UNIVERSITY OF READING

Department of Meteorology

Extratropical Lowermost Stratospheric Biases: Characteristics, Causes and Consequences

Jake Bland

A thesis submitted for the degree of Doctor of Philosophy

July 2023

Declaration

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

Jake Bland

Abstract

Predicting the weather accurately at increasingly long lead times is incredibly useful, and relies on numerical models realistically representing the atmosphere. Here the character, cause and consequence of systematic biases in humidity and temperature are investigated as improved knowledge of these biases can help ameliorate them.

Through comparison to in-situ observations, the vertical structures of an analysis moist bias and forecast cold bias in the extratropical lowermost stratosphere are determined. The moist bias has a maximum of around 170% of observed values 1km above the tropopause, and is the dominant cause of the cold bias through additional longwave cooling. Correction of the initial state is insufficient to correct forecasts, as the moisture bias is reintroduced with half-life $\approx 8 - 9$ days through a combination of vertical diffusion and advection. Despite the importance of diffusion and advection the re-moistening has negligible dependence on horizontal resolution, suggesting poor resolution of processes in the vertical is responsible.

Through comparison of corrected ensemble forecasts to a control, it is shown that a bias correction to the moisture field seen by the radiation scheme leads to significant skill increases at around three weeks, particularly over the North Atlantic and Europe, and a reduction in seasonal average error globally. The proposed mechanism for these results is via improvement of the representation of the tropopause, and consequently Rossby wave propagation along it and the jet stream.

When initially corrected, the moist bias is reintroduced by the model on a shorter timescale than that at which we see significant forecast improvements; therefore it is necessary to correct the bias both in the model and in the analysis. Modern radiosondes have value above the tropopause if their data can be assimilated. To reduce forecast error, further work is required to improve vertical diffusion and advection, through increased vertical resolution or otherwise.

Acknowledgements

A great deal of thanks is due to my supervisors Sue Gray, John Methven and Richard Forbes, for their guidance, support, invaluable insights, endless patience, and taking the time to provide feedback on so many drafts. Thank you.

Thanks also are due to the rest of the Reading Meteorology Department. The perturbation experiment of 2020-2021 in which access to the department, friends and colleagues was heavily restricted demonstrated their enormous value in making progress throughout this long project. Too many people to list: from those who were finishing as I started to those who are starting as I finish, and those who have been here since long before I and will remain long after I have left; I don't know how anyone can make it through a 4 year project without a department like this to support them.

And thanks to Jo, my rock.

Contents

Chapter 1

Introduction

The ability to accurately model the atmosphere is important for short term planning on timescales of days to weeks, longer term planning based on skill that can be obtained from seasonal forecasts, and understanding the changing climate on timescales of years. While there are many forecast models which perform very well, there is always a desire for further improvement. One of the most prominent systematic biases in many numerical weather prediction models presently is a cold bias of up to $10K$ in the extratropical lowermost stratosphere.

It has been shown that errors in extratropical lowermost stratospheric water vapour and temperature can have impacts on aspects of the global circulation such as the subtropical jets, storm tracks, Brewer-Dobson circulation and Hadley cells (e.g. Tandon *et al.* (2011), Maycock *et al.* (2013)). Such wide reaching impacts are to be expected, as changes to the structure of the tropopause will impact wave propagation along the interface and zonal wind speed, which will have subsequent consequences for the development of surface level weather systems. Furthermore, using aircraft observations the presence of a humidity bias in the extratropical lowermost stratosphere has been identified (e.g. Kunz *et al.* (2014), Dyroff *et al.* (2015), Woiwode *et al.* (2020)). The dominant effect of increased lowermost stratospheric water vapour concentration is an increased emission of radiation (Shine and Myhre 2020) and hence would be expected to result in enhanced cooling (Forster and Shine 1999). Working towards the desired future forecast model improvement, it is necessary to first quantitatively understand the character of these biases, and then to learn more about the reasons that they manifest in the model.

1.1 Aims

The four aims which will be addressed in this thesis are:

- Characterise the spatial and temporal structures of model biases in extratropical Upper Troposphere and Lower Stratosphere (UTLS) temperature, humidity and tropopause altitude
- Quantify the relationship between these biases
- Determine the reasons for the presence of these biases in models
- Quantify the impacts that these biases have on forecast skill, and hence the improvements that could be achieved by removing them

1.2 The extratropical lowermost stratosphere

The main region of focus of this thesis is the extratropical lowermost stratosphere, the area within a few kilometres above the tropopause in the midlatitudes. This lies within the "Middleworld", defined in e.g. Hoskins (1991) as the region in which isentropes intersect the tropopause but not the earth's surface. Holton *et al.* (1995) then refer to the stratospheric portion of the middleword as the "lowermost stratosphere". Transport into this part of the stratosphere comes from the "Overworld" (the stratosphere above where all isentropes lie above the tropopause), isentropic transport and mixing from the subtropical troposphere, and cross-isentropic mixing from the extratropical troposphere below. Convection drives ascent in the tropical troposphere towards the tropopause. Whether it crosses the tropopause isentropically poleward, or vertically into the tropical stratosphere, the maximum specific humidity of air masses will be determined by the temperature minimum at the tropopause and associated saturation vapour pressure. Air in the tropical lower stratosphere moves upwards, polewards and subsequently downwards in the Brewer-Dobson circulation with a timescale of years. In the extratropics a relevant mechanism of transport of air towards the tropopause is Warm Conveyor Belts (WCB), a feature of extratropical cyclones which transport warm, moist air upwards into regions of high tropopause altitude (tropopause ridges), also contributing to the expansion of these regions (e.g. Saffin *et al.* (2021)).

The lowermost stratosphere can further be divided into two strata. The layer dominated by quasi-horizontal transport and mixing from the subtropics is referred to by Bönisch *et al.* (2009) as the "free lowermost stratosphere" with transport timescales of the order of weeks to months. Below this is the extratropical transition layer (ExTL), extratropical tropopause layer or tropopause mixing layer. Using aircraft observations of ozone and carbon monoxide, Brioude *et al.* (2008) identify a mixing layer between 2-6PVU (potential vorticity units) (for reference, the "dynamic tropopause" is often considered to be at 2PVU, and the thermal tropopause often coincides with 3PVU). The PV gradient within the ExTL is much sharper than in the tropopause below or the stratosphere above. Using similar aircraft O_3 and CO measurements Pan *et al.* (2004) find that the width of the ExTL is 2-3km, centred on the thermal tropopause. Furthermore, they show that in the vicinity of the subtropical jet the tropopause layer is thicker due to the enhanced mixing activity. Due to the role of mixing, the timescales in the ExTL are a few days–much shorter than in the "free" lowermost stratosphere above.

The lowest $\approx$ 2km of the stratosphere are characterised by a change in the vertical lapse rate from tropospheric values of ≈ -6K per km to typically around $1 - 2$K per km, before becoming approximately isothermal above this (Gettelman *et al.* 2011). This Tropopause Inversion Layer (TIL) (Birner 2006) is therefore associated with relatively high static stability, and provides a barrier to vertical motion. By simulating baroclinic lifecycles with no initial TIL and no diabatic processes, Erler and Wirth (2011) show that the TIL forms due to wave breaking above anticyclonic anomalies. In the mixing layer there is more water vapour than in the stratosphere above, hence the associated radiative cooling acts to strengthen this static stability maximum. Temperature in the extratropical lowermost stratosphere is largely controlled by long-wave radiative cooling and warming from the absorption of short-wave radiation. Latent heating is not important above the troposphere, and vertical diffusion of temperature plays only a very minor role. Despite this, condensation to cloud plays a not-insignificant role in removing water vapour in the lowermost stratosphere. The oxidation of methane is an important source of water vapour in the stratosphere, but only in the overworld. The most important factor in determining lower stratospheric water vapour, as noted above, is transport. The challenge of identifying and attributing model biases in this region which has interactions with air masses above, below and equatorward is that which will be addressed in this thesis.

1.2.1 Model representations

When in this thesis I refer to "model biases", I am referring to systematic differences between the state of the atmosphere as represented by the model, as compared to the best information we have on the true state of the atmosphere from observations. Forecasts are initialised with a best guess of the state of the atmosphere created by combining a state from a previous forecast and observations in a process called data assimilation. For operational forecasts the assimilation process occurs over a finite time window (4D-Var), and the forecast run forward from this. Reforecasts are initialised from a state at a single time taken from this window after the assimilation process. Throughout this thesis I will be referring to this state representing the model's guess at the state of the atmosphere at a given time following data assimilation as the "analysis" state. As there are many high quality observations for temperature, the temperature in the analysis matches observations very closely. Radiosonde temperature measurements considered in Chapter 2 will have been used in the data assimilation. However, increments to model humidity from radiosondes are not used above the tropopause in data assimilation in the IFS (Ingleby 2017), and in the MetUM they are not used above the 5 PVU surface, or between 2.5–5 PVU if observed humidity values fall outside of the climatological range of 1–3 ppm and relative humidity $< 10\%$ RH, between 100–400 hPa (Ingleby *et al.* 2012). The extent to which humidity observations are assimilated in reanalysis products is covered in more detail in Davis *et al.* (2017). There are several difficulties with assimilating humidity data in the stratosphere, including the lack of available near-real time high quality observations with sufficient vertical resolution and global coverage. Even with radiosonde or aircraft data, there are difficulties allowing humidity increments into the stratosphere due to the sharp humidity gradient at the extratropical tropopause, as small displacements can lead to large differences between observed and background humidities (Bannister *et al.* 2020). A similar problem is found for the large gradients associated with boundary layer inversions where vertical positional errors can lead to degradation in the analysis (Fowler *et al.* 2012). The sparsity of observations also presents difficulty for assimilation, particularly again due to the sharpness of the humidity gradient across the tropopause. Furthermore, depending on the humidity variable used, allowing humidity increments in the absence of observations can lead to a moistening of the lower stratosphere as the assimilation corrects for the cool bias, due to the correlation between temperature and humidity (Dee and Da Silva 2003). It is for these reasons that the humdidity in the analysis can differ from that in observations made by radiosondes, aircraft and satellites.

Reanalysis differs from analysis in that it is not created in near-real-time with the operational model version of the time, but is often produced for a large time period at once using a single consistent model version for the whole period. A reanalysis may also make use of observations which are not available in real-time for the analysis data assimilation. A variety of reanalysis products are compared to a multi-instrument mean (MIM) from satellite limb sounders as part of the Stratosphere-troposphere Processes and their Role in Climate Reanalysis Intercomparison Project by Davis *et al.* (2017). Though the extratropical lowermost stratosphere is not the focus of Davis *et al.* (2017) all of the considered reanalyses are shown to be moister than the MIM in this region, and ERA-I (ECMWF Reanalysis Interim) performs very similarly to the other considered reanalysis products . Such an extratropical lowermost stratospheric moist bias is also shown to be present in CMIP6 and CCMI-2022 multi model means as compared to the SWOOSH satellite dataset by Charlesworth *et al.* (2023). As identified above temperature is well constrained in analyses and reanalyses, however once forecasts are no longer constrained by observations biases in temperature are free to develop. This is increasingly apparent at longer timescales as Lawrence *et al.* (2022) show, illustrating that across 16 Seasonal to sub-seasonal (S2S) models means over both high-top ("model tops at or above 0.1 hPa with several levels above 1 hPa") and low-top models show a cold bias in the extratropical lowermost stratosphere. Considering climatological mean biases over 20 years, Butchart *et al.* (2011) similarly show all considered models having some degree of cold bias at around 300 hPa. ECHAM (ECMWF Hamburg) models are shown to perform very similarly to most other climate models considered, whereas the UM versions have a slightly smaller than typical cold bias and develop a warm bias at higher altitudes towards 100hPa. This divergent behaviour will be considered further in Chapter 2. Through consideration of the radiation physics one would expect an excess of water vapour in the extratropical lowermost stratosphere to lead to excessive cooling, which will be covered in more detail in Section 2.1. In the absence of assimilation increments the reason for the moist bias could be expected to be within the model configuration. The range of possible misrepresented processes are discussed in Section 3.1. Given this knowledge, in the next section we will explore what is not known, and the gaps which I hope to fill over the course of this thesis.

1.3 Knowledge Gaps

Temperature in the analysis is well constrained, and therefore largely unbiased. It is known that in the free-running model there is a cold bias in the extratropical lowermost stratosphere (Polichtchouk *et al.* 2021). To better understand the development of this bias, the controlling factors and its impacts on medium- and extended- range forecasts, it is necessary to learn about the growth of this bias from the initial time, and how this relates to the collocated moist bias, which is present in the initial conditions. Not enough is currently known about the structures of the temperature and humidity biases, both in the vertical and how they vary spatially, therefore in order to study the relationships between the biases we must investigate their character in more detail. Additionally, as the questions pertain to the investigation of biases of the lowermost stratosphere, a further knowledge gap which must be addressed is the mechanism for the relationship between the development of tropopause-level temperature biases and tropopause altitude, and the structure of the TIL.

That there exists a moist bias in the extratropical lowermost stratosphere is known, and that it is present in both analysis and reanalysis products (e.g. Kunz *et al.* (2014), Dyroff *et al.* (2015)). As humidity increments are not applied during data assimilation in the stratosphere (Ingleby *et al.* 2012, Ingleby 2017)) one question to which the answer is not known is the extent to which the moist bias (and by extension the cold bias) could be improved in forecasts by an improved representation of water vapour in the initial conditions. Furthermore, that the model background state is biased in the absence of humidity increments would further suggest that there is an error in the model representation of the behaviour of the atmosphere resulting in a biased equilibrium state. There have been studies looking into plausible aspects, the misrepresentations of which would lead to such a moist bias. Examples of these include overshooting convection (Jensen *et al.* 2020), overly diffusive dynamics (Stenke *et al.* 2008), too warm a tropical tropopause cold-point temperature (Hardiman *et al.* 2015), which would act on different timescales and affect different spatial regions. Improvements have been made in several areas, yet the moisture bias remains, with definitive cause for and solution of this unknown.

The overarching motivation for the first two aims is to drive model improvement which will result in improvements in forecast skill. While it is understood that extratropical lowermost stratosphere cold and moist biases result in larger scale changes to the atmospheric circulation such as poleward shifted and strengthened jets (Maycock *et al.* 2013), and changes to the Hadley circulation (Tandon *et al.* 2011, Boljka and Birner 2022), there is still considerable uncertainty in the importance of correcting these biases. In qualitative terms, it is uncertain as to the sign of consequent change to the Hadley cell strength. More than this though, what is lacking is knowledge of the quantitative impacts on forecast skill of the specific biases which have been diagnosed to be present in models.

1.4 Thesis Overview

The first and second aims are addressed in Chapter 2, along with initial work on the third aim pertaining to the causes of the presence of the temperature bias. Here we take advantage of a dataset of collocated temperature and humidity observations available as a result of the NAWDEX field campaign (Schäfler *et al.* 2018) to characterise the vertical structures of the temperature and humidity biases within the upper troposphere and lower stratosphere (UTLS) and how these vary within the two-month period of September to October of 2016. A simplified single column model is also used to identify the rate of additional cooling that would be expected as a result of the observed moist bias. This Chapter is adapted from the published article "Characterising extratropical near-tropopause analysis humidity biases and their radiative effects on temperature forecasts" in the Quarterly Journal of the Royal Meteorological Society, Volume 147, Issue 741, August 2021 Bland *et al.* (2021). The main differences between this chapter and the published article are the movement of the discussion of humidity assimilation from Section 2.2.2 to Section 1.2.1, and the addition of Figure 2.3. My contributions to this paper were conceptualisation, analysis, investigation, original draft, review and editing. The contributions of the other authors to this paper were conceptualisation, supervision, review and editing. I performed the research, created the figures, wrote the article and had control of the submitted paper. The study was designed in collaboration with my supervisors, there was regular discussion around the results between all co-authors, and other authors gave advice on interpretation of findings, article structure and edited the text of the paper along with myself. Overall approximately 80% of the article was my work.

Chapter 3 continues to address the third aim, comparing hindcasts initialised with a corrected initial specific humidity field to control forecasts to determine the resultant difference in lowermost stratospheric temperature as compared to the analysis, the timescale on which the specific humidity changes following being corrected at the initial time, and the physical processes responsible for any differences in evolution between the experiment and control temperature and humidity fields. This Chapter is adapted from an article which will shortly be submitted to the Quarterly Journal of the Royal Meteorological Society as "Processes controlling extratropical near-tropopause humidity and temperature in medium-range forecasts". The distribution of contributions to this article is very similar to that for the former. The changes I made to the IFS code were built upon code originally written by co-author Richard Forbes. Again, overall approximately 80% of the article was my work.

From Chapter 3 we obtain the knowledge that the moist bias is reintroduced on a timescale of days-weeks when simply removed from the initial conditions, and that the primary impact of the moisture bias is on the temperature, with cloud processes being largely independent of lowermost stratospheric water vapour. Given this, for Chapter 4 simulations are run modifying only the temperature seen by the radiation scheme, to address the fourth aim evaluating the performance of a 25+ year set of corrected ensemble forecasts against an unperturbed control for a series of skill and error metrics applied to a range of prognostic variables.

Chapter 2

Characterising extratropical near-tropopause analysis humidity biases and their radiative effects on temperature forecasts

Abstract

A cold bias in the extratropical lowermost stratosphere in forecasts is one of the most prominent systematic temperature errors in numerical weather prediction models. Hypothesized causes of this bias include radiative effects from a collocated moist bias in model analyses. Such biases would be expected to affect extratropical dynamics and result in the misrepresentation of wave propagation at tropopause level. Here the extent to which these humidity and temperature biases are connected is quantified. Observations from radiosondes are compared to operational analyses and forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System (IFS) and Met Office Unified Model (MetUM) to determine the magnitude and vertical structure of these biases. Both operational models over-estimate lowermost stratospheric specific humidity, with a maximum moist bias around 1 km above the tropopause where humidities are around 170% of the observed values on average. This moist bias is already present in the initial conditions and changes little in forecasts over the first five days. Though temperatures are represented well in the analyses, the IFS forecasts anomalously

cool in the lower stratosphere, relative to verifying radiosonde observations, by 0.2 K day^{-1}. The IFS single column model is used to show this temperature change can be attributed to increased long-wave radiative cooling due to the lowermost stratospheric moist bias in the initial conditions. However, the MetUM temperature biases cannot be entirely attributed to the moist bias, and another significant factor must be present. These results highlight the importance of improving the humidity analysis to reduce the extratropical lowermost stratospheric cold bias in forecast models and the need to understand and mitigate the causes of the moist bias in these models.

