1)}}$ is implicitly dependent on $\mathbf{H}_{(2)}$
 3 and $\mathbf{R}_{(2)}$ (see equation (6.19)). However, a small $\mathbf{K}_{\beta_{y(1)}}$ would mean the bias coefficient
 4 analysis is almost independent of the bias-corrected observations, so we are not interested
 5 in this case.

6 **6.2.2.3 Background bias in state variables observed by both anchor and bias-**
 7 **corrected observations (figure 6.1c):**

8 If we have background bias in both \mathbf{x}_ϕ^b and \mathbf{x}_ψ^b then we come back to equation (6.41) which
 9 gives the same results as in section 6.2.1. This is that giving a higher weighting to the
 10 anchor observations will reduce the contamination of background bias on the observation
 11 bias coefficient. If there are similar background biases in state variables observed by
 12 either bias-corrected or anchor observations, then background error covariances between
 13 state variables become more important in sharing information about the background bias.

14 In sections 6.2.1 and 6.2.2 we have shown that in order for the anchor observations to
 15 have the biggest impact on reducing the effect of background bias on β^a , both anchor and
 16 bias-corrected observations need to observe states with similar background biases. If both
 17 anchor and bias-corrected observations observe the same parts of the state, the effect of
 18 the background bias is smallest when $\mathbf{D}_{p\psi}$ tends towards the identity, as was shown in
 19 section 6.2.1.1, which occurs when anchor observations are precise. If both anchor and
 20 bias-corrected observations observe background bias, but do not observe the same state
 21 variables, background error correlations with large magnitudes are important in reducing
 22 the effect of background bias on β^a , as was shown in section 6.2.1.3, as $\mathbf{D}_{p\phi\psi}$ is linearly
 23 dependent on $\mathbf{B}_{\mathbf{x}_{\phi\psi}}$, equation (6.37).

1 **6.2.3 The effect of a biased bias coefficient analysis on the state analysis**
2 **in further cycles**

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

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

$$\langle \epsilon_1^o \rangle_{t=0} = \langle \mathbf{y}_{(1)t=0} - h_{(1)}(\mathbf{x}_{t=0}^t) - c(\mathbf{x}_{t=0}^t, \beta_{t=0}^t) \rangle = \mathbf{0}, \quad (6.51)$$

$$\langle \epsilon_2^o \rangle_{t=0} = \langle \mathbf{y}_{(2)t=0} - h_{(2)}(\mathbf{x}_{t=0}^t) \rangle = \mathbf{0}, \quad (6.52)$$

$$\langle \epsilon_\beta^b \rangle_{t=0} = \mathbf{0}. \quad (6.53)$$

14 If we assume that there is a bias in the state background, which arises from a model
15 bias, then from equations (6.26) and (6.27) we have that the expected value of the analysis
16 errors at the first cycle for the state and bias coefficients respectively are,

$$\langle \epsilon_x^a \rangle_{t=0} = (\mathbf{I} - \mathbf{K}_{xy(1)} \mathbf{H}_{(1)} - \mathbf{K}_{xy(2)} \mathbf{H}_{(2)}) \langle \epsilon_x^b \rangle_{t=0}, \quad (6.54)$$

$$\langle \epsilon_\beta^a \rangle_{t=0} = (-\mathbf{K}_{\beta y(1)} \mathbf{H}_{(1)} - \mathbf{K}_{\beta y(2)} \mathbf{H}_{(2)}) \langle \epsilon_x^b \rangle_{t=0}. \quad (6.55)$$

17 Note that equation (6.55) is equivalent to equation (6.29).

18 **Cycle 2:** We assume the bias coefficient is approximately constant between cycles such

1 that $\langle \epsilon_\beta^b \rangle_{t=1} = \langle \epsilon_\beta^a \rangle_{t=0}$. We assume that the expected value of the errors in both bias-
2 corrected and anchor observations are still zero as the observation errors in equations (6.51)
3 and (6.52) are not dependent on the background state and background bias coefficient.
4 The state background is the previous state analysis evolved forward via the linearised
5 model \mathbf{M} , plus a bias increment, $\boldsymbol{\eta}\Delta t$. So the expected value of the background and
6 observation errors at the second cycle are given by,

$$\langle \epsilon_\beta^b \rangle_{t=1} = \langle \epsilon_\beta^a \rangle_{t=0} = (-\mathbf{K}_{\beta y(1)} \mathbf{H}(1) - \mathbf{K}_{\beta y(2)} \mathbf{H}(2)) \langle \epsilon_x^b \rangle_{t=0}, \quad (6.56)$$

$$\langle \epsilon_x^b \rangle_{t=1} = \mathbf{M}_{0 \rightarrow 1} (\langle \epsilon_x^a \rangle_{t=0}) + \boldsymbol{\eta}\Delta t, \quad (6.57)$$

$$\langle \epsilon_1^o \rangle_{t=1} = \langle \mathbf{y}_{(1)t=1} - h_{(1)}(\mathbf{x}_{t=1}^t) - c(\mathbf{x}_{t=1}^t, \boldsymbol{\beta}_{t=1}^t) \rangle = \mathbf{0}, \quad (6.58)$$

$$\langle \epsilon_2^o \rangle_{t=1} = \langle \mathbf{y}_{(2)t=1} - h_{(2)}(\mathbf{x}_{t=1}^t) \rangle = \mathbf{0}. \quad (6.59)$$

7 Then substituting equations (6.56)-(6.59) into equations (6.26) and (6.27) we have that
8 the expected value of the analysis errors for the state and bias coefficient at the second
9 cycle are,

$$\langle \epsilon_x^a \rangle_{t=1} = (\mathbf{I} - \mathbf{K}_{xy(1)} \mathbf{H}(1) - \mathbf{K}_{xy(2)} \mathbf{H}(2)) \langle \epsilon_x^b \rangle_{t=1} - \mathbf{K}_{xy(1)} \mathbf{C}_\beta \langle \epsilon_\beta^a \rangle_{t=0}, \quad (6.60)$$

$$\langle \epsilon_\beta^a \rangle_{t=1} = -(\mathbf{K}_{\beta y(1)} \mathbf{H}(1) + \mathbf{K}_{\beta y(2)} \mathbf{H}(2)) \langle \epsilon_x^b \rangle_{t=1} + (\mathbf{I} - \mathbf{K}_{\beta y(1)} \mathbf{C}_\beta) \langle \epsilon_\beta^a \rangle_{t=0}. \quad (6.61)$$

10 The expected value of the state analysis is now dependent on both $\langle \epsilon_x^b \rangle_{t=1}$ and $\langle \epsilon_\beta^a \rangle_{t=0}$.
11 This shows that if the bias coefficient analysis is biased at time t , then this will contaminate
12 the estimate of both the state and the bias coefficient analysis at time $t+1$. As $\langle \epsilon_\beta^b \rangle_{t=1} =$
13 $\langle \epsilon_\beta^a \rangle_{t=0}$, we see that in future cycles we cannot assume $\langle \epsilon_\beta^b \rangle = 0$, as any bias in the bias
14 coefficient analysis will become the bias in the bias coefficient background. Therefore,
15 this shows that it is important to reduce the contamination of background bias in the
16 observation bias coefficient, as $\langle \epsilon_\beta^a \rangle$ feeds into future cycles and contaminates the state
17 analysis.

6.3 Numerical results

In this section we demonstrate the theoretical results from section 6.2 using the Lorenz 96 model and data assimilation set up as described in chapter 5.

We want to demonstrate the ability of anchor observations to reduce the contamination of background bias 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 the standard deviation of the bias coefficient analysis. We can illustrate the bias in the analysed bias coefficient with the ratio

$$\frac{|\bar{\beta}^a - \beta^t|}{\sigma_{a\beta}}, \quad (6.62)$$

where $|\cdot|$ is the absolute value; $\bar{\beta}^a$ 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 to the mean value of β^a in relation to the error standard deviation. 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 to the random noise, but if the bias ratio is small, then the bias in the bias coefficient analysis will be lost within the random error and so will be insignificant. The bias ratio at 0.1 is plotted for reference as a dotted line in figures 6.2, 6.3 and 6.4 and is referred to as the “reference ratio”.

6.3.1 One Cycle Experiments (No model evolution)

In the first two experiments (sections 6.3.1.1 and 6.3.1.2) 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

1 from running a previous cycle forward using the model. The background state is given by
 2 the truth plus an error,

$$\mathbf{x}^b = \mathbf{x}^t + \mathbf{e}_x^b + \mathbf{b}^b \in \mathbb{R}^{40}, \quad (6.63)$$

$$\beta^b = \beta^t + e_\beta^b \in \mathbb{R}, \quad (6.64)$$

3 where \mathbf{e}_x^b is the random error in the state background calculated from a Gaussian distribu-
 4 tion with zero mean and from the error covariance matrix \mathbf{B}_x ; \mathbf{b}^b is the bias in the initial
 5 state background; and e_β^b is the random error in the bias coefficient background calculated
 6 from a Gaussian distribution with zero mean and error variance $\sigma_{b\beta}^2$.

7 **6.3.1.1 The effect of varying the anchor observation error variance**

8 First we present an experiment to illustrate the results found in section 6.2.1.1. The the-
 9 oretical results show that in a system fully observed by anchor observations, small anchor
 10 observation error variances reduce the contamination of model bias on β^a . Therefore in
 11 this experiment, we observe all spatial variables with both bias-corrected and anchor ob-
 12 servations, such that every spatial variable is observed twice: once by the bias-corrected
 13 observations and once by the anchor observations.

14 In figure 6.2 we plot the bias ratio as in equation (6.62) from 3000 realisations after
 15 one cycle and vary the anchor observation error standard deviation, to test the importance
 16 of the weighting given to anchor observations in reducing the effect of model bias in the
 17 analysed bias coefficient. The bias-corrected observation error standard deviations ($\sigma_{o(1)}$)
 18 are set to 1, so that they are approximately 10% of the variability of the state. \mathbf{B}_v is
 19 calculated via an ensemble, as described in sections 5.2.2.1 and 5.2.3, and is calculated with
 20 $\sigma_{o(2)} = 1$. We have used the same \mathbf{B}_v for the different values of $\sigma_{o(2)}$, as, although this gives
 21 a sub-optimal \mathbf{B}_v for the system, it isolates the effect of varying the anchor observation
 22 error variance. A background bias of 0.15 has been added to the initial conditions, as in

1 equation (6.63). The bias is approximately 1.5% of the variability of the state and was
 2 chosen so that the system could still control the background bias. $\sigma_{o(2)}$ has been varied
 3 from one tenth of $\sigma_{o(1)}$ to ten times $\sigma_{o(1)}$ (ie. between 0.1 and 10). The blue crosses are
 4 the results from the 3000 realisations and the orange line is the analytic result given by
 5 equation (6.29), to give the 'true' bias ratio for the system.

6 In figure 6.2 the bias in the observation bias coefficient becomes more significant as
 7 the anchor observation error standard deviation increases. There is some variation in the
 8 bias ratio when calculated from the realisations, which is due to the random error in the
 9 observations and background, but they follow the shape of the analytic solution. The
 10 result that the bias ratio increases with larger anchor observation error variance is in line
 11 with the results found from equation (6.32), as we showed that if the anchor observation
 12 error variance is small, then the effect of the background bias on β^a would be small. This
 13 is because the ratio \mathbf{D} in equation (6.22) will tend to the identity as $\sigma_{o(2)}$ reduces. The
 14 ratio increases as $\sigma_{o(2)}$ increases, which shows that as less weight is given to the anchor
 15 observations, the bias in the analysed bias coefficient increases. This occurs as background
 16 bias dilutes the analysis of the bias coefficient and pulls the VarBC system away from
 17 the truth. In the limit where the anchor observation error variance is large, the anchor
 18 observations receive insignificant weight in the analysis and so the biased observations are
 19 bias corrected towards the background bias, instead of to the truth.

20 **6.3.1.2 The location of the anchor observations relative to the background** 21 **bias**

22 Next we present experiments to demonstrate the results found in sections 6.2.1.3 and
 23 6.2.2, such that we have anchor and biased observations observing different state variables
 24 that do and do not have background biases. In section 6.2.2 we showed how the role of
 25 the anchor observations in reducing the effect of background bias changes depending on
 26 whether or not anchor observations observe state variables that have background bias.

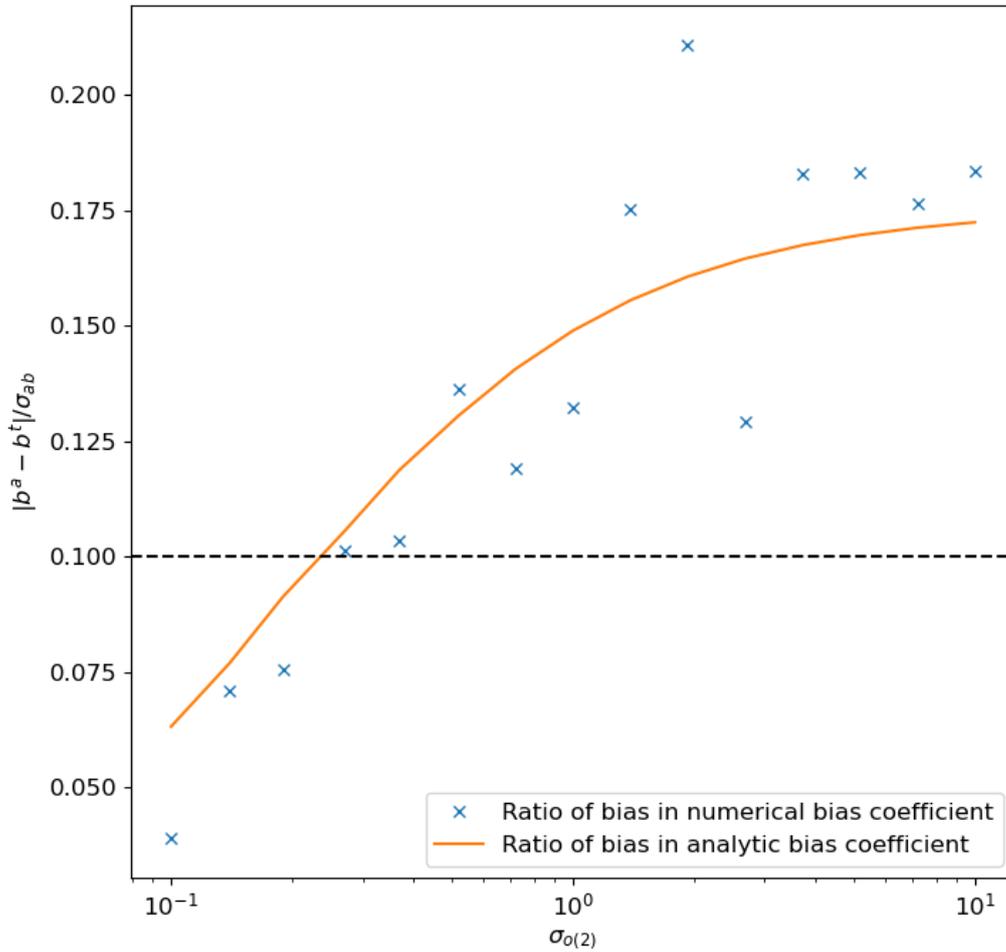


Figure 6.2: The bias ratio, equation (6.62), from running the system with different values of anchor observation error standard deviation ($\sigma_{o(2)}$). Both biased and anchor observations are at every location in the domain. The bias-corrected observation error standard deviations are set to 1. The background error covariance matrix is calculated from an ensemble, see section 5.2.2.1. A bias of 0.15 was added to all background state variables. The orange line is the analytic ratio, calculated from the linearised observation operators and error covariance matrices in the system and the blue crosses are the numerical ratios, calculated by averaging over 3000 realisations.

- 1 When anchor and biased observations observe different parts of the state, as in section
- 2 6.2.1.3, then the influence of the anchor observations is dependent on the background error

1 covariances between state variables.

2 In figure 6.3 we have plotted the ratio, equation (6.62), from 1000 realisations after
3 one cycle. The observations are spaced evenly at every other spatial variable such that the
4 biased observations observe all even state variables (x_0, x_2, x_4 etc) and anchor observations
5 observe all odd state variables (x_1, x_3, x_5 etc). The observation error standard deviations
6 are equal to 1. \mathbf{B}_x is given by the SOAR correlation function, with $\sigma_{bx}^2 = 1$, as defined
7 in equation (5.4) in section 5.2. The bias coefficient background error variance $\sigma_{b\beta}^2$, is
8 also equal to 1. To demonstrate the effect of different background error covariances, we
9 have varied \mathbf{B}_x by varying the length scale, L_b , along the x-axis in figure 6.3. When L_b is
10 small, the background error covariances between state variables are small, and when L_b
11 is increased, the background error covariances between state variables are larger. Larger
12 length scales will mean more background information is shared between parts of the state
13 that are spatially further away from each other. We have considered three cases where a
14 bias of 0.3 has been added to the background state in three different locations: in state
15 variables that are observed by biased observations; in state variables that are observed by
16 anchor observations; and in state variables that are observed by both types of observations.

17

18 **Background bias only in state variables observed by anchor observations**
19 **(orange line):** In figure 6.3, when background bias is only in state variables observed by
20 anchor observations (orange line), the bias ratio is insignificant when the error covariances
21 of \mathbf{B}_x are short. As the length scale of \mathbf{B}_x is increased, the bias ratio also increases, which
22 means β^a has a more significant bias in comparison to the random error. This reflects the
23 results found from equation (6.49), as we found that in order to reduce the contamination
24 of background bias in state variables observed by the anchor observations in $\langle \epsilon_{\beta}^a \rangle_{t=0}$, less
25 information had to be passed between the state variables observed by the anchor and
26 bias-corrected observations, ie. with smaller background error correlations. However, in
27 section 6.2.1.3, we showed that if the anchor and bias-corrected observations observed

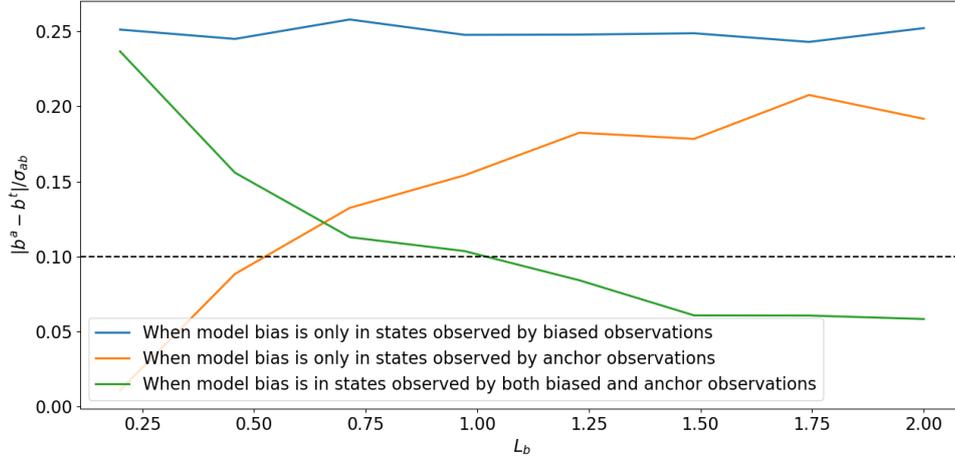


Figure 6.3: The bias ratio in equation (6.62) at one cycle when the length scale of \mathbf{B}_x has been varied. Biased observations observe all even state variables (x_0, x_2 etc), anchor observations observe all odd state variables. The dotted line is the reference ratio. A bias of 0.3 was added to the background: in state variables only observed by biased observations (blue); in state variables only observed by anchor observations (orange); in state variables observed by both types of observations (green).

1 different parts of the state and the error correlations in $\mathbf{B}_{x_{\phi\psi}}$ tended to zero, then the
 2 anchor observations would not be seen in the VarBC system. Hence this scenario would
 3 not be ideal as it would either mean that β^a is affected by background bias, or that the
 4 anchor observations are not used within the VarBC system.

5 **Background bias only in state variables not observed by anchor observations**

6 **(blue line):** In figure 6.3, when background bias is only in state variables that are observed
 7 by the biased observations (blue line), the bias in ratio remains significant regardless of
 8 the magnitude of the covariances in \mathbf{B}_x . This reflects the theoretical results, as equation
 9 (6.50) is independent of \mathbf{D} , which means that the anchor observations cannot directly
 10 reduce the effect of background bias in state variables if they do not observe the parts of
 11 the state that have background bias.

12 **Background bias in state variables observed by both anchor and bias-**

13 **corrected observations (green line):** In figure 6.3, if background bias is in state

1 variables observed by biased observations and in state variables observed by anchor ob-
 2 servations (green line), then the bias ratio is initially significant compared to the random
 3 error when the length scales of \mathbf{B}_x are small and becomes less significant as more informa-
 4 tion between background state variables is shared. As was shown in equation (6.43), figure
 5 6.3 shows that $\mathbf{B}_{x_{\phi\psi}}$ is important in reducing the effect of the background bias from parts
 6 of the state that both are and are not observed by the anchor observations. If background
 7 bias is in the system, then having strong background error correlations between the posi-
 8 tions of bias-corrected and anchor observations that observe state variables with similar
 9 background biases is the best possibility, as the anchor observations have the biggest effect
 10 in reducing the contamination of background bias on the estimate of the observation bias
 11 coefficient, whilst still being used within VarBC.

12 **6.3.2 Cycled Experiment**

13 In the experiment in section 6.3.2.1 the system is run over 50 cycles to demonstrate how the
 14 bias in the bias coefficient will accumulate and be passed into the state analysis. On the
 15 first cycle, the background state and bias coefficient are set up as in equations (6.63) and
 16 (6.64), but with zero added background bias so that $\mathbf{b}^b = 0$. This means the first few cycles
 17 will act as a spin up period to allow the background bias to settle into an equilibrium.
 18 In future cycles, the background values at cycle T are given by the analysis values at
 19 cycle $T - 1$ after they are evolved forwards in time. The bias coefficient background is
 20 evolved by the identity, ie. the previous bias coefficient analysis is taken to be the bias
 21 coefficient background. The state background is evolved forward by numerically solving
 22 the Lorenz 96 model from equation (5.1), as described in section 5.1, but replacing F with
 23 $F_{\text{biased}} = 8.8$ to add a bias to the model by changing the forcing term.

24 We set the assimilation window length to be 10 time steps, which represents approx-
 25 imately 7.5 hours in the real atmosphere, as discussed in section 5.1, to allow the back-
 26 ground errors to sufficiently grow within each window.

1 **6.3.2.1 How bias in the state analysis and bias coefficient analysis accumulates**
2 **when the system is cycled**

3 Finally, we have an experiment to demonstrate the results from section 6.2.3, which show
4 the effect of having a bias in β^a on the state analysis in the next cycle. These show
5 how the contamination of the background bias in β^a leads to a biased β^b , which in turn
6 contributes to the bias in the state analysis.

7 We run an experiment that has 100 realisations over 200 cycles. In figure 6.4 we plot
8 the bias ratio from equation (6.62) for the bias coefficient analysis (bottom panel), but
9 have included a similar ratio for the norm of the state analysis bias given by,

$$\frac{\sqrt{\frac{1}{40} \sum_{i=1}^{40} (\bar{x}_i^a - x_i^t)^2}}{\bar{\sigma}_{ax}}, \quad (6.65)$$

10 where \bar{x}_i^a is the average state analysis across all realisations for the spatial variable x_i ; x_i^t
11 is the true value of the spatial variable x_i ; 40 is the number of state variables; and $\bar{\sigma}_{ax}$
12 is the mean value of the standard deviations of the state analysis from 100 realisations,
13 where the mean is calculated over all state variables. We have both anchor and biased
14 observations at every location; the observation error standard deviations are 1; and \mathbf{B}_v
15 is the climatological background error covariance matrix, as described in sections 5.2.2.1
16 and 5.2.3. We have added a model bias by multiplying the forcing term in the model used
17 to evolve the analysis by 1.1, such that F_{biased} is 8.8, but have not added a model bias via
18 the initial conditions, as explained in section 6.3.2. Multiplying the forcing term by 1.1
19 only changes the state values by 0.01 over one cycle, allowing the model bias to still be
20 constrained. We plot the ratios over 200 cycles to show the evolution of the bias in the
21 state and the bias coefficient.

22 In figure 6.4 the ratios for the state and the bias coefficient start at approximately
23 zero and then the state ratio increases towards 0.1 and the bias coefficient ratio increases

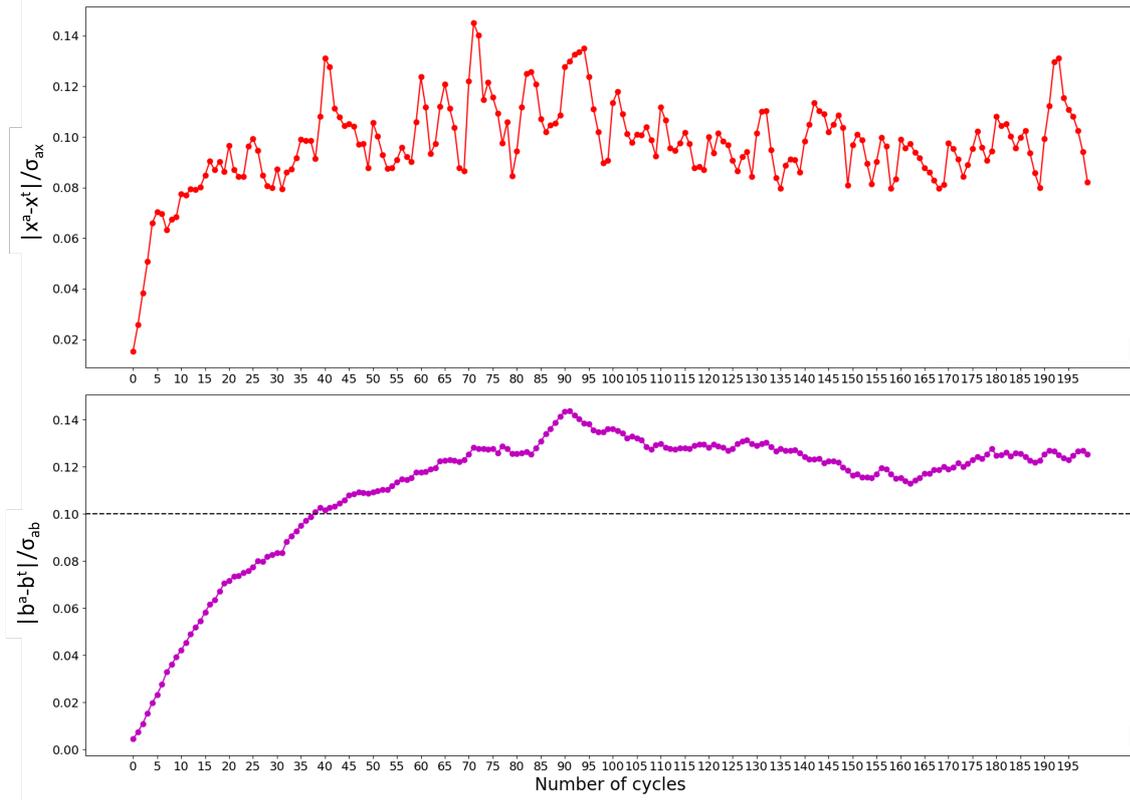


Figure 6.4: The ratios in equations (6.62) and (6.65) over 200 cycles for 100 realisations. $F_{biased} = 8.8$, error variances are 1, \mathbf{B}_v is calculated by an ensemble (section 5.2.2.1). Anchor and biased observations observe every state variable.

1 towards 0.12. When the ratios are above 0.1 then we consider the biases to be significant
 2 in comparison to the random error. Note that the magnitude of the ratios would be larger
 3 if a larger background bias were chosen, fewer state variables were observed, or there was
 4 less weighting given to the observations, which we do not show here. Therefore the 0.1
 5 line has been included in the bias ratio to show general trends, not the exact value of
 6 parameters which causes the bias to be significant. The bias in the first cycle is near zero
 7 for both the state and the bias coefficient because no background bias has been added to
 8 the initial conditions. There is only a background bias from the second cycle as model
 9 bias is added via the forcing term F_{biased} which evolves the state analysis forward, so that
 10 the state analysis of the first cycle becomes the state background of the second cycle and

1 so on. We would expect the bias in both the state and the bias coefficient to grow with
2 each cycle, as we showed in equations (6.60) and (6.61); the bias in \mathbf{x}^a and β^a are the
3 accumulation of the previous state and bias coefficient biases. In figure 6.4, the bias ratio
4 in the state appears to reach saturation after approximately 20 cycles and the bias ratio
5 in the bias coefficient reaches saturation after approximately 90 cycles. As the model bias,
6 η_i , is constant in time, the state and bias coefficient analyses reach an equilibrium between
7 the background bias and the truth from the anchor observations. We would expect an
8 equilibrium to exist, as VarBC relies on the background and the anchor observations as
9 sources of the truth, so if they are different, then the analysis will be pulled between the
10 two until it reaches an equilibrium. However, it is not straightforward to analytically
11 evaluate what the equilibrium should be.

12 **6.4 Conclusions and Discussion**

13 In order to study how background bias can contaminate the observation bias correction,
14 we have looked at the role of anchor observations in VarBC. The conclusions are based
15 on general optimal estimation theory and we have demonstrated the results with idealised
16 experiments.

17 In this study we have focused on the importance of the anchor observations in reducing
18 the contamination of the background bias in the bias coefficient analysis, as opposed to
19 reducing the effect on the state analysis. This is because if the bias coefficient analysis
20 has systematic error, then observations will be wrongly bias corrected and the systematic
21 error will filter into the state analysis, as was shown when the system was cycled in section
22 6.2.3 and figure 6.4.

23 In a theoretical world where we could have a full coverage of anchor observations, we
24 showed in equation (6.32) that the background bias can still contaminate the bias coef-
25 ficient analysis. The only way that background bias could not affect the bias coefficient

1 analysis at all, would be to have zero random error in the anchor observations. However,
2 as this is only possible theoretically, we can only look at reducing the contamination of
3 background bias on the bias coefficient analysis, not completely removing it. We showed
4 in equation (6.31) and figure 6.2 that the effect of the background bias is reduced when
5 the anchor observation error variance is smaller than the state backgrounds that they ob-
6 serve. Operationally, the anchor observation error variance needs to reflect the uncertainty
7 associated to the observations. However, it is possible to change the weighting given to
8 the anchor observations by using more anchor observations within the system, if they are
9 available. Therefore, although we have showed that more precise anchor observations will
10 reduce the contamination of background bias on the observation bias correction, this result
11 also extends to the importance of a higher spatial frequency of anchor observations, to
12 increase the weighting given to the anchor observations within the system and thus have
13 the biggest impact in reducing the contamination of background bias.

14 In equations (6.41), (6.49), (6.50) and figure 6.3 we showed that the contamination
15 of background bias on the bias coefficient analysis can only be reduced by the anchor
16 observations if the anchor observations also observe state variables that have similar back-
17 ground bias. Anchor observations have typically been from radiosondes which mostly have
18 coverage over land, but with the increased use of radio occultation (RO), the spatial and
19 temporal distribution of anchor observations in the upper troposphere and stratosphere
20 has increased. Radiosondes and RO give a good coverage of anchor observations for tem-
21 perature in the troposphere and stratosphere, but variables such as humidity and wind
22 speed are less well observed by anchor observations. This means that although the spatial
23 coverage of anchor observations is increasing, there are still significant parts of the model
24 domain where background bias could still contaminate the bias coefficient analysis.

25 We showed in equation (6.49) and figure 6.3 that if background bias only exists in state
26 variables observed by the anchor observations, such that bias-corrected observations do
27 not observe state variables with background bias, then the anchor observations will pass

1 the background bias into the bias coefficient analysis if there are strong error correlations
2 between the background state variables. This could occur for example in the lower tropo-
3 sphere, as radiosonde measurements are mostly taken over land, with little coverage over
4 sea. If there exists a temperature bias over the land but not the sea and there are strong
5 horizontal error correlations between the land and the sea, then the background bias could
6 be passed into the bias coefficients via the radiosondes.

7 In equation (6.50) and figure 6.3 we showed that, if there is no background bias in the
8 state variables that are observed by the anchor observations but there is background bias
9 in state variables observed by the bias-corrected observations, then the anchor observations
10 cannot explicitly reduce the effect of background bias in other state variables. Therefore,
11 areas that are most at risk of contamination of background bias within the observation
12 bias correction will be locations with fewer anchor observations such as humidity in the
13 upper troposphere, which is only sparsely observed by radiosondes, but is known to have
14 background biases.

15 We showed in equation (6.41) and at the end of section 6.2.2 that, if both biased
16 and anchor observations observe state variables that have background biases, then an-
17 chor observations can reduce the contamination of background bias on the bias coefficient
18 analysis. We saw in figure 6.3 that if the anchor and bias-corrected observations observe
19 different parts of the state that have similar background bias characteristics, then the
20 background error covariances become more important: larger background error correla-
21 tions will transfer more information about the background bias between state variables
22 and so will reduce the contamination of the background bias on the bias coefficient anal-
23 ysis. In an operational system, the length scales of \mathbf{B}_x will be locationally dependent, for
24 example, the length scales of \mathbf{B}_x will be larger at higher altitudes (Ingleby, 2001). There
25 will also be an element of flow dependency, such that the background error covariance
26 matrix will deform with the flow (Bannister, 2008a). This chapter shows that anchor
27 observations that observe different variables/locations to the bias-corrected observations

1 when both variables/locations have similar background biases, will have the largest impact
2 in reducing the contamination of background bias when they are in systems with larger
3 background error covariances, such as in the upper atmosphere. When similar background
4 biases are present in systems where background error covariances are small between vari-
5 ables observed by bias-corrected and anchor observations, such as between locations of
6 observations across a front, then anchor observations will only have a small impact in
7 reducing the contamination of background bias on the observation bias correction.