2.1 Introduction

The representation of specific humidity near the tropopause in numerical models has been shown to be important for the accuracy of medium-range forecasts and climate integrations. Modelling studies have demonstrated that both stratospheric and tropospheric temperatures in climate models are sensitive to stratospheric water vapour (Smith *et al.* 2001, Solomon *et al.* 2010). Many of these studies have been motivated by an observed trend of increasing stratospheric water vapour in the late 20$^{\text{th}}$ century, and show that this results in enhanced cooling of the lower stratosphere (Forster and Shine 1999, Maycock *et al.* 2011). If the water vapour concentration in the lowermost stratosphere is increased, this will increase both the emittance of the stratosphere and absorptance of upwelling radiation from the troposphere. In the strong infrared absorption bands of water vapour, the increased emission dominates over the increased absorption (Shine and Myhre 2020). Therefore an increase in water vapour in the lowermost stratosphere would lead to an increase in emission from water vapour, which would lower the lower-stratospheric temperature required for the outgoing radiation to balance the incoming radiation in equilibrium (Maycock *et al.* 2011). By increasing stratospheric water vapour in general circulation models (Joshi *et al.* 2006, Maycock *et al.* 2013) or imposing stratospheric cooling to mimic the temperature response to increased stratospheric water vapour (Tandon *et al.* 2011), it has also been shown that such changes can lead to a poleward shift in the jets and storm tracks, a strengthening of the jets and other changes to the atmospheric circulation in models. The aim of this chapter is to characterise lowermost stratosphere humidity biases in atmospheric analyses and their impact on temperature biases in numerical weather prediction forecasts.

Radiative transfer near the tropopause also plays an active role in maintaining tropopause sharpness with effects on large-scale dynamics. The abrupt drop in specific humidity immediately above the extratropical tropopause results in a peak in long-wave cooling from the troposphere immediately below the tropopause (due to lack of water vapour in the layers above). It has been shown using radiative transfer modelling (Randel *et al.* 2007) and further supported by observational analysis (Hegglin *et al.* 2009) that lower stratospheric water vapour plays an important role in maintaining the region of enhanced static stability immediately above the tropopause, often called the tropopause inversion layer (TIL) (Birner 2006). When considering large-scale dynamics, the tropopause is typically defined as an iso-surface of potential vorticity (PV) and the value of 2 PVU (PV units) is often used in mid-latitudes (Hoskins and James 2014). PV is a measure of rotation and stratification in the atmosphere and the sharp change in static stability at the tropopause is associated with a strong PV gradient and a marked change in wind shear. Since PV is materially conserved for adiabatic, frictionless flows, this definition highlights that the tropopause is approximately a material surface (this cannot be deduced from the temperature profile alone). The long-wave cooling from the tropopause level results in a dipole of diabatically-generated PV that is positive above the tropopause and negative below, and so acts to sharpen the PV gradient (Forster and Wirth 2000, Chagnon *et al.* 2013, Saffin *et al.* 2017) - an alternative description for the formation of the TIL. Additional water vapour in the lowermost stratosphere is expected to weaken the diabatic PV dipole and hence tend to reduce the PV contrast across the tropopause zone. Gray *et al.* (2014) found evidence for a marked decrease in PV gradient with lead time in global forecasts (from several operational centres) although they did not quantify the processes contributing to this decline. The unrealistic decline in PV gradient in forecasts has ramifications for Rossby waves propagating along the tropopause. Theoretical considerations have shown that smoothing PV gradients reduces Rossby wave phase speed (Harvey *et al.* 2016) and is expected to reduce the amplitude of large-scale jet meanders due to excessive PV filamentation and flux of wave activity away from the jet core (Harvey *et al.* 2018). In summary, the representation of the humidity contrast across the tropopause is expected to affect radiative heating profiles, tropopause gradients in temperature and wind, and large-scale dynamics.

As stratospheric water vapour impacts atmospheric radiative balance, its representation in simulations has been evaluated. It has been known for at least 20 years that atmospheric model analyses, re-analyses and forecasts are typically moister than observed in the extratropical lower stratosphere (Pope *et al.* 2001). This bias has been shown through comparisons to many different observational data-sets as summarised in Table 2.1. For the same reasons that a trend of increasing stratospheric water vapour would lead to a cooling of the lower stratosphere, one would expect radiative effects resulting from a moist model bias in the lowermost stratosphere in the analysis to lead to a cold bias in the extratropical lowermost stratosphere in forecasts (Stenke *et al.* 2008, Diamantakis and Flemming 2014, Shepherd *et al.* 2018). Direct measurements of temperature from radiosondes and aircraft as well as indirect measurements from satellite radiances and radio-occultation are assimilated into both models which constrain the temperature in the analyses. Dyroff *et al.* (2015) and Carminati *et al.* (2019) show the mean temperature errors of the ECMWF and MetUM analysis in the extratropical lowermost stratosphere to be within a few tenths of a Kelvin. Although the subsequent growth of a cold bias in the first days of the forecast in this region is seen in operational verification, the author is not aware of documentation of this time-range in the published literature. However, the development of the cold bias in the longer forecast range and climate of each model has been described. For example, Gates *et al.* (1999) showed a cold bias in the lower polar stratosphere in the Atmospheric Model Intercomparison Project (AMIP) ensemble compared to the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-15 reanalysis. Extratropical lower-stratospheric cold biases with respect to ERA-Interim reanalyses of up to 5 K were found more recently in 20-year AMIP-type simulations with the Met Office Unified Model (MetUM) (Hardiman *et al.* 2015, Oh *et al.* 2018) and in multiple 1-year free running simulations with the ECMWF Integrated Forecasting System (IFS) (Shepherd *et al.* 2018). The ECMWF IFS was also shown to have a more severe cold bias in forecasts in the summer hemisphere than the winter hemisphere, which may be related to the larger moist bias in ECMWF analyses in summer than in winter found by Dyroff *et al.* (2015) through the additional radiative cooling this would cause.

The aims of this study are firstly to identify and characterise humidity and temperature biases in the upper troposphere and lowermost stratosphere (UTLS) in the IFS and MetUM, and secondly, to determine the extent to which these temperature biases can be attributed to the presence of the diagnosed moist bias and explore the mechanism by which the moist bias and temperature biases may be causally related. The first aim of characterising any biases is a necessary step to determining their sources in weather forecasts. We use radiosonde observations obtained predominantly over the Eastern North Atlantic and Western Europe for the two month period of September and October 2016 to evaluate differences in specific humidity and temperature between observations and IFS and MetUM

Paper	Observation Type	Model/Analysis	Detail of humidity bias
Stenke *et al.* (2008)	UARS HALOE (Satellite)	ECHAM4.L39	Model moist bias by factor of 3–5 compared to observations in the extratropical lowermost stratosphere.
Van Thien *et al.* (2010)	Aura MLS (Satellite)	GFS and NAM model	Both models are more moist than observations over North America and nearby ocean regions between 150–250 hPa.
Oikonomou and O'Neill (2006)	MOZAIC (Aircraft)	ERA-40	ERA-40 has up to over 60% higher mixing ratios than observations in the Northern Hemisphere mid- and high-latitude lower stratosphere, as a fraction of the observed values. The wet bias is larger in summer than in winter.
Feist *et al.* (2007)	AMSOS (Aircraft)	ERA-40	ERA-40 is moister than observations in the Northern hemisphere lower stratosphere.
Kunz *et al.* (2014)	FISH (Aircraft)	ERA-Interim and ECMWF operational analyses	Analyses and Reanalyses underestimate high water vapour mixing ratios in the upper troposphere and overestimate low mixing ratios in the lower stratosphere.
Dyroff *et al.* (2015)	CARIBIC (Aircraft)	ECMWF operational analysis and short ($<$ 24 hour) forecasts	ECMWF has a moist bias in the lowermost stratosphere. This moisture excess is largest in summer with a maximum median difference around 2 km above the tropopause with the model humidity $\approx$ 220% of observed values, and smallest in winter with a maximum around 3 km above the tropopause with model humidity $\approx$ 150% of observed values.
Woiwode *et al.* (2020)	GLORIA (Aircraft)	ECMWF operational analysis and short ($<$ 12 hour) forecasts	Systematic moist bias in the polar lowermost stratosphere with model humidities on average $\approx$ 150% of the observed values.

Table 2.1: Summary of model moist biases in the upper troposphere and lower stratosphere found in other studies

operational analyses and forecasts, with a focus on tropopause relative vertical structure and the relationship between biases. The main benefit of radiosonde data over satellite data or in-situ aircraft observations made along flight tracks which is taken advantage of for this study is the high vertical resolution of the observations. This better facilitates the investigation of the vertical structures of any biases in the UTLS, and allows calculation of the observed tropopause altitude for tropopause-relative compositing, and evaluation of the model representation of the tropopause altitude. The collocated temperature and humidity measurements are also important for determining any connection between such biases. Furthermore, as radiosonde humidity observations are not assimilated above the tropopause, the lower-stratospheric humidity observations provide an independent dataset against which to assess analyses. To address the second aim of understanding the relationship between biases, the ECMWF Single Column Model (SCM) is used to investigate the radiative impact of the systematic lowermost stratosphere specific humidity biases in analyses, and to determine what proportion of the systematic temperature biases in forecasts can be attributed to this.

In Section 2.2 I describe the radiosonde data used in this comparison, the two numerical weather prediction models and the SCM. We then outline the methods used for the comparison in Section 2.3. In Section 2.4 the results of the comparison of model data to radiosonde observations are presented, followed by the results from the SCM experiments and a discussion of changes at tropopause level resultant from lowermost stratosphere humidity differences in Section 2.5. The main conclusions are then summarised in Section 2.6.

2.2 Data

2.2.1 Radiosondes

In this study I use data from 3204 radiosondes which were launched from 40 sites indicated in Figure 2.1 over the North Atlantic region (38°N–80°N, 50°W–24°E) in September and October of 2016. Of these, 2602 are of the type Vaisala RS92 (Vaisala 2013) and 602 are of the newer type Vaisala RS41 (Vaisala 2018). From 33 sites radiosondes were typically released twice per day, and from the rest once per day. The radiosonde ascents are mostly operational launches, but additional launches were also made for the NAWDEX (North

Atlantic Waveguide and Downstream Impact Experiment) field campaign: a project with aims of exploring the impact of diabatic processes on the jet stream and midlatitude weather systems. This NAWDEX period had a slightly increased frequency of surface cyclones as compared to climatology, and the dominant weather regime is Scandinavian blocking due to the anticyclone which persisted for much of the first half of October (Schäfler *et al.* 2018). The radiosondes used either reported measurements every 2 s, which corresponds to approximately every 10 m, or at significant levels (Ingleby *et al.* 2016).

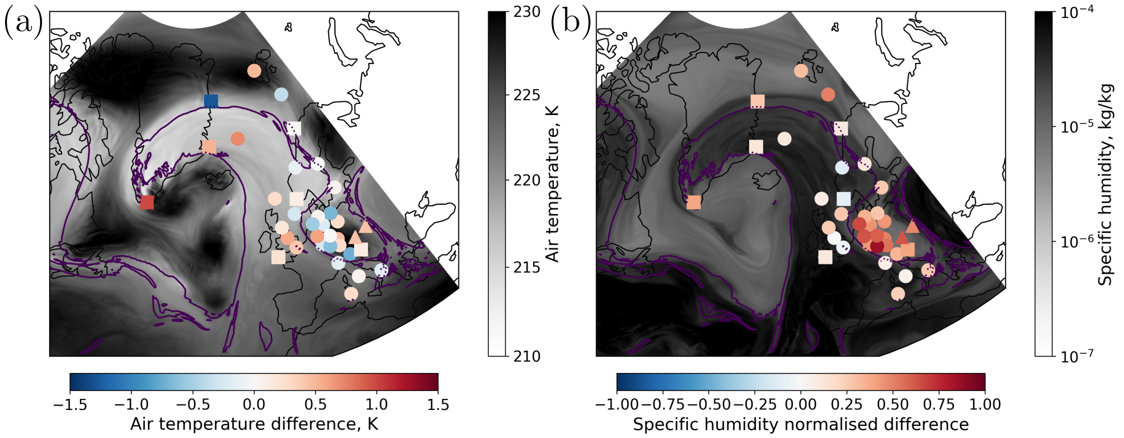

Figure 2.1: Maps of (a) air temperature, (b) specific humidity from IFS operational analyses on 3 October 2016 12 UTC at 250 hPa (greyscale colour) with the location of the dynamical tropopause shown by the 2 PVU contour (purple). Within the coloured shapes are (a) the difference between the IFS temperature and the temperature at locations as measured by radiosondes launched from these locations at the same time and pressure level, (b) the normalised difference of specific humidity between the model and the observations as defined in section 2.3.5. Squares indicate sites using the RS41 radiosonde, circles those using RS92, and triangles those using a combination of the two.

The resolution and total uncertainty of the temperature measurements made by the RS92 radiosondes are 0.1°C and 0.5°C, respectively. The resolution of the relative humidity (RH) data is 1% RH and the total uncertainty is 5% RH for temperatures > -60°C (Vaisala 2013). Total uncertainty here refers to a two standard-deviation confidence level, including repeatability and effects due to measurement conditions, response times and measurement electronics. The RS41 radiosondes have a resolution of 0.01°C and a combined uncertainty of 0.4°C for temperature measurements, and a resolution of 0.1% RH and combined uncertainty of 4% RH for humidity measurements (Vaisala 2017). For both radiosonde types the reproducibility in soundings is 2% RH. These figures are taken from Vaisala datasheets, and further information on the measurement performance can be found in Vaisala (2017). The WMO intercomparison of radiosonde systems (Nash *et al.* 2010) shows that the Vaisala RS92 radiosonde performs well in comparison to a Cryogenic Frostpoint Hygrometer (CFH) for humidity measurements, including in the upper troposphere and lower stratosphere, showing that systematic errors are less than 2% RH. The improvement in performance of the RS92 compared to earlier studies is due to improved sensor coating and correction algorithms for solar radiation and time lag, removing previously found biases (Vaisala 2020, Wang *et al.* 2013). Comparison studies show good agreement in both the temperature and humidity measurements between the RS92 and RS41 radiosondes (Jauhiainen *et al.* 2014). Although the differences in measurements are small, the RS41 demonstrates a better precision and a reduced sensitivity to solar heating (Edwards *et al.* 2014, Jensen *et al.* 2016, Motl 2013). In terms of bias between the two instruments, it can be seen in intercomparison studies (Edwards *et al.* 2014, Dirksen *et al.* 2019, Jauhiainen *et al.* 2014, Vaisala 2014) that the RS92 is $< 1.5\%$ RH drier than the RS41 in the upper troposphere and $< 1\%$ RH moister in the lower stratosphere, which is a very close agreement given the uncertainties of 4% RH and 5% RH for the two sonde types in these measurements.

Although the measured quantity is relative humidity, the humidity variable reported by radiosondes is dew point temperature. The resolution of these measurements is 0.01°C, as for temperature. The humidity quantity that which is mainly considered in this article is specific humidity, the calculation of which from the dew point temperature is detailed in section 2.3. To provide an indication of how large the measurement uncertainties and biases detailed above are as a fraction of the mean specific humidity at a given altitude, the measurement uncertainty of 5% RH as given in the RS92 datasheet corresponds to around 5–10% uncertainty in mean specific humidity in the troposphere. This increases from the top of the tropopause to around 50–100% of the mean values 2 km above the tropopause. In the lowermost stratosphere this measurement uncertainty is much larger than other sources of uncertainty such as uncertainty in temperature and precision of the dew point temperature. The mean relative humidity more than 2 km above the tropopause is below 5% RH, and therefore it is acknowledged that humidities measured at higher altitudes are associated with very large uncertainty. It is also noted in Edwards *et al.* (2014) and Nash *et al.* (2010) that there are diurnal differences in the performance of the RS92 radiosondes. By comparing data from radiosondes released at 12 UTC to those released at 00 UTC in our dataset it is found that the RS41 radiosondes exhibit negligible day/night differences, but that the RS92 radiosondes report slightly higher humidity during the day, with the difference being everywhere less than 3% of the mean specific humidity value in the troposphere, and less than 5% in the stratosphere. Measurement uncertainties in relation to the biases observed are discussed in more detail in section 2.4.1.

As noted above, the radiosonde data from this period are transmitted to the WMO Global Telecommunications System (GTS) in one of two different formats. Twenty six of the sites report the measurement data recorded every 2 seconds, giving a vertical spacing of approximately 10 m throughout the ascent. For the other eighteen sites radiosonde data is sub-sampled and transmitted only for significant levels: a set of mandatory pressure levels (1000, 925, 850, 700, 500, 400, 300, 250, 200, 150 and 100 hPa), in addition to altitudes chosen on each profile where there is a marked change in the gradient of the temperature or humidity. The profile obtained by linear interpolation between significant levels is, by design, very similar to the raw high resolution profile and the significant levels are only used to reduce data transmission from remote sites.

To make the observed data comparable to that from the models in terms of smoothness, a Gaussian kernel smoothing filter is applied to the profiles, of half width 200 m (Harvey *et al.* 2020). This smooths the observed data to a similar resolution to that of the models. The agreement between the altitudes of the tropopause from observations and the models were compared for different smoothing Gaussian half widths. Half-widths of greater than 200 m gave no improvement on the agreement and failed to resolve features of interest, while those smaller than 200 m resolved features too finely, giving an increased median difference in calculated tropopause height between model and the observations (not shown). For the radiosonde data on significant levels, the radiosonde data were first linearly interpolated between observation points to a 10 m grid before smoothing to make it comparable to the radiosonde data from the high-resolution sites. Provided vertical wind shear (particularly perpendicular to any distinct features) is small it is expected that a balloon ascent will measure a good approximation to the conditions in a vertical column of air, being advected horizontally at the same rate as the air masses (MacPherson 1995). It is acknowledged that correcting for the precise timings of the observations and any horizontal drift the radiosonde may experience during the ascent can improve forecasts (Laroche *et al.* 2013, Choi *et al.* 2015), but due to the large increased complexity this would introduce such corrections have not been implemented for this study.

2.2.2 Models

Vertical profiles from the radiosondes are compared to profiles taken from the operational IFS and MetUM forecast models. The ECMWF analysis and forecast data are interpolated to a $0.125° \times 0.125°$ latitude-longitude grid. The operational version in autumn of 2016 of the ECMWF's high resolution atmospheric model was IFS cycle 41r2 (ECMWF 2016) with horizontal resolution TCo1279 ($\sim$9 km grid spacing) (Malardel *et al.* 2016) and 137 vertical levels. The mean vertical model level spacing at the tropopause in the mid-latitudes is approximately 300 m. The Met Office analyses and forecasts from the NAWDEX period were produced using the MetUM version 10.2 in the operational global configuration GA6.1 (Walters *et al.* 2017), with a horizontal resolution of N768 ($\sim$17 km grid spacing) and 70 vertical levels. Data is linearly interpolated in the horizontal to radiosonde release sites. The vertical model level spacing at the tropopause in the mid-latitudes is approximately 550 m (Schäfler *et al.* 2020). Radiosonde data are compared to model data from the nearest six-hourly analysis or forecast. Operational radiosondes are typically launched 45 minutes before their nominal report time, so at typical ascent rates the radiosonde is expected to be close to the tropopause at 00, 06, 12 or 18 UTC.

The IFS single column model (SCM) represents the physical processes in a vertical column for a single grid-point in the horizontal. We use it here to isolate the changes in the response of these physical processes to changes in the initial vertical profiles of variables from effects due to the larger scale dynamics. For the SCM experiments version 43r3 of the IFS is used, which is a later version than used for the full model simulations but which has very similar lower stratospheric temperature errors. The SCM is run with the same 137 vertical levels as the full model. The physical processes included are as detailed in the IFS 43r3 documentation (ECMWF 2017). Further detail on how the SCM was forced for these experiments is provided in section 2.5.1, and more information about the IFS SCM can be found at, for example, Carver (2019).

2.3 Methods

In this section I outline the methods used for the calculation of specific humidity from radiosonde-reported dew-point temperature (so that these values can be compared to the model output), the method used to identify the tropopause and the tropopause-relative coordinates that are used throughout the thesis to calculate the mean properties of the extratropical lowermost stratosphere. It is then explained how I ensure that the tropopause-relative comparisons made are appropriate, and the metric used for the evaluation of the specific humidity biases.

2.3.1 Calculation of specific humidity from radiosonde ascents

Before the smoothing filter is applied to the radiosonde data, specific humidity is calculated from the dew-point temperature reported by the operational radiosonde data processing system. The saturation vapour pressure, e, is first calculated from the dew point temperature using the Sonntag numerical approximation (Sonntag 1990, Sonntag 1994), chosen because it is used in the humidity observation operators in the data assimilation for both the IFS and Met Office (Haiden *et al.* 2016, Ingleby and Edwards 2015):

$$\ln\left(\frac{e}{100}\right) = -\frac{6096.9385}{T_d} + 16.635794 - 2.711193\times10^{-2}T_d + 1.673952\times10^{-5}T_d^2 + 2.433502\ln(T_d),$$

where e is in Pa and T_d is dew-point temperature in Kelvin. The specific humidity, q, in kg kg^{-1} is subsequently calculated as

$$q = \frac{\epsilon e}{\max(p, e) - (1-\epsilon)e},$$

where p is pressure in Pa and ϵ is the ratio between the specific gas constant for dry air and the specific gas constant for water vapour, R_d/R_v, equivalent to the ratio of the molar masses of water vapour and dry air. As q is unitless, for the remainder of the thesis units of q will not be included in the text. The use of the maximum function here is to restrict the maximum value of q to 1 if, for any reason, the partial pressure, e, is calculated as being larger than the air pressure, p.

2.3.2 The thermal tropopause

As we are concerned mainly with the upper troposphere/lower stratosphere (UTLS) region, the altitude of the tropopause also needs calculating. The thermal tropopause is found using the World Meteorological Organisation definition (WMO 1957): “The lowest level at which the lapse rate decreases to 2 K/km or less, provided that the average lapse rate between this level and all higher levels within 2 km does not exceed 2 K/km", with an additional requirement that the mean specific humidity in the 1 km layer above the tropopause should be less than 4×10^{-5}. This latter threshold is applied to reduce the number of cases when the tropopause is found as a lower-level inversion, and is chosen to be sufficiently high that values of specific humidity in the stratosphere are always less than this.

2.3.3 Tropopause-relative coordinates

Once the altitude of the thermal tropopause has been calculated, using the tropopause altitude identified from the radiosonde observations, z_{rtpp}, data from all sources are interpolated to a regular 50-m grid in the shifted height coordinate $(z - z_{rtpp})$. The reference altitude, z_{rtpp}, does not affect the comparison between observation and model temperature and humidity on each profile. It only affects the composite obtained over many profiles due to the shift in each profile to the reference position, z_{rtpp}. The radiosonde-derived tropopause altitude is used because it is common to the comparisons with both the ECMWF IFS and MetUM models. This means that the sharp contrast between troposphere and stratosphere observed on most profiles is reflected in the composites and biases in the stratosphere can be clearly distinguished from those affecting the troposphere.

2.3.4 Tropopause-matching condition

Figure 2.2 shows that on average the models and radiosondes generally agree in the altitude of the tropopause using the WMO lapse rate definition. The median tropopause altitude in the models is higher than that from the observed vertical profiles, by approximately 500 m for the MetUM at all lead times and approximately 200 m in the IFS analysis increasing to 500 m in the five-day forecast. These differences are of a similar order of magnitude to the model grid spacing at these altitudes.

Figure 2.2 also shows that there are several cases where the diagnosed tropopause altitudes are very different between the model and the observations. The large differences in tropopause altitude between the observations and the analysis can be due to tropopause folds, with the lower tropopause being identified for one profile and the higher tropopause being identified for the other. In these situations, provided the model has represented the structure of the fold correctly, in agreement with observations, it is still expected that any

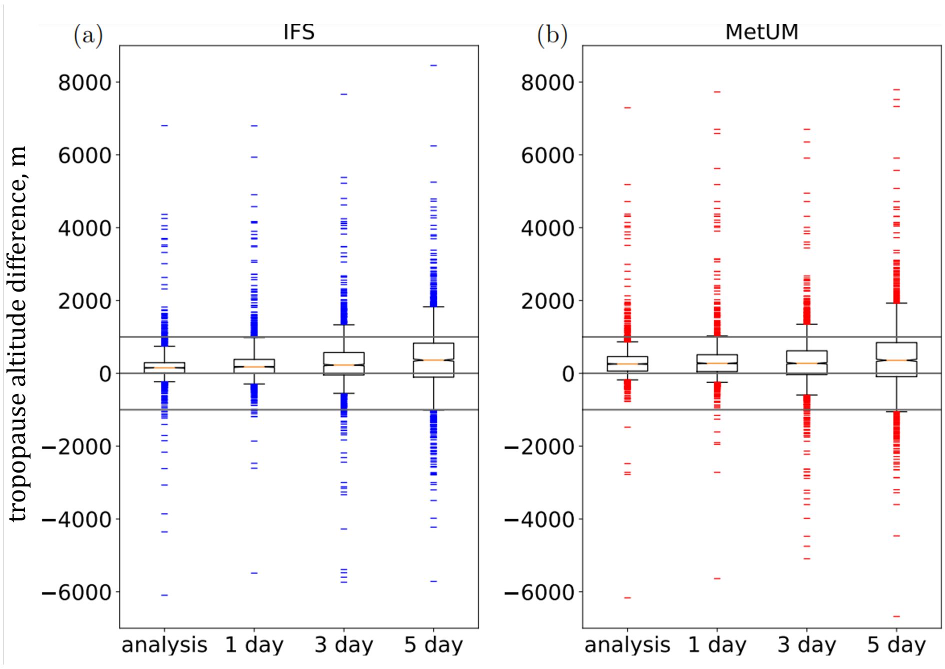

Figure 2.2: Box plots of the differences (model minus observed) in tropopause altitude as calculated from the model data and the radiosonde ascents for (a) IFS and (b) MetUM. Boxes in each panel are from left to right for analysis, 1, 3 and 5 day forecasts. Boxes denote the interquartile range with the central red line indicating the median, the whiskers the 5$^{\mathrm{th}}$ and 95$^{\mathrm{th}}$ percentiles, and individual cases outside this range are shown with blue (IFS) and red (MetUM) markers. Grey lines denote differences of ± 1km.

differences between modelled and observed profiles of temperature or humidity to represent errors in the model. The choice of z_{rtpp} used only affects the reference level used to composite the errors. Another cause of such large differences in tropopause height is feature displacement. This would occur if, for example, there was a difference in the forecast position of a Rossby wave on the tropopause at a given location such that the western side of a ridge was observed by the radiosonde, but the model profile was through the eastern side of the adjacent trough because the model had the wave slightly further east. In these situations we would also expect to see large differences in temperature and humidity for altitudes between those of the two different tropopause altitudes, as one profile would have stratospheric air here and the other tropospheric air.

One example of a large difference between tropopause altitude as calculated from radiosonde data and model data is illustrated in Figure 2.3. Observed and modelled temperatures agree fairly well, as shown by Figure 2.3 (a). However, the slight differences in lapse rate between the two result in the threshold for tropopause identification being met at approximately 8km by the radiosonde profile but not for the model data, with the

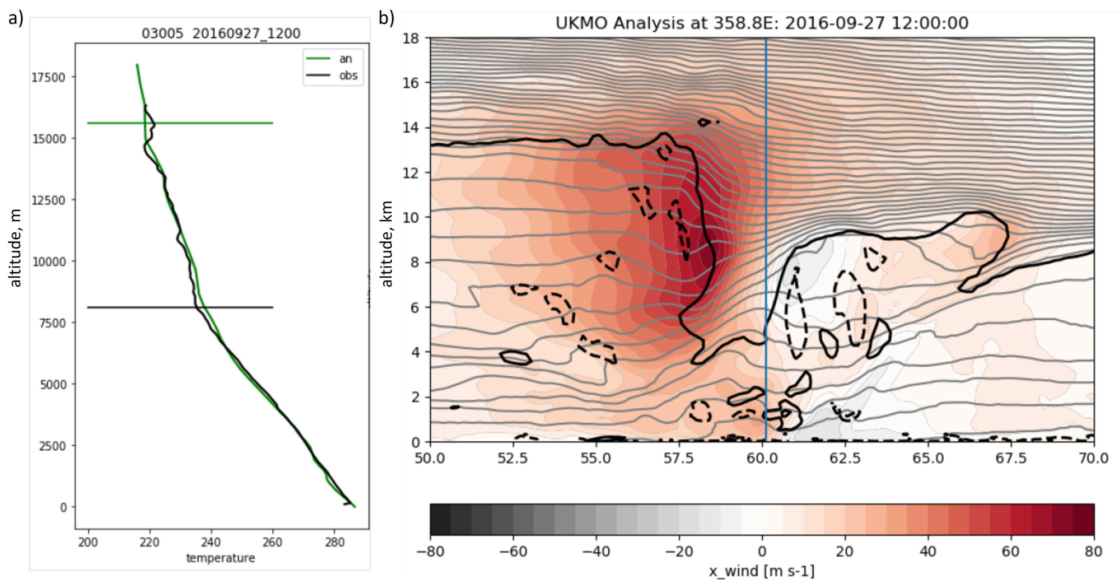

Figure 2.3: (a) Vertical profile of IFS (black) and radiosonde (green) temperature at Lerwick (60.14°N, -1.18°E) on the 27th September at 12UTC. Horizontal lines denote the altitude of the tropopause as calculated from the gradient of each respective temperature profile. (b) Meridional cross-section from the MetUM through -1.12°E on the 27th September at 12UTC. Black contours are PV = 2 (solid) and 0 (dashed), and the shaded background denotes zonal wind speed. The blue vertical line is at the location of the Lerwick radiosonde launch site (Figure 2.3 (b) from Ben Harvey (pers. comm.)).

tropopause for this being identified over 7km higher. It can be seen from Figure 2.3 (b) that at this time and location there is a tropopause fold which is likely to be a major cause of the large tropopause altitude mismatch in this highlighted example.

Such differences would not necessarily indicate an error in the model representation of the field, but rather displacement in the position of a large-scale feature, and so for the purposes of this study these cases will be removed. A cutoff is introduced such that where the difference in tropopause height is greater than 1 km the associated vertical profile is not included in these comparisons. This excludes those scenarios where the model and radiosonde profiles are through different sides of a sharp feature. This cutoff is only used in comparisons of observations to forecasts as it does not make a difference to the results of comparisons with the analysis, and including more cases allows us to produce better statistics. Data assimilation makes large feature displacement in the analysis unlikely. We can see from Figure 2.2 that the number of instances with large disagreement in tropopause altitude increases with forecast lead time, as feature displacement becomes more likely. Returning to the example of the previous paragraph: in Figure 2.3 (a) comparing the radiosonde profile to the model the temperature profiles agree well, as the observational temperature data is assimilated. If it were the case that in a 3- or 5-day forecast the tropopause fold illustrated by Figure 2.3 (b) be displaced a couple of degrees to the north, there would instead be a comparison of stratospheric air (from the radiosonde) to tropospheric air (in the displaced model) above Lerwick in the $\approx$ 7km between the two tropopauses. It is these latter cases which require removal.

2.3.5 Specific humidity normalised difference

The relative magnitude of the difference in specific humidity between model data and observations is shown using the unitless normalised difference between the model humidity and the observed humidity, calculated as

$$\frac{q_{model} - q_{sonde}}{\sqrt{q_{model}^2 + q_{sonde}^2}}.$$

The advantages of normalising the differences by the root sum squared of the modelled and observed values are that it returns a value bounded between ± 1 and is symmetric in magnitude with respect to relative differences of different sign between the model and radiosonde. Calculating the mean of this metric over a collection of ascents therefore does not give increased weight to those ascents with an overestimation of low humidity over those with an underestimation of higher humidity as would occur if normalising by the observations alone. Some fractional differences given in the text are calculated from these mean normalised differences, for ease of interpretation.

2.4 Results of comparing models with radiosonde observations

In this section, first the radiosondes are compared to the model analyses to find the magnitude and structure of any systematic bias that exists in the initial conditions for forecasts, and investigate any dependence this has on the synoptic conditions. We then consider how these biases change over the first five days of the forecast.

2.4.1 Comparison of observations to meteorological analyses

Spatial and temporal variability and structure

We begin by examining the spatial and temporal structure of the differences between the radiosonde and IFS model data. Such comparisons with the MetUM yield similar results, and the mean biases of both models will be considered in later subsections. Figure 2.1 illustrates the positions of the radiosonde sites and gives a representative picture of how model temperature and humidity fields compare to observations on both sides of the tropopause in the UTLS. Figure 2.1(a) shows that the radiosonde observations of air temperature at 250 hPa agree with the background field from the IFS to within ±1.5°C, and there is a mixture of positive and negative differences, with no obvious suggestion of a systematic difference between analysed and observed temperatures. In contrast, in Figure 2.1 (b) for specific humidity, the agreement is good in the troposphere, but the model has consistently higher humidity in the stratosphere.

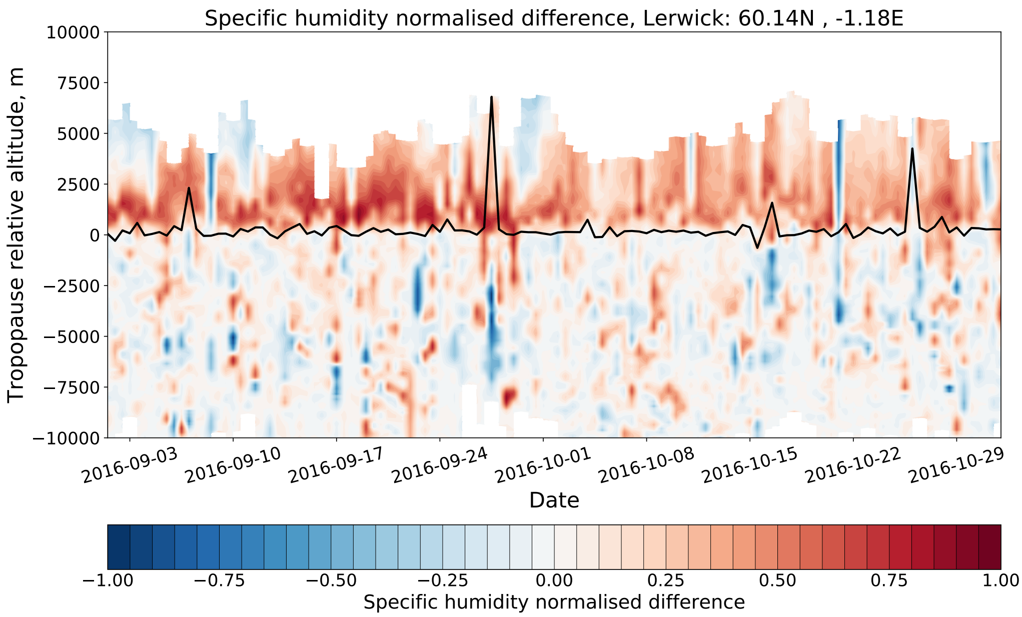

Figure 2.4: Timeseries over the two-month period from the 1 September to the 31 October 2016 at Lerwick (60.14°N, -1.18°E) of the normalised difference between IFS analysis specific humidity and radiosonde observations as a function of altitude (colour shading). Altitude is considered in a tropopause-relative framework, where zero is the altitude of the tropopause as determined by the radiosonde observations, calculated according to the WMO lapse rate definition as discussed in Section 2.3.3. The black contour is the tropopause as determined from the model data.