8 This chapter has aimed to derive new insight on the role of anchor observations for
9 mitigating the impact of model bias in VarBC. Our theoretical findings have been tested in
10 a toy system. The next steps for this work should be to understand how the assumptions
11 made hold in an operational system, for example how the bias predictors used operationally
12 allow the state and bias coefficient domains to be more clearly separated, such that the
13 bias coefficient is less affected by state background bias. In the next chapter we will extend
14 this work by studying the role of anchor observations in a 4-dimensional system, to study
15 when, in an assimilation window, anchor observations have the biggest impact in reducing
16 the contamination of model bias.

1 Chapter 7

2 The role of anchor observations in 3 VarBC in the presence of model 4 bias. Part II: The importance of 5 the timing of the anchor 6 observations

7 7.1 Introduction

8 In chapter 6 we demonstrated the role of anchor observations in 3DVar, in reducing the
9 contamination of background bias when observation bias was being corrected for using
10 VarBC. We focused on the importance of the spatial distribution of anchor observations,
11 relative to the biased observations and the importance of precise anchor observations in
12 order to answer research questions 2.1 and 2.3.

13 In this chapter, we extend chapter 6 to demonstrate the role of anchor observations in

1 a 4DVar system when observation, background and model biases are present. Studying a
 2 4DVar system allows us to show the importance of the timing of the anchor observations
 3 within an assimilation window, relative to the biased observations and therefore be able
 4 to answer research question 2.2. We demonstrate the theory of how anchor observations
 5 are used in 4DVarBC to reduce the contamination of background and model biases. We
 6 study the importance of precise anchor observations, as well as their timing in the window,
 7 relative to the biased observations.

8 In section 7.2 we extend current 4DVarBC theory to include both bias-corrected and
 9 anchor observations. We then look at the impacts of background and model biases in
 10 section 7.3, by deriving the analysis error equations when both bias-corrected and anchor
 11 observations are used and when background and model bias are present. To further our
 12 understanding of the equations, we study simple cases in sections 7.3.1, 7.3.2 and 7.3.3
 13 that only have bias-corrected and anchor observations at one time step each in the window.
 14 Finally we test our results using a simple numerical system in section 7.4.

15 **7.2 4DVarBC with two observation types**

16 We want to look at the role of anchor observations in VarBC when the observations are
 17 spread throughout the window. We therefore extend the 4DVarBC analysis equations,
 18 equations (2.75) and (2.76), in section 2.3 to include both bias-corrected observations
 19 ($\mathbf{y}_{(1)}$) and anchor observations ($\mathbf{y}_{(2)}$). Splitting \mathbf{y} into two observation types gives the
 20 analysis equations in terms of \mathbf{x}_0 and $\boldsymbol{\beta}$ as

$$\mathbf{x}_0^a = \mathbf{x}_0^b + \hat{\mathbf{K}}_{\mathbf{xy}_{(1)}} \hat{\mathbf{d}}_{v(1)}^b + \hat{\mathbf{K}}_{\mathbf{xy}_{(2)}} \hat{\mathbf{d}}_{v(2)}^b, \quad (7.1)$$

$$\boldsymbol{\beta}^a = \boldsymbol{\beta}^b + \hat{\mathbf{K}}_{\beta\mathbf{y}_{(1)}} \hat{\mathbf{d}}_{v(1)}^b + \hat{\mathbf{K}}_{\beta\mathbf{y}_{(2)}} \hat{\mathbf{d}}_{v(2)}^b, \quad (7.2)$$

- 1 where $\hat{\mathbf{d}}_{v(1)}^b$ and $\hat{\mathbf{d}}_{v(2)}^b$ are as defined in equation (2.79), but for observations that either
 2 are or are not corrected for:

$$\hat{\mathbf{d}}_{v(1)}^b = \begin{pmatrix} \mathbf{y}_{(1)0} - h_{(1)0}(\mathbf{x}_0^b) - c_0(\mathbf{x}_0^b, \boldsymbol{\beta}^b) \\ \mathbf{y}_{(1)1} - h_{(1)1}(m_0(\mathbf{x}_0^b)) - c_1(m_0(\mathbf{x}_0^b), \boldsymbol{\beta}^b) \\ \vdots \\ \mathbf{y}_{(1)N} - h_{(1)N}(m_{0 \rightarrow N}(\mathbf{x}_0^b)) - c_N(m_{0 \rightarrow N}(\mathbf{x}_0^b), \boldsymbol{\beta}^b) \end{pmatrix}, \quad (7.3)$$

$$\hat{\mathbf{d}}_{v(2)}^b = \begin{pmatrix} \mathbf{y}_{(2)0} - h_{(2)0}(\mathbf{x}_0^b) \\ \mathbf{y}_{(2)1} - h_{(2)1}(m_0(\mathbf{x}_0^b)) \\ \vdots \\ \mathbf{y}_{(2)N} - h_{(2)N}(m_{0 \rightarrow N}(\mathbf{x}_0^b)) \end{pmatrix}. \quad (7.4)$$

- 3 Note that the numbers outside the brackets represent the time step and the numbers
 4 inside the brackets represent the bias-corrected and anchor observations respectively. The
 5 terms $\hat{\mathbf{K}}_{xy(1)}$ and $\hat{\mathbf{K}}_{xy(2)}$ are the sensitivities of the state analysis to the bias-corrected and
 6 anchor observations respectively and $\hat{\mathbf{K}}_{\beta y(1)}$ and $\hat{\mathbf{K}}_{\beta y(2)}$ are the sensitivities of the bias
 7 coefficient analysis to the bias-corrected and anchor observations respectively. The terms
 8 $\hat{\mathbf{K}}_{xy(1)}$, $\hat{\mathbf{K}}_{xy(2)}$, $\hat{\mathbf{K}}_{\beta y(1)}$ and $\hat{\mathbf{K}}_{\beta y(2)}$ can be calculated by separating $\hat{\mathbf{K}}_v$ into its 4 separate
 9 parts:

$$\hat{\mathbf{K}}_v = \mathbf{B}_v \hat{\mathbf{H}}_v^T (\hat{\mathbf{H}}_v \mathbf{B}_v \hat{\mathbf{H}}_v^T + \hat{\mathbf{R}})^{-1} = \begin{pmatrix} \hat{\mathbf{K}}_{xy(1)} & \hat{\mathbf{K}}_{xy(2)} \\ \hat{\mathbf{K}}_{\beta y(1)} & \hat{\mathbf{K}}_{\beta y(2)} \end{pmatrix} \in \mathbb{R}^{(n+r) \times (m_{(1)} + m_{(2)})}. \quad (7.5)$$

- 10 Note that this takes the same form as equation (6.4) in chapter 6, but the elements
 11 may vary in time (as the observation operator is state dependent). The terms $\hat{\mathbf{K}}_{xy(1)}$,
 12 $\hat{\mathbf{K}}_{xy(2)}$, $\hat{\mathbf{K}}_{\beta y(1)}$ and $\hat{\mathbf{K}}_{\beta y(2)}$ are all time dependent, as they are dependent on the linearised
 13 observation operators for the bias-corrected and anchor observations, $\hat{\mathbf{H}}_{(1)}$, $\hat{\mathbf{H}}_{(2)}$ and $\hat{\mathbf{C}}_\beta$,

1 which are given by,

$$\hat{\mathbf{H}}_{(1)} = \begin{pmatrix} \mathbf{H}_{(1)0} \mathbf{M}_{0 \rightarrow 0} \\ \vdots \\ \mathbf{H}_{(1)N_{(1)}} \mathbf{M}_{0 \rightarrow N_{(1)}} \end{pmatrix} \in \mathbb{R}^{\sum_{i=0}^{N_{(1)}} m_{(1)i} \times n}, \quad \hat{\mathbf{H}}_{(2)} = \begin{pmatrix} \mathbf{H}_{(2)0} \mathbf{M}_{0 \rightarrow 0} \\ \vdots \\ \mathbf{H}_{(2)N_{(2)}} \mathbf{M}_{0 \rightarrow N_{(2)}} \end{pmatrix} \in \mathbb{R}^{\sum_{i=0}^{N_{(2)}} m_{(2)i} \times n},$$
(7.6)

$$\hat{\mathbf{C}}_{\beta} = \begin{pmatrix} \mathbf{C}_{\beta_0} \\ \vdots \\ \mathbf{C}_{\beta_{N_{(1)}}} \end{pmatrix} \in \mathbb{R}^{\sum_{i=0}^{N_{(1)}} m_{(1)i} \times r},$$
(7.7)

2 which are equivalent to the linearised observation operators for 4DVarBC in equation
 3 (2.73), but for $\hat{\mathbf{H}}_{(1)}$ and $\hat{\mathbf{H}}_{(2)}$ defined for the bias-corrected and anchor observations re-
 4 spectively. The dimensions $N_{(1)}$ and $N_{(2)}$ are the number of timesteps observed by bias-
 5 corrected and anchor observations; $m_{(1)i}$ and $m_{(2)i}$ are the number of bias-corrected and
 6 anchor observations in space at the i^{th} time step; n is the number of state variables; and
 7 r is the size of the observation bias coefficient vector.

8 7.3 4DVarBC with model bias that is not accounted for

9 In this section, we assume that there is a model bias at each time step which is not
 10 explicitly corrected for, as well as the observation bias that is explicitly corrected for
 11 using VarBC. We want to calculate the bias in the state and observation bias coefficient
 12 analyses to understand how the model bias contaminates the estimate of the state and
 13 observation bias coefficient and to discern how the anchor observations can mitigate this
 14 contamination. We will use a similar method for calculating the analysis error equations
 15 as we did for the 3DVar case in chapter 6, but extend it to the 4DVar scenario to include
 16 both state background and model biases. We assume that the model that evolves the state

1 from the initial time to time i is given by,

$$m_{0 \rightarrow i}(\mathbf{x}_0) = m_{i-1}^t(\dots(m_0^t(\mathbf{x}_0) + \boldsymbol{\eta}_1^t)\dots + \boldsymbol{\eta}_{i-1}^t) + \boldsymbol{\eta}_i^t := \tilde{m}_{0 \rightarrow i}^t(\mathbf{x}_0, \boldsymbol{\eta}^t), \quad (7.8)$$

2 where the true model, m^t , is acting on the state at the initial time, \mathbf{x}_0 , with an added error,
 3 $\boldsymbol{\eta}_i^t$. We assume that $\boldsymbol{\eta}_i^t$ is constant throughout the time window (we will therefore denote
 4 $\boldsymbol{\eta}_i^t$ as $\boldsymbol{\eta}^t$) and can therefore consider $\boldsymbol{\eta}^t$ as a model bias. We have denoted $\tilde{m}_{0 \rightarrow i}^t(\mathbf{x}_0, \boldsymbol{\eta}^t)$
 5 to be the model that evolves both the state and the model bias forward from the initial
 6 time step to the i^{th} time step, as defined in equation (2.82).

7 Let the errors in the backgrounds, analyses and observations be defined by,

$$\boldsymbol{\epsilon}_x^b = \mathbf{x}_0^b - \mathbf{x}_0^t, \quad \boldsymbol{\epsilon}_\beta^b = \boldsymbol{\beta}^b - \boldsymbol{\beta}^t, \quad \boldsymbol{\epsilon}_x^a = \mathbf{x}_0^a - \mathbf{x}_0^t, \quad \boldsymbol{\epsilon}_\beta^a = \boldsymbol{\beta}^a - \boldsymbol{\beta}^t, \quad (7.9)$$

$$\hat{\boldsymbol{\epsilon}}_{(1)}^o = \begin{pmatrix} \mathbf{y}_{(1)0} - h_{(1)0}(\mathbf{x}_0^t) - c_0(\mathbf{x}_0^t, \boldsymbol{\beta}^t) \\ \vdots \\ \mathbf{y}_{(1)N(1)} - h_{(1)N(1)}(m_{0 \rightarrow N(1)}^t(\mathbf{x}_0^t)) - c_{N(1)}(m_{0 \rightarrow N(1)}^t(\mathbf{x}_0^t), \boldsymbol{\beta}^t) \end{pmatrix}, \quad (7.10)$$

$$\hat{\boldsymbol{\epsilon}}_{(2)}^o = \begin{pmatrix} \mathbf{y}_{(2)0} - h_{(2)0}(\mathbf{x}_0^t) \\ \vdots \\ \mathbf{y}_{(2)N(2)} - h_{(2)N(2)}(m_{0 \rightarrow N(2)}^t(\mathbf{x}_0^t)) \end{pmatrix}, \quad (7.11)$$

8 where the true model trajectory does not contain a model bias and is given by,

$$m_{0 \rightarrow i}^t(\mathbf{x}_0) = m_{i-1}^t(\dots(m_0^t(\mathbf{x}_0))\dots). \quad (7.12)$$

9 Note that we have only included the instrument error in the observation error in equations
 10 (7.10) and (7.11), but the error that comes from evolving the state forward to the time
 11 step of the observations using the incorrect model will be considered later in equation
 12 (7.15).

13 To calculate the state and observation bias coefficient analysis errors, we first subtract

1 the true state and bias coefficient from equations (7.1) and (7.2) respectively to rewrite the
 2 differences between the state and bias coefficient analysis and the state and bias coefficient
 3 background in terms of their error equations, equation (7.9), to give,

$$\boldsymbol{\epsilon}_x^a = \boldsymbol{\epsilon}_x^b + \hat{\mathbf{K}}_{xy(1)} \hat{\mathbf{d}}_{(1)}^b + \hat{\mathbf{K}}_{xy(2)} \hat{\mathbf{d}}_{(2)}^b, \quad (7.13)$$

$$\boldsymbol{\epsilon}_\beta^a = \boldsymbol{\epsilon}_\beta^b + \hat{\mathbf{K}}_{\beta y(1)} \hat{\mathbf{d}}_{(1)}^b + \hat{\mathbf{K}}_{\beta y(2)} \hat{\mathbf{d}}_{(2)}^b. \quad (7.14)$$

4 We then manipulate the innovation vector so that it can be written in terms of the obser-
 5 vation, background and model errors. Note that we will only show the innovation vector
 6 at the i^{th} time for the bias-corrected observations for simplicity, but this can be repeated
 7 for either bias-corrected or anchor observations across the whole window. We can start to
 8 write the bias-corrected innovation vector at the i^{th} time step in terms of its errors in the
 9 following way,

$$\begin{aligned} \mathbf{y}_{(1)i} - h_{v_i}(m_{0 \rightarrow i}(\mathbf{x}_0^b), \boldsymbol{\beta}^b) &= \mathbf{y}_{(1)i} - h_{v_i}(m_{0 \rightarrow i}^t(\mathbf{x}_0^t), \boldsymbol{\beta}^t) \\ &\quad + h_{v_i}(m_{0 \rightarrow i}^t(\mathbf{x}_0^t), \boldsymbol{\beta}^t) - h_{v_i}(m_{0 \rightarrow i}(\mathbf{x}_0^t), \boldsymbol{\beta}^t) \\ &\quad + h_{v_i}(m_{0 \rightarrow i}(\mathbf{x}_0^t), \boldsymbol{\beta}^t) - h_{v_i}(m_{0 \rightarrow i}(\mathbf{x}_0^b), \boldsymbol{\beta}^b) \end{aligned} \quad (7.15)$$

10 where $h_{v_i}(m_{0 \rightarrow i}(\mathbf{x}_0), \boldsymbol{\beta}) = h_{(1)i}(m_{0 \rightarrow i}(\mathbf{x}_0)) + c_i(m_{0 \rightarrow i}(\mathbf{x}_0), \boldsymbol{\beta})$, as defined in equation (2.72).
 11 The term $\mathbf{y}_i - h_{v_i}(m_{0 \rightarrow i}^t(\mathbf{x}_0^t), \boldsymbol{\beta}^t)$ is given by $\boldsymbol{\epsilon}_{(1)i}^o$ from the i^{th} term in equation (7.10)
 12 and can be considered the instrument error; $h_{v_i}(m_{0 \rightarrow i}^t(\mathbf{x}_0^t), \boldsymbol{\beta}^t) - h_{v_i}(m_{0 \rightarrow i}(\mathbf{x}_0^t), \boldsymbol{\beta}^t)$ is the
 13 difference between the true model and the used model, so can be defined as the model
 14 error; and $h_{v_i}(m_{0 \rightarrow i}(\mathbf{x}_0^t), \boldsymbol{\beta}^t) - h_{v_i}(m_{0 \rightarrow i}(\mathbf{x}_0^b), \boldsymbol{\beta}^b)$ is the difference between the true state
 15 and observation bias coefficient with the background state and observation bias coefficient,
 16 which can be defined as the background error. We can therefore study each term separately.

17 The model error can be rearranged by writing $m_{0 \rightarrow i}(\mathbf{x}_0^t)$ in terms of the true model

1 and the model bias as in equation (7.8) to give,

$$h_{v_i}(m_{0 \rightarrow i}^t(\mathbf{x}_0^t), \boldsymbol{\beta}^t) - h_{v_i}(m_{0 \rightarrow i}(\mathbf{x}_0^t), \boldsymbol{\beta}^t) = h_{v_i}(m_{0 \rightarrow i}^t(\mathbf{x}_0^t), \boldsymbol{\beta}^t) - h_{v_i}(\tilde{m}_{0 \rightarrow i}^t(\mathbf{x}_0^t, \boldsymbol{\eta}^t), \boldsymbol{\beta}^t), \quad (7.16)$$

Equation (7.16) can be expanded to

$$h_{v_i}(m_{0 \rightarrow i}^t(\mathbf{x}_0^t), \boldsymbol{\beta}^t) - h_{v_i}(m_{0 \rightarrow i}(\mathbf{x}_0^t), \boldsymbol{\beta}^t) = h_{v_i}(m_{0 \rightarrow i}^t(\mathbf{x}_0^t), \boldsymbol{\beta}^t) - h_{v_i}(m_{i-1}^t(\dots(m_0(\mathbf{x}_0^t) + \boldsymbol{\eta}^t)\dots) + \boldsymbol{\eta}^t, \boldsymbol{\beta}^t). \quad (7.17)$$

Then approximating $\tilde{m}_{0 \rightarrow i}^t(\mathbf{x}_0^t, \boldsymbol{\eta}^t)$ about $m_{0 \rightarrow i}(\mathbf{x}_0^t)$ using the Taylor expansion gives,

$$h_{v_i}(m_{0 \rightarrow i}^t(\mathbf{x}_0^t), \boldsymbol{\beta}^t) - h_{v_i}(m_{0 \rightarrow i}(\mathbf{x}_0^t), \boldsymbol{\beta}^t) \approx h_{v_i}(m_{0 \rightarrow i}^t(\mathbf{x}_0^t), \boldsymbol{\beta}^t) - h_{v_i}(m_{0 \rightarrow i}(\mathbf{x}_0^t), \boldsymbol{\beta}^t) + \sum_{j=1}^{i-1} \mathbf{M}_{i-1} \dots \mathbf{M}_j \boldsymbol{\eta}^t + \boldsymbol{\eta}^t, \boldsymbol{\beta}^t). \quad (7.18)$$

2 Finally approximating $h_{v_i}(\cdot)$ using the Taylor expansion about \mathbf{x}_0^t and $\boldsymbol{\beta}^t$, and cancelling
3 the $h_{v_i}(m_{0 \rightarrow i}^t(\mathbf{x}_0^t), \boldsymbol{\beta}^t)$ terms gives the model error as,

$$h_{v_i}(m_{0 \rightarrow i}^t(\mathbf{x}_0^t), \boldsymbol{\beta}^t) - h_{v_i}(m_{0 \rightarrow i}(\mathbf{x}_0^t), \boldsymbol{\beta}^t) \approx -(\mathbf{H}_{(1)_i} + \mathbf{C}_{\mathbf{x}_i}) \left(\sum_{j=1}^{i-1} \mathbf{M}_{i-1} \dots \mathbf{M}_j + \mathbf{I} \right) \boldsymbol{\eta}^t. \quad (7.19)$$

4 This is the accumulation of the model biases at different time steps, due to evolving the
5 state forward to the time step of the bias-corrected observations using a biased model.

6 The background error in equation (7.15) can also be rearranged, by initially writing
7 \mathbf{x}_0^t as $\mathbf{x}_0^b - \boldsymbol{\epsilon}_x^b$, as in equation (7.9), to give,

$$h_{v_i}(m_{0 \rightarrow i}(\mathbf{x}_0^t), \boldsymbol{\beta}^t) - h_{v_i}(m_{0 \rightarrow i}(\mathbf{x}_0^b), \boldsymbol{\beta}^b) = h_{v_i}(m_{0 \rightarrow i}(\mathbf{x}_0^b - \boldsymbol{\epsilon}_x^b), \boldsymbol{\beta}^b - \boldsymbol{\epsilon}_\beta^b) - h_{v_i}(m_{0 \rightarrow i}(\mathbf{x}_0^b), \boldsymbol{\beta}^b). \quad (7.20)$$

Approximating $m_{0 \rightarrow i}(\mathbf{x}_0^b - \boldsymbol{\epsilon}_x^b)$ using the Taylor expansion about \mathbf{x}_0^b gives,

$$\begin{aligned} h_{v_i}(m_{0 \rightarrow i}(\mathbf{x}_0^t), \boldsymbol{\beta}^t) - h_{v_i}(m_{0 \rightarrow i}(\mathbf{x}_0^b), \boldsymbol{\beta}^b) \\ \approx h_{v_i}(m_{0 \rightarrow i}(\mathbf{x}_0^b) - \mathbf{M}_{i-1} \dots \mathbf{M}_0 \boldsymbol{\epsilon}_x^b, \boldsymbol{\beta}^b - \boldsymbol{\epsilon}_\beta^b) - h_{v_i}(m_{0 \rightarrow i}(\mathbf{x}_0^b), \boldsymbol{\beta}^b). \end{aligned} \quad (7.21)$$

1 Finally approximating $h_{v_i}(\cdot)$ using the Taylor expansion about \mathbf{x}_0^b and $\boldsymbol{\beta}^b$ and cancelling
 2 the $h_{v_i}(m_{0 \rightarrow i}(\mathbf{x}_0^b), \boldsymbol{\beta}^b)$ terms gives the background error as,

$$h_{v_i}(m_{0 \rightarrow i}(\mathbf{x}_0^t), \boldsymbol{\beta}^t) - h_{v_i}(m_{0 \rightarrow i}(\mathbf{x}_0^b), \boldsymbol{\beta}^b) \approx -(\mathbf{H}_{(1)_i} + \mathbf{C}_{x_i}) \mathbf{M}_{i-1} \dots \mathbf{M}_0 \boldsymbol{\epsilon}_x^b - \mathbf{C}_{\beta_i} \boldsymbol{\epsilon}_\beta^b \quad (7.22)$$

3 This is the background error (or bias if taking the expected value) after it has been evolved
 4 forward to the time step of the bias-corrected observations. It is the error that comes from
 5 initially using an imperfect \mathbf{x}^b in the observation operator and bias correction function.

6 The errors in the bias-corrected innovation vector at the i^{th} time step are given in
 7 equation (7.15) as the sum of the instrument error, the model error from evolving the
 8 state to the time step of the bias-corrected observation and the background errors from
 9 using an incorrect background state and incorrect background observation bias coefficient.

10 Therefore, the errors in the bias-corrected innovation vector are given by,

$$\mathbf{d}_{(1)_i}^b = \boldsymbol{\epsilon}_{(1)_i}^o - (\mathbf{H}_{(1)_i} + \mathbf{C}_{x_i}) \left(\sum_{j=1}^{i-1} \mathbf{M}_{i-1} \dots \mathbf{M}_j + \mathbf{I} \right) \boldsymbol{\eta}^t - (\mathbf{H}_{(1)_i} + \mathbf{C}_{x_i}) \mathbf{M}_{i-1} \dots \mathbf{M}_0 \boldsymbol{\epsilon}_x^b - \mathbf{C}_{\beta_i} \boldsymbol{\epsilon}_\beta^b. \quad (7.23)$$

11 As in section 2.3, we will assume that $\mathbf{C}_{x_i} = 0$, as, operationally, the bias correction is
 12 dependent on the broad regions of the state, so inaccuracies in the background state will
 13 have a negligible effect on the bias correction.

14 By applying the same method to the anchor observation innovation vector, the anchor

1 observation innovation vector at time i can be rewritten as,

$$\mathbf{d}_{(2)i}^b = \boldsymbol{\epsilon}_{(2)i}^o - \mathbf{H}_{(2)i} \left(\sum_{j=1}^{i-1} \mathbf{M}_{i-1} \dots \mathbf{M}_j + \mathbf{I} \right) \boldsymbol{\eta}^t - \mathbf{H}_{(2)i} \mathbf{M}_{i-1} \dots \mathbf{M}_0 \boldsymbol{\epsilon}_x^b. \quad (7.24)$$

2 The errors in the anchor observation innovation vector are also the sum of the instrument
 3 error, the model bias from evolving the state to the time-step of the anchor observations
 4 using a biased model, and the background error from using the wrong state background.

5 By combining equations (7.13), (7.14), (7.23) and (7.24), the analysis error equations
 6 are given by,

$$\boldsymbol{\epsilon}_x^a = \boldsymbol{\epsilon}_x^b + \hat{\mathbf{K}}_{xy(1)} (\hat{\boldsymbol{\epsilon}}_{(1)}^o - \hat{\mathbf{H}}_{(1)} \boldsymbol{\epsilon}_x^b - \hat{\mathbf{C}}_{\beta} \boldsymbol{\epsilon}_{\beta}^b - \hat{\boldsymbol{\eta}}_{\mathbf{H}(1)}) + \hat{\mathbf{K}}_{xy(2)} (\hat{\boldsymbol{\epsilon}}_{(2)}^o - \hat{\mathbf{H}}_{(2)} \boldsymbol{\epsilon}_x^b - \hat{\boldsymbol{\eta}}_{\mathbf{H}(2)}), \quad (7.25)$$

$$\boldsymbol{\epsilon}_{\beta}^a = \boldsymbol{\epsilon}_{\beta}^b + \hat{\mathbf{K}}_{\beta y(1)} (\hat{\boldsymbol{\epsilon}}_{(1)}^o - \hat{\mathbf{H}}_{(1)} \boldsymbol{\epsilon}_x^b - \hat{\mathbf{C}}_{\beta} \boldsymbol{\epsilon}_{\beta}^b - \hat{\boldsymbol{\eta}}_{\mathbf{H}(1)}) + \hat{\mathbf{K}}_{\beta y(2)} (\hat{\boldsymbol{\epsilon}}_{(2)}^o - \hat{\mathbf{H}}_{(2)} \boldsymbol{\epsilon}_x^b - \hat{\boldsymbol{\eta}}_{\mathbf{H}(2)}), \quad (7.26)$$

7 where $\hat{\boldsymbol{\eta}}_{\mathbf{H}(1)}$ and $\hat{\boldsymbol{\eta}}_{\mathbf{H}(2)}$ have been denoted by

$$\hat{\boldsymbol{\eta}}_{\mathbf{H}(1)} = \begin{pmatrix} \mathbf{0} \\ \mathbf{H}_{(1)1} \boldsymbol{\eta} \\ \mathbf{H}_{(1)2} (\mathbf{M}_1 \boldsymbol{\eta} + \boldsymbol{\eta}) \\ \vdots \\ \mathbf{H}_{(1)N(1)} \left(\sum_{j=1}^{N(1)-1} \mathbf{M}_{N(1)} \dots \mathbf{M}_{j+1} \boldsymbol{\eta} + \boldsymbol{\eta} \right) \end{pmatrix}, \quad (7.27)$$

$$\hat{\boldsymbol{\eta}}_{\mathbf{H}(2)} = \begin{pmatrix} \mathbf{0} \\ \mathbf{H}_{(2)1} \boldsymbol{\eta} \\ \mathbf{H}_{(2)2} (\mathbf{M}_1 \boldsymbol{\eta} + \boldsymbol{\eta}) \\ \vdots \\ \mathbf{H}_{(2)N(2)} \left(\sum_{j=1}^{N(2)-1} \mathbf{M}_{N(2)} \dots \mathbf{M}_{j+1} \boldsymbol{\eta} + \boldsymbol{\eta} \right) \end{pmatrix}, \quad (7.28)$$

8 where $\hat{\boldsymbol{\eta}}_{\mathbf{H}(1)}$ and $\hat{\boldsymbol{\eta}}_{\mathbf{H}(2)}$ are the model biases that accumulate from evolving the state to
 9 the time step of the bias-corrected and anchor observations respectively.

1 If we take the expected value of equations (7.25) and (7.26) it will give us the bias in
2 both the state analysis and the observation bias coefficient analysis. We can assume that
3 the anchor observations have no bias and that the bias-corrected observations have no bias
4 when they are corrected by the true bias correction, such that $\langle \epsilon_{(1)}^o \rangle = \langle \epsilon_{(2)}^o \rangle = 0$. For
5 simplicity, we assume that the bias coefficient background has no bias such that $\langle \epsilon_{\beta}^b \rangle = 0$,
6 which could be true for the first cycle. Then the expected value of the analysis errors is
7 reduced to,

$$\langle \epsilon_x^a \rangle_{t=0} = -(-\mathbf{I} + \hat{\mathbf{K}}_{xy(1)} \hat{\mathbf{H}}_{(1)} + \hat{\mathbf{K}}_{xy(2)} \hat{\mathbf{H}}_{(2)}) \langle \epsilon_x^b \rangle - \hat{\mathbf{K}}_{xy(1)} \langle \hat{\boldsymbol{\eta}}_{\mathbf{H}(1)} \rangle - \hat{\mathbf{K}}_{xy(2)} \langle \hat{\boldsymbol{\eta}}_{\mathbf{H}(2)} \rangle, \quad (7.29)$$

$$\langle \epsilon_{\beta}^a \rangle_{t=0} = -(\hat{\mathbf{K}}_{\beta y(1)} \hat{\mathbf{H}}_{(1)} + \hat{\mathbf{K}}_{\beta y(2)} \hat{\mathbf{H}}_{(2)}) \langle \epsilon_x^b \rangle - \hat{\mathbf{K}}_{\beta y(1)} \langle \hat{\boldsymbol{\eta}}_{\mathbf{H}(1)} \rangle - \hat{\mathbf{K}}_{\beta y(2)} \langle \hat{\boldsymbol{\eta}}_{\mathbf{H}(2)} \rangle, \quad (7.30)$$

8 which we have denoted as $\langle \epsilon_x^a \rangle_{t=0}$ and $\langle \epsilon_{\beta}^a \rangle_{t=0}$ to mean the expected value of the analysis
9 errors at the first cycle. Equation (7.30) is equivalent to the bias coefficient analysis
10 equation in 3DVar, equation (6.28), in chapter 6, but in the 4DVar case there are now also
11 additional model biases that come from evolving the state forward in the window using a
12 biased model.

13 As $\hat{\mathbf{K}}_v$ takes the same form as \mathbf{K}_v in equation (6.4) of chapter 6, we can also rewrite
14 $\hat{\mathbf{K}}_{\beta y(2)}$ in terms of $\hat{\mathbf{K}}_{\beta y(1)}$ as in equation (6.20) to give,

$$\hat{\mathbf{K}}_{\beta y(2)} = -\hat{\mathbf{K}}_{\beta y(1)} \hat{\mathbf{H}}_{(1)} \mathbf{B}_x \hat{\mathbf{H}}_{(2)}^T (\hat{\mathbf{H}}_{(2)} \mathbf{B}_x \hat{\mathbf{H}}_{(2)}^T + \hat{\mathbf{R}}_{(2)})^{-1}. \quad (7.31)$$

15 We can therefore rewrite $\langle \epsilon_{\beta}^a \rangle_{t=0}$, so that it is no longer explicitly a function of $\hat{\mathbf{K}}_{\beta y(2)}$ to
16 give,

$$\langle \epsilon_{\beta}^a \rangle_{t=0} = -\hat{\mathbf{K}}_{\beta y(1)} [\hat{\mathbf{H}}_{(1)} (\mathbf{I} - \hat{\mathbf{D}} \hat{\mathbf{H}}_{(2)}) \langle \epsilon_x^b \rangle + \langle \hat{\boldsymbol{\eta}}_{\mathbf{H}(1)} \rangle - \hat{\mathbf{H}}_{(1)} \hat{\mathbf{D}} \langle \hat{\boldsymbol{\eta}}_{\mathbf{H}(2)} \rangle] \quad (7.32)$$

1 where we have written

$$\hat{\mathbf{D}} = \mathbf{B}_x \hat{\mathbf{H}}_{(2)}^T (\hat{\mathbf{H}}_{(2)} \mathbf{B}_x \hat{\mathbf{H}}_{(2)}^T + \hat{\mathbf{R}}_{(2)})^{-1} \quad (7.33)$$

2 for simplicity. The term $\hat{\mathbf{D}}$ is the ratio between the background error covariance matrix
 3 and the sum of the background and anchor observation error covariance matrices. Other
 4 than $\hat{\mathbf{K}}_{\beta_{Y(1)}}$ and $\langle \hat{\boldsymbol{\eta}}_{\mathbf{H}(2)} \rangle$, $\hat{\mathbf{D}}$ is the only term in equation (7.32) that is dependent on the
 5 linearised anchor observation operator and the anchor observation error covariance matrix.
 6 Just as in chapter 6, although we want to know when the bias in the bias coefficient
 7 analysis is reduced, we are not interested in reducing $\hat{\mathbf{K}}_{\beta_{Y(1)}}$, as it would mean reducing
 8 the sensitivity of the bias coefficient analysis to the bias-corrected observations.

9 Equation (7.32) shows how the state background and state model biases contaminate
 10 the estimate of the observation bias coefficient analysis. It is similar to the bias coef-
 11 ficient analysis error equation for 3DVarBC, equation (6.29), derived in chapter 6, but
 12 also includes the model biases that accumulate to reach the timing of the bias-corrected
 13 observations ($\langle \hat{\boldsymbol{\eta}}_{\mathbf{H}(1)} \rangle$) and anchor observations ($\langle \hat{\boldsymbol{\eta}}_{\mathbf{H}(2)} \rangle$).