The situation illustrated in Figure 2.1 of a moist bias in the lower stratospheric regions is representative of the entire two-month observation period considered. This persistence is shown in Figure 2.4 for a single observation site at Lerwick which used the RS92 radiosondes, though the results are consistent across the other sites considered. Below the tropopause there is variability in the specific humidity normalised difference, but no significant bias. In contrast, in the first few kilometres above the tropopause the IFS has a systematic moist bias compared to the observations. The model tropopause is generally within a few hundred metres of that observed, with only a few outliers, for example, on the 27 September the model tropopause is much higher than observed as shown in Figure 2.3 and discussed in the previous section. This sounding would be removed from the data used in the comparisons of the forecasts.

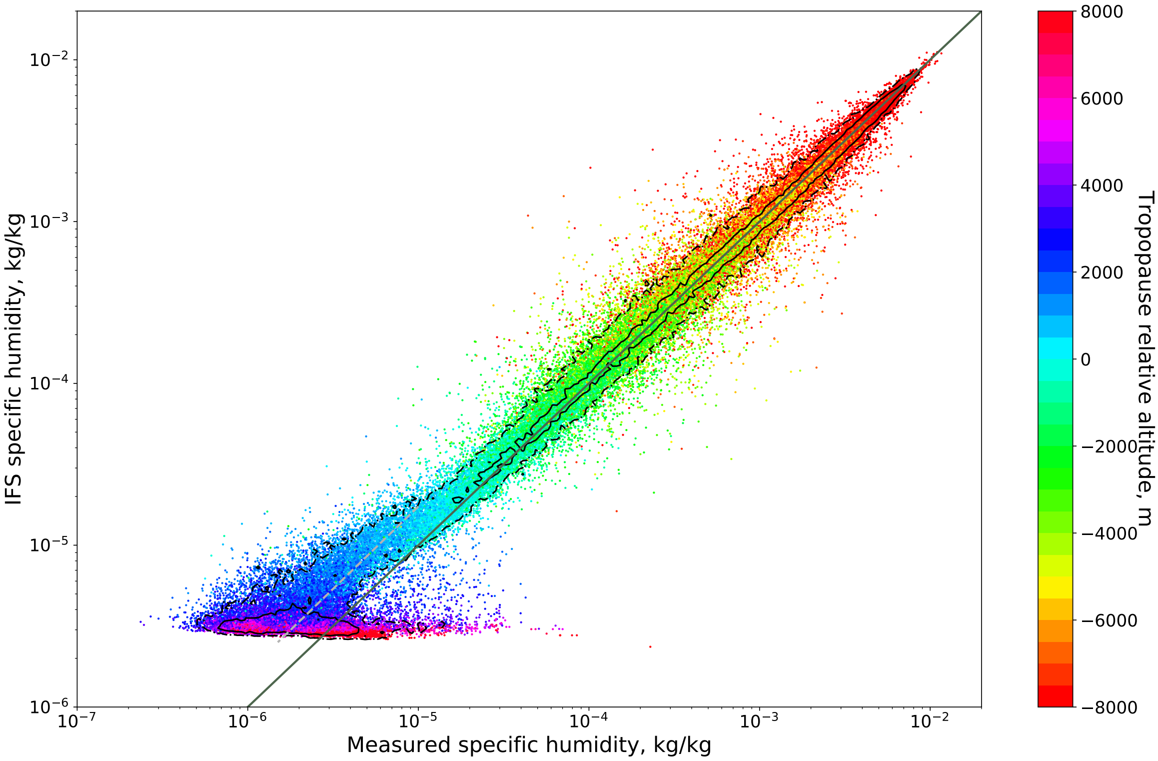

Figure 2.5: Average values of specific humidity in 500m bins as measured by radiosondes (x-axis) against specific humidity at the same locations in IFS analyses (y-axis). Dots are coloured according to their tropopause-relative altitude. The solid grey line indicates a 1:1 relationship, and the dashed light grey line indicates the 0.57:1 relationship found considering a linear regression of points between 1 and 3km above the tropopause. The black contours provide an indication of point density: points are five times more dense within the smaller solid contoured regions containing almost half of the points, than within the dot-dashed larger contour containing almost 90% of the points.

Considering now data from all locations and all times during the two-month period, in the scatter of observed against IFS model humidities (Figure 2.5) we see that in the troposphere for the majority of places and times the humidities in both the model and the observations agree very closely (i.e. follow the 1:1 line). For the lowermost stratosphere, on the other hand, the model has a clear positive humidity bias compared to the observations. Taking those measurements made between 1 and 3 km above the tropopause and performing a linear regression on these we recover a slope of 0.57, indicating that the IFS is 175% as moist as is observed, in the mean in this region. A further notable feature of Figure 2.5 is that values for the specific humidity in the model at altitudes above around 6 km above the tropopause seem to have a minimum at $\approx 3 \times 10^{-6}$. This is not as a result of an artificially-imposed minimum value in the model, as it has been found that the IFS is capable of sustaining lower values of specific humidity than this in forecasts (not shown). Rather, as humidity increments are not applied above the tropopause, stratospheric humidity values will be dominated by the model background state or "model climatology". Further inhibiting the occurrence of very low humidities in the model is the use of a quasi-monotone filter in the horizontal. This is applied following the interpolation in the dynamics, and prevents the humidity in a grid point from being lower than the minimum value from the horizontally adjacent points at the previous time-step.

Mean vertical structure

We now consider the mean vertical structure of the model analysis humidity bias in an Eulerian frame of reference over the North Atlantic in the two month period and in a tropopause-relative altitude coordinate system. For this analysis, data from the MetUM model in addition to that from the IFS are now also used. Additionally, as observations made by radiosondes of two different types with different uncertainty characteristics are being used, here comparisons of the models to these are composited separately. Figure 2.6(a) and (c) shows that analyses from both models represent the specific humidity well in the troposphere, where observations are assimilated, but have mean moist biases in the stratosphere that increase in magnitude from small values at the tropopause to a maximum 1–2 km above the tropopause. There is also a sharp decrease across the tropopause in the observed profiles of specific humidity that is not replicated in the models. This decrease in the observations but not the models is still seen (though slightly less sharply) when compositing profiles relative to the model tropopause altitude (not shown) as opposed to the radiosonde tropopause altitude as shown here. This indicates that the difference between modelled and observed specific humidity across the tropopause is not caused by compositing model profiles with slightly different tropopause altitudes but rather is a robust feature. These panels also give an indication of the data availability as a function of tropopause-relative altitude and show that this drops off with height in the stratosphere. In order to make a comparison relative to the tropopause altitude, all radiosonde profiles necessarily reach this level (ascents which do not cannot be used) and at some altitude above this the radiosonde balloons must burst.

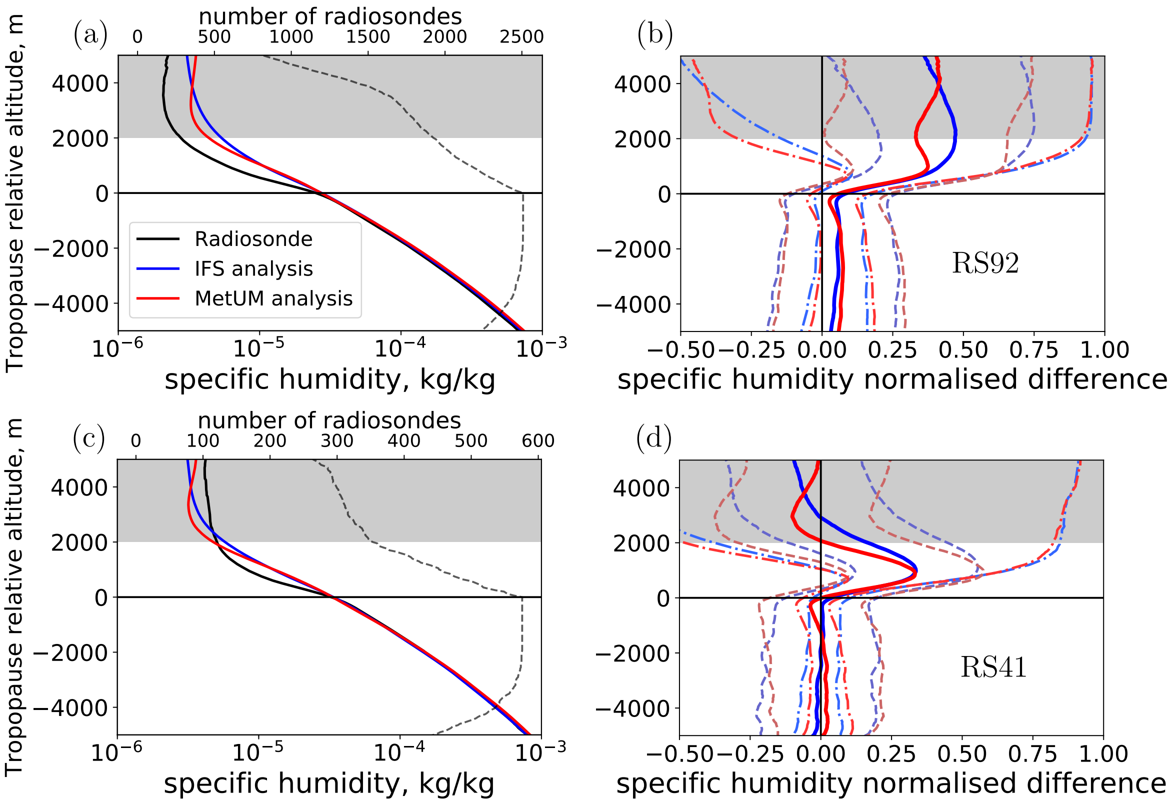

Figure 2.6: Mean tropopause-relative vertical profiles of (a, c) specific humidity and (b, d) the normalised difference of specific humidity. (a, b) show composites over data from sites using the RS92 radiosonde, whereas (c, d) show composites over data from the type RS41. The black lines in panels (a) and (c) are for the radiosonde data, blue lines IFS analyses, and red lines MetUM analyses. The top x-axis is the scale for the number of points for which there are radiosonde data as a function of altitude, and hence the number of points contributing to the means at each tropopause-relative altitude, shown by the dashed grey line. In panels (b) and (d) dot-dashed lines show the measurement uncertainty as described in the text, and the dashed lines show one standard deviation of our sample of profiles of specific humidity normalised differences. Tropopause relative altitudes above 2km are greyed-out to indicate the large uncertainties in measurements of humidity here.

Figure 2.6 illustrates how the measurement uncertainty of the radiosondes as given in section 2.2.1 in terms of relative humidity is related to uncertainty in specific humidity using the following method: for each ascent the relative humidity is calculated, upper and lower bounds on this are found according to the manufacturer-specified combined uncertainties quoted above, and these bounds are then converted back to specific humidity and averaged in the same way as specific humidity. As the mean over thousands of data points is being considered, the standard error of the mean is very small, assuming that systematic errors have been corrected for and the remaining measurement errors are random. The standard deviation shown here therefore does not indicate a lack of confidence in the magnitude of the bias, but rather illustrates the spread. Compared to the RS92 radiosondes in Figure 2.6(b), the analyses at ≈1 km above the tropopause have a maximum mean normalised difference of ≈0.37. Compared to the RS41 radiosondes in Figure 2.6(d), analyses have a maximum mean normalised difference of ≈0.34 at ≈1km above the tropopause. A mean normalised difference of 0.34 means that the mean specific humidity in analyses is 166% of the mean observed value.

In the troposphere the RS41 radiosondes agree with model humidities, while observations from RS92 radiosondes are slightly drier. This is in agreement with the discussion of sensor differences in section 2.2.1. Similarly we might expect measurements from RS92 radiosondes to be slightly moister than from RS41 in the lower stratosphere, resulting in a slightly reduced difference between the RS92 measurements and the model. In the lowermost stratosphere the bias found through comparison to both radiosonde types is very similar, and even considering the uncertainty in the observations the lower limit of the normalised difference still indicates a moist bias 1 km above the tropopause in the analyses from both models. The character of the normalised specific humidity difference is different for the two radiosonde types higher up in the stratosphere. However, as is noted in section 2.2.1, at altitudes above 2 km above the tropopause the instrument uncertainty is large. Furthermore, as the number of measurements at these altitudes is relatively small, and the RS41 and RS92 radiosondes are typically launched from different observing sites, from the available data we are unable to draw any conclusions regarding the comparison at these higher altitudes. Though values in this region are still plotted for consistency with the later consideration of temperature biases, for plots using humidity observations these regions are shaded grey. As the two radiosonde types perform similarly in our region of interest between the tropopause and 2 km above it, data from the two radiosonde types are combined in the remainder of this chapter. It should be noted that as there are around five times more observations from RS92 than RS41 radiosondes, these dominate subsequent statistics.

Meteorological dependence of the vertical structure

To consider how the vertical structures and magnitudes of these biases vary under different atmospheric conditions similar figures to Figure 2.6 can be produced, but instead taking the mean over only profiles that satisfy certain criteria. This partitioning has been done conditioning on profiles that satisfy the following criteria: the presence of a low static stability layer within a certain distance below the tropopause; clouds within a certain distance of the tropopause; whether the vertical profile is taken through a ridge or a trough, and ridge profiles for which the air motion is northward (roughly equivalent to the western region of the ridge). The application of most of these conditions resulted in no notable systematic difference between the composite vertical profiles of humidity (not shown) and are not discussed further here; the exception is when separating profiles in ridges and troughs. Ridges are identified relative to the mean height in ERA-Interim of the tropopause at each radiosonde site over the months of September and October from 2005-2017. The mean ERA-Interim tropopause height is calculated from the height of the 2 PVU surface which, due to the sharp PV gradient between the troposphere and stratosphere, is a convenient identifier of the tropopause in the extratropics. A profile is classified to be in a ridge if the calculated tropopause altitude is larger than one standard deviation above this mean height. The standard deviation used is that of the set of 26 monthly means (from two months in each of 13 years). Troughs are identified similarly, but using one standard deviation below this mean height. Of the 3204 vertical profiles, 1605 of these were classified as ridges, and 772 as troughs.

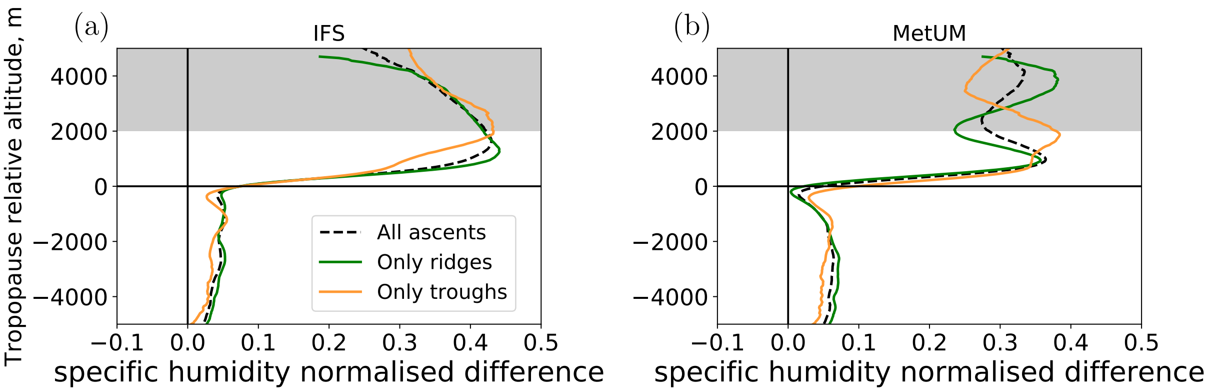

Figure 2.7: Mean tropopause-relative vertical profiles of the normalised difference of specific humidity for (a) IFS and (b) MetUM analyses. Black lines show results over all ascents, green only those in ridges, and orange only those in troughs. Grey shading is as for Figure 2.6.

The humidity biases have similar vertical structures in ridges and troughs, but with a larger vertical length scale for troughs, and smaller for ridges (Figure 2.7). This results in the maximum difference being found at ≈1 km above the tropopause in ridges, but ≈2 km above in troughs. Additionally, although in the IFS the moist biases in the troughs and ridges are of very similar magnitude, there is a more pronounced difference in the MetUM with the troughs having a slightly larger and much deeper systematic bias.

2.4.2 Comparison of forecasts to observations

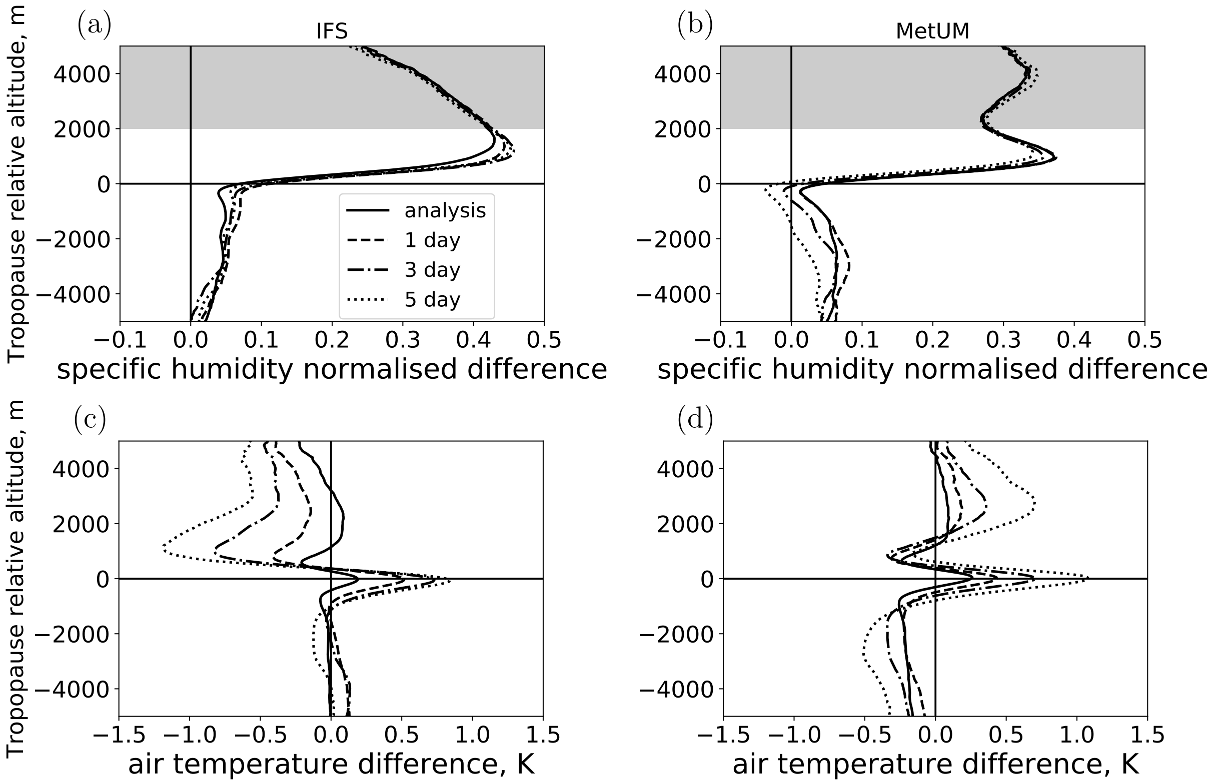

Figure 2.8: Mean tropopause-relative vertical profiles of (a, b) specific humidity normalised difference and (c, d) air temperature difference for (a, c) IFS and (b, d) MetUM. Solid lines denote differences with model analyses and the other lines differences with forecasts: dashed lines one-day, dot-dashed lines three-day, and dotted lines five-day forecasts. Grey shading is as for figure 2.6.

Having considered the comparison of these radiosonde observations to model analyses, I now also compare to forecasts out to five days. The specific humidity bias is large in the analyses, and it can be seen from Figures 2.8(a) and (b) that the mean analysis value of the normalised difference in the lowermost stratosphere is much larger than any changes that occur in the first five days of the forecasts. In the IFS the mean moist bias 1 km above the tropopause increases slightly over the forecast period, whereas in the MetUM it decreases slightly.

Additionally, although the temperature in the model analyses is very similar to that from the radiosonde observations, it diverges from the observations during the first five days of a forecast in both the IFS and the MetUM. Initially in both models there is a slight cold bias in the lowermost stratosphere of $\approx$0.2 K with respect to the radiosondes, and a slight warm bias at the tropopause of similar magnitude. Subsequently, the difference from observations of the lowermost stratosphere temperature evolves differently in the IFS and MetUM (compare Figure 2.8(c) and (d)). For the IFS a cold bias develops in the lowermost stratosphere at a rate of $\approx$0.2 K day^{-1}; the largest bias is initially 600 m above the tropopause, but increases to 1200 m above the tropopause during the five day forecast. In the MetUM, the slight mean cold bias at $\approx$1 km height grows at a slower rate than for the IFS, and a mean warm bias develops at an altitude of $\approx$3 km above the tropopause; this bias develops at an accelerating rate, reaching a magnitude of $\approx$0.7 K after five days. Another feature in Figure 2.8(c) and (d) is a peak in the air temperature difference at the tropopause. This sharp peak at the tropopause is not present when compositing differences relative to the model tropopause (not shown). If tropopause altitudes between model and observations differ the consequence of compositing model temperature profiles relative to the observed tropopause is that the transition from tropospheric to stratospheric lapse rates at the tropopause will be smoothed in the vertical, and the temperature in the narrow region about the tropopause in the model will be slightly increased. There is however instead in a model-tropopause relative comparison a slight warm bias developing at a rate of approximately 0.1 K day^{-1} in a much deeper region below the model tropopause.

It is shown in Saffin *et al.* (2017) that the behaviours of physical processes affecting PV at the tropopause are different in ridges and troughs. Separating the air temperature difference profiles from Figure 2.8(c) and (d) into ridges and troughs yields Figure 2.9. The greater number of profiles through ridges compared to troughs (more than twice as many) contribute to the mean profiles in Figure 2.8(c) and (d) bearing a closer resemblance to those shown for ridges than troughs. For the IFS, the magnitude of the peak temperature difference is very similar in ridges and troughs in the analyses and one- and three-day forecasts ($\approx$ 0.85 K at 1 km above the tropopause). Recall that the specific humidity normalised difference was similar between ridges and troughs for the IFS (Figure 2.7 (a)). However, whereas the maximum value of the normalised difference in specific humidity was at a higher tropopause-relative altitude for troughs than ridges, the altitude of the

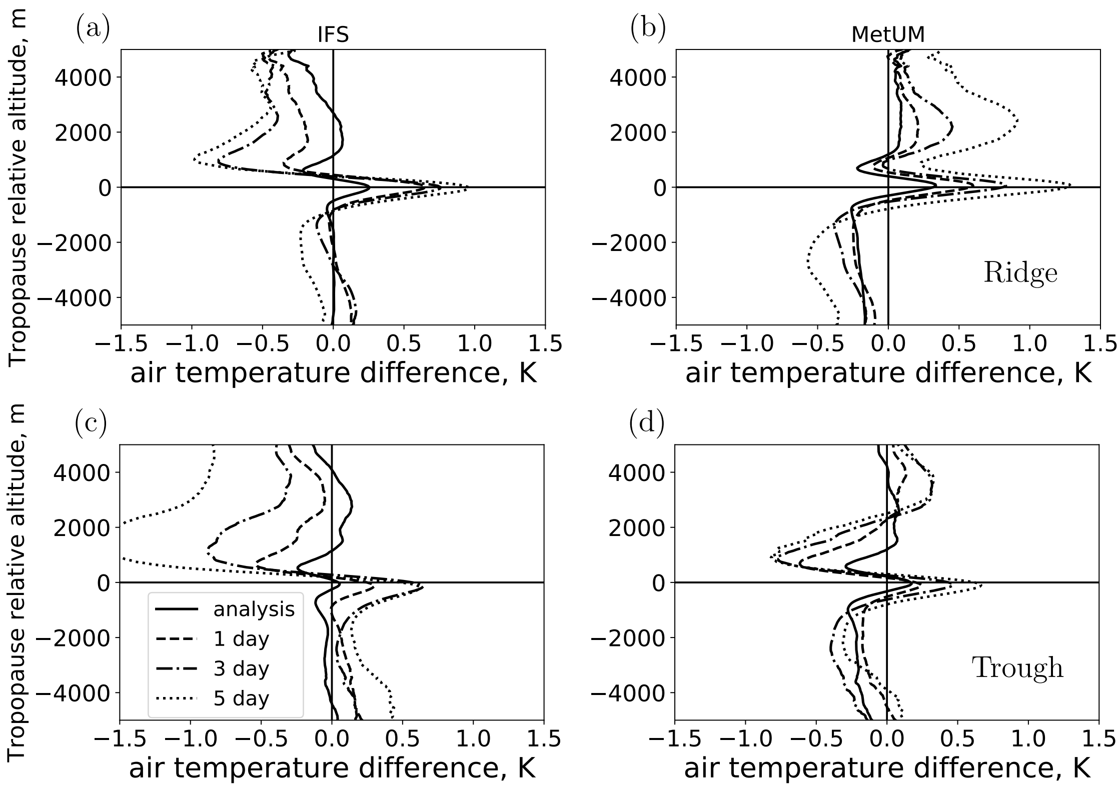

Figure 2.9: As for Figure 2.8 (c, d): mean tropopause-relative vertical profiles of air temperature difference for (a, c) IFS and (b, d) UKMO. Panels (a) and (b) show means over only those ascents in ridges and (c) and (d) only those in troughs.

maximum cold bias is the same in both ridges and troughs, but the cold bias extends further up into the stratosphere in troughs.

In the MetUM there are larger differences when separating profiles into those in ridges and troughs, with a cold bias of around 0.8 K at 1 km above the tropopause increasing over the first three days of the forecast in troughs, as with the IFS, but a warm bias developing with lead time above ridges at 2–3 km above the tropopause.

Now we will consider briefly the sensitivity of these results to the choice of tropopause-matching condition (not shown). Results for humidity are not sensitive to the threshold chosen, nor are results for IFS temperature in the analysis and 1- and 3-day forecasts. There is however an increase in the cold bias from ≈ 0.2 to ≈ 0.3 K day^{-1} in IFS 5 day forecasts with the inclusion of cases where the model tropopause is between 1–1.5 km above that observed. MetUM temperature forecasts are also sensitive to the choice of tropopause matching condition. Regardless of the chosen threshold there is the development of a warm bias 3 km above the tropopause in ridges and a cold bias 1 km above the tropopause in troughs, however in troughs without the condition the mean cold bias is markedly larger at both 3 and 5 days, and in the mean over all profiles when the matching condition is not applied there is a reduced warm bias at 3 km, and there is instead a larger cold bias at 1 km. It is for this reason that we are able to find a systematic tropopause-relative warm bias in the lowermost stratosphere of the MetUM when using this tropopause matching condition, while in climatological comparisons it is known that the MetUM has a cold bias in the extratropical lower stratosphere (Hardiman *et al.* 2015). It is shown in figure 2.2 that there are more cases where the model tropopause is higher than observed excluded by the matching condition than lower, increasing with lead time, and these differences when removing the tropopause matching condition are consistent with this.The relation between the biases and tropopause altitude will be discussed further in Section 2.5.2.

2.5 Attributing temperature biases to the humidity bias

In section 2.1 the mechanism by which a moist bias in the lowermost stratosphere can lead to a collocated cold bias through changes in long-wave radiative transfer was discussed. In Section 2.4, the locations and magnitudes of both a moist bias in the model analyses and the development of a temperature bias in forecasts were identified. In this section I examine what proportion of the forecast temperature biases can be attributed to changes in radiative cooling, as a consequence of the presence of the moist bias in the analyses, and consequent changes in the structure of the tropopause.

2.5.1 Attribution of temperature biases using the IFS Single Column Model

The magnitude of the cold bias attributable to the radiative response to the identified moist bias is quantified using the IFS single column model (SCM). The SCM simulations are used to provide information on the typical heating rates from parametrized physical processes, the contributions of which are output as individual rates of change of temperature, referred to as “process tendencies”. Although in the full 3D forecast models air masses are advected around, such that over a five day forecast they will experience heating or cooling in different regions, the SCM just represents the impact of the physical processes in a column with no advection. Two simulations are run for 9 days initialised with identical temperature profiles, but humidity profiles representative of the IFS analysis state and the observed state respectively. The differences in temperature between these are therefore as a result of the differing initial humidity profiles. The SCM simulation representative of the IFS model state (labeled IFS-prof in Figure 2.10) is initialised with a humidity profile equivalent to the mean over the humidity profiles from the IFS analyses at all radiosonde launch sites and times. This profile is similar to that shown in Figure 2.6(a) and (c), but the mean is taken in a ground-relative sense in pressure coordinates, as opposed to a tropopause-relative sense in altitude coordinates. The SCM is also initialised with a similarly produced temperature profile, zero wind speed, and a surface pressure of one atmosphere: 101325 Pa. The surface boundary condition is zero sensible and latent heat flux. It is acknowledged that although this zero forcing approach provides the most isolated view of the temperature changes, a consequence is that the profile cools continuously as can be seen from Figure 2.10 (c) comparing the initial profile (grey) to the simulated values (black, blue) at 5 days which are much cooler (roughly proportional to the tropopause-altitude-relative mean temperature tendencies from longwave radiation shown later in Figure 3.6 (a)).

The simulation representative of the atmosphere as observed by the radiosondes (labelled Obs-prof in Figure 2.10) is initialised with the same temperature as those used for the IFS simulation, but with the humidity profile changed only in the lowermost stratosphere, as indicated in Figure 2.10(a). In section 2.4 the mean tropopause-relative specific humidity normalised difference over all considered radiosonde ascents are identified, illustrated by the solid line in Figure 2.8(a). The observation-representative humidity profile is created such that the tropopause-relative normalised difference between Obs-prof and IFS-prof between the tropopause and 2 km above is the same as that difference found in section 2.4. This is achieved by removing the appropriate amount from the IFS-prof averaged humidity profile. This layer is chosen to isolate the effect of a moist bias in the lowermost stratosphere where there is the highest confidence in the presence of a moist bias. The number of radiosonde measurements above this level is small and the measurement uncertainty is much greater, so a humidity bias cannot be robustly confirmed above this level. Below the tropopause, and above 3 km above the tropopause, the imposed normalised difference profile between Obs-prof and IFS-prof is set to be zero, and it varies linearly between the mean value at 2 km and zero at 3 km. A linear reduction of the normalised difference in this layer is imposed to prevent any effects arising from a sharp boundary. Such a layer is not needed below the tropopause as the normalised difference here is already approximately zero. Results from SCM simulations initialised with these idealised mean profiles are representative of results from SCM simulations initialised using profiles from individual times and locations.

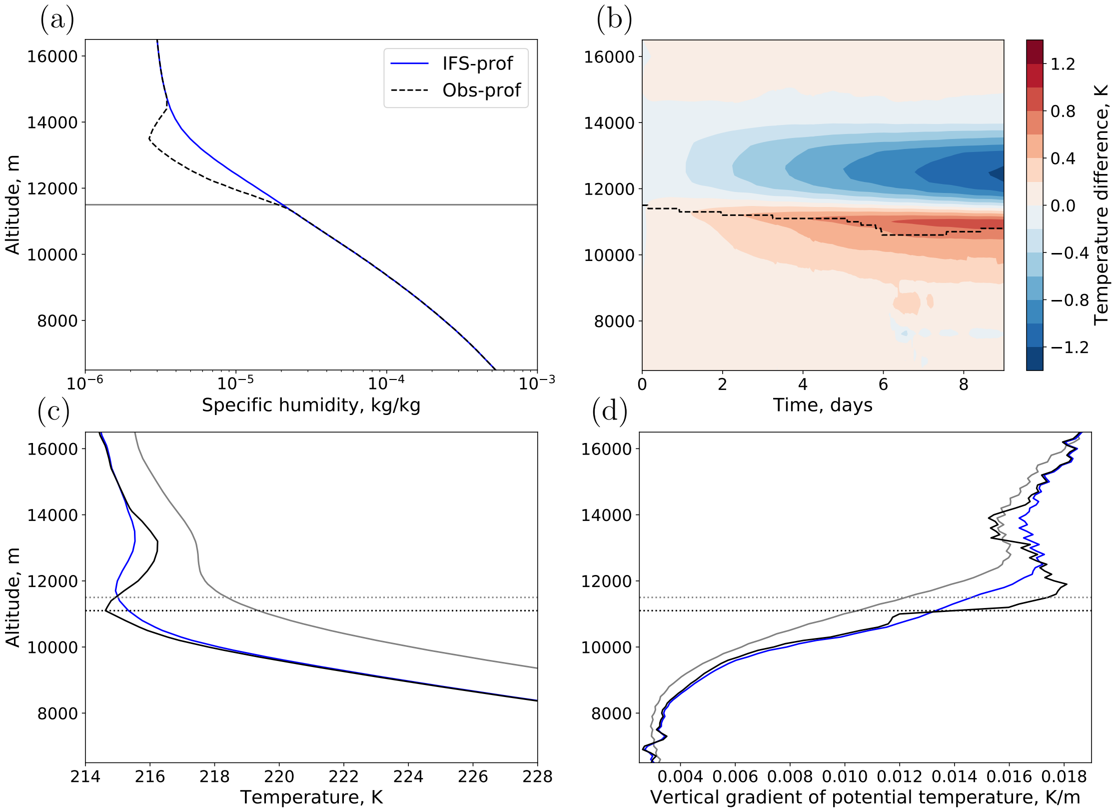

Figure 2.10: (a) Idealised vertical profiles of specific humidity representative of the IFS (blue) and radiosonde observations (black dashed) used for initial conditions of the SCM simulations, as described in the text. The horizontal grey line is the tropopause as calculated from the mean temperature profile using the WMO definition. (b) The temperature in the simulation with IFS humidity minus that with the humidity modified towards that observed, for the nine days of the forecast. The black dotted line is the tropopause as calculated from the temperature profiles of the Obs-prof simulation. Vertical profiles of (c) temperature and (d) the vertical gradient of potential temperature at five days for IFS (blue, IFS-prof) and radiosonde (black, Obs-prof) representative SCM simulations, and the profile at initial time (grey).

Figure 2.10(b) illustrates the difference in temperature fields between the two SCM simulations. From examination of the temperature tendencies (not shown) it is clear that while it is not the only parameterisation present in the SCM, the vast majority of this difference in heating rates is attributable to the radiation scheme. A dipole of temperature difference emerges over the nine days of forecasts shown, with a lowermost stratosphere cold bias and near-tropopause warm bias. The additional cooling of IFS-prof relative to Obs-prof is -0.175 K day^{-1} at day three, reducing to -0.16 K day^{-1} at day five and continuing to slow for the remaining four days of the simulation. If an alternative scenario is assumed where the moist bias extends upward and humidity is reduced to match observations in the lowest 5 km of the stratosphere, and linearly to no difference at 7.5 km (not shown) then the cooling rate slightly decreases to ≈ -0.14 K day^{-1} at day five. The warm bias in the vicinity of the tropopause, and just below, is shallower and of a lower magnitude than the cold bias above, growing at a similar rate of $\approx +0.15$ K day^{-1}.