14 We want to understand how the timing of anchor observations impacts their ability
 15 to reduce the contamination of background and model biases in 4DVarBC. In order to
 16 simplify equations (7.29) and (7.30), we study three simple cases where the anchor and
 17 bias-corrected observations vary in time relative to each other. This reduces the time-
 18 dependent matrices as each observation type will only be valid at one time step.

19 **7.3.1 Bias-corrected observations at $t = 1$, anchor observations at $t = 2$**

20 In our first example, the anchor observations are set to be later than the bias-corrected
 21 observations to test the anchor observations' ability to reduce the contamination of back-
 22 ground and model biases when they are later than the bias-corrected observations. If
 23 the analysis is calculated at $t = 0$, then assuming that the bias-corrected observations

1 observe states at $t = 1$ and anchor observations observe states at $t = 2$, then the linearised
 2 observation operators from equation (7.6) are given by,

$$\hat{\mathbf{H}}_{(1)} = \begin{pmatrix} \mathbf{0} \\ \mathbf{H}_{(1)1}\mathbf{M}_0 \\ \mathbf{0} \end{pmatrix}, \quad \hat{\mathbf{H}}_{(2)} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{H}_{(2)2}\mathbf{M}_1\mathbf{M}_0 \end{pmatrix}, \quad (7.34)$$

3 where again the number in brackets denotes the bias-corrected or anchor observations and
 4 the numbers without brackets denote the time step. The model biases to evolve the state
 5 to the biased and anchor observations from equations (7.27) and (7.28) respectively will
 6 be,

$$\hat{\boldsymbol{\eta}}_{\mathbf{H}(1)} = \begin{pmatrix} \mathbf{0} \\ \mathbf{H}_{(1)1}\boldsymbol{\eta} \\ \mathbf{0} \end{pmatrix}, \quad \hat{\boldsymbol{\eta}}_{\mathbf{H}(2)} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{H}_{(2)2}(\mathbf{M}_1\boldsymbol{\eta} + \boldsymbol{\eta}) \end{pmatrix}. \quad (7.35)$$

7 As we have anchor observations at the second time step, we can denote

$$\mathbf{D}_2 = \mathbf{B}_x\mathbf{M}_0^T\mathbf{M}_1^T\mathbf{H}_{(2)2}^T(\mathbf{H}_{(2)2}\mathbf{M}_1\mathbf{M}_0\mathbf{B}_x\mathbf{M}_0^T\mathbf{M}_1^T\mathbf{H}_{(2)2}^T + \mathbf{R}_{(2)2})^{-1}\mathbf{H}_{(2)2}. \quad (7.36)$$

8 Note that we have included $\mathbf{H}_{(2)2}$ at the end of \mathbf{D}_2 , but not in the original $\hat{\mathbf{D}}$ in equation
 9 (7.33). The term $\mathbf{H}_{(2)2}$ does not appear in equation (7.33) because the linearised anchor
 10 observation operator appears in the model bias term in equations (7.25) and (7.26) instead
 11 and cannot be easily separated. We have chosen to include $\mathbf{H}_{(2)2}$ in equation (7.36) (instead
 12 of writing $\mathbf{D}_2\mathbf{H}_{(2)2}$) so that all of the terms that are explicitly dependent on the anchor
 13 observations are within one term (\mathbf{D}_2).

14 Therefore substituting equations (7.34), (7.35) and (7.36) into equation (7.32) gives
 15 the bias in the estimate of the observation bias coefficient analysis when bias-corrected

1 observations are at $t = 1$ and anchor observations are at $t = 2$ as,

$$\langle \epsilon_{\beta}^a \rangle_{t=0} = -[\hat{\mathbf{K}}_{\beta_{Y(1)}}]_{t=1} (\mathbf{H}_{(1)1} \mathbf{M}_0 (\mathbf{I} - \mathbf{D}_2 \mathbf{M}_1 \mathbf{M}_0) \langle \epsilon_x^b \rangle + \mathbf{H}_{(1)1} \langle \boldsymbol{\eta} \rangle - \mathbf{H}_{(1)1} \mathbf{M}_0 \mathbf{D}_2 (\mathbf{M}_1 \langle \boldsymbol{\eta} \rangle + \langle \boldsymbol{\eta} \rangle)), \quad (7.37)$$

2 where $[\hat{\mathbf{K}}_{\beta_{Y(1)}}]_{t=1}$ is the value of $\hat{\mathbf{K}}_{\beta_{Y(1)}}$ at the first time step (as at the other time steps it
3 will have values of $\mathbf{0}$). $\langle \epsilon_x^b \rangle$ is the background bias that comes from the state. The term
4 $\mathbf{H}_{(1)1} \langle \boldsymbol{\eta} \rangle$ comes from the model bias when the system has evolved from the initial to the
5 first time step to reach the bias-corrected observations. The term $\mathbf{H}_{(1)1} \mathbf{M}_0 \mathbf{D}_2 (\mathbf{M}_1 \langle \boldsymbol{\eta} \rangle +$
6 $\langle \boldsymbol{\eta} \rangle)$ is the weighting of the model biases from the first and second time steps that occur
7 when the system is evolved forward to the time step of the anchor observations. The
8 linearised anchor observation operator and the anchor observation error covariance matrix
9 are both only defined within \mathbf{D}_2 . Therefore, the characteristics of the anchor observations
10 and their role in reducing the contamination of model bias is described through \mathbf{D}_2 .

11 Equation (7.37) can be rearranged such that the model biases from the first and second
12 time steps are factorised separately to give,

$$\langle \epsilon_{\beta}^a \rangle_{t=0} = -[\hat{\mathbf{K}}_{\beta_{Y(1)}}]_{t=1} \mathbf{H}_{(1)1} [\mathbf{M}_0 (\mathbf{I} - \mathbf{D}_2 \mathbf{M}_1 \mathbf{M}_0) \langle \epsilon_x^b \rangle + (\mathbf{I} - \mathbf{M}_0 \mathbf{D}_2 \mathbf{M}_1) \langle \boldsymbol{\eta} \rangle - \mathbf{M}_0 \mathbf{D}_2 \langle \boldsymbol{\eta} \rangle]. \quad (7.38)$$

13 If the left hand side of equation (7.38) is zero, then the bias coefficient analysis has zero
14 bias, which would mean the state background and model biases would not have contami-
15 nated the observation bias coefficient analysis. In order for the matrix multiplied by $\langle \epsilon_x^b \rangle$
16 in equation (7.38) to tend to zero, $\mathbf{D}_2 \mathbf{M}_1 \mathbf{M}_0$ would have to tend to the identity. In a
17 theoretical case, this would occur when $\mathbf{D}_2 = \mathbf{M}_0^{-1} \mathbf{M}_1^{-1}$, which would happen when $\mathbf{R}_{(2)2}$
18 tends to $\mathbf{0}$ and $\mathbf{H}_{(2)2}$, \mathbf{M}_0 and \mathbf{M}_1 are full-rank and are invertible, although this is not
19 a generalisable result as we would not necessarily expect $\mathbf{H}_{(2)2}$ to be invertible. In the
20 scalar (theoretical) limit, this case would be when the anchor observations have zero ran-
21 dom error. If $\mathbf{D}_2 = \mathbf{M}_0^{-1} \mathbf{M}_1^{-1}$, then the term $\mathbf{I} - \mathbf{M}_0 \mathbf{D}_2 \mathbf{M}_1$ would also be zero and hence

1 there would be no contribution from the model bias from the first time step. However, if
2 the anchor observation error variance was zero such that $\mathbf{D}_2 = \mathbf{M}_0^{-1}\mathbf{M}_1^{-1}$, then $\mathbf{M}_0\mathbf{D}_2\langle\boldsymbol{\eta}\rangle$
3 in (7.38) would tend to $-\mathbf{M}_1^{-1}\langle\boldsymbol{\eta}\rangle$. As \mathbf{M}_1^{-1} is nonzero, $\langle\boldsymbol{\epsilon}_\beta^a\rangle$ would also be nonzero. This
4 shows that the bias in $\langle\boldsymbol{\epsilon}_\beta^a\rangle$ can never fully be removed unless $\hat{\mathbf{K}}_{\beta y(1)} = 0$ or $\mathbf{H}_{(1)1} = 0$,
5 which are both trivial cases. However, as $\boldsymbol{\eta}$ is multiplied by the inverse of the linearised
6 model at the first time step in the limit when the anchor observation error variance tends
7 to zero, then any part of $\boldsymbol{\eta}$ that is projected onto growing modes will decay, and any part of
8 $\boldsymbol{\eta}$ that is projected onto decaying modes will grow. If the model has growing modes, then
9 the inverse of the linearised model multiplied by the model bias would decay the model
10 bias. In general, atmospheric models have errors that grow (Lorenz, 2005; Simmons et
11 al., 1995), therefore although the term $-\mathbf{M}_1^{-1}\langle\boldsymbol{\eta}\rangle$ would not be zero in the limit with zero
12 anchor observation error variance, the model bias would be reduced.

13 In the more realistic case where the anchor observation error variance is small (instead
14 of $\mathbf{0}$), the term $\mathbf{D}_2\mathbf{M}_1\mathbf{M}_0$ would only tend towards the identity. This is because $\mathbf{D}_2\mathbf{M}_1\mathbf{M}_0$
15 is approximately the ratio between the background error variances and the sum of the
16 background and anchor observation error variances. Therefore, if the anchor observations
17 observe states at a time step later than the bias-corrected observations, then the anchor
18 observations can reduce the contamination of state background bias and model bias from
19 the first time step, but the anchor observations will detrimentally introduce a model bias
20 into the bias coefficient analysis. This additional model bias occurs when the system is
21 evolved from the time step of the bias-corrected observations to the time step of the anchor
22 observations, but would be small if the anchor observations are precise for model bias that
23 is projected onto growing modes.

24 **7.3.1.1 Anchor and bias-corrected observations observe different spatial states**

25 Next we will split the spatial domain into sections: \mathbf{x}_ϕ and \mathbf{x}_ψ , as we did for the 3DVar
26 case in section 6.2.1.2. This will test how well the anchor observations can reduce the

1 contamination of background and model biases in the observation bias coefficient, where
 2 the biases come from states the anchor observations do not observe. We will assume bias-
 3 corrected observations can observe anywhere in the domain, but anchor observations only
 4 observe states in \mathbf{x}_ψ . Then the observation operators become,

$$\mathbf{H}_{(1)1} = \begin{pmatrix} \mathbf{H}_{(1)1_\phi} & \mathbf{H}_{(1)1_\psi} \end{pmatrix}, \quad (7.39)$$

$$\mathbf{H}_{(2)2} = \begin{pmatrix} \mathbf{0} & \mathbf{H}_{(2)2_\psi} \end{pmatrix}, \quad (7.40)$$

5 where $\mathbf{H}_{(1)1_\phi}$ is related to bias-corrected observations of \mathbf{x}_ϕ at time $t = 1$ and $\mathbf{H}_{(1)1_\psi}$
 6 and $\mathbf{H}_{(2)2_\psi}$ are related to observations of \mathbf{x}_ψ at times $t = 1$ and $t = 2$ respectively. The
 7 background error covariance matrix which describes the relationships between \mathbf{x}_ϕ^b and \mathbf{x}_ψ^b
 8 is,

$$\mathbf{B}_x = \begin{pmatrix} \mathbf{B}_{x_\phi} & \mathbf{B}_{x_{\phi\psi}} \\ \mathbf{B}_{x_{\phi\psi}}^\top & \mathbf{B}_{x_\psi} \end{pmatrix}. \quad (7.41)$$

9 The tangent linear models for the first two time steps are given by,

$$\mathbf{M}_0 = \begin{pmatrix} \mathbf{M}_{0_\phi} & \mathbf{M}_{0_{\phi\psi}} \\ \mathbf{M}_{0_{\psi\phi}} & \mathbf{M}_{0_\psi} \end{pmatrix}, \quad \mathbf{M}_1 = \begin{pmatrix} \mathbf{M}_{1_\phi} & \mathbf{M}_{1_{\phi\psi}} \\ \mathbf{M}_{1_{\psi\phi}} & \mathbf{M}_{1_\psi} \end{pmatrix}. \quad (7.42)$$

10 As the matrix \mathbf{D}_2 (equation 7.36)) determines how much control the anchor observations
 11 have over the observation bias coefficient, we calculate it for this case. It will be given by,

$$\mathbf{D}_2 = \begin{pmatrix} \mathbf{B}_{x_\phi} & \mathbf{B}_{x_{\phi\psi}} \\ \mathbf{B}_{x_{\phi\psi}}^\top & \mathbf{B}_{x_\psi} \end{pmatrix} \begin{pmatrix} \mathbf{M}_{0_\phi}^\top & \mathbf{M}_{0_{\psi\phi}}^\top \\ \mathbf{M}_{0_{\phi\psi}}^\top & \mathbf{M}_{0_\psi}^\top \end{pmatrix} \begin{pmatrix} \mathbf{M}_{1_\phi}^\top & \mathbf{M}_{1_{\psi\phi}}^\top \\ \mathbf{M}_{1_{\phi\psi}}^\top & \mathbf{M}_{1_\psi}^\top \end{pmatrix} \begin{pmatrix} \mathbf{0} \\ \mathbf{H}_{(2)2_\psi}^\top \end{pmatrix} [\dots]^{-1} \begin{pmatrix} \mathbf{0} & \mathbf{H}_{(2)2_\psi} \end{pmatrix}, \quad (7.43)$$

1 where we have denoted $[\dots]^{-1}$ to be the inverse matrix in equation (7.36). We can write
 2 equation (7.43) in block element form as,

$$\mathbf{D}_{2\phi} = \mathbf{0} \quad (7.44)$$

$$\begin{aligned} \mathbf{D}_{2\phi\psi} = & [(\mathbf{B}_{x\phi} \mathbf{M}_{0\phi}^T + \mathbf{B}_{x\phi\psi} \mathbf{M}_{0\phi\psi}^T) \mathbf{M}_{1\psi\phi}^T \mathbf{H}_{(2)2\psi}^T \\ & + (\mathbf{B}_{x\phi} \mathbf{M}_{0\psi\phi}^T + \mathbf{B}_{x\phi\psi} \mathbf{M}_{0\psi}^T) \mathbf{M}_{1\psi}^T \mathbf{H}_{(2)2\psi}^T] [\dots]^{-1} \mathbf{H}_{(2)2\psi} \end{aligned} \quad (7.45)$$

$$\mathbf{D}_{2\psi\phi} = \mathbf{0} \quad (7.46)$$

$$\begin{aligned} \mathbf{D}_{2\psi} = & [(\mathbf{B}_{x\psi} \mathbf{M}_{0\phi}^T + \mathbf{B}_{x\psi} \mathbf{M}_{0\phi\psi}^T) \mathbf{M}_{1\psi\phi}^T \mathbf{H}_{(2)2\psi}^T \\ & + (\mathbf{B}_{x\phi\psi} \mathbf{M}_{0\psi\phi}^T + \mathbf{B}_{x\psi} \mathbf{M}_{0\psi}^T) \mathbf{M}_{1\psi}^T \mathbf{H}_{(2)2\psi}^T] [\dots]^{-1} \mathbf{H}_{(2)2\psi} \end{aligned} \quad (7.47)$$

3 where for simplicity, we have denoted the ϕ and ψ blocks of elements in \mathbf{D}_2 to be,

$$\mathbf{D}_2 = \begin{pmatrix} \mathbf{D}_{2\phi} & \mathbf{D}_{2\phi\psi} \\ \mathbf{D}_{2\psi\phi} & \mathbf{D}_{2\psi} \end{pmatrix}. \quad (7.48)$$

4 Now that we have calculated \mathbf{D}_2 for the case when anchor observations are at $t = 2$ and
 5 bias-corrected observations are at $t = 1$ when we have split the spatial state into ϕ and ψ ,
 6 we can use this to determine where the anchor observations can reduce the contamination
 7 of state background and state model biases, relative to the spatial domain. In equation
 8 (7.38) the state background bias is pre-multiplied by $\mathbf{I} - \mathbf{D}_2 \mathbf{M}_1 \mathbf{M}_0$. As $\mathbf{D}_{2\phi}$ and $\mathbf{D}_{2\psi\phi}$ are
 9 both zero, then calculating $\mathbf{I} - \mathbf{D}_2 \mathbf{M}_1 \mathbf{M}_0$ gives,

$$\mathbf{I} - \mathbf{D}_2 \mathbf{M}_1 \mathbf{M}_0 = \begin{pmatrix} \mathbf{I} - \mathbf{D}_{2\phi\psi} \mathbf{M}_{1\psi\phi} \mathbf{M}_{0\phi} - \mathbf{D}_{2\phi\psi} \mathbf{M}_{1\psi} \mathbf{M}_{0\psi\phi} & -\mathbf{D}_{2\phi\psi} \mathbf{M}_{1\psi\phi} \mathbf{M}_{0\phi\psi} - \mathbf{D}_{2\phi\psi} \mathbf{M}_{1\psi} \mathbf{M}_{0\psi} \\ -\mathbf{D}_{2\psi} \mathbf{M}_{1\psi\phi} \mathbf{M}_{0\phi} - \mathbf{D}_{2\psi} \mathbf{M}_{1\psi} \mathbf{M}_{0\psi\phi} & \mathbf{I} - \mathbf{D}_{2\psi} \mathbf{M}_{1\psi\phi} \mathbf{M}_{0\phi\psi} - \mathbf{D}_{2\psi} \mathbf{M}_{1\psi} \mathbf{M}_{0\psi} \end{pmatrix}. \quad (7.49)$$

1 If the state background bias is split into its two spatial parts such that it is given by,

$$\langle \boldsymbol{\epsilon}_x^b \rangle = \begin{pmatrix} \langle \boldsymbol{\epsilon}_{x_\phi}^b \rangle \\ \langle \boldsymbol{\epsilon}_{x_\psi}^b \rangle \end{pmatrix}, \quad (7.50)$$

2 then

$$\begin{aligned} (\mathbf{I} - \mathbf{D}_2 \mathbf{M}_1 \mathbf{M}_0) \langle \boldsymbol{\epsilon}_x^b \rangle = \\ \left(\begin{aligned} &(\mathbf{I} - \mathbf{D}_{2\phi\psi} \mathbf{M}_{1\psi\phi} \mathbf{M}_{0\phi} - \mathbf{D}_{2\phi\psi} \mathbf{M}_{1\psi} \mathbf{M}_{0\psi\phi}) \langle \boldsymbol{\epsilon}_{x_\phi}^b \rangle - (\mathbf{D}_{2\phi\psi} \mathbf{M}_{1\psi\phi} \mathbf{M}_{0\phi\psi} + \mathbf{D}_{2\phi\psi} \mathbf{M}_{1\psi} \mathbf{M}_{0\psi}) \langle \boldsymbol{\epsilon}_{x_\psi}^b \rangle \\ &-(\mathbf{D}_{2\psi} \mathbf{M}_{1\psi\phi} \mathbf{M}_{0\phi} + \mathbf{D}_{2\psi} \mathbf{M}_{1\psi} \mathbf{M}_{0\psi\phi}) \langle \boldsymbol{\epsilon}_{x_\phi}^b \rangle + (\mathbf{I} - \mathbf{D}_{2\psi} \mathbf{M}_{1\psi\phi} \mathbf{M}_{0\phi\psi} - \mathbf{D}_{2\psi} \mathbf{M}_{1\psi} \mathbf{M}_{0\psi}) \langle \boldsymbol{\epsilon}_{x_\psi}^b \rangle \end{aligned} \right) \end{aligned} \quad (7.51)$$

3 Equation (7.51) shows that both the state background bias that is only observed by the
4 bias-corrected observations ($\langle \boldsymbol{\epsilon}_{x_\phi}^b \rangle$) and the state background bias that is observed by both
5 kinds of observations ($\langle \boldsymbol{\epsilon}_{x_\psi}^b \rangle$) are somewhat controlled by the anchor observations through
6 the operator \mathbf{D}_2 . We calculated an equivalent equation for when the anchor observations
7 do not observe the whole state in 3DVarBC in equation (6.41), chapter 6. In the 3DVar
8 case, $\langle \boldsymbol{\epsilon}_{x_\phi}^b \rangle$ was not multiplied by any elements of \mathbf{D} and as a result, $\langle \boldsymbol{\epsilon}_{x_\phi}^b \rangle$ could only be
9 reduced implicitly by the anchor observations through the background error correlations
10 between the states \mathbf{x}_ϕ and \mathbf{x}_ψ . In the 4DVar case, more information is spread across the
11 domain by the model, which means that anchor observations will have more control over
12 state background biases in states that are only observed by the bias-corrected observations,
13 not the anchor observations. It should be noted however, that this spreading is limited
14 to the advection velocity of the model and that anchor observations may still not be able
15 to control background biases from states observed by bias-corrected observations that are
16 far away from the anchor observations.

17 Unlike in 3DVar, model biases from different spatial states in the domain will also
18 contaminate the observation bias coefficient analysis. In equation (7.38) the model bias
19 $\boldsymbol{\eta}$ has been left-multiplied by $\mathbf{I} - \mathbf{M}_0 \mathbf{D}_2 \mathbf{M}_1$ and $\mathbf{M}_0 \mathbf{D}_2$ respectively. In the case where

1 anchor observations do not observe the state ϕ , the term $\mathbf{I} - \mathbf{M}_0\mathbf{D}_2\mathbf{M}_1$ is given by,

$$\mathbf{I} - \mathbf{M}_0\mathbf{D}_2\mathbf{M}_1 = \begin{pmatrix} \mathbf{I} - \mathbf{M}_{0\phi}\mathbf{D}_{2\phi\psi}\mathbf{M}_{1\psi\phi} - \mathbf{M}_{0\phi\psi}\mathbf{D}_{2\psi}\mathbf{M}_{1\psi\phi} & -\mathbf{M}_{0\phi}\mathbf{D}_{2\phi\psi}\mathbf{M}_{1\psi} - \mathbf{M}_{0\phi\psi}\mathbf{D}_{2\psi}\mathbf{M}_{1\psi} \\ -\mathbf{M}_{0\psi\phi}\mathbf{D}_{2\phi\psi}\mathbf{M}_{1\psi\phi} - \mathbf{M}_{0\psi}\mathbf{D}_{2\psi\phi}\mathbf{M}_{1\psi\phi} & \mathbf{I} - \mathbf{M}_{0\psi\phi}\mathbf{D}_{2\phi\psi}\mathbf{M}_{1\psi} - \mathbf{M}_{0\psi}\mathbf{D}_{2\psi\phi}\mathbf{M}_{1\psi} \end{pmatrix}. \quad (7.52)$$

2 Therefore, if the model bias is split into its spatial parts such that it is given by,

$$\langle \boldsymbol{\eta} \rangle = \begin{pmatrix} \langle \boldsymbol{\eta}_\phi \rangle \\ \langle \boldsymbol{\eta}_\psi \rangle \end{pmatrix}, \quad (7.53)$$

3 then

$$\begin{aligned} (\mathbf{I} - \mathbf{M}_0\mathbf{D}_2\mathbf{M}_1)\langle \boldsymbol{\eta} \rangle = \\ \begin{pmatrix} (\mathbf{I} - \mathbf{M}_{0\phi}\mathbf{D}_{2\phi\psi}\mathbf{M}_{1\psi\phi} - \mathbf{M}_{0\phi\psi}\mathbf{D}_{2\psi}\mathbf{M}_{1\psi\phi})\langle \boldsymbol{\eta}_\phi \rangle - (\mathbf{M}_{0\phi}\mathbf{D}_{2\phi\psi}\mathbf{M}_{1\psi} + \mathbf{M}_{0\phi\psi}\mathbf{D}_{2\psi}\mathbf{M}_{1\psi})\langle \boldsymbol{\eta}_\psi \rangle \\ -(\mathbf{M}_{0\psi\phi}\mathbf{D}_{2\phi\psi}\mathbf{M}_{1\psi\phi} + \mathbf{M}_{0\psi}\mathbf{D}_{2\psi\phi}\mathbf{M}_{1\psi\phi})\langle \boldsymbol{\eta}_\phi \rangle + (\mathbf{I} - \mathbf{M}_{0\psi\phi}\mathbf{D}_{2\phi\psi}\mathbf{M}_{1\psi} - \mathbf{M}_{0\psi}\mathbf{D}_{2\psi\phi}\mathbf{M}_{1\psi})\langle \boldsymbol{\eta}_\psi \rangle \end{pmatrix}. \end{aligned} \quad (7.54)$$

4 As for the background bias in equation (7.51), equation (7.54) shows that the model bias in
5 the first time step from both the states observed only by bias-corrected observations ($\langle \boldsymbol{\eta}_\phi \rangle$)
6 and states observed by both bias-corrected and anchor observations ($\langle \boldsymbol{\eta}_\psi \rangle$) are controlled
7 by anchor observations through the variable \mathbf{D}_2 . Just as for the state background bias,
8 the model spreads information between states that are and are not observed by the anchor
9 observations, which allows the anchor observations to be able to control model biases in
10 some of the states that they do not observe, as long as the dynamics of the model expand
11 far enough.

12 In equation (7.38), the term that comes from the model bias at the second time step
13 is given by,

$$\mathbf{M}_0\mathbf{D}_2\langle \boldsymbol{\eta} \rangle = \begin{pmatrix} (\mathbf{M}_{0\phi}\mathbf{D}_{2\phi\psi} + \mathbf{M}_{0\phi\psi}\mathbf{D}_{2\psi})\langle \boldsymbol{\eta}_\psi \rangle \\ (\mathbf{M}_{0\psi\phi}\mathbf{D}_{2\phi\psi} + \mathbf{M}_{0\psi}\mathbf{D}_{2\psi\phi})\langle \boldsymbol{\eta}_\psi \rangle \end{pmatrix} \quad (7.55)$$

14 In equation (7.55) there is no $\langle \boldsymbol{\eta}_\phi \rangle$ term. This is because equation (7.55) is the model bias

1 that comes from reaching the time step of the anchor observations ($t = 2$). Therefore,
 2 as the anchor observations do not observe \mathbf{x}_ϕ and the bias-corrected observations do not
 3 observe states later than at $t = 1$, there is no model bias associated with the state \mathbf{x}_ϕ at
 4 time $t = 2$. This means that any bias associated with a state that is not observed by the
 5 anchor observations will not be propagated further than the time step they are observed
 6 (ie. the time step of the bias-corrected observations).

7 **7.3.2 Anchor observations at $t = 1$, bias-corrected observations at $t = 2$**

8 In this next section, we set the anchor observations to observe states earlier in the window
 9 than the bias-corrected observations, such that we have anchor observations at $t = 1$ and
 10 bias-corrected observations at $t = 2$ (the analysis time will still be at $t = 0$). Then the
 11 linearised observation operators are given by,

$$\hat{\mathbf{H}}_{(1)} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{H}_{(1)2} \mathbf{M}_1 \mathbf{M}_0 \end{pmatrix}, \quad \hat{\mathbf{H}}_{(2)} = \begin{pmatrix} \mathbf{0} \\ \mathbf{H}_{(2)1} \mathbf{M}_0 \\ \mathbf{0} \end{pmatrix}, \quad (7.56)$$

12 and the model biases that come from evolving the state to the time step of the biased and
 13 anchor observations respectively are given by,

$$\hat{\boldsymbol{\eta}}_{\mathbf{H}(1)} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{H}_{(1)2} (\mathbf{M}_1 \boldsymbol{\eta} + \boldsymbol{\eta}) \end{pmatrix}, \quad \hat{\boldsymbol{\eta}}_{\mathbf{H}(2)} = \begin{pmatrix} \mathbf{0} \\ \mathbf{H}_{(2)1} \boldsymbol{\eta} \\ \mathbf{0} \end{pmatrix}. \quad (7.57)$$

14 As the anchor observations are at $t = 1$, we denote,

$$\mathbf{D}_1 = \mathbf{B}_x \mathbf{M}_0^T \mathbf{H}_{(2)1}^T (\mathbf{H}_{(2)1} \mathbf{M}_0 \mathbf{B}_x \mathbf{M}_0^T \mathbf{H}_{(2)1}^T + \mathbf{R}_{(2)1})^{-1} \mathbf{H}_{(2)1}. \quad (7.58)$$

1 Substituting equations (7.56), (7.57) and (7.58) into equation (7.32), we find the expected
 2 value of the bias coefficient analysis error when the anchor observations are at $t = 1$ and
 3 the bias-corrected observations are at $t = 2$ is given by,

$$\langle \epsilon_{\beta}^a \rangle_{t=0} = -[\hat{\mathbf{K}}_{\beta_{y(1)}}]_{t=2} \mathbf{H}_{(1)2} [\mathbf{M}_1 \mathbf{M}_0 (\mathbf{I} - \mathbf{D}_1 \mathbf{M}_0) \langle \epsilon_x^b \rangle + \mathbf{M}_1 \langle \boldsymbol{\eta} \rangle + \langle \boldsymbol{\eta} \rangle - \mathbf{M}_1 \mathbf{M}_0 \mathbf{D}_1 \langle \boldsymbol{\eta} \rangle], \quad (7.59)$$

4 where $[\hat{\mathbf{K}}_{\beta_{y(1)}}]_{t=2}$ is the value of $\hat{\mathbf{K}}_{\beta_{y(1)}}$ at the second time step (as at the other time
 5 steps it will have values of $\mathbf{0}$). The term $\langle \epsilon_x^b \rangle$ is the state background bias that comes
 6 into play when comparing the biased background with both the bias-corrected and anchor
 7 observations. The terms $\mathbf{M}_1 \langle \boldsymbol{\eta} \rangle$ and $\langle \boldsymbol{\eta} \rangle$ are the model biases associated with evolving the
 8 initial state to the time step of the bias-corrected observations and the term $\mathbf{M}_1 \mathbf{M}_0 \mathbf{D}_1 \langle \boldsymbol{\eta} \rangle$
 9 is the model bias associated with evolving the state from the initial time step to the
 10 time step of the anchor observations. Note that, as we are calculating the bias in the
 11 observation bias coefficient, the model bias associated with the anchor observations has
 12 been multiplied by $\mathbf{M}_1 \mathbf{M}_0$ as it has been evolved to the time step of the bias-corrected
 13 observations, in order to compare with the other biases at the same time. The linearised
 14 anchor observation operator and the anchor observation error variance only appear in \mathbf{D}_1 ,
 15 hence the anchor observation statistics will only directly impact \mathbf{D}_1 .

16 Equation (7.59) can be rearranged to give,

$$\langle \epsilon_{\beta}^a \rangle_{t=0} = -[\hat{\mathbf{K}}_{\beta_{y(1)}}]_{t=2} \mathbf{H}_{(1)2} [\mathbf{M}_1 \mathbf{M}_0 (\mathbf{I} - \mathbf{D}_1 \mathbf{M}_0) \langle \epsilon_x^b \rangle + \mathbf{M}_1 (\mathbf{I} - \mathbf{M}_0 \mathbf{D}_1) \langle \boldsymbol{\eta} \rangle + \langle \boldsymbol{\eta} \rangle] \quad (7.60)$$

17 where we have factorised the model biases from the same time step together, such that
 18 the term $\mathbf{M}_1 (\mathbf{I} - \mathbf{M}_0 \mathbf{D}_1) \langle \boldsymbol{\eta} \rangle$ is the model bias associated with the first time step and the
 19 term $\langle \boldsymbol{\eta} \rangle$ is the model bias associated with the second time step. If $\mathbf{D}_1 \mathbf{M}_0$ tends to the
 20 identity, then the dependency of $\langle \epsilon_{\beta}^a \rangle$ on $\langle \epsilon_x^b \rangle$ will tend to zero. If $\mathbf{D}_1 \mathbf{M}_0$ tends to the
 21 identity and \mathbf{M}_0 and \mathbf{D}_2 are full rank, then $\mathbf{M}_0 \mathbf{D}_1$ will also tend to the identity, reducing
 22 the second term ($\mathbf{M}_1 (\mathbf{I} - \mathbf{M}_0 \mathbf{D}_1) \langle \boldsymbol{\eta} \rangle$) to zero. Hence if the anchor observations reduce

1 the contamination of state background bias, they will also reduce the contamination of
 2 model bias that comes from evolving the system to the first time step. The terms $\mathbf{D}_1\mathbf{M}_0$
 3 and $\mathbf{M}_0\mathbf{D}_1$ will tend to the identity when $\mathbf{H}_{(2)1}$ and \mathbf{M}_1 are full rank and the anchor
 4 observation error variance is very small, i.e. when we have very precise anchor observations.
 5 Therefore, the more precise the anchor observations, the smaller the contamination of state
 6 background bias and model bias from the first time step. However, there is a $\langle\boldsymbol{\eta}\rangle$ term,
 7 caused by evolving the state to the time step of the bias-corrected observations, that is left
 8 behind in this reduction, which is only multiplied by $\hat{\mathbf{K}}_{\beta_{Y(1)}}$ and $\mathbf{H}_{(1)2}$. The contamination
 9 of this final $\langle\boldsymbol{\eta}\rangle$ on $\langle\boldsymbol{\epsilon}_\beta^a\rangle$ could only be reduced if either $\hat{\mathbf{K}}_{\beta_{Y(1)}}$ or $\mathbf{H}_{(1)2}$ are small, which
 10 would be trivial cases. Therefore, if the bias-corrected observations observe states that are
 11 later in the time window than the states observed by the anchor observations, then there
 12 will be a model bias caused by evolving the state to the bias-corrected observations, that
 13 the anchor observations cannot explicitly reduce the effect of. This additional model bias
 14 is worse than the additional model bias in equation (7.37) when the anchor observations
 15 were later in the window, as, in equation (7.60), $\boldsymbol{\eta}$ is not multiplied by \mathbf{D}_1 , so the anchor
 16 observations have no control over it, whereas in equation (7.37) the additional model bias
 17 is multiplied by $-\mathbf{M}_0\mathbf{D}_2$, which, as we showed in section 7.3.1, will decay with more precise
 18 anchor observations.

19 **7.3.2.1 Anchor and bias-corrected observations observe different spatial states**

20 We will again split the spatial domain into two sections: \mathbf{x}_ϕ and \mathbf{x}_ψ , where \mathbf{x}_ϕ and \mathbf{x}_ψ
 21 do not overlap, to test how well anchor observations can reduce the contamination of bias
 22 in states they do not observe. We will assume bias-corrected observations can observe
 23 states in both \mathbf{x}_ϕ and \mathbf{x}_ψ , but anchor observations can only observe states in \mathbf{x}_ψ , to test
 24 the ability of anchor observations to reduce the contamination of bias when they do not

1 observe every state. Then the observation operators become,

$$\mathbf{H}_{(1)2} = \begin{pmatrix} \mathbf{H}_{(1)2\phi} & \mathbf{H}_{(1)2\psi} \end{pmatrix}, \quad (7.61)$$

$$\mathbf{H}_{(2)1} = \begin{pmatrix} \mathbf{0} & \mathbf{H}_{(2)1\psi} \end{pmatrix}, \quad (7.62)$$

2 where $\mathbf{H}_{(1)2\phi}$ is related to bias-corrected observations of \mathbf{x}_ϕ at time $t = 2$ and $\mathbf{H}_{(1)2\psi}$ and
 3 $\mathbf{H}_{(2)1\psi}$ are related to observations of \mathbf{x}_ψ at times $t = 2$ and $t = 1$ respectively. If we use
 4 \mathbf{B}_x and \mathbf{M}_0 from equations (7.41) and (7.42), then \mathbf{D}_1 becomes,

$$\mathbf{D}_1 = \begin{pmatrix} \mathbf{B}_{x_\phi} & \mathbf{B}_{x_{\phi\psi}} \\ \mathbf{B}_{x_{\phi\psi}}^\top & \mathbf{B}_{x_\psi} \end{pmatrix} \begin{pmatrix} \mathbf{M}_{0\phi}^\top & \mathbf{M}_{0\psi\phi}^\top \\ \mathbf{M}_{0\phi\psi}^\top & \mathbf{M}_{0\psi}^\top \end{pmatrix} \begin{pmatrix} \mathbf{0} \\ \mathbf{H}_{(2)1\psi}^\top \end{pmatrix} [\dots]^{-1} \begin{pmatrix} \mathbf{0} & \mathbf{H}_{(2)1\psi} \end{pmatrix}, \quad (7.63)$$

5 where we have denoted $[\dots]^{-1}$ to be the inverse matrix in equation (7.58). Multiplying the
 6 matrices in equation (7.63) together gives,

$$\mathbf{D}_1 = \begin{pmatrix} \mathbf{0} & (\mathbf{B}_{x_\phi} \mathbf{M}_{0\psi\phi}^\top \mathbf{H}_{(2)1\psi}^\top + \mathbf{B}_{x_{\phi\psi}} \mathbf{M}_{0\psi}^\top \mathbf{H}_{(2)1\psi}^\top) [\dots]^{-1} \mathbf{H}_{(2)1\psi} \\ \mathbf{0} & (\mathbf{B}_{x_{\phi\psi}}^\top \mathbf{M}_{0\psi\phi}^\top \mathbf{H}_{(2)1\psi}^\top + \mathbf{B}_{x_\psi} \mathbf{M}_{0\psi}^\top \mathbf{H}_{(2)1\psi}^\top) [\dots]^{-1} \mathbf{H}_{(2)1\psi} \end{pmatrix}. \quad (7.64)$$

7 We can denote each block element of \mathbf{D}_1 as,

$$\mathbf{D}_1 =: \begin{pmatrix} \mathbf{D}_{1\phi} & \mathbf{D}_{1\phi\psi} \\ \mathbf{D}_{1\psi\phi} & \mathbf{D}_{1\psi} \end{pmatrix} \quad (7.65)$$

8 for simplicity.

9 From equation (7.60), the state background bias is pre-multiplied by $\mathbf{I} - \mathbf{D}_1 \mathbf{M}_0$, which
 10 is given by,

$$\mathbf{I} - \mathbf{D}_1 \mathbf{M}_0 = \begin{pmatrix} \mathbf{I} - \mathbf{D}_{1\phi} \mathbf{M}_{0\psi\phi} & -\mathbf{D}_{1\phi\psi} \mathbf{M}_{0\psi} \\ -\mathbf{D}_{1\psi\phi} \mathbf{M}_{0\psi\phi} & \mathbf{I} - \mathbf{D}_{1\psi} \mathbf{M}_{0\psi} \end{pmatrix}. \quad (7.66)$$

1 If the state background bias is split into \mathbf{x}_ϕ and \mathbf{x}_ψ parts as in equation (7.50), then the
 2 contamination of state background bias on the observation bias coefficient analysis is given
 3 by,

$$(\mathbf{I} - \mathbf{D}_1 \mathbf{M}_0) \langle \boldsymbol{\epsilon}_{\mathbf{x}}^b \rangle = \begin{pmatrix} (\mathbf{I} - \mathbf{D}_{1\phi\psi} \mathbf{M}_{0\psi\phi}) \langle \boldsymbol{\epsilon}_{\mathbf{x}_\phi}^b \rangle - \mathbf{D}_{1\phi\psi} \mathbf{M}_{0\psi} \langle \boldsymbol{\epsilon}_{\mathbf{x}_\psi}^b \rangle \\ -\mathbf{D}_{1\psi} \mathbf{M}_{0\psi\phi} \langle \boldsymbol{\epsilon}_{\mathbf{x}_\phi}^b \rangle + (\mathbf{I} - \mathbf{D}_{1\psi} \mathbf{M}_{0\psi}) \langle \boldsymbol{\epsilon}_{\mathbf{x}_\psi}^b \rangle \end{pmatrix}. \quad (7.67)$$

4 Equation (7.67) shows that both the state background bias that comes from the states
 5 only observed by bias-corrected observations ($\langle \boldsymbol{\epsilon}_{\mathbf{x}_\phi}^b \rangle$) and the state background bias that
 6 comes from the states observed by both observations ($\langle \boldsymbol{\epsilon}_{\mathbf{x}_\psi}^b \rangle$) are somewhat controlled by
 7 the anchor observations, as they are both multiplied by an element of \mathbf{D}_1 . This is a similar
 8 result to the result we found in section 7.3.1.1: anchor observations that are later than
 9 the initial time step have some control over the state background bias in states they do
 10 not observe, as some information can be spread across the domain by the model, despite
 11 the anchor observations only observing a subsection of the domain. It should again be
 12 noted that the model can only spread information at the rate of the advection velocities,
 13 so anchor observations could only control the contamination of the state background bias
 14 in regions near to where they are observing.

15 Model biases from different spatial states in the domain will also contaminate the
 16 observation bias coefficient analysis. In equation (7.60), the model bias that comes from
 17 evolving the state to the first time step (the time step of the anchor observations) is pre-
 18 multiplied by $\mathbf{I} - \mathbf{M}_0 \mathbf{D}_1$, which, when the anchor observations only observe \mathbf{x}_ψ , is given
 19 by,

$$\mathbf{I} - \mathbf{M}_0 \mathbf{D}_1 = \begin{pmatrix} \mathbf{I} & -\mathbf{M}_{0\phi} \mathbf{D}_{1\phi\psi} - \mathbf{M}_{0\phi\psi} \mathbf{D}_{1\psi} \\ \mathbf{0} & \mathbf{I} - \mathbf{M}_{0\psi\phi} \mathbf{D}_{1\phi\psi} - \mathbf{M}_{0\psi} \mathbf{D}_{1\psi} \end{pmatrix}. \quad (7.68)$$

20 When the model bias is split into the \mathbf{x}_ϕ and \mathbf{x}_ψ parts, as in equation (7.53), the contam-
 21 ination of model bias from the first time step on the observation bias coefficient analysis

1 is given by,

$$(\mathbf{I} - \mathbf{M}_0 \mathbf{D}_1) \langle \boldsymbol{\eta} \rangle = \begin{pmatrix} \langle \boldsymbol{\eta}_\phi \rangle - (\mathbf{M}_{0_\phi} \mathbf{D}_{1_{\phi\psi}} + \mathbf{M}_{0_{\phi\psi}} \mathbf{D}_{1_\psi}) \langle \boldsymbol{\eta}_\psi \rangle \\ (\mathbf{I} - \mathbf{M}_{0_\psi} \mathbf{D}_{1_{\phi\psi}} - \mathbf{M}_{0_\psi} \mathbf{D}_{1_\psi}) \langle \boldsymbol{\eta}_\psi \rangle \end{pmatrix}. \quad (7.69)$$

2 In equation (7.69), the model biases from both the states observed by only the bias-
 3 corrected observations ($\langle \boldsymbol{\eta}_\phi \rangle$) and from states observed by both bias-corrected and an-
 4 chor observations ($\langle \boldsymbol{\eta}_\psi \rangle$) appear. However, the model bias that is only observed by the
 5 bias-corrected observations is not multiplied by \mathbf{D}_1 , which means that, when anchor ob-
 6 servations are earlier than the bias-corrected observations, the anchor observations will
 7 not explicitly be able to control the contamination of the model bias, which comes from
 8 evolving the model to the time step of the anchor observations, from states they do not
 9 observe. This is more similar to the 3DVar case in equation (6.41), chapter 6, as the
 10 model bias that is not observed by the anchor observations will only be able to implicitly
 11 be reduced by the anchor observations through the background error correlations between
 12 \mathbf{x}_ϕ and \mathbf{x}_ψ , that appear in $\mathbf{D}_{1_{\phi\psi}}$ and \mathbf{D}_{1_ψ} .

13 7.3.3 Anchor and bias-corrected observations at the same time

14 If the analysis is calculated at $t = 0$ and both anchor and bias-corrected observations
 15 observe states at $t = 1$ such that the linearised observation operators are given by,

$$\hat{\mathbf{H}}_{(1)} = \begin{pmatrix} \mathbf{0} \\ \mathbf{H}_{(1)1} \mathbf{M}_0 \end{pmatrix}, \quad \hat{\mathbf{H}}_{(2)} = \begin{pmatrix} \mathbf{0} \\ \mathbf{H}_{(2)1} \mathbf{M}_0 \end{pmatrix}, \quad (7.70)$$

1 then the model biases that come from evolving the state to the bias-corrected and anchor
 2 observations respectively are,

$$\hat{\boldsymbol{\eta}}_{\mathbf{H}(1)} = \begin{pmatrix} \mathbf{0} \\ \mathbf{H}_{(1)1} \boldsymbol{\eta} \end{pmatrix}, \quad \hat{\boldsymbol{\eta}}_{\mathbf{H}(2)} = \begin{pmatrix} \mathbf{0} \\ \mathbf{H}_{(2)1} \boldsymbol{\eta} \end{pmatrix}. \quad (7.71)$$

3 The expected value of the bias coefficient analysis error, equation (7.32), is therefore given
 4 by

$$\langle \boldsymbol{\epsilon}_{\beta}^a \rangle_{t=0} = -[\hat{\mathbf{K}}_{\beta y(1)}]_{t=1} \mathbf{H}_{(1)1} [\mathbf{M}_0(\mathbf{I} - \mathbf{D}_1 \mathbf{M}_0) \langle \boldsymbol{\epsilon}_x^b \rangle + (\mathbf{I} - \mathbf{M}_0 \mathbf{D}_1) \langle \boldsymbol{\eta} \rangle], \quad (7.72)$$

5 where \mathbf{D}_1 is as defined in equation (7.58). If $\mathbf{D}_1 \mathbf{M}_0$ tends to the identity, then $\mathbf{M}_0 \mathbf{D}_1$ will
 6 also tend to the identity, so $\langle \boldsymbol{\epsilon}_{\beta}^a \rangle$ will tend to zero. The terms $\mathbf{D}_1 \mathbf{M}_0$ or $\mathbf{M}_0 \mathbf{D}_1$ tend to the
 7 identity when the anchor observation error variance tends to zero. Therefore, if anchor
 8 observations and bias-corrected observations observe states at the same time step, then,
 9 although a bias is added when the system is evolved to the time step of the observations,
 10 the contamination of both state background bias and model bias will be reduced with
 11 more precise anchor observations. Reducing the contamination of model bias in VarBC
 12 when anchor and bias-corrected observations are at the same time step gives an equivalent
 13 result to reducing the contamination of model bias in VarBC in a 3DVar system, as was
 14 shown in chapter 6.

15 **7.3.3.1 Anchor and bias-corrected observations observe different spatial states**

16 We will again split the spatial domain into two sections: \mathbf{x}_{ϕ} and \mathbf{x}_{ψ} , where \mathbf{x}_{ϕ} and \mathbf{x}_{ψ}
 17 do not overlap, to test how well the anchor observations can reduce biases in states they
 18 do not observe. We will assume bias-corrected observations can observe states in both
 19 \mathbf{x}_{ϕ} and \mathbf{x}_{ψ} , but anchor observations can only observe states in \mathbf{x}_{ψ} . Then the linearised

1 observation operators become,

$$\mathbf{H}_{(1)1} = \begin{pmatrix} \mathbf{H}_{(1)1\phi} & \mathbf{H}_{(1)1\psi} \end{pmatrix}, \quad (7.73)$$

$$\mathbf{H}_{(2)1} = \begin{pmatrix} \mathbf{0} & \mathbf{H}_{(2)1\psi} \end{pmatrix}, \quad (7.74)$$

2 where $\mathbf{H}_{(1)1\phi}$ is related to bias-corrected observations of \mathbf{x}_ϕ at time $t = 1$ and $\mathbf{H}_{(1)1\psi}$ and
 3 $\mathbf{H}_{(2)1\psi}$ are related to observations of \mathbf{x}_ψ at time $t = 1$.

4 In equation (7.72), $\mathbf{I} - \mathbf{D}_1\mathbf{M}_0$ is multiplied by $\langle \epsilon_x^b \rangle$ and $\mathbf{I} - \mathbf{M}_0\mathbf{D}_1$ is multiplied by
 5 $\langle \boldsymbol{\eta} \rangle$. \mathbf{D}_1 can be calculated explicitly as in equation (7.64), as the anchor observations are
 6 again at the first time step. Therefore, the term $(\mathbf{I} - \mathbf{D}_1\mathbf{M}_0)\langle \epsilon_x^b \rangle$ has also already been
 7 calculated in equation (7.67), where we found that the anchor observations could control
 8 background biases in states they do not observe, as long as the states are close enough to
 9 the anchor observations so that information can be transferred by the model.

10 We have also already calculated $(\mathbf{I} - \mathbf{M}_0\mathbf{D}_1)\langle \boldsymbol{\eta} \rangle$ in equation (7.69). We found that
 11 the model bias that came from states the anchor observations did not observe ($\langle \boldsymbol{\eta}_\phi \rangle$)
 12 was not explicitly controlled by the anchor observations, so the only way to reduce the
 13 contamination of model bias in those areas would be if the background error correlations
 14 between states \mathbf{x}_ϕ and \mathbf{x}_ψ are strong enough to pass information about the model biases
 15 between them, as was the case for 3DVar in chapter 6.

16 7.3.4 Summary of theoretical results

17 By studying the three simple theoretical cases for anchor observations at different times
 18 to the bias-corrected observations, we found:

- 19 • When the anchor observations were later than the bias-corrected observations, more
 20 precise anchor observations were able to reduce the contamination of background
 21 bias and model bias associated with reaching the time step of the bias-corrected
 22 observations, but an additional model bias was introduced that came from evolving

1 the state to the time step of the anchor observations. This additional model bias
2 was weighted by the inverse of the model when the anchor observations were more
3 precise, so growing errors in the model would be reduced, but would never completely
4 disappear.

- 5 • When the anchor observations were earlier than the bias-corrected observations,
6 more precise anchor observations were able to reduce the contamination of back-
7 ground bias and model bias associated with reaching the time step of the anchor
8 observations, but an additional model bias was introduced in evolving the state to
9 the time step of the bias-corrected observations, which could not be reduced with
10 more precise anchor observations.
- 11 • When the anchor and bias-corrected observations were at the same time, more precise
12 anchor observations could reduce both the background bias and the model bias.

13 This therefore suggests that to reduce the most contamination of bias on the observation
14 bias coefficient analysis, it is best to have the bias-corrected and anchor observations at
15 the same time step.

16 **7.4 Numerical Experiments**

17 In order to demonstrate our theoretical results of how anchor observations could reduce the
18 contamination of model bias in observation bias correction in 4DVarBC, we demonstrate
19 the theory by using a simple numerical system that has a multi-variate spatial domain
20 and can be run over many time steps.

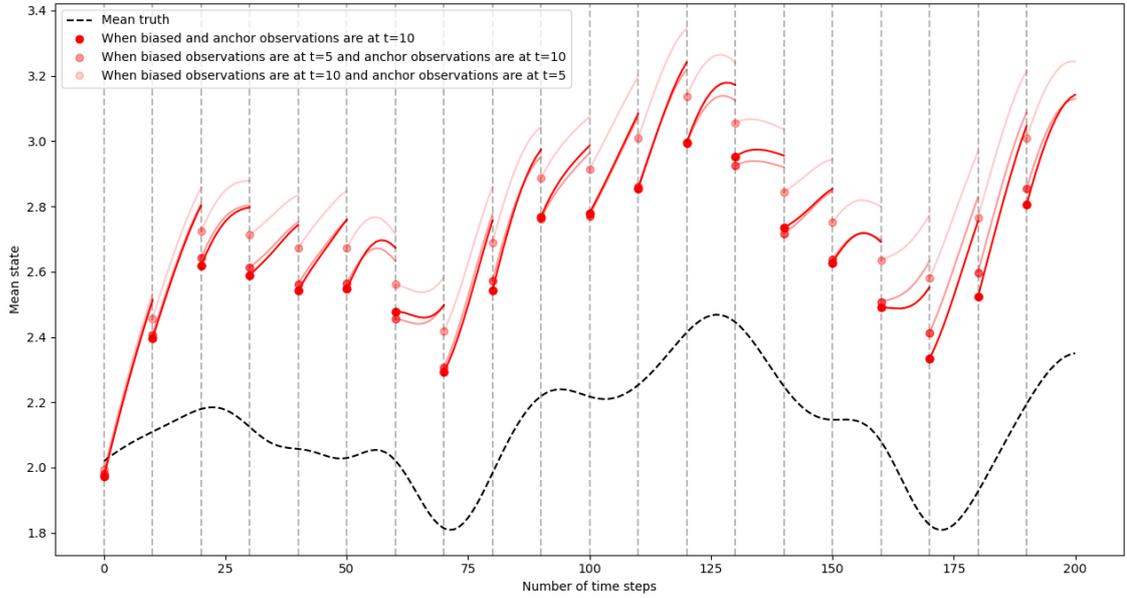
21 We use the Lorenz 96 model (Lorenz, 1996), as described in chapter 5, to create the
22 data assimilation system. The observations are generated from the true model with a
23 random error of error variance 1 (unless stated otherwise) and the biased observations
24 have a bias of 0.5. Note that as all the observations have the same bias, only one bias

1 coefficient is used, so β is scalar, as described in section 5.2.3. There are bias-corrected
2 and anchor observations directly observing the state at every spatial location, but three
3 different temporal sampling of the observations are considered: when the bias-corrected
4 observations are at $t = 5$ and the anchor observations are at $t = 10$; when the anchor
5 observations are at $t = 5$ and the bias-corrected observations are at $t = 10$; and when the
6 anchor and bias-corrected observations are both at the end of the window ($t = 10$). This
7 is analogous to the theoretical cases described above in sections 7.3.1, 7.3.2 and 7.3.3. A
8 bias is added to the model which is used to evolve the system forward in time, by changing
9 the forcing parameter F to 12 ($= F_{\text{biased}}$). The observation and model biases are chosen
10 so that they cause a similar bias in the state analysis when only one bias is present. The
11 initial cycle has no added background bias, but background bias will naturally accumulate
12 in the later cycles if the analysis of the previous cycle is biased. The assimilation length
13 is 10 time steps, which gives the biased model enough time to sufficiently evolve away
14 from the true model; we present the results over 20 windows. We have repeated the
15 experiments for 1000 realisations, which are all initialised with different random errors in
16 the background and observations and then averaged over all realisations.

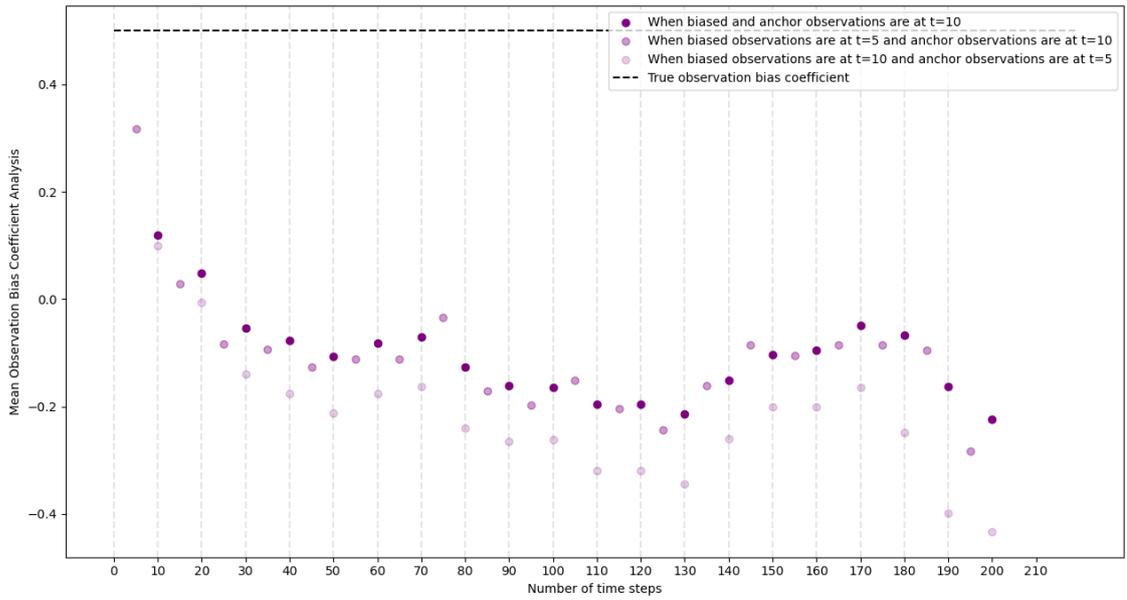
17 In figures 7.1a and 7.2a we have plotted the mean state analysis across all realisations
18 and all states: the circles are the mean state analyses at the beginning of each window,
19 and the tails are the analysis trajectories across the windows. The dashed black line is
20 the mean true trajectory across all states which uses the true model ($F = 8$). In figures
21 7.1b and 7.2b we have plotted the mean observation bias coefficient analysis across all
22 realisations. The observation bias coefficient analysis has been plotted at the time of the
23 biased observations, but is considered to be constant across the window.

24 In figure 7.1 we have plotted the mean state analysis and the mean observation bias co-
25 efficient analysis when the anchor observation error variance is equal to the bias-corrected
26 observation error variance ($= 1$). We have shown three timings of the bias-corrected and
27 anchor observations relative to each other, which are shown by the different transparencies

of the circles. It is obvious in figure 7.1a that the state analysis has been contaminated



(a) Mean state analysis



(b) Mean observation bias coefficient analysis

Figure 7.1: VarBC to correct for observation bias in a 4DVar system, with anchor and bias-corrected observations at different times in the window. $\sigma_{o(2)}^2 = 1$.

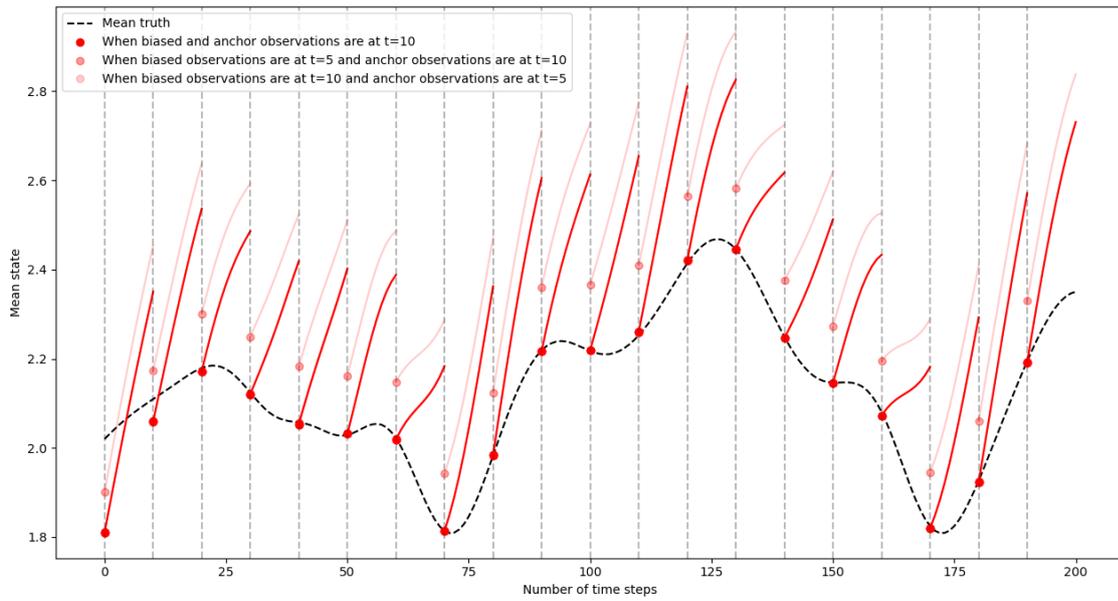
1 by model bias as, not only is there a bias of between approximately 0.4 and 0.7 over all
2 cycles, but the analysis trajectories also do not follow the true trajectories. This bias
3 in the state analysis has occurred despite having anchor observations at every spatial lo-
4 cation. The state analysis bias trajectories increase over the window as the model bias
5 is pulling the state analysis away from the truth. This gives an additional bias in the
6 background in the following cycle, so that VarBC is now being contaminated with both
7 background bias and model bias. The analysis bias is initially close to zero as there is no
8 background bias in the initial cycle. The observation bias coefficient in figure 7.1b has
9 been underestimated by between approximately 0.5 and 0.8 due to the contamination of
10 model bias. The observation bias coefficient is also predicted to be negative, whereas the
11 true observation bias is positive ($=0.5$), which means that VarBC is correcting observation
12 bias in the opposite direction. However, it should be noted that this change of sign is due
13 to the small magnitude of the true observation bias, so when the system underestimates
14 the observation bias coefficient, the estimate is pulled below zero. If the true observation
15 bias were bigger, then underestimating it would not necessarily lead to a change in sign
16 in the estimate of the observation bias. The observation bias was chosen to be 0.5 here as
17 we wanted the model and observation biases to have the same effect on the analysis.

18 In the theory, we showed that when the anchor observations were earlier in the window
19 than the bias-corrected observations, there was an additional model bias term that could
20 not be controlled by the anchor observations, so the analysis would have a larger bias
21 than if the bias-corrected observations were earlier in the window. We can see that the
22 observation bias coefficient has a larger bias when the anchor observations are at $t = 5$
23 (the lightest dots) in figure 7.1b as the observation bias coefficient is further away from
24 the true observation bias than the other cases. The state analysis also has a larger bias
25 when the anchor observations are earlier in the window as the lighter mean state analysis
26 in figure 7.1a is further away from the truth. There is not much difference between when
27 the bias-corrected observations are at $t = 5$ (earlier than the anchor observations) and

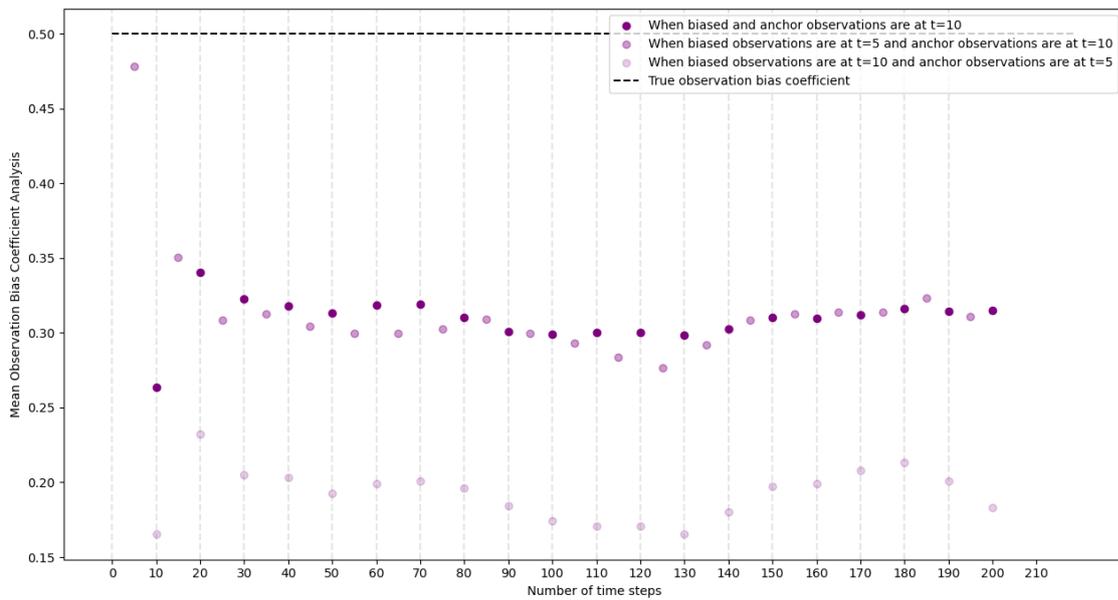
1 when the bias-corrected observations are at $t = 10$ (the same time as the anchor observa-
2 tions), as both the mean state analysis and mean observation bias coefficient analysis give
3 similar results for both cases. However, in general there is a small increase in bias when
4 anchor observations are later in the window than the bias-corrected observations, which
5 reflects the additional bias added when reaching the anchor observation time step that
6 we saw in equation (7.38). This difference is however negligible in figure 7.1a. Overall,
7 these numerical results agree with our theoretical results, as they show that when anchor
8 observations are earlier in the window than the bias-corrected observations, model bias
9 will more strongly contaminate the observation bias correction, leading to a larger state
10 analysis bias. Ideally, we would want the anchor observations to be at the same time (or
11 slightly later if necessary) as the bias-corrected observations, in order to reduce the most
12 contamination of background and model bias in the analysis.

13 Figure 7.2 is similar to figure 7.1, but we have now plotted the mean state analysis
14 and mean observation bias coefficient analysis when the anchor observation error variance
15 is $\sigma_{o(2)}^2 = 0.1$ and the bias-corrected observation error variance is $\sigma_{o(1)}^2 = 1$, to test the
16 impact of more precise anchor observations. Again, the three timings of the anchor and
17 bias-corrected observations are shown by the transparency of the circles.

18 In figure 7.2a the mean state analysis bias has substantially reduced for all timings of
19 the observations, which shows that more precise anchor observations has reduced the bias
20 in the mean state analysis, regardless of the timing of the anchor observations. This is
21 especially true when the anchor observations are later in the window, as the state analysis
22 now lies almost exactly on the truth. In the first two cycles there is now a negative bias,
23 but this disappears by the third cycle. This suggests that VarBC has over-compensated
24 for the model bias in the first two cycles and takes a few cycles to find the balance between
25 the anchor observations and the model. Interestingly, the bias in the forecast at the end
26 of the window is a similar magnitude to the bias at the beginning of the window when the
27 anchor observations were less precise (figure 7.1). This suggests that, as the model bias



(a) Mean state analysis



(b) Mean observation bias coefficient analysis

Figure 7.2: As in figure 7.1 but with more precise anchor observations ($\sigma_{o(2)}^2 = 0.1$).

- 1 has not been corrected for, the model is trying to reach its (biased) climatology, but is
- 2 always pulled back by the anchor observations in the analysis at the next cycle. This is

1 important as, although the analysis at the beginning of the window is much closer to the
2 truth when we use more precise anchor observations, the forecast throughout the window
3 is still biased.

4 In figure 7.2b the mean observation bias coefficient analysis is still biased, but has a
5 much smaller bias than when the anchor observations were less precise (figure 7.1b). The
6 bias in the observation bias coefficient analysis is now approximately 0.2, compared to
7 approximately 0.6 when $\sigma_{o(2)}^2 = 1$. This shows that more precise anchor observations has
8 improved the estimate of the observation bias coefficient, as it has reduced the contami-
9 nation of model bias in the observation bias coefficient analysis.

10 Overall figure 7.2 agrees with our theoretical results as more precise anchor observations
11 has reduced the contamination of state background and model biases on the state and
12 observation bias coefficient analyses. More precise anchor observations have removed the
13 most bias in the state and bias-coefficient analyses when anchor observations are later in
14 the window. There is a larger bias when anchor observations are earlier than the bias-
15 corrected observations, as was predicted from the additional model bias from the bias-
16 corrected observations in equation (7.32), which the anchor observations had no control
17 over.

18 **7.5 Conclusions and Discussion**

19 We have demonstrated the role of anchor observations in reducing the contamination of
20 model bias in a 4DVarBC system. We initially examined the theory about how the timing
21 of the anchor observations can reduce the contamination of model bias on the estimate
22 of observation bias in VarBC and then demonstrated these results in a simple numerical
23 system.

24 We calculated the bias coefficient analysis error equation for a 4DVarBC system when
25 both anchor and bias-corrected observations were used. Equation (7.32) showed that

1 some amount of model bias will always contaminate the estimate of the observation bias
2 in VarBC, as having anchor observations with zero random error would not be possible
3 operationally.