The increased long-wave cooling in the lowermost stratosphere associated with the moist bias leads to the growth of the negative temperature bias in these forecasts locally. However, the increased down-welling long-wave emission from the lowermost stratosphere, will be absorbed in the upper troposphere (Birner and Charlesworth 2017) leading to increased warming in the upper troposphere with respect to a profile without a moist bias. Consequently, we expect the upper tropospheric warm bias to grow in tandem with the lower stratospheric cold bias as seen in the SCM (Figure 2.10).

The SCM-derived cooling rate above the tropopause is consistent with the -0.2 K day^{-1} found in the mean through comparison of operational forecasts to radiosonde temperature observations in Figure 2.8, given the idealised nature of the SCM simulations. This consistency strongly suggests that the cold bias that is found to develop in the lowermost stratosphere of the IFS is largely a result of the moist bias in the lowermost stratosphere analysis.

The growth of a warm bias is not seen in operational models to the same extent as in the SCM. This difference in the temperature bias between the SCM and operational model suggests there are other processes acting in the full IFS forecasts, such as advection and mixing, that modify the warm bias in the region of the tropopause.

2.5.2 Changes in static stability near the tropopause

Water vapour in the lowermost stratosphere is also shown to affect tropopause altitude and sharpness. The temperature profile of the Obs-prof SCM simulation, initialised with a drier lowermost stratosphere, evolves to have a sharp transition from positive to negative lapse rates at the tropopause, above which the temperature then increases to a local maximum $\approx$ 2 km above the tropopause at the top of the TIL, before becoming approximately isothermal for several kilometres (Figure 2.10(c)). This structure is similar to that of temperature profiles from the observations and analysis (not shown). However, with the moist bias in IFS-prof there is a weaker lapse-rate in the upper troposphere, a smoother transition from positive to negative lapse rates, the local minimum in temperature is at

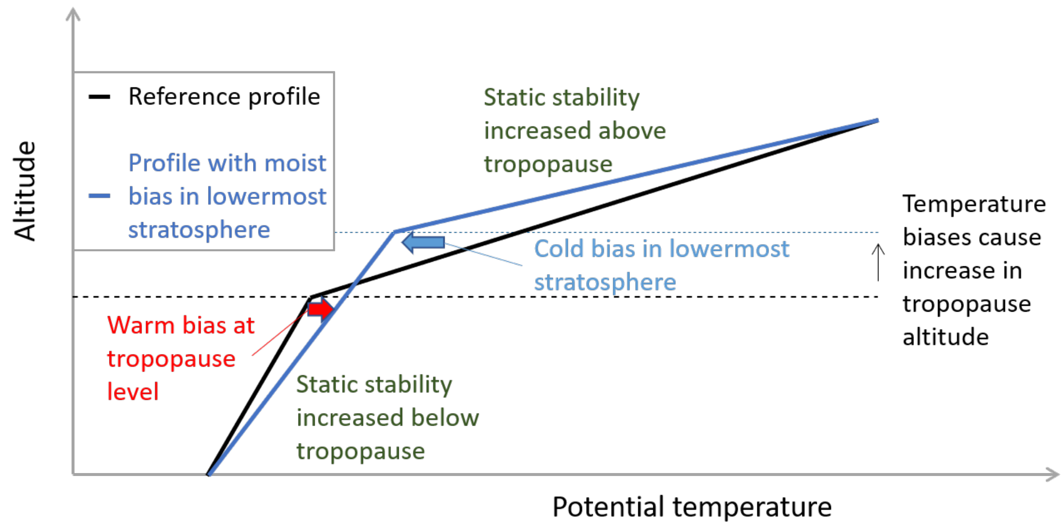

Figure 2.11: Schematic illustration of the effect of the lowermost stratospheric moist bias on an idealised vertical profile of potential temperature, where the black line is an idealised reference profile. The blue line represents the profile following the effects of the anomalous radiative heating dipole that results from the presence of a moist bias. The horizontal dashed lines indicate the tropopause altitudes for the respective profiles.

a higher altitude and the magnitude of the local temperature maximum at the top of the TIL is reduced. This, on the other hand, is similar to that of temperature profiles from the operational five-day forecasts. That these similarities are evident, despite the SCM being initialised with a smoothed temperature profile, demonstrates the dependence of the structure of the equilibrium model temperature profile at the tropopause on initial lowermost stratospheric humidity.

The schematic in Figure 2.11 illustrates how the radiative heating dipole in response to the moist bias is expected to influence an idealised atmospheric profile where static stability is piecewise uniform. To first order, the tropopause can be represented as a sharp change in static stability from typical values in the troposphere to a higher value in the lower stratosphere (approximately double). The air cools radiatively where there is the moist bias and potential temperature increases below (near the tropopause). Local turbulent mixing tends to maintain uniform static stability in the troposphere and the blue curve in Figure 2.11 is obtained by satisfying the radiative changes in potential temperature and piecewise uniform static stability. A necessary result is an increase in tropopause height, as well as increased static stability above the level of peak anomalous cooling, as seen in Figure 2.10(d). Static stability must also decrease in the model between the level of the observed tropopause and the peak of anomalous cooling. This tropopause smoothing is illustrated by the SCM results in Figure 2.10 at the level of the Obs-prof tropopause, showing that in models with a lowermost stratosphere moist bias in the initial conditions there is a reduction of the sharp static stability gradient.

This can explain plausibly why the tropopause altitude calculated from model data is on average higher than in radiosonde observations, and that this difference increases with forecast lead time (Figure 2.2(a)).

2.5.3 Unattributed temperature biases

Finally, the temperature biases that developed in the MetUM forecasts are considered. In contrast to the findings for the IFS operational forecasts, the SCM experiments indicate that the radiative response to a moist bias cannot completely explain the temperature biases in the MetUM operational forecasts. Additional SCM simulations were run using idealised humidity profiles representative of the mean over those from the MetUM as well as taken from individual ascents. Through comparisons of temperature fields from these simulations (not shown) it is evident that the lowermost stratosphere cold bias develops due to the effects of long-wave radiation in a similar way to the IFS, although with a smaller magnitude. In Figure 2.8(d) and Figure 2.9(b) and (d) the temperature difference 1 km above the tropopause is more negative than that 3 km above the tropopause, which we would expect from radiative effects of the moist bias. The warm bias 3 km above the tropopause in ridges, in conjunction with the smaller negative temperature bias below this, may then be the result of a secondary factor causing a warming throughout the stratosphere in ridges. It is hypothesised that another of the dominant stratospheric trace gases, i.e. ozone or carbon dioxide, is responsible for this additional anomalous warming. Both the IFS and the MetUM use an ozone climatology that does not take into account the varying tropopause height. As a result of this both have higher concentrations of ozone in the lower stratosphere above ridges than observed at the same altitudes by ozonesondes (WOUDC 2020) or the AIRS satellite instrument (AIRS Science Team and Teixeira 2013). At most latitudes this difference is notably larger than the difference between the climatologies used by the two models. So, although one might expect a positive bias in lower stratospheric ozone concentrations in ridges in the MetUM to cause anomalous warming there, it is unclear why this would would result in a warm bias in the MetUM (at 3km above the tropopause), but not in the IFS. As mentioned previously, this temperature bias will be affected by the history of the air parcels as well as local radiative processes.

2.6 Summary and Conclusions

An accurate representation of temperature and humidity in the extratropical lower stratosphere is important for global weather forecast and climate models. Many models have been shown to have significant biases in this region and there has been little progress in reducing these biases. To improve the models, it is imperative to gain a better understanding of the errors and their sources and this study is a step forward in that direction. The specific aims of this study are to identify and characterise humidity and temperature biases in the upper troposphere and lowermost stratosphere in the IFS and MetUM weather forecast models using radiosonde observations, quantify the temperature bias growth attributable to a diagnosed moist bias, and explore the influences of these biases on other tropopause level features. The main conclusions are summarised below.

It is found that both the IFS and MetUM have a mean moist bias in the lowermost stratosphere with a maximum approximately 1 km above the tropopause with the model humidities approximately 170% of the observed values. The magnitude of this bias found through comparison with radiosonde data is largely consistent with those found by the previous studies listed in Table 2.1 derived from comparison with aircraft and satellite observations, and the altitude is consistent with the autumn tropopause-relative comparison from Dyroff *et al.* (2015). When considering only radiosonde ascents through tropopause-level troughs this maximum value of the specific humidity normalised difference occurs around 2 km above the tropopause; for ridges it occurs at around 1 km, the same height as in the mean. In the IFS, the tropopause-relative vertical structure of the moist bias is very similar between ridges and troughs whereas in the MetUM features in troughs have a larger vertical length scale. The moist bias is not found to be systematically dependant on the presence of cloud, layers of low static stability, or position within a ridge, and the magnitude of the moist bias changes very little during five-day forecasts in both models, being dominated instead by the bias present initially.

The temperature fields in model analyses agree with radiosonde observations to within ± 0.2 K. However, a cold bias develops in the lowermost stratosphere in the IFS operational forecasts, also with a maximum approximately 1 km above the tropopause and growing at a rate of ≈ 0.2 K day^{-1}. There is little difference in this growth rate between troughs and ridges in the first three days, although the cold bias in troughs extends deeper into the lowermost stratosphere. Using the IFS single column model (SCM) it is shown that the growth of this lowermost stratosphere cold bias is consistent with the additional longwave radiative cooling calculated as a result of the presence of the lowermost stratosphere moist bias. The SCM simulations also show that the moist bias would result in a warming around the tropopause level, extending below into the upper troposphere, with a smaller depth and growth rate than the lowermost stratospheric cooling. A warming is also seen in the operational forecasts as a sharp feature at the tropopause but this does not extend into the upper troposphere in the same way as the warming in the SCM. This is likely to be due to advection and mixing processes that are not represented in the SCM but modify the warming feature in the 3-D model. This feature in the composites is however also sensitive to the specification of tropopause-relative coordinates.

As is the case for the IFS, the MetUM operational forecasts also develop a cold bias at 1km above the tropopause, but with a smaller magnitude than that in the IFS and not present when considering only profiles in ridges. For ridges instead a warm bias develops with a maximum at around 2.5 km above the tropopause. This warm bias is seen too in the mean over all profiles, corresponding to an additional warming of ≈ 0.1 K day^{-1}. We cannot explain the MetUM lower stratosphere warm bias as a radiative response to the moisture bias. Indeed, additional long-wave cooling would be expected due to the moist bias in both models. Although the static ozone climatology would be expected to generate a warm bias above large-scale tropopause ridges, this does not explain why the responses of the MetUM and IFS differ. Therefore, further investigation would be required to understand the development of the warm bias in the full MetUM model. In addition to temperature biases, it is demonstrated that the presence of a lowermost stratosphere moist bias results in a higher tropopause, a less pronounced TIL, and a smoother cross-tropopause static stability gradient than in an atmospheric profile with a drier lowermost stratosphere (summarised in Figure 2.11). As discussed in section 2.1, such changes to the structure of the tropopause would be expected to systematically affect wave propagation on this sharp gradient, and through this other aspects of forecast development in the troposphere.

A limitation of the results regarding the magnitude and structure of the found biases is that data were only considered over the eastern North Atlantic and western Europe region, in a two-month period. These data were chosen to more easily facilitate future studies of the particular impact of these biases on the development of extratropical cyclones in this region, and consequently on forecast quality. Due to the NAWDEX campaign the high resolution radiosonde data were readily available. However, as is shown by Dyroff *et al.* (2015) , the lowermost stratosphere moist bias in the IFS varies seasonally with a maximum in summer and minimum in winter. To further understand the sources of error and their seasonal variations temperature and humidity biases could be similarly analysed in other seasons. Furthermore, although the temperature data are reliable at all considered altitudes, due to the measurement uncertainty in the instruments there is only confidence in our assessment of the moist bias in the lowest 2 km of the stratosphere. This is why complementary studies using other observation techniques such as Krüger *et al.* (2022) are valuable. Krüger *et al.* (2022) corroborate the results presented in this chapter in the mean, and in contrast to the above discussion of relatively large spatial and temporal scales allow a better understanding of the model representation of humidity on finer spatial scales through the use of LiDAR along flight tracks.

Previous work has shown that a moist bias in the extratropical lowermost stratosphere would be expected to cause a collocated cold bias (Forster and Shine 2002, Maycock *et al.* 2011). This study has found that the growth rate of this lowermost stratosphere cold bias in the ECMWF operational forecast model is quantitatively consistent with the additional radiative cooling rate calculated using the magnitude of the moist bias found in the IFS analysis. We have shown that the moist bias is dominated by the bias in the initial conditions, and therefore to reduce this cold bias in forecasts, reduction of the moist bias in the analysis is required. However, as increments are not currently applied to humidity fields in this region during data assimilation, as discussed in Section 2.2, the analysis humidity field is not constrained by observations. Whether or not humidity observations are assimilated, it is important to understand and reduce the sources of the moist bias in the forecast model. It is likely that there are contributions from excessive diffusion or transport of water vapour across the hygropause from the high water vapour values in the troposphere due to errors in numerical or physical processes. The characterisation of the magnitude and vertical structure of the model moist bias presented in this chapter facilitates the more detailed follow-on investigation of model processes controlling humidity in this region required to address this problem presented in Section 3.

Chapter 3

Processes controlling extratropical near-tropopause humidity and temperature in medium-range forecasts

Abstract

Accurate representation of near-tropopause fields is important for medium-range forecast skill, yet it is here that the largest systematic forecast biases are found. While extratropical near-tropopause humidity, temperature and wind model biases are known, more knowledge of their causes is required to improve models. Typically, a moist bias is present in the lowermost stratosphere in the analyses used to initialise forecasts, and a cold bias grows here over the course of forecasts. Experiments are conducted with the ECMWF forecast system where the humidity in a layer 0–4 km above the tropopause is reduced to correct the moist bias in the initial conditions. In this experiment the lowermost stratosphere re-moistens, returning to typical analysis values with a half-life of around 8–9 days, and the cold bias growth is halved compared to the control. The reduction in lowermost-stratospheric cooling is due to a reduction in long-wave radiative emission from the water vapour above the tropopause. The main contributors to the re-moistening are resolved advective transport and parametrized diffusion in the model, the cloud microphysical

process rates being similar in the modified and control experiments. The biases in near-tropopause moisture transport are almost independent of horizontal resolution. These results show that the extratropical lower-stratospheric temperature bias can be reduced by correcting the collocated moist bias, but the moist bias cannot be fixed by solely correcting the initial conditions. The moist bias arises from a combination of advective transport and mixing, and not simply explicit vertical diffusion locally across the tropopause. The insensitivity to horizontal resolution indicates that it is vertical resolution that hinders the representation of long-range transport of dry layers and their mixing into the surroundings. Therefore, to reduce forecast error, further work is required to improve vertical diffusion and advection.

3.1 Introduction

There is a cold bias in the lowermost stratosphere of many forecast models (Wu and Reichler 2020), which impacts not only the accurate forecasting of stratospheric events such as Sudden Stratospheric Warmings (SSW) but also weather in the troposphere. It has further been shown that there is a coincident moist bias present in analyses and reanalyses (e.g., Dyroff *et al.* (2015), Oikonomou and O'Neill (2006)). Results from the single column modelling in Chapter 2 indicate that the cooling of the forecasts with respect to the analysis is consistent with the additional long-wave radiative cooling one would expect from the excessive water vapour in the lowermost stratosphere of the model, compared to the observed atmosphere. The impact of cooling and/or moistening at the tropopause level on global circulation have been investigated in climate models either through the increase of stratospheric water vapour (Joshi *et al.* 2006, Maycock *et al.* 2013) or direct imposition of stratospheric or tropopause-level cooling (Tandon *et al.* 2011, Boljka and Birner 2022). It has been shown that such biases result in a strengthened and poleward shifted subtropical jet, in addition to influence on the strength and width of the Hadley circulation: Tandon *et al.* (2011) imposed $\approx 5K$ cooling to mimic an increase in water vapour and find a weakened Hadley cell, however Boljka and Birner (2022) used $\approx 60K$ cooling to strengthen the Tropopause Inversion Layer (TIL) and instead find a strengthened Hadley cell, therefore it is clear that both the magnitude and positioning of these biases are important in determining the response of the general circulation.

A key unanswered problem, which is the focus of this study, is the cause of the lowermost-stratospheric moist bias in model analyses and how this can be corrected. Cloud processes can influence stratospheric water vapour in the sub-tropics and mid-latitudes. Deep convection overshooting the tropopause can transport water to the lower stratosphere, where it can evaporate or sublimate and increase local humidity, as has been identified in studies such as that by Jensen *et al.* (2020) over North America. Furthermore, Müller *et al.* (2015) find evidence of ice crystals from cirrus clouds formed in the troposphere being transported into the extratropical lowermost stratosphere, which has the potential to provide an additional mechanism for lowermost-stratospheric moistening. Another process that is a source of water vapour in the stratosphere is the oxidation of methane which occurs in the upper branch of the Brewer-Dobson circulation; however, while methane oxidation is present in the model this process is less relevant to the lowermost stratosphere than at higher altitudes (Noël *et al.* 2018) and hence will be omitted for the present chapter.

Another important factor controlling moisture in the stratosphere is transport from the tropics (Gettelman *et al.* 2010, Bönisch *et al.* 2009). As a part of the Brewer-Dobson circulation there is large-scale ascent in the tropics. There are then two pathways by which air is transported to the extratropical stratosphere: entering the tropical lower stratosphere and moving up towards the stratopause as it moves polewards before descending at midlatitudes, or moving polewards isentropically entering the stratosphere in the subtropics. The coldest part of the troposphere is in the region of the tropical tropopause, and air transported through this region will be dried due to water or ice deposition, reducing the vapour pressure to the low saturation vapour pressure associated with such cold temperatures. Tropospheric air can then be mixed into the extratropical lowermost stratosphere as a result of mesoscale mixing such as Kelvin-Helmholtz instability across the tropopause (Kunkel *et al.* 2019), or the creation of filaments of tropospheric air by Rossby-wave breaking events, which are then stretched to increasingly fine scales (Vaughan and Timmis 1998, O'Connor *et al.* 1999, Bradshaw *et al.* 2002). It has been identified by Hardiman *et al.* (2015) that in the MetUM a warm Tropical Tropopause Layer (TTL) results in a moist bias of the stratosphere, as a warmer TTL has a higher saturation vapour pressure, and hence more water vapour can be transported across the tropopause. The IFS, however, has a cold bias at the tropical tropopause (Polichtchouk *et al.* 2019b) so, while polewards transport is still an important factor when considering extratropical lowermost-stratospheric water vapour, the moist bias cannot necessarily be explained in the same way across all models.