4 In order to better understand the role of anchor observations in time, we have looked
5 at three simplified cases: when the bias-corrected observations observe states before the
6 anchor observations; when the bias-corrected observations observe states after the anchor
7 observations; and when the bias-corrected and anchor observations observe states at the
8 same time step.

9 In section 7.3.1 we showed that when anchor observations observe states after the bias-
10 corrected observations, more precise anchor observations will reduce the contamination of
11 the background bias and the model bias associated with evolving the state to the time step
12 of the bias-corrected observations. We found that there was an additional model bias that
13 was associated with evolving the state to the time step of the anchor observations, that
14 could not completely be removed even with very precise anchor observations. However,
15 when the anchor observations were more precise, we showed that this term would be
16 reduced as it became a product of the model bias and the inverse of the linearised model.
17 As atmospheric dynamics are described by models with growing errors (Lorenz, 2005;
18 Simmons et al., 1995), this shows that the additional model bias would be reduced. We
19 illustrated this result in the medium transparency line in figures 7.1 and 7.2, as increasing
20 the precision of the anchor observations reduced the bias in both the state analysis and
21 the observation bias coefficient analysis.

22 In section 7.3.2 we showed that when anchor observations were earlier than the bias-
23 corrected observations, more precise anchor observations could reduce the background bias
24 and model bias associated with evolving the state to the time step of the anchor observa-
25 tions, but that there was an additional model bias associated with evolving the state to the
26 time step of the bias-corrected observations that the anchor observations could not con-
27 trol. This meant that, unlike for the additional model bias when the anchor observations

1 were later than the bias-corrected observations, more precise anchor observations would
2 not reduce the additional model bias term when bias-corrected observations were later in
3 the window than the anchor observations. We demonstrated these results in the lightest
4 transparency line in figures 7.1 and 7.2, as using more precise anchor observations reduced
5 the bias in the state and observation bias coefficient analyses. However, the lightest line
6 had the most bias compared to the other timings of the anchor observations. This reflected
7 our theoretical results when the anchor observations were earlier than the bias-corrected
8 observations, as the bias in the lightest line showed the additional model bias that could
9 not be controlled by the anchor observations.

10 Finally, in section 7.3.3 we showed that if the anchor and bias-corrected observations
11 observe states at the same time step, then the contamination of state background bias from
12 all observable states can be reduced by more precise anchor observations. More precise
13 anchor observations will also reduce the contamination of model bias that comes from
14 evolving the state to the observed time step, but only in spatial states that are observed
15 by the anchor observations. We demonstrated these results in figure 7.2 as we found that
16 reducing the error variance of the anchor observations reduced the bias in both the state
17 and observation bias coefficient analyses when the anchor and bias-corrected observations
18 were at the same time.

19 We have shown that if bias-corrected observations are later in the window than the
20 anchor observations, an extra model bias will contaminate the estimate of the observation
21 bias correction, regardless of the precision of the anchor observations. It is therefore safer
22 for anchor observations to be later in the assimilation window, so that there are fewer
23 bias-corrected observations after the anchor observations. In practice, the network of
24 satellite radiance observations from polar orbiting satellites provide data at a particular
25 point at least once every 6 hours, but this could be early or late in the window, depending
26 on where the satellite is giving data for (*Met Office NRT Quality Monitoring, 2023*). In
27 contrast, anchor observations such as radiosondes only provide data every 6 or 12 hours

1 (*ECMWF Geographical Coverage*, 2023), so usually give information in the middle of
2 a 6 hour window. Therefore, future developments of anchor observations should look
3 to develop observations that are available at more frequent times, for example in the
4 continued expansion of radio occultation instruments. If changing the timing of the anchor
5 observations is not possible, data assimilation centres could look into changing the timing
6 of the assimilation window (Milan et al., 2020), so that the available anchor observations
7 fall at the end of the window.

8 At ECMWF, a quasi-continuous data assimilation system is used, whereby new ob-
9 servations are used in regular updates to the analysis within an inner loop, so that more
10 observations can be used within the window, as the final analysis update uses less com-
11 puting power and therefore the cut off time for observations is shorter (Lean et al., 2021).
12 This would likely not change the amount of radiosonde data available for each window,
13 as they are so infrequent, but could change the amount of satellite data available. This
14 could be beneficial if later radio occultation data is available as we have shown that later
15 anchor observations are better at reducing the contamination of model bias, but could
16 have a detrimental impact if more bias-corrected observations are used at the end of the
17 window and if there is also a model bias present in the system, as we have shown that
18 bias-corrected observations later in the window can be associated with model biases that
19 anchor observations cannot reduce.

20 This chapter has aimed to derive insight into the role of anchor observations in a 4DVar
21 data assimilation system, when both observation and model biases are present and when
22 observation bias is corrected for, in order to answer research question 2. This extended
23 the work of chapter 6 to test the importance of the timing of the anchor observations, as
24 well as their spatial distribution and precision, as was discussed for 3DVarBC. The next
25 steps for this work would be to test the results on a system that more closely relates to an
26 operational set up where the anchor observations only observe small parts of the domain
27 and where the observation or model biases are spatially variant so that the bias spatial

1 structures between the observation and model biases are different. VarBC could also be
2 tested on systems that only have model bias present and do not have observation bias
3 present, to test whether VarBC would try to correct for the wrong type of bias.

4 **7.6 Summary**

5 In chapter 7 we have derived analysis error equations for the state and observation bias co-
6 efficient when both anchor and bias-corrected observations are used and when background
7 and model biases are present. Using these equations we have discussed the importance
8 of the timing of the anchor observations in 4DVarBC as well as how the precision of the
9 anchor observations impacts their ability to reduce both state background and state model
10 biases. In the next chapter we will consider the role of anchor observations when correcting
11 for model bias in the presence of observation bias. We will then compare the importance
12 of anchor observations when using VarBC, WC4DVar or both simultaneously.

1 Chapter 8

2 The interaction between VarBC 3 and WC4DVar and the role of 4 anchor observations

5 8.1 Introduction

6 In chapters 6 and 7 we studied the characteristics of anchor observations needed to ef-
7 fectively reduce the contamination of model bias on the observation bias correction. In
8 chapter 3.2 we discussed how model bias can be also corrected for, using a technique known
9 as weak-constraint 4DVar (WC4DVar). Laloyaux et al. (2020a) demonstrated that biased
10 observations would contaminate the estimate of the model bias in WC4DVar, unless an-
11 chor observations were also present. This work did not, however, study the characteristics
12 needed in anchor observations to effectively reduce the contamination of observation bias
13 in the presence of model bias. As we showed in section 2.2, anchor observations can be
14 sparse in both space and time, especially in comparison to satellite radiance observations.
15 Therefore, in this chapter we will study the interaction between correcting for observation

1 and model biases simultaneously, and the role anchor observations play in mitigating the
2 contamination of bias from one source (model/observation) on the correction of the other
3 to answer research question 3.

4 In section 8.2 we will initially study the theory of how observation bias contaminates the
5 estimate of the model bias, and how anchor observations can mitigate this contamination,
6 when only model bias is corrected for. Although in practice WC4DVar would not be used
7 without also correcting for observation bias, we isolate the contamination of observation
8 bias on the correction of model bias to mirror chapters 6 and 7, and to obtain theoretical
9 results for a simpler problem. Together with the previous chapters, this can be used to
10 provide insight into how WC4DVar and VarBC interact, which is more difficult to study
11 theoretically. We will study three simple cases in sections 8.2.1.1, 8.2.1.2 and 8.2.1.3,
12 where the biased and anchor observations are only at one time step each in the window.
13 We will then demonstrate our theoretical findings in a simple numerical system in section
14 8.2.2. Finally in section 8.3, we will compare the numerical experiments for the VarBC
15 results from chapter 7 with the WC4DVar case, as well as with numerical experiments that
16 correct for both observation and model biases respectively, to compare the role of anchor
17 observations in each bias correction technique. Overall conclusions will be discussed in
18 section 8.4.

19 **8.2 WC4DVar with uncorrected observation bias**

20 We want to study the role of anchor observations in WC4DVar when both observation and
21 model biases are present, but only model biases are being accounted for. To accomplish
22 this, we split up the observations in the analysis error equations for WC4DVar, equation
23 (2.95), such that we have biased observations $\mathbf{y}_{(1)}$ and anchor observations $\mathbf{y}_{(2)}$, as we did
24 for the VarBC cases in chapters 6 and 7. We therefore have two observation operators for
25 both observation types: $h_{(1)\mathbf{p}}(\mathbf{p})$ and $h_{(2)\mathbf{p}}(\mathbf{p})$, where \mathbf{p} is the vector containing the state,

1 \mathbf{x} , and model bias parameter, $\boldsymbol{\eta}$, defined in equation (2.83), and the numbers in brackets
2 represent the biased or anchor observations. The biased observations will not be corrected
3 for. The linearised observation operator, equation (2.90), is split into two observation
4 types (denoted by $\hat{\mathbf{H}}_{(1)\text{p}}$ and $\hat{\mathbf{H}}_{(2)\text{p}}$) such that $\hat{\mathbf{H}}_{\text{p}}$ from equation (2.90) is given by,

$$\hat{\mathbf{H}}_{\text{p}} = \begin{pmatrix} \hat{\mathbf{H}}_{(1)\text{x}} & \hat{\mathbf{H}}_{(1)\eta} \\ \hat{\mathbf{H}}_{(2)\text{x}} & \hat{\mathbf{H}}_{(2)\eta} \end{pmatrix} \quad (8.1)$$

5 where

$$\hat{\mathbf{H}}_{(k)\text{x}} = \begin{pmatrix} \mathbf{H}_{(k)\text{x},0} \\ \vdots \\ \mathbf{H}_{(k)\text{x},i} \mathbf{M}_{i-1} \dots \mathbf{M}_0 \\ \vdots \\ \mathbf{H}_{(k)\text{x},N} \mathbf{M}_{N-1} \dots \mathbf{M}_0 \end{pmatrix}, \quad \hat{\mathbf{H}}_{(k)\eta} = \begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{H}_{(k)\text{x},i} (\sum_{j=1}^{i-1} \mathbf{M}_{i-1} \dots \mathbf{M}_j + \mathbf{I}) \\ \vdots \\ \mathbf{H}_{(k)\text{x},N} (\sum_{j=1}^{N-1} \mathbf{M}_{N-1} \dots \mathbf{M}_j + \mathbf{I}) \end{pmatrix}. \quad (8.2)$$

6 Note that equation (8.2) is equivalent to equations (2.91) and (2.92), but the (k) denotes
7 the biased or anchor observation type. The observation error covariance matrix for the
8 two observation types becomes

$$\hat{\mathbf{R}} = \begin{pmatrix} \hat{\mathbf{R}}_{(1)} & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{R}}_{(2)} \end{pmatrix} \quad (8.3)$$

9 where we have assumed that the error covariances between the biased and anchor obser-
10 vations are zero. Then the state and model bias analyses can be extended from equation
11 (2.95) to give,

$$\mathbf{x}_0^{\text{a}} = \mathbf{x}_0^{\text{b}} + \hat{\mathbf{K}}_{\text{xy}(1)} \hat{\mathbf{d}}_{(1)\text{p}}^{\text{b}} + \hat{\mathbf{K}}_{\text{xy}(2)} \hat{\mathbf{d}}_{(2)\text{p}}^{\text{b}}, \quad (8.4)$$

$$\boldsymbol{\eta}^{\text{a}} = \boldsymbol{\eta}^{\text{b}} + \hat{\mathbf{K}}_{\eta\text{Y}(1)} \hat{\mathbf{d}}_{(1)\text{p}}^{\text{b}} + \hat{\mathbf{K}}_{\eta\text{Y}(2)} \hat{\mathbf{d}}_{(2)\text{p}}^{\text{b}}, \quad (8.5)$$

1 where $\hat{\mathbf{K}}_{xy(1)}$, $\hat{\mathbf{K}}_{xy(2)}$, $\hat{\mathbf{K}}_{\eta y(1)}$ and $\hat{\mathbf{K}}_{\eta y(2)}$ are calculated by expanding,

$$\begin{pmatrix} \hat{\mathbf{K}}_{xy(1)} & \hat{\mathbf{K}}_{xy(2)} \\ \hat{\mathbf{K}}_{\eta y(1)} & \hat{\mathbf{K}}_{\eta y(2)} \end{pmatrix} = \begin{pmatrix} \mathbf{B}_x & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{H}}_{(1)x}^T & \hat{\mathbf{H}}_{(2)x}^T \\ \hat{\mathbf{H}}_{(1)\eta}^T & \hat{\mathbf{H}}_{(2)\eta}^T \end{pmatrix} \times \\ \left[\begin{pmatrix} \hat{\mathbf{H}}_{(1)x} & \hat{\mathbf{H}}_{(1)\eta} \\ \hat{\mathbf{H}}_{(2)x} & \hat{\mathbf{H}}_{(2)\eta} \end{pmatrix} \begin{pmatrix} \mathbf{B}_x & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{H}}_{(1)x}^T & \hat{\mathbf{H}}_{(2)x}^T \\ \hat{\mathbf{H}}_{(1)\eta}^T & \hat{\mathbf{H}}_{(2)\eta}^T \end{pmatrix} + \begin{pmatrix} \hat{\mathbf{R}}_{(1)} & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{R}}_{(2)} \end{pmatrix} \right]^{-1}, \quad (8.6)$$

2 where \mathbf{B}_x is the state background error covariance matrix and \mathbf{Q} is the model error co-
3 variance matrix, as defined in equation (2.84). As the inverse matrix in equation (8.6)
4 is symmetric, we can calculate the inverse as in equation (4.2) in Lu and Shiou (2002)
5 to explicitly calculate $\hat{\mathbf{K}}_{xy(1)}$, $\hat{\mathbf{K}}_{xy(2)}$, $\hat{\mathbf{K}}_{\eta y(1)}$ and $\hat{\mathbf{K}}_{\eta y(2)}$. Therefore, following the same
6 method to calculate $\mathbf{K}_{xy(1)}$, $\mathbf{K}_{xy(2)}$, $\mathbf{K}_{\eta y(1)}$ and $\mathbf{K}_{\eta y(2)}$ in equations (6.16) - (6.19), $\hat{\mathbf{K}}$ can
7 be separated as follows

$$\hat{\mathbf{K}}_{xy(1)} = \mathbf{B}_x \hat{\mathbf{H}}_{(1)x}^T (\mathbf{W} - \mathbf{XZ}^{-1} \mathbf{X}^T)^{-1} - \mathbf{B}_x \hat{\mathbf{H}}_{(2)x}^T \mathbf{Z}^{-1} \mathbf{X}^T (\mathbf{W} - \mathbf{XZ}^{-1} \mathbf{X}^T)^{-1}, \quad (8.7)$$

$$\hat{\mathbf{K}}_{xy(2)} = -\mathbf{B}_x \hat{\mathbf{H}}_{(1)x}^T (\mathbf{W} - \mathbf{XZ}^{-1} \mathbf{X}^T) \mathbf{XZ}^{-1} + \mathbf{B}_x \hat{\mathbf{H}}_{(2)x}^T (\mathbf{Z}^{-1} + \mathbf{Z}^{-1} \mathbf{X}^T (\mathbf{W} - \mathbf{XZ}^{-1} \mathbf{X}^T) \mathbf{XZ}^{-1}), \quad (8.8)$$

$$\hat{\mathbf{K}}_{\eta y(1)} = \mathbf{Q} \hat{\mathbf{H}}_{(1)\eta}^T (\mathbf{W} - \mathbf{XZ}^{-1} \mathbf{X}^T)^{-1} - \mathbf{Q} \hat{\mathbf{H}}_{(2)\eta}^T \mathbf{Z}^{-1} \mathbf{X}^T (\mathbf{W} - \mathbf{XZ}^{-1} \mathbf{X}^T)^{-1}, \quad (8.9)$$

$$\hat{\mathbf{K}}_{\eta y(2)} = -\mathbf{Q} \hat{\mathbf{H}}_{(1)\eta}^T (\mathbf{W} - \mathbf{XZ}^{-1} \mathbf{X}^T)^{-1} \mathbf{XZ}^{-1} + \mathbf{Q} \hat{\mathbf{H}}_{(2)\eta}^T (\mathbf{Z}^{-1} + \mathbf{Z}^{-1} \mathbf{X}^T (\mathbf{W} - \mathbf{XZ}^{-1} \mathbf{X}^T)^{-1} \mathbf{XZ}^{-1}), \quad (8.10)$$

8 where we have denoted \mathbf{W} , \mathbf{X} and \mathbf{Z} to be the element blocks from the matrix to be
9 inverted in equation (8.6),

$$\mathbf{W} = \hat{\mathbf{H}}_{(1)x} \mathbf{B}_x \hat{\mathbf{H}}_{(1)x}^T + \hat{\mathbf{H}}_{(1)\eta} \mathbf{Q} \hat{\mathbf{H}}_{(1)\eta}^T + \hat{\mathbf{R}}_{(1)}, \quad (8.11)$$

$$\mathbf{X} = \hat{\mathbf{H}}_{(1)x} \mathbf{B}_x \hat{\mathbf{H}}_{(2)x}^T + \hat{\mathbf{H}}_{(1)\eta} \mathbf{Q} \hat{\mathbf{H}}_{(2)\eta}^T, \quad (8.12)$$

$$\mathbf{Z} = \hat{\mathbf{H}}_{(2)x} \mathbf{B}_x \hat{\mathbf{H}}_{(2)x}^T + \hat{\mathbf{H}}_{(2)\eta} \mathbf{Q} \hat{\mathbf{H}}_{(2)\eta}^T + \hat{\mathbf{R}}_{(2)}. \quad (8.13)$$

1 \mathbf{W} , \mathbf{X} and \mathbf{Z} come from the $\mathbf{HB}_p\mathbf{H}^T + \mathbf{R}$ term, where, \mathbf{W} and \mathbf{Z} correspond to the biased
2 and anchor observations respectively, and \mathbf{X} corresponds to the cross term. If \mathbf{X} is $\mathbf{0}$, then
3 $\hat{\mathbf{K}}_{xy(1)}$ and $\hat{\mathbf{K}}_{\eta y(1)}$ would only depend on the biased observations and $\hat{\mathbf{K}}_{xy(2)}$ and $\hat{\mathbf{K}}_{\eta y(2)}$
4 would only depend on the anchor observations.

5 The error equations for \mathbf{x}^a and $\boldsymbol{\eta}^a$ can be extended from equations (2.104) and (2.105)
6 to include two observation types and are given by,

$$\boldsymbol{\epsilon}_x^a = \boldsymbol{\epsilon}_x^b + \hat{\mathbf{K}}_{xy(1)}(\hat{\boldsymbol{\epsilon}}_{(1)}^o - \hat{\mathbf{H}}_{(1)x}\boldsymbol{\epsilon}_x^b - \hat{\mathbf{H}}_{(1)\eta}\boldsymbol{\epsilon}_\eta^b) + \hat{\mathbf{K}}_{xy(2)}(\hat{\boldsymbol{\epsilon}}_{(2)}^o - \hat{\mathbf{H}}_{(2)x}\boldsymbol{\epsilon}_x^b - \hat{\mathbf{H}}_{(2)\eta}\boldsymbol{\epsilon}_\eta^b) \quad (8.14)$$

$$\boldsymbol{\epsilon}_\eta^a = \boldsymbol{\epsilon}_\eta^b + \hat{\mathbf{K}}_{\eta y(1)}(\hat{\boldsymbol{\epsilon}}_{(1)}^o - \hat{\mathbf{H}}_{(1)x}\boldsymbol{\epsilon}_x^b - \hat{\mathbf{H}}_{(1)\eta}\boldsymbol{\epsilon}_\eta^b) + \hat{\mathbf{K}}_{\eta y(2)}(\hat{\boldsymbol{\epsilon}}_{(2)}^o - \hat{\mathbf{H}}_{(2)x}\boldsymbol{\epsilon}_x^b - \hat{\mathbf{H}}_{(2)\eta}\boldsymbol{\epsilon}_\eta^b) \quad (8.15)$$

7 where the biased and anchor observation error vectors in time are given by,

$$\hat{\boldsymbol{\epsilon}}_{(1)}^o = \begin{pmatrix} \mathbf{y}_{(1)0} - h_{(1)p}(\mathbf{x}_0^t, \boldsymbol{\eta}^t) \\ \vdots \\ \mathbf{y}_{(1)i} - h_{(1)p}(\tilde{m}_{0 \rightarrow i}(\mathbf{x}_0^t, \boldsymbol{\eta}^t)) \\ \vdots \\ \mathbf{y}_{(1)N(1)} - h_{(1)p}(\tilde{m}_{0 \rightarrow N(1)}(\mathbf{x}_0^t, \boldsymbol{\eta}^t)) \end{pmatrix}, \quad \hat{\boldsymbol{\epsilon}}_{(2)}^o = \begin{pmatrix} \mathbf{y}_{(2)0} - h_{(2)p}(\mathbf{x}_0^t, \boldsymbol{\eta}^t) \\ \vdots \\ \mathbf{y}_{(2)i} - h_{(2)p}(\tilde{m}_{0 \rightarrow i}(\mathbf{x}_0^t, \boldsymbol{\eta}^t)) \\ \vdots \\ \mathbf{y}_{(2)N(2)} - h_{(2)p}(\tilde{m}_{0 \rightarrow i}(\mathbf{x}_0^t, \boldsymbol{\eta}^t)) \end{pmatrix}, \quad (8.16)$$

8 where $\tilde{m}_{0 \rightarrow i}(\mathbf{x}_0^t, \boldsymbol{\eta}^t)$ is the model that evolves both the state and the model bias to the i^{th}
9 time step, as defined in equation (2.82).

10 To calculate the bias in the model bias parameter, we take the expected value of the
11 errors in equation (8.15). Assuming that the state background bias and the model bias
12 parameter background bias are zero ($\langle \boldsymbol{\epsilon}_x^b \rangle = 0$, $\langle \boldsymbol{\epsilon}_\eta^b \rangle = 0$), which could be true for the initial
13 cycle; and the anchor observations are unbiased ($\langle \hat{\boldsymbol{\epsilon}}_{(2)}^o \rangle = 0$), so that the only expected
14 value of error term, that is nonzero, is from the biased observations, then the expected
15 value of the state and model bias analysis error equations, equations (8.14) and (8.15)

1 respectively, become,

$$\langle \epsilon_x^a \rangle_{t=0} = \hat{\mathbf{K}}_{xy(1)} \langle \hat{\epsilon}_{(1)}^o \rangle \quad (8.17)$$

$$\langle \epsilon_\eta^a \rangle_{t=0} = \hat{\mathbf{K}}_{\eta y(1)} \langle \hat{\epsilon}_{(1)}^o \rangle. \quad (8.18)$$

2 Therefore, the amount of observation bias that contaminates either the estimate of the
 3 state or the model bias parameter is dependent on the sensitivity of the state or model bias
 4 parameter to the biased observations. Note that equations (8.17) and (8.18) would look
 5 the same if we only had biased observations, but when we have both observation types, the
 6 terms $\hat{\mathbf{K}}_{xy(1)}$ and $\hat{\mathbf{K}}_{\eta y(1)}$ are dependent on both the anchor and the biased observations.
 7 Therefore, the following section will focus on how the anchor observations vary $\hat{\mathbf{K}}_{xy(1)}$ and
 8 $\hat{\mathbf{K}}_{\eta y(1)}$ in order to demonstrate how the anchor observations can reduce the contamination
 9 of observation bias in WC4DVar.

10 In the following section, we investigate the role of anchor observations when the biased
 11 and anchor observations observe one time step each, which could be the same or different
 12 time steps to each other. In order to simplify the system further, we observe a scalar
 13 system, so that only one state is observed between the biased and anchor observations.

14 **8.2.1 Scalar system with anchor and biased observations at one time** 15 **step**

16 To simplify the equations for $\hat{\mathbf{K}}_{xy(1)}$ and $\hat{\mathbf{K}}_{\eta y(1)}$ we study a scalar system with observations
 17 only observing one time step. If the system is scalar such that there is only one state that is
 18 observed by both the anchor and bias-corrected observations, and both of the observations
 19 are only at one time step (this could be the same or different time steps with respect to
 20 each other) then the observation error covariance matrices from equation (8.3) will now
 21 be given by the variances ($\sigma_{o(k)}^2$) at the time steps they are observed and 0 everywhere

1 else. The background and model error variances for the scalar system will be denoted by,

$$\mathbf{B}_x = \sigma_{bx}^2, \quad \mathbf{Q} = \sigma_q^2. \quad (8.19)$$

2 We assume that the observation operators for both observations are equal and given
3 by the identity as if we have direct observations, such that they are given by

$$h_{(k)_i}(x_i) = x_i, \quad (8.20)$$

4 where the (k) denotes the biased or anchor observations respectively and i is the time step.
5 Operationally, this assumption is unlikely as biased observations would often be satellite
6 radiance observations, which require a more complicated observation operator. However,
7 we make this assumption to simplify equations (8.7) and (8.9), in order to understand the
8 importance of the anchor observation error variance.

9 We define the model to be a simple linear model given by,

$$x_{i+1} = m(x_i) + \eta = \alpha x_i + \eta, \quad (8.21)$$

10 and choose α to be a constant between 0 and 2. When α is less than 1, the errors in the
11 model decay and so the errors can be considered stable, and when α is greater than 1, the
12 errors in the model grow and so the errors can be considered unstable.

13 If the observations can either be at time $t = 1$ or $t = 2$, then the observation operator
14 for observations at time 1 and the observation operator for observations at time 2 are
15 given by,

$$h_{(k)_1}(x_1) = x_1 = m_0(x_0) + \eta = \alpha x_0 + \eta, \quad (8.22)$$

$$h_{(k)_2}(x_2) = x_2 = m_1(x_1) + \eta = m_1(m_0(x_0) + \eta) + \eta = \alpha^2 x_0 + (\alpha + 1)\eta. \quad (8.23)$$

1 In our experiments we will set $\sigma_{\text{bx}}^2 = 1$, $\sigma_{\text{q}}^2 = 0.1$ and $\sigma_{\text{o}(1)}^2 = 1$, so that the biased
2 observations and background have the same error variances and the model error variance
3 is ten times smaller than the background error variance. We want the model error variance
4 to be smaller than the background error variance as the model error variance only reflects
5 the error at one time step, whereas the background error is the accumulation of error
6 across the window.

7 In sections 8.2.1.1, 8.2.1.2, 8.2.1.3 we will plot $K_{\text{xy}(1)}$ and $K_{\eta\text{y}(1)}$ for five values of α :
8 ranging from 0.5 to 1.5, varying the anchor observation error variance, unless specified
9 otherwise. Changing α alters the stability of the errors in the model and changing the
10 anchor observation error variance alters the precision of the anchor observations. Note
11 that in a non-scalar system, changing the precision of the anchor observations could also
12 refer to changing the overall weighting given to the anchor observations, which accounts
13 not only for the error in a single observation, but also the combined effect of several anchor
14 observations which are spatially close to each other. We vary α to ensure we have a wider
15 variety of model conditions when making our conclusions. We vary the anchor observation
16 error variance to test the role of the anchor observations in reducing the contamination of
17 observation bias on the estimate of the state and the model bias. As we are only varying
18 the anchor observation error variance (not the biased observation error variance or the
19 background error variance), the analysis error variance for the system will change with
20 different $\sigma_{\text{o}(2)}^2$, and so the background error variance should also change in subsequent
21 cycles. However, in this simple system, we want to demonstrate the extent that the
22 anchor observations can reduce the contamination of observation bias in both the model
23 bias parameter and the state analysis. Therefore, we only reduce the anchor observation
24 error variance to demonstrate the role of the anchor observations. In reality, the anchor
25 observations often make up a small portion of the observation data and so varying the
26 anchor observation error variance would not have a large impact on the analysis error
27 variance. It is therefore reasonable not to change the background error variance when the

1 anchor observation error variance is altered.

2 In the following sections we present explicit cases of the anchor and biased observations
 3 observing the state at either the first or second time step.

4 **8.2.1.1 Biased and anchor observations at $t = 1$**

5 If both biased and anchor observations are at $t = 1$, then the linearised observation
 6 operators for the biased and anchor observations would be,

$$\hat{H}_{(1)x} = \begin{pmatrix} 0 \\ H_{(1)x,1} M_0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ \alpha \\ 0 \end{pmatrix}, \quad \hat{H}_{(1)\eta} = \begin{pmatrix} 0 \\ H_{(1)x,1} \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad (8.24)$$

$$\hat{H}_{(2)x} = \begin{pmatrix} 0 \\ H_{(2)x,1} M_0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ \alpha \\ 0 \end{pmatrix}, \quad \hat{H}_{(2)\eta} = \begin{pmatrix} 0 \\ H_{(2)x,1} \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}. \quad (8.25)$$

7 The linearised observation operators with respect to η are independent of α as both of the
 8 observations are only at the first time step, so they are not dependent on the linearised
 9 model.

10 Therefore by substituting equations (8.24) and (8.25) into equations (8.7) and (8.9),
 11 the sensitivities of the state and model bias parameter analyses for the scalar case when
 12 anchor and biased observations are at the same time, are given by,

$$[K_{xy(1)}]_{t=1} = \frac{\alpha \sigma_{bx}^2 \sigma_{o(2)}^2}{\alpha^2 \sigma_{bx}^2 \sigma_{o(1)}^2 + \alpha^2 \sigma_{bx}^2 \sigma_{o(2)}^2 + \sigma_q^2 \sigma_{o(1)}^2 + \sigma_q^2 \sigma_{o(2)}^2 + \sigma_{o(1)}^2 \sigma_{o(2)}^2}, \quad (8.26)$$

$$[K_{\eta y(1)}]_{t=1} = \frac{\sigma_q^2 \sigma_{o(2)}^2}{\alpha^2 \sigma_{bx}^2 \sigma_{o(1)}^2 + \alpha^2 \sigma_{bx}^2 \sigma_{o(2)}^2 + \sigma_q^2 \sigma_{o(1)}^2 + \sigma_q^2 \sigma_{o(2)}^2 + \sigma_{o(1)}^2 \sigma_{o(2)}^2}, \quad (8.27)$$

13 where $[K_{xy(1)}]_{t=1}$ and $[K_{\eta y(1)}]_{t=1}$ are the sensitivities at $t = 1$ (at the other times the
 14 sensitivities will be 0). The numerator in equation (8.26) shows that the sensitivity of the

1 state to the biased observations is proportional to the stability of the errors in the model
 2 (α), the state background error variance (σ_{bx}^2), and the anchor observation error variance
 3 ($\sigma_{\text{o}(2)}^2$). If either the state background error variance or the anchor observation error
 4 variance is large (ie. less weight is given to the state background or anchor observation),
 5 then the state estimate will be more dependent on the biased observations. The same
 6 is true for the numerator in equation (8.27), which is proportional to the model error
 7 variance (σ_{q}^2) and the anchor observation error variance ($\sigma_{\text{o}(2)}^2$): if either the model error
 8 variance or the anchor observation error variance is large, then the model bias parameter
 9 estimate will be more sensitive to the biased observations. These two equations show
 10 that giving more weight to the anchor observations or bias-corrected model would likely
 11 mean less weight is given to the biased observations and hence less observation bias would
 12 contaminate the estimate of the state and the model bias parameter. As the denominator
 13 of equations (8.26) and (8.27) is more complicated than the numerators, to understand how
 14 the denominators of equations (8.26) and (8.27) impact the sensitivities of the state and
 15 model bias parameter to the biased observations, we will run simple numerical experiments
 16 with given parameters, as described in the introduction of section 8.2.1.

17 In figure 8.1 we have plotted $K_{\text{xy}(1)}$ (red) and $K_{\eta\text{y}(1)}$ (blue) whilst varying the anchor
 18 observation error variance, which ranges from 10 times smaller to 10 times bigger the
 19 biased observation error variance ($\sigma_{\text{o}(1)}^2 = 1$). They have been plotted for 10 values of
 20 α , from 0.5 to 1.5, shown by the increasing transparency of the lines (where the darkest
 21 lines are the smallest values of α and the lightest lines are the largest values of α). $K_{\eta\text{y}(1)}$
 22 is always smaller than $K_{\text{xy}(1)}$ for the corresponding values of α , which suggests that the
 23 observation bias would have a larger impact on the state analysis, than on the model bias
 24 coefficient analysis. This makes sense as $K_{\text{xy}(1)}$ and $K_{\eta\text{y}(1)}$ are proportional to σ_{bx}^2 and σ_{q}^2
 25 respectively in equations (8.26) and (8.27) and we have set σ_{q}^2 to be much smaller than
 26 σ_{bx}^2 , as the model error variance is only valid at one time step, whereas the background
 27 error variance would be the error accumulation across a whole window. Increasing the

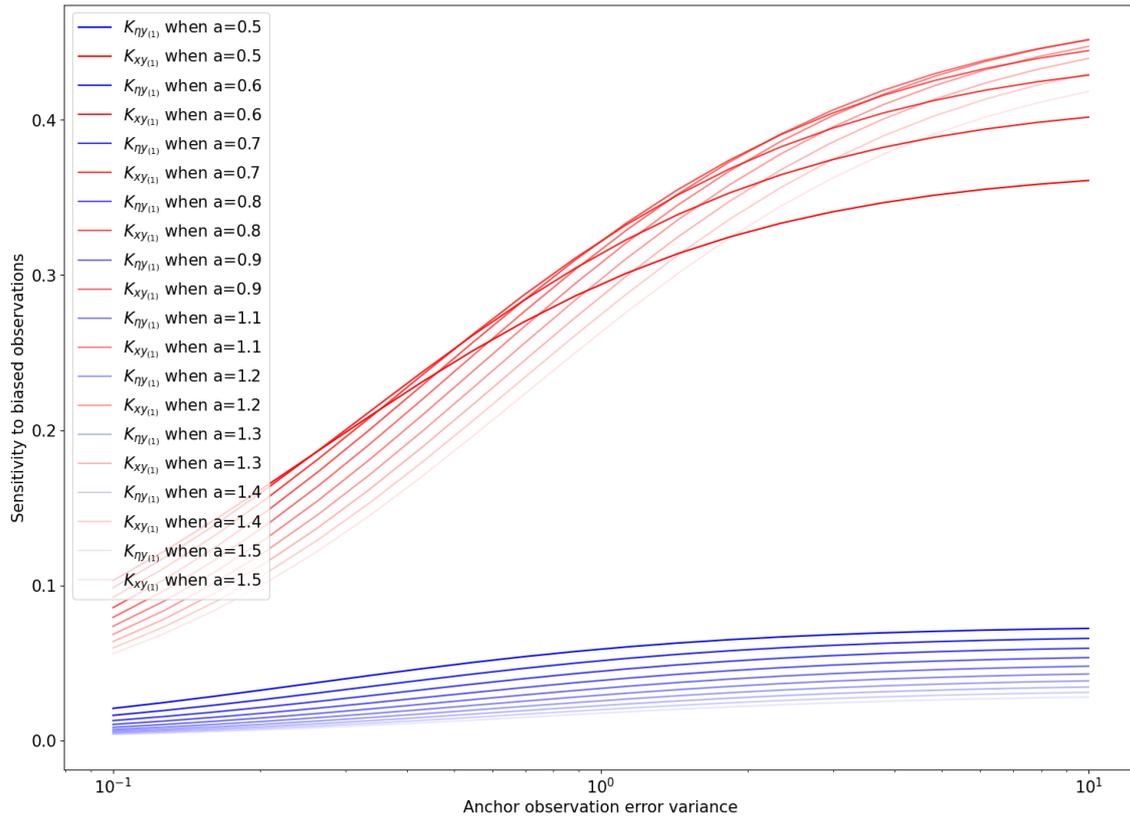


Figure 8.1: $K_{\eta_{Y(1)}}$ (blue) and $K_{X_{Y(1)}}$ (red) when biased observations at $t = 1$, anchor observations at $t = 1$ for different values of α .

1 anchor observation error variance has increased the sensitivity of both the state and the
 2 model bias parameter to the biased observations, as was predicted in equations (8.26)
 3 and (8.27), because $K_{\eta_{Y(1)}}$ and $K_{X_{Y(1)}}$ are both proportional to $\sigma_{o(2)}^2$. We would expect
 4 that the sensitivities of the state and model bias parameter to the biased observations
 5 increases with increased anchor observation error variance, as putting less trust into the
 6 anchor observations would suggest more trust needs to be put into the biased observations.
 7 Importantly, this shows that more precise anchor observations will reduce both $K_{X_{Y(1)}}$ and
 8 $K_{\eta_{Y(1)}}$ when biased and anchor observations are at the same time step. As the biases in the
 9 state and model bias parameter are directly proportional to $K_{X_{Y(1)}}$ and $K_{\eta_{Y(1)}}$, as shown in
 10 equations (8.17) and (8.18), the state and model bias parameter will be less contaminated

1 by the observation bias.

2 **8.2.1.2 Biased observations at $t = 1$, anchor observations at $t = 2$**

3 In the next case, we set the biased observations to be at time $t = 1$ and the anchor
 4 observations to be at time $t = 2$, again in a scalar system. The linearised observation
 5 operators would be given by,

$$\hat{H}_{(1)x} = \begin{pmatrix} 0 \\ H_{(1)x,1}M_0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ \alpha \\ 0 \end{pmatrix}, \quad \hat{H}_{(1)\eta} = \begin{pmatrix} 0 \\ H_{(1)x,1} \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad (8.28)$$

$$\hat{H}_{(2)x} = \begin{pmatrix} 0 \\ 0 \\ H_{(2)x,2}M_1M_0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \alpha^2 \end{pmatrix}, \quad \hat{H}_{(2)\eta} = \begin{pmatrix} 0 \\ 0 \\ H_{(2)x,2}(M_1 + 1) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \alpha + 1 \end{pmatrix}. \quad (8.29)$$

6 Substituting equations (8.28), (8.29) into equations (8.7) and (8.9) gives the sensitiv-
 7 ities of the state and model bias parameter analyses to the biased observations when the
 8 biased observations are earlier than the anchor observations as,

$$[K_{xy(1)}]_{t=1} = \frac{\alpha\sigma_{\text{bx}}^2((\alpha + 1)\sigma_{\text{q}}^2 + \sigma_{\text{o}(2)}^2)}{\alpha^2\sigma_{\text{bx}}^2(\sigma_{\text{q}}^2 + \sigma_{\text{o}(2)}^2) + \alpha^2\sigma_{\text{o}(1)}^2 + \sigma_{\text{q}}^2(\sigma_{\text{o}(2)}^2 + (\alpha + 1)^2\sigma_{\text{o}(1)}^2) + \sigma_{\text{o}(1)}^2\sigma_{\text{o}(2)}^2}, \quad (8.30)$$

$$[K_{\eta y(1)}]_{t=1} = \frac{\sigma_{\text{q}}^2(\sigma_{\text{o}(2)}^2 - \alpha^3\sigma_{\text{bx}}^2)}{\alpha^2\sigma_{\text{bx}}^2(\sigma_{\text{q}}^2 + \sigma_{\text{o}(2)}^2) + \alpha^2\sigma_{\text{o}(1)}^2 + \sigma_{\text{q}}^2(\sigma_{\text{o}(2)}^2 + (\alpha + 1)^2\sigma_{\text{o}(1)}^2) + \sigma_{\text{o}(1)}^2\sigma_{\text{o}(2)}^2}. \quad (8.31)$$

10 Equations (8.30) and (8.31) are immediately more complicated than equations (8.26) and
 11 (8.27), when the anchor and biased observations were at the same time step. However,
 12 $K_{xy(1)}$ is still directly proportional to σ_{bx}^2 and $K_{\eta y(1)}$ is directly proportional to σ_{q}^2 . This
 13 shows that, as more weight is given to the background and model respectively, the sen-
 14 sitivity of the state and model bias parameter to the biased observations will be smaller,
 15 and hence there will be less contamination of observation bias on the state and model

1 bias parameter estimates. As it is less clear how the precision of the anchor observations
 2 impacts $K_{xy(1)}$ and $K_{\eta y(1)}$, we will test how $K_{xy(1)}$ and $K_{\eta y(1)}$ vary with given parameters.

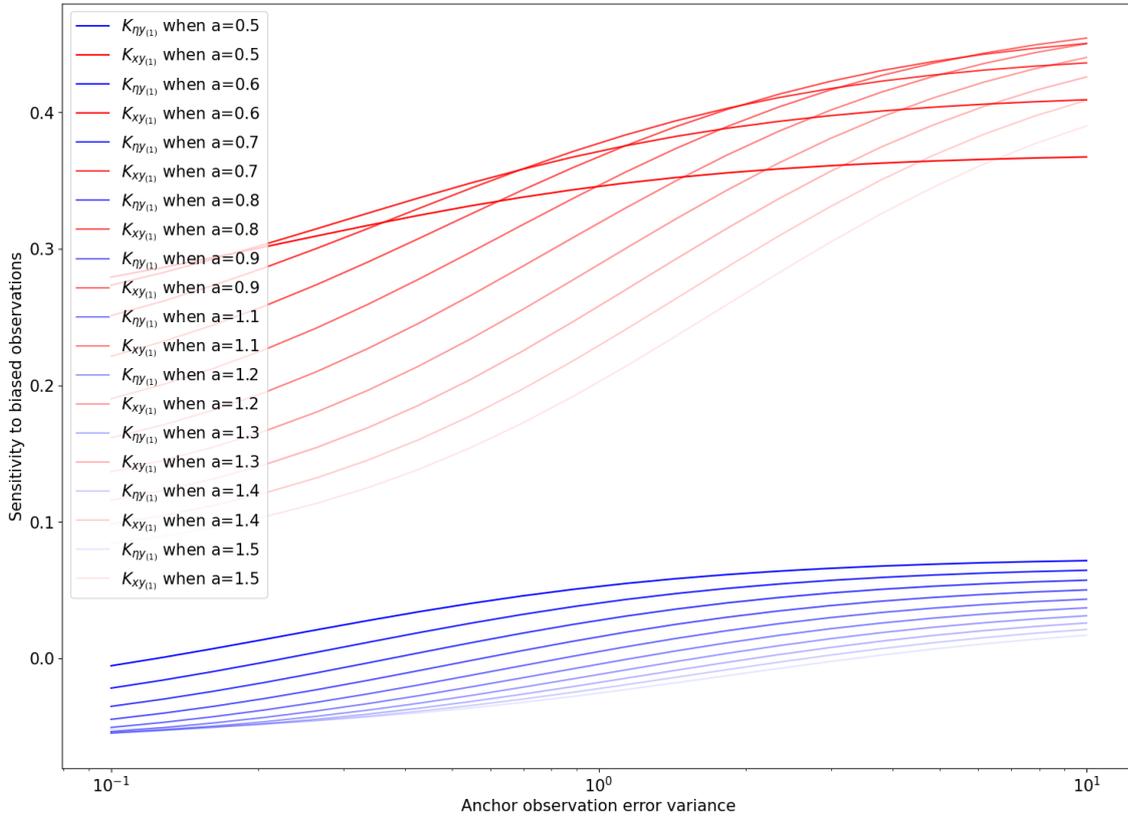


Figure 8.2: $K_{\eta y(1)}$ (blue) and $K_{xy(1)}$ (red) with biased observations at $t = 1$ and anchor observations at $t = 2$ for varied α .

3 In figure 8.2 we have plotted $K_{xy(1)}$ (red) and $K_{\eta y(1)}$ (blue), when we have varied the
 4 anchor observation error variance between 10 times smaller and 10 times larger the biased
 5 observation error variance ($\sigma_{o(1)}^2 = 1$), as in figure 8.1. This has been plotted for 10
 6 values of α between 0.5 and 1.5 (with increasing transparency as α gets larger). $K_{\eta y(1)}$ is
 7 negative when the anchor observation error variance is small, which means that the model
 8 bias parameter analysis would be updated by the observations in the opposite sign to the
 9 innovation vector.

10 $K_{xy(1)}$ is larger than $K_{\eta y(1)}$ for the corresponding value of α for all values of the anchor

1 observation error variance, which is what we expect from equations (8.30) and (8.31) as
 2 $K_{xy(1)}$ is directly proportional to σ_{bx}^2 and $K_{\eta y(1)}$ is directly proportional to σ_q^2 , and we have
 3 defined the model error variance to be much smaller than the background error variance.
 4 This suggests that the observation bias will more greatly contaminate the state analysis,
 5 than it will contaminate the model bias parameter analysis. When the anchor observation
 6 error variance is increased, so does the sensitivity of both the state and the model bias
 7 parameter to the biased observation. This is as we would expect as putting less weight
 8 into the anchor observations would mean more weight is put into the biased observations,
 9 but this was not so obvious from equations (8.30) and (8.31) alone. This suggests that
 10 more precise anchor observations will have a larger impact on reducing the contamination
 11 of observation bias in both the state analysis and the model bias parameter analysis. As
 12 α increases, the gradient of $K_{xy(1)}$ becomes steeper for increasing anchor observation error
 13 variance. This means that as the errors in the model become more unstable, the value of
 14 $K_{xy(1)}$ becomes more sensitive to the anchor observations.

15 8.2.1.3 Biased observations at $t = 2$, anchor observations at $t = 1$

16 In our final experiment we set the biased observations to be at $t = 2$ and the anchor
 17 observations to be at $t = 1$. Therefore, in our scalar system the linearised observation
 18 operators are given by,

$$\hat{H}_{(1)x} = \begin{pmatrix} 0 \\ 0 \\ H_{(1)x,2} M_1 M_0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \alpha^2 \end{pmatrix}, \quad \hat{H}_{(1)\eta} = \begin{pmatrix} 0 \\ 0 \\ H_{(1)x,2} M_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \alpha + 1 \end{pmatrix}, \quad (8.32)$$

$$\hat{H}_{(2)x} = \begin{pmatrix} 0 \\ H_{(2)x,1} M_0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ \alpha \\ 0 \end{pmatrix}, \quad \hat{H}_{(2)\eta} = \begin{pmatrix} 0 \\ H_{(2)x,1} \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}. \quad (8.33)$$

1 Substituting equations (8.32), (8.33) into equations (8.7) and (8.9) gives the sensitivi-
 2 ties of the state and model bias parameter analyses when the biased observations are later
 3 than the anchor observations as,

$$[K_{xy(1)}]_{t=2} = \frac{\alpha^2 \sigma_{bx}^2 (\alpha(2\alpha - 1)\sigma_q^2 + \sigma_{o(2)}^2)}{\sigma_q^2(2\alpha^3\sigma_q^2 - \alpha^2\sigma_q^2 + \sigma_{o(1)}^2 - (\alpha + 1)^4\sigma_q^2) + \sigma_{o(1)}^2(\alpha^2\sigma_{bx}^2 + \sigma_{o(2)}^2)}, \quad (8.34)$$

$$[K_{\eta y(1)}]_{t=2} = \frac{\sigma_q^2(\alpha^2\sigma_{bx}^2 + \alpha\sigma_q^2 + (\alpha + 1)\sigma_{o(2)}^2)}{\sigma_q^2(2\alpha^3\sigma_q^2 - \alpha^2\sigma_q^2 + \sigma_{o(1)}^2 - (\alpha + 1)^4\sigma_q^2) + \sigma_{o(1)}^2(\alpha^2\sigma_{bx}^2 + \sigma_{o(2)}^2)} \quad (8.35)$$

4
 5 where $[K_{xy(1)}]_{t=2}$ and $[K_{\eta y(1)}]_{t=2}$ are the sensitivities at $t=2$ (at the other times the sen-
 6 sitivities will be 0). The terms $K_{xy(1)}$ and $K_{\eta y(1)}$ are again proportional to σ_{bx}^2 and σ_q^2
 7 respectively, which suggests that the sensitivities of the state and model bias parameter
 8 estimates are strongly dependent on the state background and model error variances re-
 9 spectively. As the relationship between the sensitivities and the precision of the anchor
 10 observation is still unclear from equations (8.34) and (8.35), we will plot $K_{xy(1)}$ and $K_{\eta y(1)}$
 11 for different values of α and $\sigma_{o(2)}^2$.

12 In figure 8.3 we have plotted $K_{xy(1)}$ (red) and $K_{\eta y(1)}$ (blue) whilst varying the an-
 13 chor observation error variance for different values of α (lines are more transparent as
 14 α increases). There are now negative values of $K_{xy(1)}$ when the anchor observation error
 15 variance is very small and when the model is stable, which would mean that the state anal-
 16 ysis would be updated by the observations with the opposite sign to the biased innovation
 17 vector.

18 Unlike in the previous two cases, $K_{\eta y(1)}$ is not always smaller than $K_{xy(1)}$ for the
 19 corresponding value of α . This is unexpected from equations (8.34) and (8.35), but is due
 20 to the difference term on the numerator of $K_{xy(1)}$, which did not exist for the previous two
 21 cases. This suggests that the contamination of observation bias on the state analysis may
 22 be smaller in some cases than the contamination of observation bias on the model bias
 23 parameter analysis. This is detrimental to the system, as the more that the observation

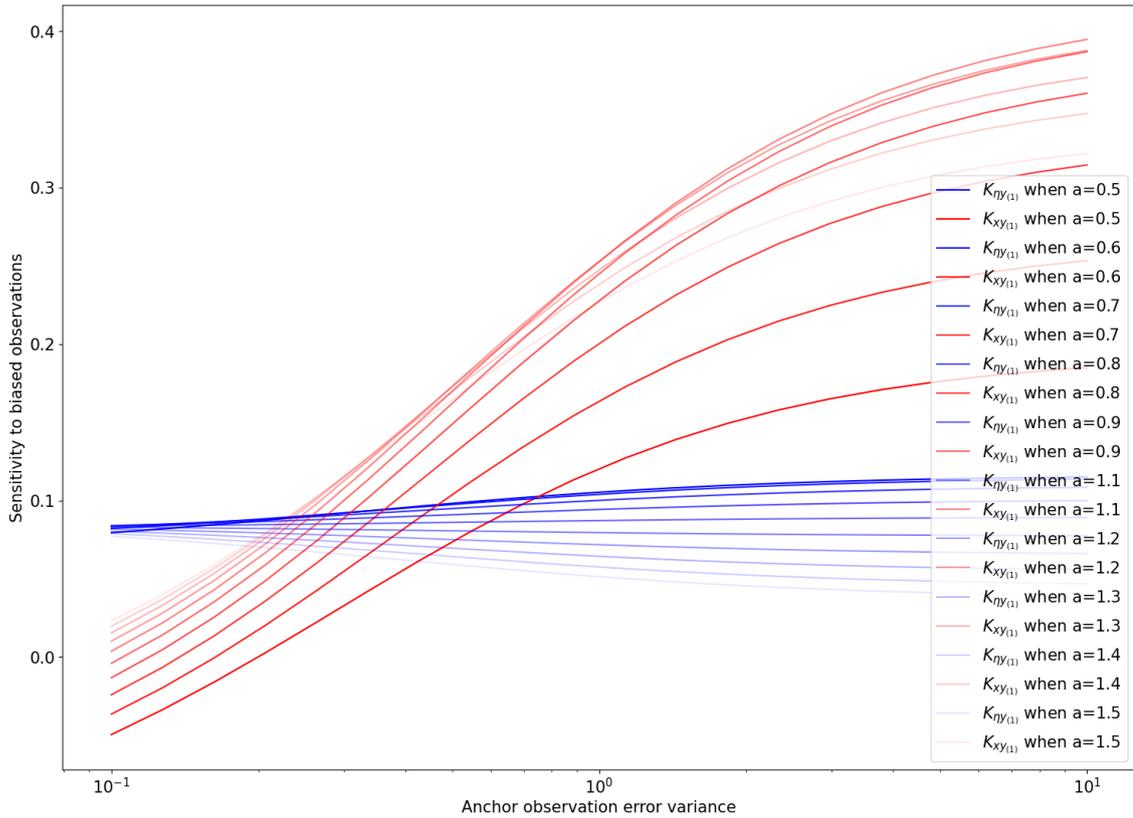


Figure 8.3: $K_{\eta y_{(1)}}$ (blue) and $K_{xy_{(1)}}$ (red) when biased observations at $t = 2$, anchor observations at $t = 1$, for varied α .

1 bias contaminates the model bias parameter analysis, the worse the model bias correction
 2 will be. Bias in the model bias parameter analysis will then filter into the state analysis
 3 in subsequent cycles, increasing the state analysis bias further.

4 When the anchor observation error variance is increased, the sensitivity of the state to
 5 the biased observations increases as we would expect, however the sensitivity of the model
 6 bias parameter to the biased observations either increases or decreases depending on the
 7 stability of the model. As α increases, the gradient of $K_{\eta y_{(1)}}$ changes between positive,
 8 to negative, then back to positive again (positive when $\alpha > 1.6$, not shown here). In
 9 general, $K_{\eta y_{(1)}}$ is not very sensitive to the anchor observation error variance, regardless of
 10 α , which suggests that the anchor observations only have a small impact in reducing the

1 contamination of observation bias in the model bias correction when they are earlier in
 2 the window than the biased observations.

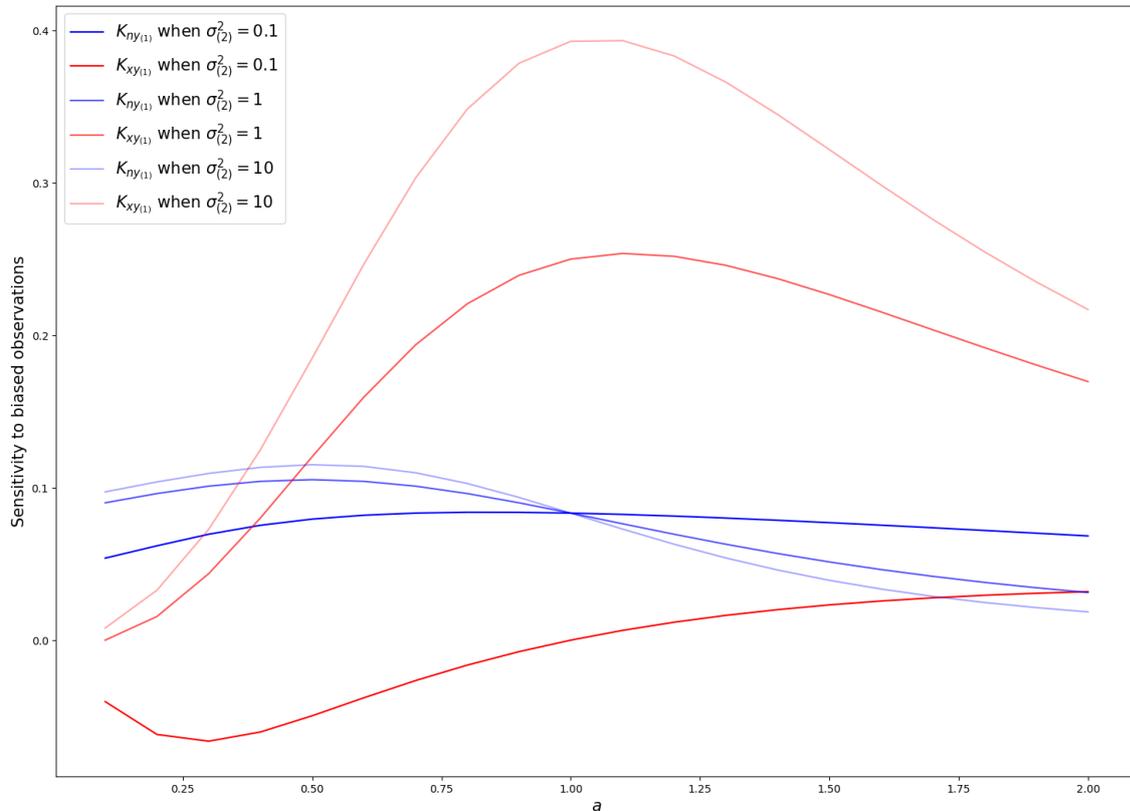


Figure 8.4: $K_{\eta Y(1)}$ (blue) and $K_{XY(1)}$ (red) when biased observations at $t = 2$, anchor observations at $t = 1$ for different values of α and 3 values of $\sigma_{o(2)}$.

3 To test how the sensitivities of the state and the model bias parameter to the biased
 4 observations are dependent on the stability of the model, for the case when anchor ob-
 5 servations are earlier than the biased observations, we have varied α over a larger range.
 6 In figure 8.4 we have plotted $K_{\eta Y(1)}$ (blue) and $K_{XY(1)}$ (red) when we vary α for 3 val-
 7 ues of $\sigma_{o(2)}^2$: 0.1, 1 and 10, shown with increasing levels of transparency for larger $\sigma_{o(2)}^2$.
 8 Each curve of $K_{XY(1)}$ increases with increasing anchor observation error variance, as we
 9 saw in figure 8.3. However, for $K_{\eta Y(1)}$, the lines for the different anchor observation error
 10 variances cross each other. When α is less than 1, $K_{\eta Y(1)}$ increases with increasing anchor

1 observation error variance as expected, but when α is greater than 1, $K_{\eta Y(1)}$ decreases with
2 increasing anchor observation error variance. This shows that when the anchor observa-
3 tions are earlier in the window than the biased observations, the anchor observations do
4 not necessarily perform as we would expect and increasing their precision does not always
5 reduce the contamination of observation bias on the model bias parameter. The ability
6 of the anchor observations to reduce the contamination of observation bias on the model
7 bias parameter is dependent on the stability of the model. It should be noted however,
8 that these numbers are dependent on the other parameters chosen (e.g. non-varied er-
9 ror variances), which means that the impact of having anchor observations earlier in the
10 window and their ability to reduce observation biases needs to be researched further in a
11 non-scalar system in order to come to a more robust conclusion.

12 **8.2.2 Numerical experiments in a multi-variable system**

13 Our theoretical results have mostly focused on a scalar system, in order to somewhat
14 simplify the analysis error equations for the state and model bias parameter, equations
15 (8.17) and (8.18). In order to understand the role of anchor observations when correcting
16 for model bias in the presence of observation bias in a non-scalar system, we test the
17 importance of the timing and the precision of the anchor observations in a simple numerical
18 system, with the same parameters as described in section 7.4 for the 4DVarBC case, but
19 details will be repeated here for the convenience of the reader.

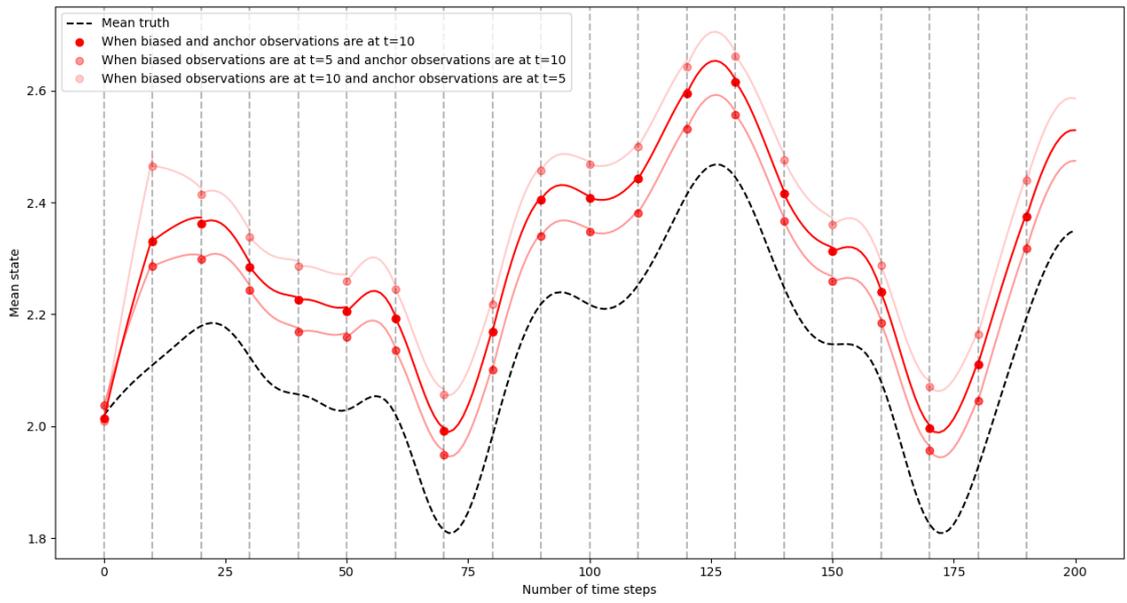
20 We use the Lorenz 96 model (Lorenz, 1996), as described in chapter 5 to create the
21 data assimilation system. The climatological background error covariance matrix is used,
22 as described in section 5.2.2.1, and, as the model bias parameter is scalar, \mathbf{Q} is scalar
23 and is defined as 10% of the mean state background error variances. A bias is added
24 to the forecast model by changing the forcing parameter, such that $F_{\text{biased}} = 12$. The
25 anchor and biased observations are generated from the true model with a random error of
26 error variance 1 (unless stated otherwise). The biased observations also have a constant

1 bias of 0.5 added to the random error. There are biased and anchor observations directly
2 observing the state at every spatial location, but the observations vary in time: when
3 the anchor and biased observations are both at the end of the window ($t = 10$); when
4 the biased observations are at $t = 5$ and the anchor observations are at $t = 10$; when
5 the anchor observations are at $t = 5$ and the biased observations are at $t = 10$. This is
6 analogous to the theoretical cases described above in sections 8.2.1.1, 8.2.1.2 and 8.2.1.3.
7 The observation and model biases are chosen so that they cause a similar bias in the
8 state analysis when only one bias is present. The initial cycle has no added background
9 bias, but background bias will naturally accumulate in the later cycles if the analysis of
10 the previous cycle is biased. The window length is 10 time steps, which gives the biased
11 model enough time to sufficiently evolve away from the true model; we present the results
12 over 20 windows. We have repeated the experiments for 1000 realisations, which are all
13 initialised with different random errors in both the background and observations, and then
14 average the results over all realisations. The model bias parameter η is scalar as the model
15 bias is spatially and temporally invariant.

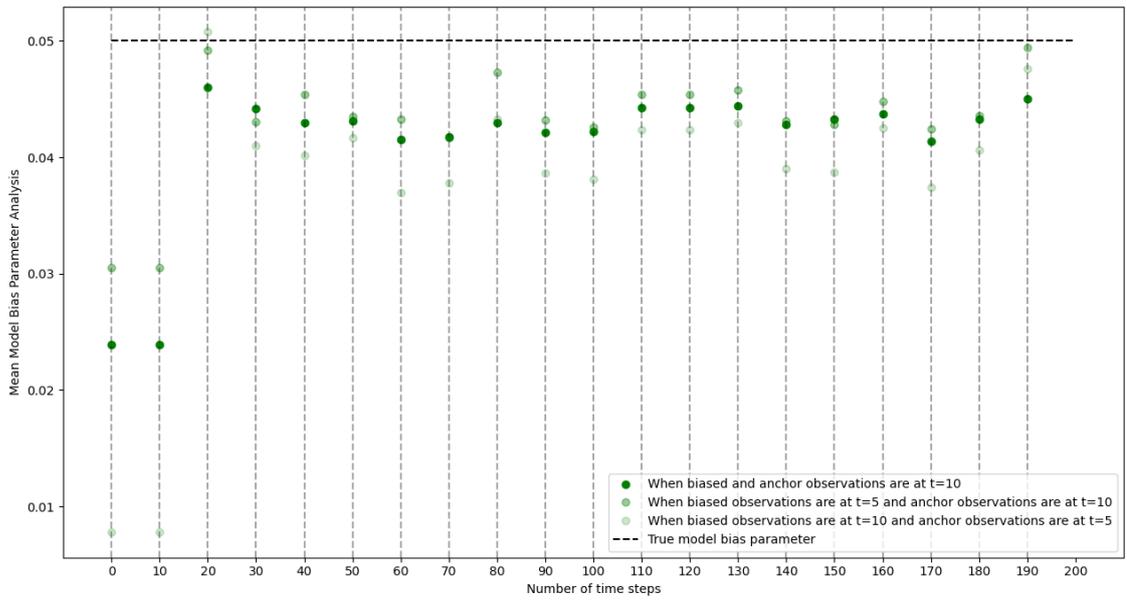
16 In figures 8.5a and 8.6a we have plotted the mean state analysis across all realisations
17 and all states for two different anchor observation error variances: the circles are the mean
18 state analyses at the beginning of each window, and the tails are the trajectories across
19 the windows. The dashed black line is the mean true trajectory across all states which
20 uses the true model ($F = 8$). In figures 8.5b and 8.6b we have plotted the mean model
21 bias parameter analysis across all realisations. As the model bias parameter analysis is
22 constant throughout each window, it is plotted as one dot at the beginning of the window.

23 In figure 8.5 we have plotted the mean state analysis and the mean model bias param-
24 eter analysis when the anchor and biased observations have the same error variance (as
25 described above). The biased and anchor observation times vary for the three experiments
26 as explained above, which are shown by the varying transparency of the lines.

27 In figure 8.5a, the state analysis has been shifted vertically away from the truth for



(a) Mean state analysis



(b) Mean model bias parameter analysis

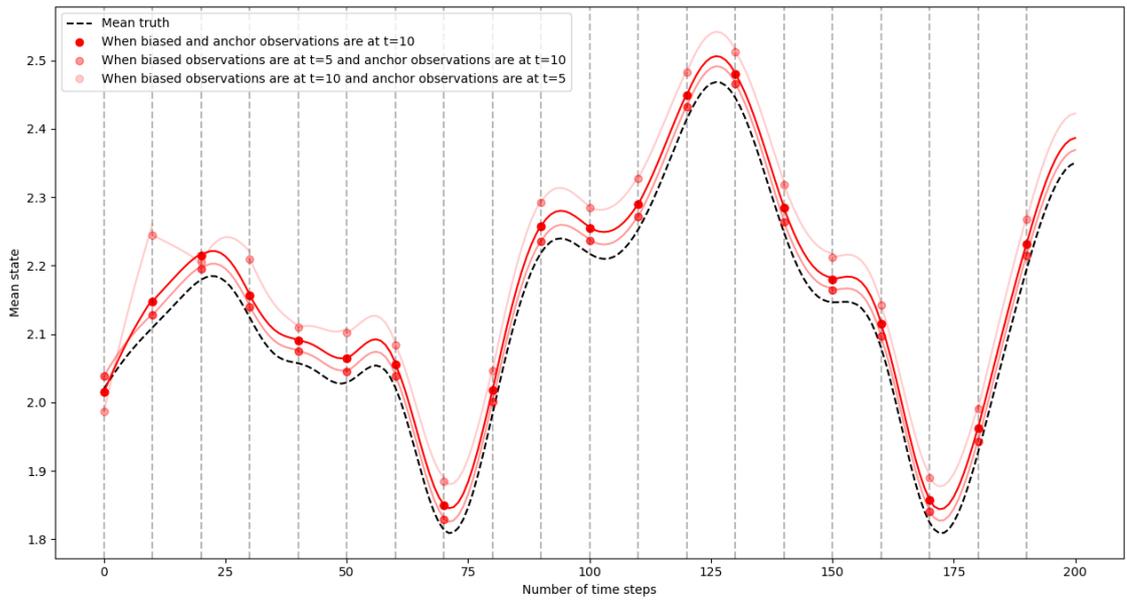
Figure 8.5: WC_4DVar to correct for model bias only, with anchor and biased observations at different locations in time. $\sigma_{o(2)}^2 = 1$.

1 all three cases, but the analysis trajectory closely follows the true trajectory (albeit also
2 shifted away from the truth), which suggests that most of the model bias has been suc-
3 cessfully removed and the bias in the state analysis is caused by the observation bias. The
4 mean state analysis initially starts at the truth as there is no background bias in the first
5 window.