A method for investigating the model processes responsible for changes to the lowermost-stratospheric humidity field is by considering process tendencies. This separation of changes made by a forecast model to the prognostic fields into the sum of parts associated with the model representation of different processes has been utilised previously. Rodwell and Palmer (2007) considered the forecast tendencies in the model that evolve the state prior to the application of the analysis increment in the assimilation cycle in their initial tendency methodology. For the analysis presented in this study we consider process tendencies accumulated over a much longer period of five days, which allows for the investigation of the time variation of the tendencies as the background fields adjust.

It was established in Chapter 2 that the moist bias present in the initial conditions for operational ECMWF Integrated Forecasting System (IFS) and Met Office (MetUM) forecasts changes very little during a five day forecast. Noted also in Chapter 2 are the difficulties associated with the assimilation of humidity observations in the stratosphere. Humidity increments are not applied above the tropopause during data assimilation in the IFS (Ingleby 2017). With no contribution from observations we hypothesise that the moist-biased analysis state is a consequence of a bias in the model, leading to what will be referred to in the remainder of this chapter as a "moist-biased model climatology" for the lowermost stratosphere. To investigate this bias, in the present study IFS reforecasts are run with a correction applied to the initial humidity field to bring this as close as practicably possible to the available observations. The sensitivity to the initial lowermost-stratospheric humidity is then used to infer possible factors contributing to the IFS temperature bias.

Following from this, the aims of this study are as follows:

1. Determine the evolution of the vertical structure of moisture in the Upper Troposphere and Lower Stratosphere (UTLS) region as the lowermost-stratospheric moisture bias re-establishes itself after correction.

2. Determine the vertical structure and generation timescale of the temperature bias in the UTLS region due to the analysis bias in lowermost stratospheric moisture, and the associated impact on tropopause altitude.

3. Attribute the moisture and temperature biases in the UTLS region to model processes.

4. Determine the sensitivity of the findings to model resolution.

In Section 3.2 we provide some information on the models and data used to carry out this study. The methods used to address our aims are outlined in Section 3.3, before detailing the impacts of the initial humidity reduction on humidity and then temperature in the forecasts in Section 3.4. In Section 3.5 we initially examine the physical parameterisations which dominate the evolution of UTLS temperature, and subsequently discuss what the process tendencies tell us about the model treatment of water vapour. Section 3.6 expands upon this by also considering such tendencies in forecasts run at different horizontal resolutions, and finally this study is summarised and concluded in Section 3.7.

3.2 Sources of model output and observational data

To address the aims presented in Section 3.1, we use data from forecasts that were run with the ECMWF IFS cycle 47R1.1 (ECMWF 2020) which was operational between June 2020 and May 2021. One of the major improvements of 47R1 over previous model cycles is the introduction of quintic vertical interpolation instead of cubic (Polichtchouk *et al.* 2019a). Prior to this, forecasts at higher horizontal resolutions would generate waves at too small a scale to be resolved in the vertical, leading to unphysical grid-scale oscillations in the vertical which caused excessive cooling in the stratosphere. The use of this IFS cycle means that our analysis considers the lowermost-stratospheric bias that remains after this correction.

Forecasts were run at three different horizontal resolutions: the current resolution for operational numerical weather prediction with a grid spacing of approximately 9 km (TCo1279), the standard resolution for medium-range research experiments with a grid spacing of approximately 25 km (TCo399), and the resolution used for longer-range forecasts to assess model climate with a grid spacing of approximately 50 km (TCo199). The "TCo" prefix here stands for Triangular Cubic-octahedral, a detailed explanation of which can be found in Malardel *et al.* (2016). All forecasts have 137 model levels in the vertical, which corresponds to a vertical spacing of around 250 m in the vertical near the tropopause.

We consider sets of 30 forecasts initialised at 00 UTC each day from the 15 September to 14 October 2016. This period is chosen to coincide with the NAWDEX field campaign (Schäfler *et al.* 2018), from which observations were used in Chapter 2 to characterise UTLS temperature and humidity biases. As one of the aims of this study is to investigate the effect of the IFS moisture bias, we have chosen the same period as that over which it was quantified in Chapter 2. Forecasts are run for lengths of both 5 and 15 days. Preliminary year-long experiments run at TCo199 indicated that 15 days is a good timescale to observe the difference in development of humidity and temperature between experiments. For tendencies we want to look at the characteristics of their evolution, and differences between experiments both during the initial adjustment over the first ∼1 day, and subsequent adjustment over the next few days, while working within the constraint of the number of times for which it is feasible to archive the relatively large suite of process tendencies. Therefore forecasts outputting process tendencies are run for 5 days, and archived 12 hourly for TCo199 and TCo399 and 24 hourly for TCo1279.

Observations used to constrain the humidity correction are taken from both radiosondes and the Aura Microwave Limb Sounder (MLS). Radiosonde data from the NAWDEX campaign used here are from the European Meteorological Services Network (EUMETNET) as described in Schäfler *et al.* (2018). Radiosonde models used are Vaisala RS92 (Vaisala 2013) and RS41 (Vaisala 2018). A more in-depth discussion of these can be found in Chapter 2. MLS data are obtained from the Goddard Earth Sciences Data and Information Services Center (GES DISC) (Lambert *et al.* 2020), and used only between $40 - 60°N$ at pressures below 316 hPa. The satellite data used have been binned daily onto a regular grid ("Level 3" data), considering only bins with ≥ 1 valid data points. MLS observations are compared to IFS analysis fields at 12 UTC, with the IFS analysis tropopause altitude used for tropopause-relative composition when required, as for Figure 3.1 below in Section 3.3. Further details for the MLS data can be found in Livesey *et al.* (2020), with information on water vapour retrieval in their section 3.9, and the Level 3 datasets in their section 4.

3.3 Methods

In the following subsections we outline the method used to reduce the lowermost-stratospheric humidity to realistic values for experimental forecasts, the use of tropopause-relative coordinates, the Eulerian tendencies used to attribute changes in temperature and humidity to model parametrisations, and methods used for compositing and masking of data.

3.3.1 Initial humidity reduction

To investigate the effects of the lowermost-stratospheric moist bias on the model behaviour we run one set of forecasts with the humidity field matching the analysis (termed CTRL), and another set of forecasts in which the specific humidity is reduced in the lowermost stratosphere (termed QMOD). The humidity is modified during the first time-step (only), on a grid point basis in the same way as parameterised physical processes. The structure and magnitude of this reduction is based on the tropopause-altitude-relative humidity bias identified in Chapter 2 through comparison of operational analyses to radiosonde humidity measurements, which are shown also in Figure 3.1 (a). Specific humidity normalised difference between two humidity profiles q_A and q_B is calculated as

$$\frac{q_A - q_B}{\sqrt{q_A^2 + q_B^2}},$$

as in Chapter 2. The equation for the reduction of the humidity takes the form

$$q_{mod}(p,z) = q_{an}(p,z) - q_{an}(p,z) \times f(p,z), \tag{3.1}$$

where q_{mod} is the modified humidity field which is used to initialise the QMOD forecasts, q_{an} the analysis humidity field, p and z the model pressure in Pa and altitude in m, respectively, and f the factor by which the humidity is reduced. This factor can be expressed as the product $f(p,z) = f_{max} \times f_1(z) \times f_2(p)$, where $0 < f_{max}, f_1, f_2 < 1$ for all p, z.

The radiosonde observations can be used to define f_1 for altitudes $z > z_{trop}$, where z_{trop} is the altitude of the model tropopause in a column at a given latitude and longitude as calculated using the standard WMO lapse rate definition as defined in Section 2.3.2. We define

$$f_1(z) = \begin{cases} 0 & z < z_{trop} \\ \frac{z - z_{trop}}{500} & z_{trop} < z < z_{trop} + 500 \\ 1 & z_{trop} + 500 \leq z \leq z_{trop} + 2000 \end{cases} \tag{3.2}$$

where z is in units of m. The linear reduction between the tropopause and 500 m above covers two or three model levels, and moderates the sharpness of the gradient we are introducing here. As we have less confidence in the radiosonde observations for $z > z_{trop} + 2000$ m, we consider also a comparison of the model analysis humidity field to observations from the AURA MLS (Microwave Limb Sounder) satellite instrument (Lambert *et al.* 2020). Despite the coarser resolution of these observations in the vertical, compared to the radiosonde observations, the MLS observations show that the moist bias in analyses is largely confined to the lowermost stratosphere within 3.5 km of the tropopause (brown solid line in Figure 3.1a). Humidity reduction factors that dry the atmosphere at higher levels were shown to worsen the agreement with these MLS observations. To avoid the introduction of a sharp humidity gradient at the top of our dried layer the modification function declines linearly with height:

$$f_1(z) = \begin{cases} \frac{3500+z_{trop}-z}{1500} & z_{trop}+2000 \leq z \leq z_{trop}+3500 \\ 0 & z > z_{trop}+3500. \end{cases} \tag{3.3}$$

The function $f_1(z)$ is illustrated in Figure 3.1 (b). Furthermore, the model moist bias that we are investigating is confined to the extratropics. To prevent humidity reduction in the tropics, while also avoiding the introduction of an unphysical sharp meridional humidity gradient, we take advantage of the fact that the pressure at the tropopause generally decreases towards the equator, and we do not have evidence of a moist bias at lower pressures higher in the stratosphere. Therefore we have

$$f_2(p) = \begin{cases} 0 & p < 10000 \\ \frac{p-10000}{5000} & 10000 \leq p \leq 15000 \\ 1 & p > 15000 \end{cases} \tag{3.4}$$

where p is in units of Pa. Finally, we use a maximum reduction factor $f_{max} = 0.55$, selected after trying values within the range identified in Chapter 2 to ensure that the reduction factor obtained through the product of the combination of the above choices of f_{max}, f_1 and f_2 gives the greatest agreement with the available observations that we can achieve while keeping the method fairly simple.

Figure 3.1 illustrates that the normalised difference between the adjusted initial conditions, q_{mod}, and the radiosonde observations is near zero on average between the tropopause and 2 km above (pink dashed curve). Above this level (where radiosonde humidity measurements are uncertain), the normalised difference between q_{mod} and the MLS observations is near zero (brown dashed curve). Therefore, it can be concluded that the modified field q_{mod} is a better representation of the atmospheric state and closer to the best observational measurements at each level than the operational analysis.

By making the adjustment in a tropopause-relative sense we expect that the correction will be appropriate for every profile. However, it was noted in Chapter 2 that the vertical tropopause-relative structure of the moist bias differs from the mean in tropopause trough

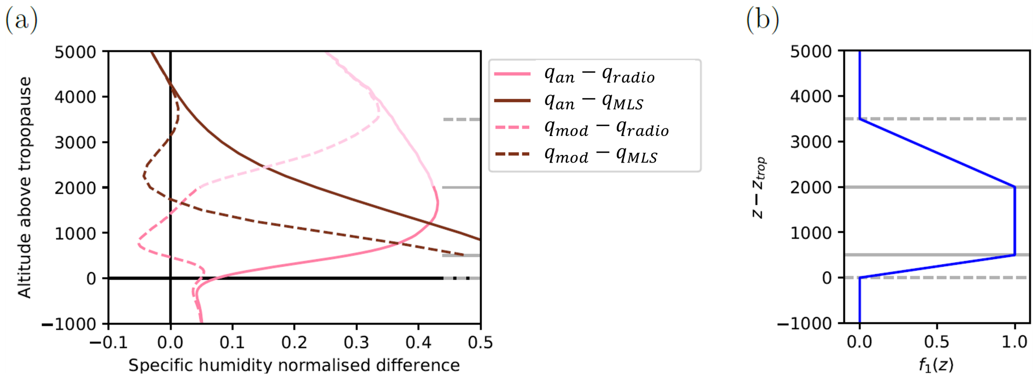

Figure 3.1: (a) Normalised difference in specific humidity comparing the analysis humidity (q_{an}, solid) and modified initial conditions from QMOD experiment (q_{mod}, dashed) to the radiosonde data (q_{radio}, pink curves) and MLS satellite data (q_{MLS}, brown curves). It is noted in Chapter 2 that the radiosonde observations are less reliable above 2 km above the tropopause, and so the lines are faded above this point to indicate this. The MLS satellite retrievals are less reliable within 1.5 km of the tropopause. (b) Illustration of the vertical structure of $f_1(z)$ with horizontal lines to indicate altitudes defining the humidity modification function. Similar lines are included on the right of panel (a).

regions. Such differences for regions of low tropopause altitude have also been observed by Krüger *et al.* (2022). A humidity modification function $f_1(z, z_{trop})$ that varied not only with altitude within a given profile but also according to tropopause altitude, and hence yielded a different structure when the tropopause altitude was low, was also trialled (not shown), but performed so similarly to the simpler formula presented above that the additional complexity was not justified.

3.3.2 Tropopause-relative coordinates

We are considering fields in a close proximity to the tropopause, with particular interest in the dependence of the values of these fields on their relative position to the tropopause. We want to visualise these fields and take composites in a tropopause-relative framework. However, issues arise as a result of the tropopause being an interface associated with sharp gradients, and hence markedly different properties on either side, which varies in time, space and also between profiles at the same spatial and temporal location in differing data sets.

We re-grid fields to an altitude coordinate relative to a tropopause altitude that is the same, at a given latitude, longitude and time, between all pertinent data sources in any given comparison. This re-gridding is because we want to compare like-for-like regions of the atmosphere on a profile-by-profile basis in a way that is not sensitive to the particular thresholds chosen for the definition of the tropopause. Doing so avoids taking the differences between fields in regions which are vertically displaced from one another by more than differences in altitude at model levels between data sources. The method used is to re-grid fields in the vertical from model level coordinates (of the 137 IFS model levels) to $\bar{z} - \bar{z}_{trop}$ coordinates, where z are the altitudes of the model levels considered, z_{trop} is the altitude of the tropopause, and the overbar refers to the arithmetic mean between the values in the two data sets being compared, which is typically a model forecast and analysis. It is necessary to take the mean between the tropopause altitudes from the two data sets as comparing fields from two data sets relative to the tropopause determined from just one of these data sets can lead to sizeable systematic errors in composites. This error occurs even if there are no systematic differences in the altitude or tropopause altitude between the two data sets.

3.3.3 Cumulative tendency budget

To determine the model processes responsible for changes to the fields of temperature and humidity over the course of the forecast, additional variables are introduced which are hereafter referred to as cumulative Eulerian tendencies, or just tendencies. These cumulative Eulerian tendencies are each incremented by changes to temperature or humidity at a grid point by an isolated component of the model, such as a subroutine for the parameterisation of a physical process, or the update to the field following the solution of the dynamics for a timestep. The total changes to these fields are partitioned into tendencies such that the sum of these cumulative tendencies for a field in a region over a time period is equal to the total change of that field during that time in the model.

For better visualisation in a tropopause-relative sense, these whole forecast cumulative Eulerian tendencies are decomposed into twelve-hourly cumulative tendencies such that

$$x_{12hrly}^{tend}(t) = x_{cumul.}^{tend}(t) - x_{cumul.}^{tend}(t-12), \tag{3.5}$$

where t is time in hours, and $x_{cumul.}^{tend}$ and x_{12hrly}^{tend} are, respectively, the whole forecast cumulative (from times zero to t) and twelve-hourly cumulative (from times $t-12$ to t) tendencies of the variable x (either temperature or specific humidity) due to some process *tend*. Consequently, as $x_{12hrly}^{tend}(t)$ is the change resulting from the process *tend* centred on time $t-6$, it does not make sense to use $(\bar{z}(t) - \bar{z}_{trop}(t))$ as our tropopause-relative altitude coordinate as defined in the above subsection. Therefore instead we use

$$\frac{(\bar{z}(t) - \bar{z}_{trop}(t)) + (\bar{z}(t-12) - \bar{z}_{trop}(t-12))}{2} \approx (\bar{z}(t-6) - \bar{z}_{trop}(t-6)). \tag{3.6}$$

Additionally, the specific humidity twelve-hourly cumulative tendencies will be primarily expressed as a fraction of the mean humidity background state in the centre of this window, $q^{tend}_{12hrly}(t)/\bar{q}(t-6)$, where $\bar{q}(t-6)$ is the arithmetic mean of $\bar{q}(t)$ and $\bar{q}(t-12)$ as for z above.

3.3.4 Compositing and masking

In addition to the processing described above, data are also masked for regions where $\bar{z} < 0$ before compositing. For composites, an area-weighted arithmetic mean is taken over fields and differences between fields over the latitude band 40–75°N, and then a further mean is taken over data from each date and time at which a forecast was initialised. Where frequency distributions are considered no spatial or temporal means are taken, and all latitudes within the band are given equal weighting.

3.4 The control of lowermost-stratospheric humidity on the state of the UTLS in 15-day forecasts

In this section we outline the main differences in specific humidity, temperature and tropopause altitude between the set of QMOD forecasts that have the lowermost-stratospheric humidity initially corrected, the set of control (CTRL) forecasts, and the analysis.