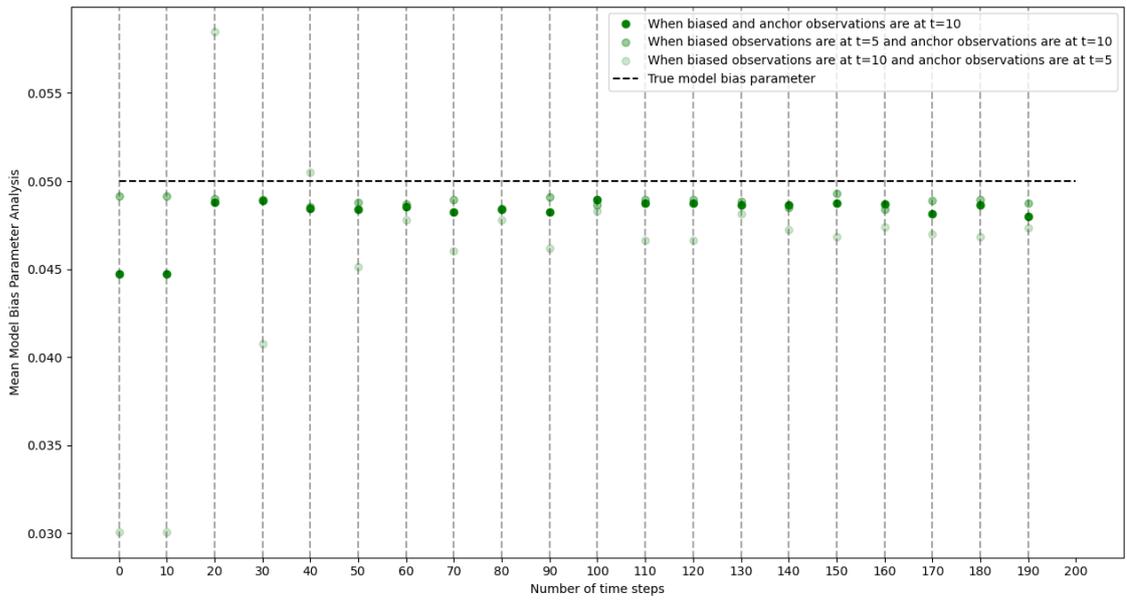
6 In figure 8.5a it is clear that the state analysis has the most bias when the anchor
7 observations are earlier in the window than the biased observations. This shows that
8 anchor observations should be later in the window to have the biggest impact on reducing
9 the contamination of observation bias. The case that causes the smallest bias in the
10 state analysis is when the biased observations are earlier in the window than the anchor
11 observations. This makes sense as equation (8.17) showed that the contamination of
12 observation bias in the state analysis was controlled by the sensitivity of the state to the
13 biased observations. Observations later in the window will have a larger weighting in the
14 data assimilation system, so the analysis will have a larger sensitivity to later observations.
15 Therefore to reduce the contamination of observation bias, it is better to have a smaller
16 weighting given to the biased observations (by having them earlier in the window) and a
17 larger weighting given to the anchor observations (by having them later in the window).

18 In figure 8.5b, there is only a small contamination of observation bias in the mean
19 model bias parameter analysis. This was as we predicted in the theory in sections 8.2.1.1,
20 8.2.1.2 and 8.2.1.3 as we showed $K_{\eta_{Y(1)}}$ was generally much smaller than $K_{xy(1)}$ for the
21 scalar case, which implied that less observation bias would contaminate the model bias
22 parameter than the state analysis. In general, the mean model bias parameter analysis has
23 the largest bias when the anchor observations are earlier in the window than the biased
24 observations (lightest circle) and it has the smallest bias when the biased observations are
25 earlier in the window than the anchor observations (middle circle).

26 In figure 8.6 we have again plotted the mean state analysis and the mean model bias
27 parameter analysis for different observation times in the window. However, the anchor



(a) Mean state analysis



(b) Mean model bias parameter analysis

Figure 8.6: As in figure 8.5, but with more precise anchor observations. $\sigma_{o(2)}^2 = 0.1$.

- 1 observation error variance is now ten times smaller than the biased observations and is
- 2 given by $\sigma_{o(2)}^2 = 0.1$, to test the impact of using more precise anchor observations.

1 Figure 8.6a shows that with more precise anchor observations, the bias in the state
2 analysis has been reduced for all three cases, but a small bias still remains. This is in line
3 with our theory in sections 8.2.1.1, 8.2.1.2 and 8.2.1.3, as we showed that more precise
4 anchor observations would reduce $\mathbf{K}_{xy(1)}$, which would in turn reduce the contamination
5 of observation bias on the state analysis error.

6 In figure 8.6b the bias in the model bias parameter analysis has been reduced when
7 more precise anchor observations are used in comparison to figure 8.5b, especially when
8 the anchor observations are later in the window. In our theoretical results we showed that,
9 when the anchor observations were before the biased observations, increasing the precision
10 of the anchor observations did not necessarily reduce the sensitivity of the model bias
11 parameter to the biased observations and hence reduce the contamination of observation
12 bias on the model bias parameter. In figure 8.6b the bias in the mean model bias parameter,
13 when the anchor observations are earlier than the biased observations, has in general
14 been reduced (but not as much as when anchor observations are later), although it takes
15 the system several cycles for the model bias parameter to reach an equilibrium with a
16 small bias. As our results in sections 8.2.1.1 - 8.2.1.3 were dependent on the stability of
17 the errors in the model, our multivariate numerical experiments should be tested on a
18 different system, perhaps on an operational model with simulated observations, to test
19 whether theoretically increasing the precision of the anchor observations ever increases
20 the bias in the model bias parameter, when the anchor observations are earlier in the
21 window. Alternatively, to avoid this odd case, it would be better to avoid having anchor
22 observations earlier in the window than the biased observations.

23 **8.2.3 Summary of WC4DVar when not correcting for observation bias**

24 In this section we have looked at how observation bias contaminates the estimate of model
25 bias in WC4DVar. In equation (8.18) we showed that the contamination of observation
26 bias on the model bias parameter can be constrained by $\hat{\mathbf{K}}_{\eta y(1)}$, the sensitivity of the

1 model bias parameter to the biased observations, which is implicitly dependent on the
2 anchor observations through the linearised anchor observation operator and the anchor
3 observation error variance. In sections 8.2.1.1, 8.2.1.2 and 8.2.1.3 and figures 8.5 and
4 8.6 we demonstrated the role of anchor observations in reducing the contamination of
5 observation bias by studying three simple cases: when the anchor and biased observations
6 observe the same time step; when the biased observations observe a state earlier than
7 the anchor observations; and when the biased observations observe a state later than the
8 anchor observations.

9 For all three cases, the sensitivity of the state to the biased observations increased
10 as the anchor observation error variance was increased. This is what we would expect,
11 as giving less weighting to the anchor observations would suggest that more sensitivity
12 needs to go to the other sources of information. In figure 8.6a, we demonstrated that
13 increasing the precision of the anchor observations reduced the bias in the state analysis
14 compared to figure 8.5a. Therefore, combining our theoretical and numerical results show
15 that more precise anchor observations will reduce the sensitivity of the state analysis to
16 the biased observations and therefore reduce the contamination of observation bias on the
17 state analysis.

18 When the anchor observations were later in the window (either at the same time step
19 or later than the biased observations), we showed in sections 8.2.1.1 and 8.2.1.2 that
20 reducing the anchor observation error variance also reduced the sensitivity of the model
21 bias parameter to the biased observations, which would mean less of the observation bias
22 could contaminate the model bias parameter analysis. However, when anchor observations
23 were earlier in the window in section 8.2.1.3, reducing the anchor observation error variance
24 did not always correlate to a reduction in the sensitivity of the model bias parameter
25 to the biased observations. In our numerical results we found in figures 8.5b and 8.6b
26 that reducing the anchor observation error variance reduced the bias in the model bias
27 parameter for all three cases, even when the anchor observations were earlier in the window

1 than the biased observations. However, when the anchor observations were earlier, it took
2 more cycles for the model bias parameter to reach an equilibrium between the observation
3 bias and the anchor observations.

4 In our numerical results in figure 8.5a, we found that having biased observations earlier
5 than the anchor observations reduced the most contamination of observation bias. This
6 was also true for the model bias parameter in figure 8.5b, where the contamination of ob-
7 servation bias on the model bias estimate was most reduced when the anchor observations
8 were later in the window than the biased observations, although the differences between
9 the three cases were less significant than for the state analysis. This result makes sense
10 as equation (8.18) showed that observation bias contaminates the model bias estimate via
11 the sensitivity of the model bias parameter to the biased observations. In a data assimi-
12 lation system, observations later in the window have a larger impact on the analysis than
13 observations earlier in the window. Therefore, the earlier the biased observations in the
14 window, the less impact they will have on the state and model bias parameter analyses
15 and so there will be less contamination of observation bias.

16 **8.3 Comparison of the role of anchor observations when cor-** 17 **recting for observation bias, model bias, or both**

18 In chapter 7 we studied the role of anchor observations in 4DVarBC in the presence of
19 model bias. We found that having anchor observations at the same time step or later than
20 the bias-corrected observations reduced the most contamination of model bias, as, if the
21 bias-corrected observations were later in the window, more model bias could accumulate
22 which the anchor observations could not control. In our WC4DVar experiments in the
23 presence of observation bias, we similarly found that having the anchor observations later
24 in the window than the biased observations meant that there was less of a contamination of
25 observation bias on the model bias correction. In both techniques we found that using more

1 precise anchor observations reduced the contamination of bias in the system. In this section
 2 we will demonstrate the ability of the anchor observations to reduce the contamination of
 3 model and observation bias when both are being corrected for simultaneously. As we found
 4 similar results for the ability of the anchor observations to reduce the contamination of bias
 5 in both 4DVarBC and WC4DVar, we would expect that anchor observations would reduce
 6 the most contamination of both observation and model biases on the other correction,
 7 when they are precise and later than the biased observations.

8 As we described in section 3.2, Lorente-Plazas and Hacker (2017) compared correcting
 9 for observation and/or model bias in a simple numerical system to test which method
 10 reduced the most root mean square error in the analysis. However, they did not include
 11 experiments that used anchor observations, which we have shown to play an important
 12 role when both observation and model biases are present. In this section we will extend
 13 the results of Lorente-Plazas and Hacker (2017) by comparing our numerical results for
 14 VarBC and WC4DVar in the presence of both observation and model biases when anchor
 15 observations are available. We will also compare the separate use of the correction methods
 16 with their simultaneous use, in order to correct for both observation and model biases
 17 simultaneously. We will not study the theoretical equations for combining VarBC and
 18 WC4DVar, as they are too complicated to be insightful. All of the numerical results come
 19 from the same system as described in sections 7.4 and 8.2.2, with observation and model
 20 biases that cause similar state analysis biases.

21 In table 8.1, the table of values shows the mean difference between the state analysis
 22 and the truth, where the mean is taken over all 40 state variables, all 1000 realisations
 23 and over all time and is calculated by the following,

$$\frac{1}{20} \sum_{t=1}^{20} \left(\frac{1}{40} \sum_{i=1}^{40} \left(\frac{1}{1000} \sum_{j=1}^{1000} (x_{t,i,j}^a - x_{t,i}^t) \right) \right), \quad (8.36)$$

24 where j is each realisation; i is each spatial state; and t is each cycle. We have only

1 calculated the differences at the beginning of each window (to give the difference in the
2 analysis bias, not its trajectory in the window), which is why t is between 1 and 20, where
3 20 is the number of cycles. We will refer to these mean differences in equation (8.36) as
4 the mean state analysis bias.

State bias	$t_{\text{biased}} = t_{\text{unbiased}} = 10$			$t_{\text{biased}} = 5, t_{\text{unbiased}} = 10$			$t_{\text{biased}} = 10, t_{\text{unbiased}} = 5$		
	$\sigma_{\alpha(2)}^2 = 0.1$	$\sigma_{\alpha(2)}^2 = 1$	$\sigma_{\alpha(2)}^2 = 10$	$\sigma_{\alpha(2)}^2 = 0.1$	$\sigma_{\alpha(2)}^2 = 1$	$\sigma_{\alpha(2)}^2 = 10$	$\sigma_{\alpha(2)}^2 = 0.1$	$\sigma_{\alpha(2)}^2 = 1$	$\sigma_{\alpha(2)}^2 = 10$
No correction	-	0.324	-	-	-	-	-	-	-
VarBC	-0.014	0.477	-	-0.014	0.492	-	0.117	0.609	-
WC4DVar	0.033	0.169	0.278	0.017	0.118	0.289	0.064	0.226	0.291
Both	-	-0.008	-0.019	-	-0.007	-0.014	-	-0.012	-0.020

Table 8.1: Table of mean state analysis biases when observation and model biases are present and no/one/both bias correction techniques are used. 9 cases are presented: when biased and anchor observations are at different times in the window and when the anchor observation error variance is varied. All values rounded to 3 decimal places.

5 In table 8.1 we have presented the mean state analysis bias when there is both ob-
6 servation and model bias present, but when: neither bias has been corrected for; obser-
7 vation bias has been corrected for using VarBC; model bias has been corrected for using
8 WC4DVar; and both biases have been corrected for using VarBC and WC4DVar simulta-
9 neously. Within each of these correction techniques we have varied where the observations
10 are in the window: both the anchor and biased observations at $t = 10$; biased observa-
11 tions at $t = 5$, anchor observations at $t = 10$; and biased observations at $t = 10$, anchor
12 observations at $t = 5$. For each of these we have also varied the anchor observation error
13 variance between 0.1 and 10. The results for VarBC and WC4DVar are equivalent to
14 those shown in sections 7.4 and 8.2.2. Where we have left results blank, we have not done
15 the experiments, as we felt they would not give any extra information to our results for
16 the computing power needed to complete them. For example, the bias would either be
17 significantly larger or smaller than the other values, so were considered unimportant.

18 The first thing to note from table 8.1 is that in all cases, there is still a bias within
19 the state analysis, albeit very small when both bias correction techniques are used. This
20 is important, as in this system we have complete spatial coverage of anchor observations

1 that directly observe the state. Operationally, we would expect a less extensive coverage
2 of anchor observations (Eyre et al., 2020) and yet full coverage of anchor observations still
3 leads to an analysis state that is biased.

4 Comparing the experiments that correct for one or two biases with the experiment
5 that did not correct for any biases (rows 2,3,4 compared with row 1), when the anchor
6 observation error variance is 1 and the observations are at the same time, we see that
7 either correcting for model bias or both model and observation biases reduces the state
8 analysis bias from 0.324 to 0.169 and -0.008 respectively. However, only correcting for
9 observation bias using VarBC increases the state analysis bias from 0.324 to 0.477. The
10 result for using VarBC in the presence of model bias reflects the results in Dee (2005), who
11 suggested that correcting for only one type of bias could be worse than performing “bias-
12 blind” assimilation and also reflects the results in Eyre (2016) who showed that correcting
13 for observation bias can exacerbate the effect of model bias. Operational centres have
14 found that correcting for observation bias using VarBC is vital in order to use radiance
15 observations, so it is important to note here that in this experiment we have specifically
16 chosen the observation and model biases to cause similar magnitudes of bias in the state
17 analysis, whereas usually one bias would dominate the other. Also, predictors are used
18 in operational VarBC, which have been designed to mitigate the contamination of model
19 bias in the observation bias correction (Cameron & Bell, 2018; Harris & Kelly, 2001), so
20 we would expect operational VarBC to perform better than not correcting for observation
21 bias at all.

22 Next we compare the experiments that have small anchor observation error variance
23 (the green columns) with the experiments with anchor observation error variance equal to
24 1 (the yellow columns). In table 8.1, more precise anchor observations have reduced the
25 bias in the state analysis when using either VarBC or WC4DVar compared to using anchor
26 observations with the same error variance as the biased observations. This is particularly
27 true for VarBC where more precise anchor observations have reduced the bias in the state

1 from 0.477 to -0.014 (rounded from -0.0141), 0.492 to -0.014 (rounded from -0.0137) and
2 0.609 to 0.117 respectively for the three cases of observations in time, which shows the
3 importance of precise anchor observations, particularly when correcting for model bias
4 in VarBC. This is an extension of Eyre (2016), who demonstrated the need for anchor
5 observations in VarBC in the presence of model bias in a scalar system, whereas here we
6 have shown the importance of precise anchor observations in a 40-variable 4DVar system.

7 Now we look to the three groups of columns outlined with the bold lines; when the
8 biased and anchor observations are at different times in the window. The mean state
9 analysis bias in table 8.1 has the largest magnitude between equivalent experiments when
10 the anchor observations are earlier in the window, as we could clearly see from figures
11 7.1a and 8.5a. This suggests that, for anchor observations to have the largest impact in
12 reducing the contamination of bias when correcting for only one type of bias, they need
13 to be towards the end of the window (or at least at the same/later time step to the biased
14 observations). This reflects our results in the theory for VarBC and WC4DVar in sections
15 7.3 and 8.2 respectively.

16 Lorente-Plazas and Hacker (2017) suggested that correcting for both biases simulta-
17 neously was better than only correcting for one. However, in our table 8.1, we see that
18 there is an exception to this, which is when the observation bias has been corrected using
19 VarBC with very precise anchor observations. In this case the mean state analysis biases
20 are -0.014 (rounded from -0.0141), -0.014 (rounded from -0.0137) and 0.117 for the three
21 observation times respectively, which is lower when compared to correcting for both with
22 very imprecise anchor observations, which have mean state analysis biases: -0.019, -0.014
23 and -0.012 for the three respective observation times. Therefore, this suggests that cor-
24 recting for VarBC with precise anchor observations could reduce the state analysis bias
25 more than correcting for both with less precise (or fewer) anchor observations. This is im-
26 portant as implementing both correction techniques operationally requires a lot of time to
27 set up, as well as a lot of prior knowledge about the system in order to create both model

1 bias parameter and observation bias coefficient background error covariance matrices.

2 In table 8.2 the table of values for both the mean observation bias coefficient bias and
 3 the mean model bias parameter bias are given for the same experiments as in table 8.1.

4 The mean observation bias coefficient bias is defined as,

$$\frac{1}{20} \sum_{t=1}^{20} \left(\frac{1}{1000} \sum_{j=1}^{1000} (\beta_{t,j}^a - \beta_t^t) \right), \quad (8.37)$$

5 and the mean model bias parameter bias is defined as,

$$\frac{1}{20} \sum_{t=1}^{20} \left(\frac{1}{1000} \sum_{j=1}^{1000} (\eta_{t,j}^a - \eta_t^t) \right), \quad (8.38)$$

6 which are the means of the biases in the observation bias coefficient and model bias pa-
 7 rameter over all realisations and all cycles. Studying the mean observation bias coefficient
 8 and model bias parameter analyses demonstrates how having both model and observation
 biases can contaminate the estimate of the other. In table 8.2a, all of the values of the

β^a bias	$t_{\text{biased}} = t_{\text{unbiased}} = 10$			$t_{\text{biased}} = 5, t_{\text{unbiased}} = 10$			$t_{\text{biased}} = 10, t_{\text{unbiased}} = 5$		
	$\sigma_{o(2)}^2 = 0.1$	$\sigma_{o(2)}^2 = 1$	$\sigma_{o(2)}^2 = 10$	$\sigma_{o(2)}^2 = 0.1$	$\sigma_{o(2)}^2 = 1$	$\sigma_{o(2)}^2 = 10$	$\sigma_{o(2)}^2 = 0.1$	$\sigma_{o(2)}^2 = 1$	$\sigma_{o(2)}^2 = 10$
VarBC	-0.193	-0.589	-	-0.178	-0.588	-	-0.309	-0.694	-
Both	-	0.008	0.020	-	0.004	0.013	-	0.013	0.022

(a) Table of mean observation bias coefficient bias. True observation bias coefficient = 0.5.

η^a bias	$t_{\text{biased}} = t_{\text{unbiased}} = 10$			$t_{\text{biased}} = 5, t_{\text{unbiased}} = 10$			$t_{\text{biased}} = 10, t_{\text{unbiased}} = 5$		
	$\sigma_{o(2)}^2 = 0.1$	$\sigma_{o(2)}^2 = 1$	$\sigma_{o(2)}^2 = 10$	$\sigma_{o(2)}^2 = 0.1$	$\sigma_{o(2)}^2 = 1$	$\sigma_{o(2)}^2 = 10$	$\sigma_{o(2)}^2 = 0.1$	$\sigma_{o(2)}^2 = 1$	$\sigma_{o(2)}^2 = 10$
WC4DVar	-0.002	-0.009	-0.014	-0.001	-0.007	-0.018	-0.004	-0.012	-0.015
Both	-	-4×10^{-4}	3×10^{-5}	-	-6×10^{-4}	-7×10^{-7}	-	-2×10^{-4}	8×10^{-5}

(b) Table of mean model bias parameter bias. True model bias parameter = 0.05.

Table 8.2: As in figure 8.1 but for the mean observation bias coefficient bias (when using VarBC) and the mean model bias parameter bias (when using WC4DVar).

9

10 observation bias coefficient bias are larger in magnitude than the equivalent values of the
 11 model bias parameter bias in table 8.2b. This is as expected as the true observation bias
 12 coefficient is ten times larger than the true model bias parameter. However, the bias in
 13 the observation bias coefficient is more than ten times larger the bias in the model bias

1 parameter. This demonstrates that observation bias correction is more largely impacted
2 by model bias than model bias correction is impacted by observation bias. It should be
3 noted that the value by which the observations/model are contaminated, is dependent on
4 the model/observation bias present in the particular system, and the biases in our ex-
5 periments, have been chosen so that both the observation and model biases cause similar
6 biases in the state analysis.

7 The conclusions for the precision and timing of the anchor observations are mostly
8 the same as in table 8.1, except for the model bias parameter when correcting for both
9 observation and model biases. The bias in the model bias parameter is most reduced when
10 the anchor observations are earlier in the window than the biased observations. As this
11 is the opposite result for both the state and the observation bias coefficient analyses, this
12 suggests that the model bias parameter has taken in some of the bias from the system
13 in order to reduce the bias in the state and/or observation bias coefficient. As correcting
14 for both biases simultaneously has reduced the bias in the model bias parameter to a
15 value close to zero for all three cases of observations in time, we will regard this case as
16 unimportant, but this may present a problem if the biases in the model bias parameter
17 were significantly larger.

18 Correcting for both observation and model bias gives the most accurate estimates of
19 the biases compared to using only one or neither bias correction technique in both tables
20 8.1 and 8.2. The biases in the observation bias coefficient and model bias parameter are
21 very small, which suggests that biases have been correctly attributed to their sources.
22 However, we wanted to confirm that this is not because we have full spatial coverage of
23 anchor observations and therefore the system has a good knowledge of the truth. The
24 orange columns in tables 8.1, 8.2a and 8.2b represent when the anchor observations have a
25 very large error variance ($\sigma_{o(2)}^2 = 10$) in comparison to the biased observations ($\sigma_{o(1)}^2 = 1$),
26 so can be regarded as practically unused in the system. We can see that the biases have
27 increased for all times of the observations, but the biases remain negligible (from -0.008 to

1 -0.019, -0.007 to -0.014 and -0.012 to -0.020 for the mean state analysis bias). This shows
2 that, in our experiments, correcting for both observation and model biases simultaneously
3 reduces the bias in the state analysis, regardless of the use of anchor observations, which
4 is consistent with the Lorente-Plazas and Hacker (2017) results, which did not consider
5 anchor observations at all. This could be because, although the model and observation
6 biases have similar spatial structures (i.e. they are both constant over the spatial domain),
7 they have different temporal structures, so this may be how the system is able to separate
8 them. We tested this hypothesis by reducing the window to two time steps, so that the
9 model and observation biases would contaminate the system at two and one time steps
10 respectively (ie. their temporal structures are more similar). The system was still mostly
11 able to correct for both biases simultaneously, leading to a negligible bias in the state
12 analysis of approximately 0.015 when the observations were at the same time step. This
13 is just under a doubling of the magnitude of the bias when the window length was ten
14 time steps (-0.008), so does not disprove this hypothesis, but as the bias is still small, it
15 suggests that the different structures between the observation and model biases is not the
16 only cause.

17 Dee (2005) suggested that biases can only be correctly attributed to their source if the
18 structure of the biases is known. In our experiments, the system knows that the bias in
19 both the model and the observation is just a constant added to the data. Operationally,
20 the form of the bias may not be known and is only an assumption. Therefore, it is likely
21 that correcting for both biases simultaneously performed well in our experiments because
22 the structure of the biases was known. Correcting for both biases simultaneously should
23 thus be tested on a system where the form of the bias is unknown, or mis-specified, in
24 order to understand the effect of assuming the bias correction function. This additional
25 work is beyond the scope of this project.

1 8.3.1 Model and observation biases acting in different directions

2 So far we have only demonstrated the results of experiments where both the observation
 3 and model biases pull the state analysis in the same (positive) direction. In this next
 4 section we use a negative observation bias of $\beta^t = -0.5$ so that the observation and model
 5 biases pull the state analysis in opposite directions. We have set the observation error
 6 variances to be equal (=1) and the anchor and biased observations are both at the end of
 the window ($t = 10$).

State bias		β^a bias		η^a bias	
No correction	-0.018				
VarBC	0.476	VarBC	-0.585		
WC4DVar	-0.185			WC4DVar	0.008
Both	-0.001	Both	0.011	Both	-5×10^{-4}

Table 8.3: The mean state, observation bias coefficient and model bias parameter analysis biases when the observation and model biases act in different directions. $t_{biased} = t_{unbiased} = 10$. $\sigma_{o(2)}^2 = 1$.

7

8 In table 8.3, we have presented the mean state analysis bias, the mean observation bias
 9 coefficient analysis and the mean model bias parameter analysis when the observation and
 10 model biases act in opposite directions. When no bias correction technique is used, the
 11 mean state analysis bias is -0.018 (compared to 0.324 when β^t is positive). This is close
 12 to zero and so shows that the two biases naturally cancel each other out, without having
 13 to use either bias correction technique. When only one bias correction technique is used
 14 (i.e. either VarBC or WC4DVar), the state analysis bias is significantly larger than when
 15 no bias correction technique is used, giving a mean state analysis bias of 0.476 (using
 16 VarBC only) and -0.185 (using WC4DVar only). These values are similar in magnitude to
 17 when the observation bias was positive (table 8.1), which shows that the observation bias
 18 and model bias interact in a similar way when correcting for only one bias, regardless of
 19 whether the observation and model biases act in the same or opposite directions. This is
 20 also supported by the fact that the observation bias coefficient bias and the model bias

1 parameter biases have similar magnitudes to when β^t was positive. Finally, if both model
2 and observation biases are corrected for, then the mean state analysis bias is the smallest
3 at -0.001, which is smaller than correcting for one type of bias or correcting for neither.
4 Therefore, this gives the same results as in Lorente-Plazas and Hacker (2017), that if the
5 observation and model biases act in opposite directions, then our numerical experiments
6 give smaller biases when either both or neither of the biases are corrected for, compared
7 to when only one is corrected for. As previously discussed, more research is needed in a
8 numerical system that mis-specifies the form of the biases, to check whether correcting for
9 both would still perform so well in a more realistic system.

10 **8.4 Conclusions**

11 The aim of this chapter has been to understand the role of anchor observations in the in-
12 teraction of 4DVarBC and WC4DVar. This was approached by first studying theoretically
13 how uncorrected observation bias contaminates WC4DVar, to mirror the studies in chapter
14 7, where theoretical results were discussed for how uncorrected model bias contaminates
15 4DVarBC. We then compared numerical experiments that corrected for observation bias,
16 model bias, or both simultaneously to compare how well the anchor observations were
17 able to disentangle the observation and model biases when either one or both biases were
18 corrected for.

19 In equation (8.17) and figures 8.5a and 8.6a we demonstrated that using more precise
20 anchor observations reduced more of the contamination of uncorrected observation bias
21 on the state analysis in WC4DVar. This mirrors the conclusion of chapter 7, that more
22 precise anchor observations reduced more of the contamination of uncorrected model bias
23 in 4DVarBC. However, in table 8.1, we showed that anchor observations had a larger im-
24 pact in reducing the contamination of uncorrected model bias in VarBC, than in reducing
25 the contamination of uncorrected observation bias in WC4DVar. When we have referred

1 to more precise anchor observations, we have used it to mean a smaller anchor observation
2 error variance such that they are given more weight in the analysis. The precision of
3 observations can also be reduced by having a larger number of anchor observations in one
4 area (which is easier to achieve in practice than having more precise individual observa-
5 tions). As we showed in chapter 2.2, anchor observations, particularly radiosondes, are
6 quite sparse. Therefore areas with less coverage of anchor observations, for example in the
7 southern hemisphere in the troposphere will be more susceptible to biases contaminating
8 the model or observation bias corrections. Therefore, this work shows that future anchor
9 observing networks should aim to either reduce the anchor observation error variance, or
10 as this can be difficult in practice, aim to increase the frequency of observations, in order to
11 increase their weighting within the data assimilation system. Further studies in this area
12 should also test correcting for observation/model biases in systems with inhomogeneously
13 spaced anchor observations, to study how well the anchor observations can reduce the
14 contamination of bias when they are not evenly spaced throughout the domain, as would
15 occur in operational observations. Presumably, the ability of the anchor observations to
16 reduce the contamination of biases in locations that are unobserved would reduce, as we
17 found for VarBC when anchor observations did not observe the whole domain in chap-
18 ters 6 and 7, where the background error correlations between states and the model were
19 given greater importance. However, this should be tested thoroughly on a more realistic
20 numerical system and so is beyond the scope of this thesis.

21 In figure 8.5 we demonstrated that having anchor observations later in the window
22 and biased observations earlier in the window significantly reduced the contamination of
23 uncorrected observation bias when correcting for model bias, compared to having the ob-
24 servations at other times. This mirrors the conclusion of chapter 7 which also found that
25 anchor observations later in the window reduced more of the contamination of uncorrected
26 model bias in 4DVarBC. In fact we showed in table 8.1 that, when correcting for either
27 or both observation and/or model bias, having anchor observations later in the window

1 reduced the most bias in the state analysis. Therefore, regardless of the correction tech-
2 nique used, in order to ensure the least contamination of bias on the analysis, it is safest
3 to have anchor observations later in the window so that biased observations are less likely
4 to be later than the anchor observations. Therefore, as previously discussed in chapter 7,
5 as radiosondes tend to only provide data every 6 or 12 hours (*ECMWF Geographical Cov-*
6 *erage*, 2023), continued work should be put into developing anchor observations that can
7 provide more regular data updates, such as in the expansion of radio occultation satellite
8 observations (e.g. Harnisch et al., 2013; Cucurull et al., 2018), or alternatively, operational
9 centres could look to change the timing of the assimilation windows so that the anchor
10 observations occur at the end of the window.

11 In table 8.1 we found that correcting for both observation and model biases simulta-
12 neously significantly reduced the bias in the state analysis, compared to only correcting
13 for one, regardless of the weighting given to the anchor observations. This reflected the
14 results found in Lorente-Plazas and Hacker (2017), who did not consider anchor observa-
15 tions in their experiments. This tentatively suggests that, if both observation and model
16 biases are corrected for, then fewer anchor observations would be required to achieve an
17 equally accurate analysis (to only correcting for observation bias). However, in order to
18 more robustly demonstrate that correcting for both biases simultaneously is better than
19 only correcting for one and to test when anchor observations become important in disen-
20 tangling the biases, the numerical results should be repeated in a numerical system where
21 the structure of the biases is mis-specified as Dee (2005) predicted that biases could only
22 be attributed to their correct sources if the forms of the biases were known.

23 **8.5 Summary**

24 In chapter 8 we have demonstrated the impact that anchor observations can make in a
25 system correcting for model bias in the presence of observation bias and then compared

1 the numerical results to systems that only correct for observation bias, or that correct
2 for both biases simultaneously. We have found that, in answer to research question 2.3,
3 precise anchor observations are important in reducing the contamination of bias, but they
4 are particularly important when correcting for observation bias in the presence of model
5 bias, answering research question 3. In answer to research question 2.1, the timing of the
6 anchor observations significantly impacted their ability to reduce the contamination of
7 bias: we concluded that, when correcting for model bias, it is more important for anchor
8 observations to be later than the biased observations, and when correcting for observation
9 bias, it is important for the observations to be at the same time. In the next chapter we
10 will discuss the overall conclusions for the whole thesis, discussing the implications of the
11 results and where the results fall short.

1 Chapter 9

2 Conclusions

3 In this thesis we have discussed how the presence of observation and model biases in a data
4 assimilation system can be detrimental to our estimate of the state, despite an attempt
5 to correct for one or both of the biases using VarBC (for observation bias) or WC4DVar
6 (for model bias).

7
8 In section 3.3 we defined research question 1 as:

- 9 • RQ1: What are the consequences of mis-specifying the background error statistics
10 in VarBC?

11 In chapter 3 (section 3.1) we discussed the work of Eyre and Hilton (2013), who demon-
12 strated that the analysis error variances in variational data assimilation could be larger
13 than the corresponding background error variances if the background error covariance ma-
14 trix was mis-specified. They defined this scenario as the “danger zone”, as it meant that
15 performing data assimilation would degrade the estimate of the state from the background.

16 In chapter 4 we extended this work to a VarBC system, where both the state and bias
17 coefficient background error covariance matrices could be mis-specified, in order to answer
18 RQ1. We found that the consequences of mis-specifying the background error statistics

1 was the presence of danger zones for both the state and the bias-coefficient analyses. If
2 the state analysis fell into the danger zone, then the assimilation of observations would
3 degrade the state analysis from the state background. Similarly, if the bias coefficient anal-
4 ysis fell into the danger zone, the assimilation of observations would degrade the estimate
5 of the observation bias correction from the observation bias coefficient background. In a
6 scalar example, we showed that the state analysis was more likely to fall into the danger
7 zone when the state background error variance was overestimated and the bias coefficient
8 background error variance was underestimated. The opposite was true for the bias coef-
9 ficient: the bias coefficient analysis was more likely to fall into the danger zone when the
10 state background error variance was underestimated and the bias coefficient background
11 error variance was overestimated. This means that if one danger zone is avoided, then the
12 other danger zone is more likely. Operationally, the state and bias coefficient background
13 error covariance matrices are usually estimated, based on prior knowledge of the system,
14 without knowing what the 'true' background error covariance matrices are. Therefore it
15 would be difficult to know whether the background error covariance matrices are being
16 under or overestimated. However, this work has aimed to derive insight into why systems
17 may be performing unexpectedly when using VarBC, rather than suggesting to purposely
18 under or overestimate the background error covariance matrices in the future.

19 Although we derived vector equations for the analysis error covariance matrices when
20 the background error covariance matrices were mis-specified, the majority of the results
21 from chapter 4 were based on a scalar system. In a non-scalar system, both the variances
22 for each state variable and the covariances between state variables could potentially be
23 mis-specified. Therefore, future work should study the implications of mis-specifying the
24 background error statistics within VarBC in a vector system, to understand the impact
25 of mis-specifying both the variances and the covariances. Eyre and Hilton (2013) studied
26 both a scalar and vector system when observation bias correction was not used. They
27 found that an example of the danger zone also existed for a realistic non-scalar case, but

1 the primary cause of the danger zone was due to mis-specifying the variances, rather than
2 the correlations. Therefore, as the analysis error covariance matrix equations for VarBC
3 have a similar structure to the analysis error covariance matrix equation when VarBC
4 is not used, we would hypothesise that, in VarBC, mis-specifying the background error
5 variances would be the primary cause of the danger zone, rather than mis-specifying the
6 background error correlations between the state variables or between the bias coefficient
7 variables.

8

9 In section 3.3 we defined research question 2 as:

- 10 • RQ2: What criteria are needed in the anchor observations to successfully reduce
11 bias in the analysis when both model and observation biases are present but only
12 observation bias is accounted for?

13 In order to tackle this problem, we broke RQ2 down into three sub questions:

- 14 • RQ 2.1: Where are anchor observations most effective in reducing the contamination
15 of bias?
- 16 • RQ 2.2: When are anchor observations most effective in reducing the contamination
17 of bias?
- 18 • RQ 2.3: Does the quality of the anchor observations matter in reducing the contam-
19 ination of bias?

20 In chapter 3 (section 3.2) we discussed how uncorrected biases can contaminate the data
21 assimilation system, as well as contaminating the estimate of the bias correction. If both
22 model and observation biases are present, then Eyre (2016) showed in a 3-dimensional,
23 scalar system that anchor (unbiased) observations are vital in preventing the observation
24 bias correction from drifting towards the model bias.

1 In chapter 6 we extended Eyre (2016)'s work to understand the role of anchor obser-
2 vations in a 3-dimensional, non-scalar VarBC system in the presence of model bias (in the
3 form of background bias), in order to answer RQ2. In answer to RQ 2.1, we found that
4 anchor observations were most effective at reducing the contamination of background bias
5 in the observation bias correction when they observed the same states as the bias-corrected
6 observations. Therefore, this work showed that the regions most at risk of contamination
7 from model bias are regions where anchor observations are sparse, such as temperature
8 observations in the upper atmosphere (above the stratosphere) or the lower troposphere in
9 the Southern Hemisphere. However, we found that if the anchor and bias-corrected obser-
10 vations observed different regions or variables, the background error correlations between
11 these states became more important:

- 12 • If the anchor and bias-corrected observations observed states with similar back-
13 ground biases, then large background error correlations successfully allowed the an-
14 chor observations to reduce the contamination of bias in states they did not observe.
- 15 • If the anchor observations observed states with background biases, but the bias-
16 corrected observations did not observe states with background biases, then large
17 background error correlations would detrimentally transfer biased information from
18 the anchor observations to the bias-corrected observations.
- 19 • If the anchor observations did not observe states with background biases, but the
20 bias-corrected observations observed states with background biases, then, regardless
21 of the background error correlations, the anchor observations could not reduce the
22 contamination of bias on the states they did not observe.

23 Operationally, we cannot control whether the background error correlations are large or
24 small, as the correlations will be dependent on the system and the model. Therefore,
25 regions with large background error correlations, such as along a front, will transfer more

1 information about model biases from the anchor observations to the bias-corrected obser-
2 vations, than regions with small background error correlations, for example across vertical
3 layers separated by a temperature inversion that prevents mixing. As we have shown,
4 this can have either a favourable or detrimental impact on the amount of model bias that
5 can contaminate a system, so can be used to derive insight into why VarBC systems may
6 give less accurate results than expected. In answer to RQ 2.3, we found that more precise
7 anchor observations were more able to reduce the contamination of model bias on the ob-
8 servation bias correction, due to the anchor observations being given a greater weighting
9 in the analysis. A greater weighting of anchor observations could be achieved either with
10 individual instruments that are more precise (although this is difficult to achieve in prac-
11 tice), or by increasing the number of anchor observations in a particular area. Increasing
12 the number of anchor observations is more achievable with the continued expansion of
13 radio occultation instruments, as they can give a global coverage of data (although this
14 is limited to temperature measurements in the stratosphere and upper troposphere, or
15 humidity measurements in the lower troposphere).

16 In chapter 7 we extended chapter 6 and Eyre (2016) to understand the role of anchor
17 observations in a 4-dimensional, non-scalar VarBC system in the presence of model bias.
18 In answer to RQ 2.2, we found that anchor observations were most effective at reducing
19 the contamination of model bias in the observation bias correction when they were at the
20 same time step as the bias-corrected observations in the assimilation window. In answer to
21 RQ 2.3, when the anchor and bias-corrected observations were at the same time step, we
22 found that more precise anchor observations were able to reduce the contamination of both
23 background and model biases on the estimate of the state and observation bias coefficient
24 analyses, as we found for background biases in the 3DVar case in chapter 6. If anchor
25 observations were later than the bias-corrected observations, then an additional model bias
26 term would occur in the analysis that came from evolving the state to the time step of the
27 anchor observations. This was somewhat reduced with more precise anchor observations,

1 and in our theoretical analysis we found that the reduction of this bias, when precise
2 anchor observations were available, depended on using a model that had growing errors.
3 If anchor observations were earlier in the assimilation window than the bias-corrected
4 observations, then an additional model bias term would occur in the analysis that came
5 from evolving the state to the time step of the bias-corrected observations. The anchor
6 observations had no control over the additional model bias and the model bias would
7 not reduce, regardless of the precision of the anchor observations. As satellite radiance
8 observations tend to give data across a time window (Lean et al., 2021), it would be safer
9 for anchor observations to be at the end of the window, so that they are either at the same
10 time or later than the bias-corrected observations. Changing the timing of the anchor
11 observations is not always possible operationally, as observations such as radiosondes tend
12 to have set times. However, increasing the number of radio occultation instruments would
13 allow more anchor observations of the stratosphere and upper troposphere to be available
14 across the time window, which would therefore provide more anchor observations at the
15 end of the window. Alternatively, the timing of the assimilation windows could be shifted
16 so that anchor observations such as radiosondes would be forced to the end of the window.

17 We presented the theoretical results of chapter 6 and 7 for a general vector system, but
18 we only demonstrated the numerical results using our simple Lorenz 96 system, which only
19 had one bias coefficient and had observations at every location. Operationally, predictors
20 are used in the bias correction function, which have been chosen to reduce the contamina-
21 tion of model bias (Harris & Kelly, 2001; Cameron & Bell, 2018). Therefore, our numerical
22 experiments should be extended into a more realistic system, where the bias correction
23 function is defined by a number of predictors, as well as having inhomogeneously spaced
24 anchor observations. This would test how the spatial distribution affects the role of anchor
25 observations in reducing the contamination of model bias in a realistic system, especially
26 when there are large regions that are not observed by anchor observations.

27

1 In section 3.3 we defined research question 3 as:

- 2 • RQ3: How important are anchor observations when correcting for model and/or
3 observation bias?

4 In chapter 3 (section 3.2) we discussed a comparison study of the analysis RMSE (root
5 mean square error) when correcting for observation and/or model bias when both biases
6 were present (Lorente-Plazas & Hacker, 2017). Lorente-Plazas and Hacker found that, in
7 the presence of both biases, only correcting for model bias reduced the state RMSE more
8 than only correcting for observation bias, but correcting for both observation and model
9 biases simultaneously had the biggest improvement on the state RMSE. However, these
10 experiments were undertaken without the use of anchor observations, which have been
11 shown to be vital when correcting for observation or model bias in the presence of both
12 biases (Eyre, 2016; Laloyaux et al., 2020a).

13 In chapter 8 we extended Lorente-Plazas and Hacker (2017) by comparing the role of
14 anchor observations when correcting for observation and/or model bias when both biases
15 were present in order to answer RQ3. We found that anchor observations were important
16 when correcting for observation or model bias, but particularly made an impact when
17 correcting for observation bias using VarBC. Our results when using anchor observations
18 conflicted with the results in Lorente-Plazas and Hacker (2017), who did not use anchor
19 observations, as we found that using precise anchor observations in VarBC reduced the
20 state analysis bias more than using precise anchor observations in WC4DVar. Furthermore,
21 we found that using precise anchor observations in VarBC reduced the state analysis
22 bias more than using very imprecise anchor observations when correcting for both biases
23 simultaneously. This is important operationally, as most operational centres use VarBC to
24 correct for observation bias, but few use WC4DVar to correct for model bias (Gustafsson
25 et al., 2018). Therefore, these results show that it is more important to focus on improving
26 the precision of anchor observations (by increasing the precision of individual instruments

1 or increasing the number of instruments), rather than necessarily moving to a WC4DVar
2 system.

3 However, in line with the results in Lorente-Plazas and Hacker (2017), we found in
4 chapter 8 that correcting for both biases simultaneously reduced the state analysis bias
5 more than only correcting for one, for each equivalent case (precision of anchor observa-
6 tions and/or timing of anchor observations). This result was true even when the anchor
7 observations had very low precision, which suggested that correcting for both observation
8 and model biases simultaneously performed very well, without the need for anchor obser-
9 vations, as was the case in Lorente-Plazas and Hacker (2017). This is counter-intuitive
10 as we would expect the system to perform badly if there is no 'truthful' reference infor-
11 mation to anchor the system. However, we believe this may be due to two factors, which
12 would need to be tested further. The first is that, although the observation and model
13 biases had similar spatial structures, their temporal structures were different, as the model
14 bias accumulates across the assimilation window, whereas the observation bias only occurs
15 where there are biased observations. This means that the system is able to disentangle
16 the two biases and therefore correct them separately. The second factor is that the spatial
17 structures of the observation and model biases was known, which means that the system
18 could more easily identify each bias individually.

19 In chapters 7 and 8 we studied the theory behind model bias contaminating VarBC,
20 and observation bias contaminating WC4DVar. However, we did not study the theory
21 of combining VarBC and WC4DVar, due to the complexity of the equations. Therefore,
22 to have more robust conclusions about the effectiveness of correcting for both biases si-
23 multaneously, the numerical experiments should be repeated in a more realistic system,
24 where the anchor observations are more inhomogeneously spaced, and the structure of the
25 biases have been mis-specified. This would test whether using both VarBC and WC4DVar
26 simultaneously really would significantly reduce the state analysis bias, with the anchor
27 observations that we have available.

1 References

- 2 Ao, C. O., Meehan, T., Hajj, G., Mannucci, A., & Beyerle, G. (2003). Lower troposphere
3 refractivity bias in GPS occultation retrievals. *Journal of Geophysical Research:*
4 *Atmospheres*, 108(D18).
- 5 Auligné, T., McNally, A. P., & Dee, D. P. (2007). Adaptive bias correction for satellite
6 data in a numerical weather prediction system. *Quarterly Journal of the Royal*
7 *Meteorological Society: A journal of the atmospheric sciences, applied meteorology*
8 *and physical oceanography*, 133(624), 631–642.
- 9 Bacmeister, J. T. (1993). Mountain-wave drag in the stratosphere and mesosphere inferred
10 from observed winds and a simple mountain-wave parameterization scheme. *Journal*
11 *of Atmospheric Sciences*, 50(3), 377–399.
- 12 Bannister, R. N. (2008a). A review of forecast error covariance statistics in atmospheric
13 variational data assimilation. I: Characteristics and measurements of forecast error
14 covariances. *Quarterly Journal of the Royal Meteorological Society*, 134(637), 1951–
15 1970.
- 16 Baoding, L. (2007, Jul). *Uncertainty theory. 2. ed.* doi: 10.1007/978-3-540-73165-8
- 17 Bennett, A. F., Chua, B. S., & Leslie, L. (1996). Generalized inversion of a global numerical
18 weather prediction model. *Meteorology and Atmospheric Physics*, 60(1), 165–178.
- 19 Bland, J., Gray, S., Methven, J., & Forbes, R. (2021). Characterising extratropical
20 near-tropopause analysis humidity biases and their radiative effects on temperature

1 forecasts. *Quarterly Journal of the Royal Meteorological Society*, 147(741), 3878–
2 3898.

3 Booton, A., Bell, W., & Atkinson, N. (2014). An improved bias correction for SSMIS. In
4 *Proceedings of the international tovs study conference (itsc)* (Vol. 19).

5 Bouttier, F. (1996). Application of Kalman filtering to numerical weather prediction. In
6 *Proceeding 1996 ECMWF seminar on data assimilation and workshop on non-linear*
7 *aspects of data assimilation. ECMWF, Reading, UK* (pp. 61–90).

8 Bouttier, F., & Courtier, P. (2002). Data assimilation concepts and methods March 1999.
9 *Meteorological training course lecture series. ECMWF*, 718, 59.

10 Brajard, J., Carrassi, A., Bocquet, M., & Bertino, L. (2020). Combining data assimilation
11 and machine learning to emulate a dynamical model from sparse and noisy observa-
12 tions: a case study with the Lorenz 96 model. *Journal of Computational Science*,
13 44, 101171.

14 Brayshaw, D. J., Hoskins, B., & Blackburn, M. (2009). The basic ingredients of the
15 North Atlantic storm track. part I: Land–sea contrast and orography. *Journal of the*
16 *Atmospheric Sciences*, 66(9), 2539–2558.

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

19 Cardinali, C. (2009). Monitoring the observation impact on the short-range forecast.
20 *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric*
21 *sciences, applied meteorology and physical oceanography*, 135(638), 239–250.

22 Carminati, F., Migliorini, S., Ingleby, B., Bell, W., Lawrence, H., Newman, S., . . . Smith,
23 A. (2019). Using reference radiosondes to characterise NWP model uncertainty for
24 improved satellite calibration and validation. *Atmospheric Measurement Techniques*,
25 12(1), 83–106.

26 Courtier, P., Thépaut, J.-N., & Hollingsworth, A. (1994). A strategy for operational
27 implementation of 4d-var, using an incremental approach. *Quarterly Journal of the*

- 1 *Royal Meteorological Society*, 120(519), 1367–1387.
- 2 Cucurull, L., Anthes, R., & Tsao, L.-L. (2014). Radio occultation observations as anchor
3 observations in numerical weather prediction models and associated reduction of bias
4 corrections in microwave and infrared satellite observations. *Journal of Atmospheric
5 and Oceanic Technology*, 31(1), 20–32.
- 6 Cucurull, L., Atlas, R., Li, R., Mueller, M., & Hoffman, R. (2018). An observing system
7 simulation experiment with a constellation of radio occultation satellites. *Monthly
8 Weather Review*, 146(12), 4247–4259.
- 9 Dee, D. P. (2004). Variational bias correction of radiance data in the ECMWF system.
10 In *Proceedings of the ECMWF workshop on assimilation of high spectral resolution
11 sounders in NWP, Reading, UK* (Vol. 28, pp. 97–112).
- 12 Dee, D. P. (2005). Bias and data assimilation. *Quarterly Journal of the Royal Meteorolog-
13 ical Society: A journal of the atmospheric sciences, applied meteorology and physical
14 oceanography*, 131(613), 3323–3343.
- 15 Dee, D. P., & Uppala, S. (2009). Variational bias correction of satellite radiance data in the
16 ERA-Interim reanalysis. *Quarterly Journal of the Royal Meteorological Society: A
17 journal of the atmospheric sciences, applied meteorology and physical oceanography*,
18 135(644), 1830–1841.
- 19 Derber, J. C. (1989). A variational continuous assimilation technique. *Monthly weather
20 review*, 117(11), 2437–2446.
- 21 Di Tomaso, E., & Bormann, N. (2011). *Assimilation of ATOVS radiances at ECMWF:
22 First year EUMETSAT fellowship report* (Tech. Rep.). European Centre for
23 Medium-Range Weather Forecasts.
- 24 Dyroff, C., Zahn, A., Christner, E., Forbes, R., Tompkins, A. M., & van Velthoven, P. F.
25 (2015). Comparison of ECMWF analysis and forecast humidity data with CARIBIC
26 upper troposphere and lower stratosphere observations. *Quarterly Journal of the
27 Royal Meteorological Society*, 141(688), 833–844.

- 1 *ECMWF geographical coverage*. (2023). Retrieved from [https://charts.ecmwf.int/](https://charts.ecmwf.int/catalogue/packages/monitoring/products/dcover)
2 [catalogue/packages/monitoring/products/dcover](https://charts.ecmwf.int/catalogue/packages/monitoring/products/dcover) (Accessed on 14th February
3 2023)
- 4 Efron, B. (2013). Bayes' theorem in the 21st century. *Science*, *340*(6137), 1177–1178.
- 5 Elvidge, A. D., Sandu, I., Wedi, N., Vosper, S. B., Zadra, A., Boussetta, S., ... Ujiie,
6 M. (2019). Uncertainty in the representation of orography in weather and climate
7 models and implications for parameterized drag. *Journal of Advances in Modeling*
8 *Earth Systems*, *11*(8), 2567–2585.
- 9 English, S., McNally, T., Bormann, N., Salonen, K., Matricardi, M., Moranyi, A., ... Geer,
10 A. (2013). *Impact of satellite data* (Tech. Rep.). European Centre for Medium-Range
11 Weather Forecasts.
- 12 Eyre, J. (1991). A fast radiative transfer model for satellite sounding systems. *ECMWF*
13 *Tech. Memo 176*.
- 14 Eyre, J. (1997). Variational assimilation of remotely-sensed observations of the atmo-
15 sphere. *Journal of the Meteorological Society of Japan*, *75*(1B), 331–338.
- 16 Eyre, J. (2000). Planet Earth seen from space: Basic concepts. In *Proceedings of the*
17 *ECMWF seminar on the exploitation of the new generation of satellite instruments*
18 *for numerical weather forecasts* (pp. 5–20).
- 19 Eyre, J. (2016). Observation bias correction schemes in data assimilation systems: A
20 theoretical study of some of their properties. *Quarterly Journal of the Royal Mete-*
21 *orological Society*, *142*(699), 2284–2291.
- 22 Eyre, J., Bell, W., Cotton, J., English, S. J., Forsythe, M., Healy, S. B., & Pavelin, E. G.
23 (2022). Assimilation of satellite data in numerical weather prediction. part II: Recent
24 years. *Quarterly Journal of the Royal Meteorological Society*, *148*(743), 521–556.
- 25 Eyre, J., English, S. J., & Forsythe, M. (2020). Assimilation of satellite data in numer-
26 ical weather prediction. Part I: The early years. *Quarterly Journal of the Royal*
27 *Meteorological Society*, *146*(726), 49–68.

- 1 Eyre, J., & Hilton, F. (2013). Sensitivity of analysis error covariance to the mis-
2 specification of background error covariance. *Quarterly Journal of the Royal Me-*
3 *teorological Society*, 139(671), 524–533.
- 4 Eyre, J., & Woolf, H. M. (1988). Transmittance of atmospheric gases in the microwave
5 region: a fast model. *Applied Optics*, 27(15), 3244–3249.
- 6 Fertig, E. J., Harlim, J., & Hunt, B. R. (2007). A comparative study of 4D-VAR and
7 a 4D ensemble Kalman filter: Perfect model simulations with Lorenz-96. *Tellus A:*
8 *Dynamic Meteorology and Oceanography*, 59(1), 96–100.
- 9 Forster, P. M. d. F., & Shine, K. (2002). Assessing the climate impact of trends in
10 stratospheric water vapor. *Geophysical research letters*, 29(6), 10–1.
- 11 Francis, D., Fowler, A., Lawless, A., Eyre, J., & Migliorini, S. (2023). The effective use
12 of anchor observations in variational bias correction in the presence of model bias.
13 *Quarterly Journal of the Royal Meteorological Society*, 149(754), 1789–1809.
- 14 Gleisner, H., Ringer, M. A., & Healy, S. B. (2022). Monitoring global climate change
15 using GNSS radio occultation. *npj Climate and Atmospheric Science*, 5(1), 6.
- 16 Gratton, S., & Tshimanga, J. (2009). An observation-space formulation of variational as-
17 similation using a restricted preconditioned conjugate gradient algorithm. *Quarterly*
18 *Journal of the Royal Meteorological Society: A journal of the atmospheric sciences,*
19 *applied meteorology and physical oceanography*, 135(643), 1573–1585.
- 20 Gustafsson, N., Janjić, T., Schraff, C., Leuenberger, D., Weissmann, M., Reich, H., ...
21 others (2018). Survey of data assimilation methods for convective-scale numerical
22 weather prediction at operational centres. *Quarterly Journal of the Royal Meteoro-*
23 *logical Society*, 144(713), 1218–1256.
- 24 Han, W., & Bormann, N. (2016). *Constrained adaptive bias correction for satellite radiance*
25 *assimilation in the ECMWF 4D-Var system* (Tech. Rep.). European Centre for
26 Medium Range Weather Forecasts Reading, UK.
- 27 Han, W., & McNally, A. P. (2010). The 4D-Var assimilation of ozone-sensitive infrared

- 1 radiances measured by IASI. *Quarterly Journal of the Royal Meteorological Society*,
2 *136*(653), 2025–2037.
- 3 Harnisch, F., Healy, S., Bauer, P., & English, S. (2013). Scaling of GNSS radio occultation
4 impact with observation number using an ensemble of data assimilations. *Monthly*
5 *Weather Review*, *141*(12), 4395–4413.
- 6 Harris, B. A., & Kelly, G. (2001). A satellite radiance-bias correction scheme for data
7 assimilation. *Quarterly Journal of the Royal Meteorological Society*, *127*(574), 1453–
8 1468.
- 9 He, Q., Wang, Z., & He, J. (2016). Bias correction for retrieval of atmospheric parameters
10 from the microwave humidity and temperature sounder onboard the Fengyun-3C
11 satellite. *Atmosphere*, *7*(12), 156.
- 12 Healy, S. B., & Thépaut, J.-N. (2006). Assimilation experiments with CHAMP GPS radio
13 occultation measurements. *Quarterly Journal of the Royal Meteorological Society*,
14 *132*(615), 605–623.
- 15 Houtekamer, P. L., Lefavre, L., Derome, J., Ritchie, H., & Mitchell, H. L. (1996). A sys-
16 tem simulation approach to ensemble prediction. *Monthly Weather Review*, *124*(6),
17 1225–1242.
- 18 Huang, X., & Ding, A. (2021). Aerosol as a critical factor causing forecast biases of
19 air temperature in global numerical weather prediction models. *Science Bulletin*,
20 *66*(18), 1917–1924.
- 21 Ingleby, B. (2001). The statistical structure of forecast errors and its representation in
22 the Met. Office global 3-D variational data assimilation scheme. *Quarterly Journal*
23 *of the Royal Meteorological Society*, *127*(571), 209–231.
- 24 Ingleby, B. (2017). *An assessment of different radiosonde types 2015/2016* (Tech. Rep.).
- 25 Jean ThiéBaux, H., & Morone, L. L. (1990). Short-term systematic errors in global fore-
26 casts: Their estimation and removal. *Tellus A: Dynamic Meteorology and Oceanog-*
27 *raphy*, *42*(2), 209–229.

- 1 Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal*
2 *of Basic Engineering*, *82*(1), 35-45.
- 3 Karimi, A., & Paul, M. R. (2010). Extensive chaos in the Lorenz-96 model. *Chaos: An*
4 *interdisciplinary journal of nonlinear science*, *20*(4), 043105.
- 5 Krüger, K., Schäfler, A., Wirth, M., Weissmann, M., & Craig, G. C. (2022). Vertical struc-
6 ture of the lower-stratospheric moist bias in the ERA5 reanalysis and its connection
7 to mixing processes. *Atmospheric Chemistry and Physics*, *22*(23), 15559–15577.
- 8 Kursinski, E., Hajj, G., Schofield, J., Linfield, R., & Hardy, K. R. (1997). Observing
9 earth’s atmosphere with radio occultation measurements using the global positioning
10 system. *Journal of Geophysical Research: Atmospheres*, *102*(D19), 23429–23465.
- 11 Kwon, I.-H., English, S., Bell, W., Potthast, R., Collard, A., & Ruston, B. (2018). Assess-
12 ment of progress and status of data assimilation in numerical weather prediction.
13 *Bulletin of the American Meteorological Society*, *99*(5), ES75–ES79.
- 14 Laloyaux, P., Bonavita, M., Chrust, M., & Gürol, S. (2020a). Exploring the potential and
15 limitations of weak-constraint 4D-Var. *Quarterly Journal of the Royal Meteorological*
16 *Society*, *146*(733), 4067–4082.
- 17 Laloyaux, P., Bonavita, M., Dahoui, M., Farnan, J., Healy, S., Hólm, E., & Lang, S. T. K.
18 (2020b). Towards an unbiased stratospheric analysis. *Quarterly Journal of the Royal*
19 *Meteorological Society*, *146*(730), 2392–2409.
- 20 Lean, P., Hólm, E., Bonavita, M., Bormann, N., McNally, A., & Järvinen, H. (2021).
21 Continuous data assimilation for global numerical weather prediction. *Quarterly*
22 *Journal of the Royal Meteorological Society*, *147*(734), 273–288.
- 23 Liu, Y., Zhang, L., & Lian, Z. (2018). Conjugate gradient algorithm in the four-
24 dimensional variational data assimilation system in GRAPES. *Journal of Meteo-*
25 *rological Research*, *32*(6), 974–984.
- 26 Lorenc, A. C. (1986). Analysis methods for numerical weather prediction. *Quarterly*
27 *Journal of the Royal Meteorological Society*, *112*(474), 1177–1194.

- 1 Lorenc, A. C., & Marriott, R. T. (2014). Forecast sensitivity to observations in the Met
2 Office global numerical weather prediction system. *Quarterly Journal of the Royal*
3 *Meteorological Society*, *140*(678), 209–224.
- 4 Lorente-Plazas, R., & Hacker, J. P. (2017). Observation and model bias estimation in
5 the presence of either or both sources of error. *Monthly Weather Review*, *145*(7),
6 2683–2696.
- 7 Lorenz, E. N. (1996). Predictability: A problem partly solved. In *Proc. seminar on*
8 *predictability* (Vol. 1).
- 9 Lorenz, E. N. (2005). Designing chaotic models. *Journal of the atmospheric sciences*,
10 *62*(5), 1574–1587.
- 11 Lu, T.-T., & Shiou, S.-H. (2002). Inverses of 2×2 block matrices. *Computers & Mathe-*
12 *matics with Applications*, *43*(1-2), 119–129.
- 13 Maycock, A. C., Shine, K. P., & Joshi, M. M. (2011). The temperature response to
14 stratospheric water vapour changes. *Quarterly Journal of the Royal Meteorological*
15 *Society*, *137*(657), 1070–1082.
- 16 *Met Office NRT quality monitoring*. (2023). Retrieved from [https://nwp-saf.eumetsat](https://nwp-saf.eumetsat.int/site/monitoring/nrt-monitoring/met-office-nrt/)
17 [.int/site/monitoring/nrt-monitoring/met-office-nrt/](https://nwp-saf.eumetsat.int/site/monitoring/nrt-monitoring/met-office-nrt/) (Accessed on 26th
18 May 2023)
- 19 Milan, M., Macpherson, B., Tubbs, R., Dow, G., Inverarity, G., Mittermaier, M., ...
20 others (2020). Hourly 4D-Var in the Met Office UKV operational forecast model.
21 *Quarterly Journal of the Royal Meteorological Society*, *146*(728), 1281–1301.
- 22 Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J., Corazza, M., ... Yorke,
23 J. A. (2004). A local ensemble Kalman filter for atmospheric data assimilation.
24 *Tellus A: Dynamic Meteorology and Oceanography*, *56*(5), 415–428.
- 25 Parrish, D. F., & Derber, J. C. (1992). The National Meteorological Center’s spectral
26 statistical-interpolation analysis system. *Monthly Weather Review*, *120*(8), 1747–
27 1763.

- 1 Prill, F., Reinert, D., Rieger, D., & Zängl, G. (2022). ICON tutorial. *ICON*.
- 2 Rutherford, I. D. (1972). Data assimilation by statistical interpolation of forecast error
3 fields. *Journal of Atmospheric Sciences*, *29*(5), 809–815.
- 4 Saha, S. (1992). Response of the NMC MRF model to systematic-error correction within
5 integration. *Monthly weather review*, *120*(2), 345–360.
- 6 Sasaki, Y. (1970). Numerical variational analysis with weak constraint and application to
7 surface analysis of severe storm gust. *Monthly Weather Review*, *98*(12), 899–910.
- 8 Saunders, R., Blackmore, T. A., Candy, B., Francis, P. N., & Hewison, T. J. (2013). Moni-
9 toring satellite radiance biases using NWP models. *IEEE transactions on geoscience
10 and remote sensing*, *51*(3), 1124–1138.
- 11 Saunders, R., Hocking, J., Turner, E., Rayer, P., Rundle, D., Brunel, P., . . . others (2018).
12 An update on the RTTOV fast radiative transfer model (currently at version 12).
13 *Geoscientific Model Development*, *11*(7), 2717–2737.
- 14 Saunders, R., Matricardi, M., & Brunel, P. (1999). An improved fast radiative transfer
15 model for assimilation of satellite radiance observations. *Quarterly Journal of the
16 Royal Meteorological Society*, *125*(556), 1407–1425.
- 17 Simmons, A., Mureau, R., & Petroliaqis, T. (1995). Error growth and estimates of
18 predictability from the ECMWF forecasting system. *Quarterly Journal of the Royal
19 Meteorological Society*, *121*(527), 1739–1771.
- 20 Simonin, D., Ballard, S. P., & Li, Z. (2014). Doppler radar radial wind assimilation
21 using an hourly cycling 3D-Var with a 1.5 km resolution version of the Met Office
22 unified model for nowcasting. *Quarterly Journal of the Royal Meteorological Society*,
23 *140*(684), 2298–2314.
- 24 Slingo, J., & Palmer, T. (2011). Uncertainty in weather and climate prediction. *Philosoph-
25 ical Transactions of the Royal Society A: Mathematical, Physical and Engineering
26 Sciences*, *369*(1956), 4751–4767.
- 27 Staniforth, A., & Wood, N. (2008). Aspects of the dynamical core of a nonhydrostatic,

- 1 deep-atmosphere, unified weather and climate-prediction model. *Journal of Compu-*
2 *tational Physics*, 227(7), 3445–3464.
- 3 Stenke, A., Grewe, V., & Ponater, M. (2008). Lagrangian transport of water vapor
4 and cloud water in the ECHAM4 GCM and its impact on the cold bias. *Climate*
5 *dynamics*, 31(5), 491–506.
- 6 Sun, B., Reale, A., Schroeder, S., Seidel, D. J., & Ballish, B. (2013). Toward improved cor-
7 rections for radiation-induced biases in radiosonde temperature observations. *Jour-*
8 *nal of geophysical research: Atmospheres*, 118(10), 4231–4243.
- 9 Thépaut, J.-N., Hoffman, R. N., & Courtier, P. (1993). Interactions of dynamics and ob-
10 servations in a four-dimensional variational assimilation. *Monthly Weather Review*,
11 121(12), 3393–3414.
- 12 Van Leeuwen, P. J., & Evensen, G. (1996). Data assimilation and inverse methods in
13 terms of a probabilistic formulation. *Monthly weather review*, 124(12), 2898–2913.
- 14 Vidard, P. A., Piacentini, A., & Le Dimet, F.-X. (2004). Variational data analysis with
15 control of the forecast bias. *Tellus A: Dynamic Meteorology and Oceanography*,
16 56(3), 177–188.
- 17 Waller, J. A., Ballard, S. P., Dance, S. L., Kelly, G., Nichols, N. K., & Simonin, D.
18 (2016). Diagnosing horizontal and inter-channel observation error correlations for
19 SEVIRI observations using observation-minus-background and observation-minus-
20 analysis statistics. *Remote sensing*, 8(7), 581.
- 21 Walters, D., Boutle, I., Brooks, M., Melvin, T., Stratton, R., Vosper, S., ... others
22 (2017). The Met Office unified model global atmosphere 6.0/6.1 and JULES global
23 land 6.0/6.1 configurations. *Geoscientific Model Development*, 10(4), 1487–1520.
- 24 Wark, D. Q. (1993). Adjustment of TIROS Operational Vertical Sounder data to a vertical
25 view.
- 26 Watts, P., & McNally, A. (1988). The sensitivity of a minimum variance retrieval scheme
27 to the value of its principal parameters. In *Proc. fourth int. TOVS study conf* (pp.

1 399–412).

2 Wergen, W. (1992). The effect of model errors in variational assimilation. *Tellus A*, 44(4),
3 297–313.

4 Zupanski, M. (1993). Regional four-dimensional variational data assimilation in a quasi-
5 operational forecasting environment. *Monthly weather review*, 121(8), 2396–2408.