3.4.1 Differences in humidity

We first consider the humidity field, as it is differences in this that we anticipate will force the behaviour of other model variables. Figure 3.2 (a) shows the mean difference between specific humidity in the mid-latitude northern hemisphere in forecasts and the analysis as a function of altitude relative to the tropopause using the method detailed in section 3.3. The difference for QMOD at day one is very similar to the difference which we impose in the lowermost stratosphere (as expected given that the difference has not had much time to change), and is therefore the same at all three horizontal resolutions considered, although only the mid-resolution results are shown in panel (a).

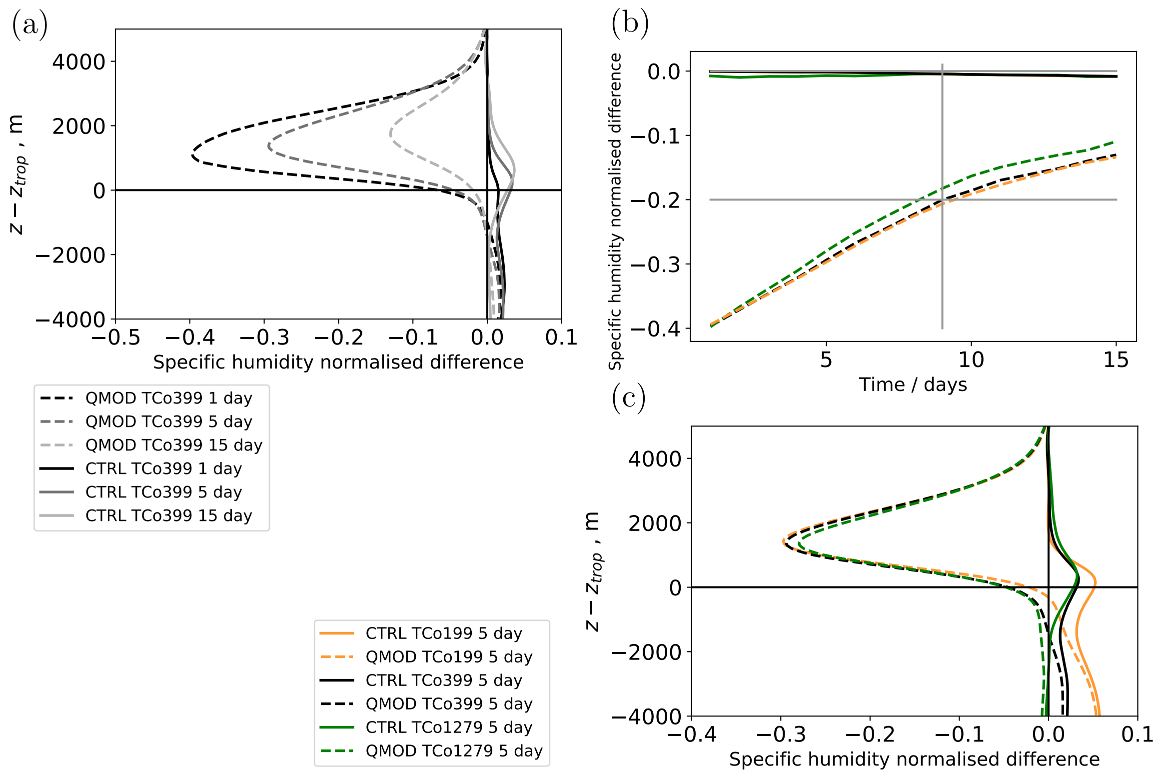

Figure 3.2: (a) Tropopause-relative vertical profiles of specific humidity normalised difference between forecast experiments (resolution TCo 399) and analysis at lead times of 1 (black), 5 (grey) and 15 (light grey) days into the forecast, with CTRL in solid lines and QMOD dashed. (b) The minimum normalised difference in 12 hourly mean vertical profiles as a function of lead time showing experiments with resolution TCo 199 in orange, 399 black and 1279 green. Additional grey lines mark a normalised difference of -0.2, and day 9. (c) Vertical profiles, as in (a), at 5-day lead time for each of these three horizontal resolutions.

Looking further at the mean difference in humidity between the analysis and the five- and fifteen-day forecasts, the difference between the QMOD forecast and analysed humidity in the lowermost stratosphere decreases with increasing forecast lead time, whereas the humidity in the CTRL forecasts remains fairly similar to the analysis, and hence the model climatology, throughout. This result means that following initialisation with a humidity field which has been reduced to more realistic values in a layer above the tropopause, the humidity of the lowermost stratosphere increases in the QMOD forecasts (relative to that in the CTRL) to bring the field back towards values in the analysis. It can be seen from Figure 3.2 (b) that the initial humidity difference halves within the first 8–9 days. Starting at a mean normalised difference of -0.4, this corresponds to an increase of around 2.5% (of the background value) per day. As the QMOD humidity field becomes closer to the analysis, the rate at which the difference decreases slows.

The specific humidity remoistening in the lowermost stratosphere is very similar between the three investigated resolutions, as can be seen from Figure 3.2 (c). In the lowermost stratosphere, QMOD forecasts for the highest resolution of TCo1279 are closest to the analysis after five days (i.e., the peak negative normalised difference has the smallest magnitude). However, at the altitude of this peak difference the corresponding CTRL forecasts are furthest from the analysis (in the opposite direction) so the differences between the QMOD and CTRL forecasts are similar for all three resolutions.

3.4.2 Impacts on temperature and tropopause altitude

We now look at the response of temperature and tropopause altitude to the removal of moisture in the lowermost stratosphere. The development of an extratropical cold bias in the lowermost stratosphere in forecasts is well known, and evidence is discussed in Sect. 3.1. This cold bias development in forecasts is illustrated in Figure 3.3 (a) for the mid-resolution forecasts, with the temperature in the CTRL forecast becoming colder than that in the analysis by around $-0.15\ \mathrm{K\,day^{-1}}$ at the altitude where the magnitude of the difference peaks. From the single column experiments of Chapter 2 it was concluded that this rate of cold bias growth was consistent with, if slightly in excess of, that which would be expected to result solely from additional long-wave cooling the presence of the positive moist bias in the lowermost stratosphere.

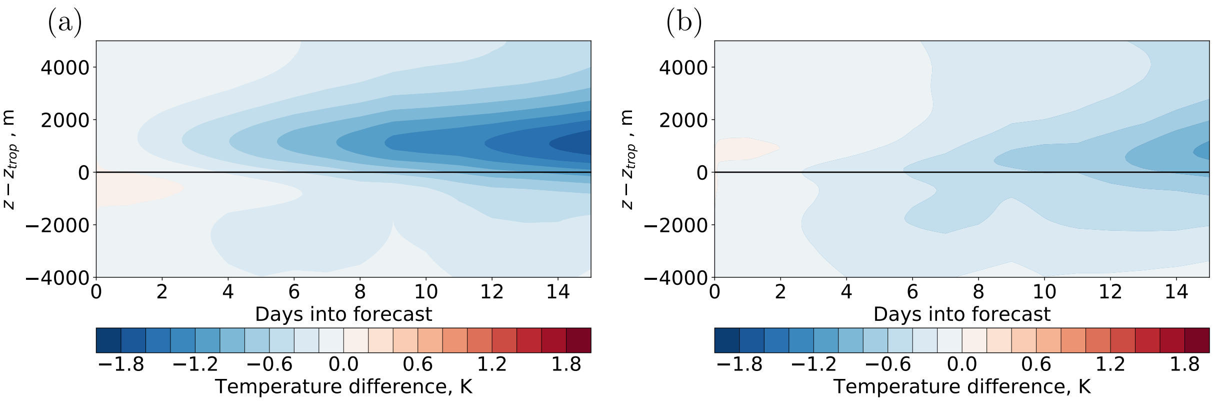

Figure 3.3: Timeseries of the mean difference in temperature for experiments (a) CTRL, (b) QMOD (at resolution TCo399) relative to the operational analysis.

Here, 30 pairs of simulations using the full IFS model, rather than a simplified model, give a fuller picture of the sensitivity of the mean temperature to lowermost-stratospheric humidity (Figure 3.3). Figure 3.3 (b) shows the evolution of the difference between QMOD forecast and analysis temperature in the 15-day forecasts. In this panel the temperature bias develops at an approximately halved rate of around -0.07 K day^{-1}. The difference in cooling between CTRL and QMOD is shown again in Figure 3.4 (a) for snapshots at 1, 5 and 15 days into the forecasts, and the approximately constant respective peak cooling rates throughout the forecasts are shown in Figure 3.4 (b).

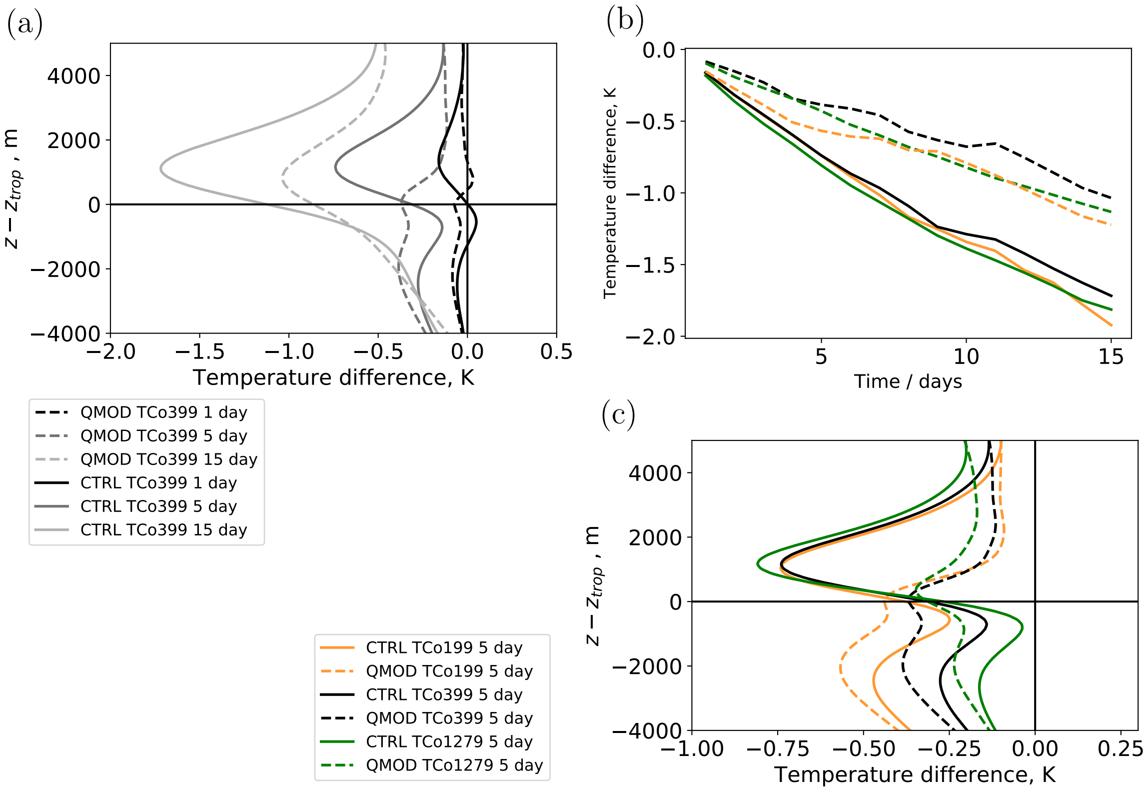

Figure 3.4: As for Figure 3.2, but showing the temperature difference from operational analyses instead of specific humidity difference.

Similarly to the specific humidity profiles, there is very little difference in temperature bias development in the lowermost-stratospheric forecasts with respect to the analysis between varying horizontal resolutions (see Figure 3.4 (c) which shows vertical profiles of the mean temperature difference from analysis at day 5 and Figure 3.4 (b) which tracks the maximum magnitude of this profile through time). The resolution independence of the rates of growth of the differences between forecast an analysed temperatures (of ≈ -0.15 and ≈ -0.07 K day^{-1} for CTRL and QMOD, respectively) is illustrated in Figure 3.4 (b). Another similarity to Figure 3.2 (b) is that over the 15 days the rate of change of the difference begins to slow, particularly evident here in the CTRL forecasts. One may hypothesise that as the specific humidity field tends towards an equilibrium state the temperature tends towards an equilibrium with this humidity field from which it would be kept in operational forecasts by increments during data assimilation. We note here that although these results will be somewhat influenced by the gradual reintroduction of moisture to the lowermost stratosphere identified in the previous subsection, in simulations for which the humidity field itself was not directly modified, but instead only modified when calculating the radiative increment to the temperature at each timestep (not shown) results are qualitatively similar to QMOD, with very slightly closer temperatures to analysis at longer times (by no means is the cold bias development completely prevented).

From the idealised experiments of Chapter 2.5, the expected radiative response of the near-tropopause temperature to an over-abundance of water vapour in the lowermost stratosphere is one of both excessive cooling in the lower stratosphere and also of warming in the upper troposphere. However, in the comparisons of operational forecasts to the observations in Chapter 2 such a tropospheric warming bias in the forecasts was not present when compared against the radiosonde observations. Here too in the upper troposphere the CTRL forecasts are not systematically warmer than the analysis. With the reduction of lowermost-stratospheric humidity in QMOD, not only is there the expected reduction in development of the stratospheric cold bias, but there is also additional cooling in the upper troposphere (Figure 3.4 (a) and (c)). This is consistent with what we would expect from the introduction of a drier layer above the upper troposphere (Cau *et al.* 2005). However, the difference between the upper troposphere in CTRL and QMOD is less than that in the lowermost stratosphere, as we expect the shorter mixing timescales of the troposphere to more rapidly dissipate any radiatively induced temperature differences. Nevertheless it would seem that there may be some compensating errors in the model's upper troposphere, and if the lowermost-stratospheric humidity values are brought closer to the observed values, in addition to the reduction of the lower stratospheric cold bias, a cold bias may develop in the upper troposphere.

Despite their similarity in the lowermost stratosphere there is a notable difference between the three investigated model resolutions is in the troposphere (Figure 3.4 (c)) with colder temperatures at lower resolutions. A similar difference can be seen in Figure 3.2 (c), with moister temperatures at lower resolutions at day 5. The tropopause in the lower resolution forecasts is also moister than the analysis at day one (not shown), which is likely to be a result of initialising these with fields interpolated from a higher-resolution analysis, and consequent adjustment within the first day, and therefore not relevant to the aims of this study.

We are defining the tropopause based on the temperature lapse rate, therefore systematic changes in temperature will also be expected to result in systematic changes in the identified tropopause altitude. It has been explained in Chapter 2 how the structure of the identified temperature bias resulting from the lowermost-stratospheric moist bias would lead to a systematic positive bias in the tropopause altitude in forecasts. Figure 3.5 (a) illustrates how the reduction of lowermost-stratospheric humidity in QMOD shifts the distribution of the tropopause altitude difference from the analysis towards lower values as compared to the CTRL. In the CTRL forecasts the tropopause is systematically at a slightly higher altitude than the analysis, increasing with lead time, while this bias is reduced for QMOD. The difference between the five-day QMOD and CTRL frequency distributions for TCo399 from Figure 3.5 (a) are shown in Figure 3.5 (b) along with those from TCo199 and TCo1279 further highlighting this shift and indicating that it too is largely independent of horizontal resolution.

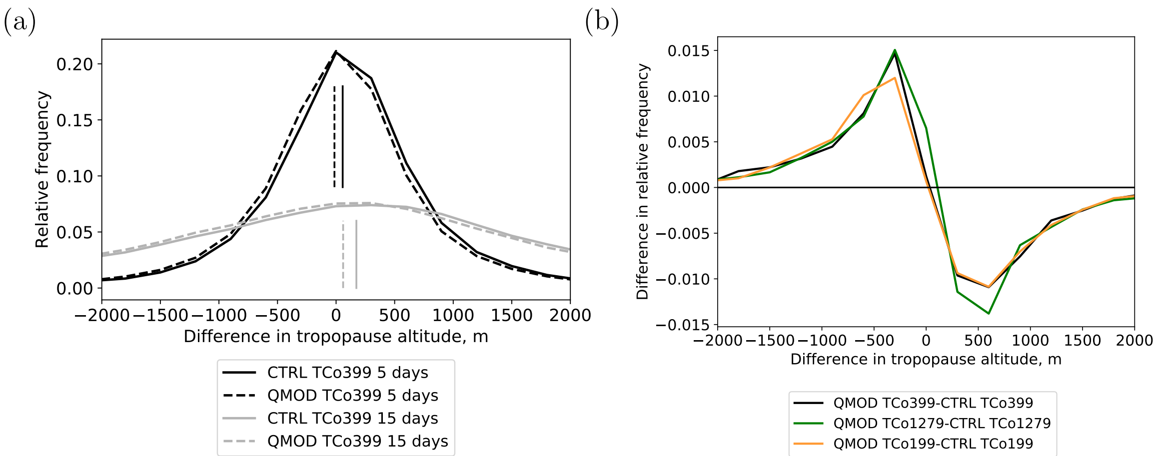

Figure 3.5: (a) Relative frequencies of difference in tropopause altitude between TCo399 forecast and analysis using 300 m bins 5- (black) and 15- (grey) days into the forecast for both CTRL (solid) and QMOD (dashed). Vertical lines indicate the means for each distribution. (b) QMOD minus CTRL differences in the relative frequencies as in panel (a) at day 5 for TCo199 (orange), 399 (black) and 1279 (green).

3.5 Model processes controlling the differences in forecast evolution over the first five days

In this section we make use of the partition of the changes to the prognostic model fields into Eulerian tendencies as introduced in Section 3.3.3. Considering these tendencies we investigate both the factors influencing UTLS temperature, which lead to the development of the forecast lowermost stratospheric cold bias, and those controlling water vapour, to elucidate mechanisms that are most likely to contribute to the presence of the model extratropical lowermost stratospheric analysis moist bias. The physical processes affecting both temperature and specific humidity shown in the following subsections are vertical diffusion, total cloud processes and convection. For both temperature and humidity, also shown are the cumulative changes to the temperature and humidity fields following the solution of the model dynamics in each time step. For temperature we additionally consider the effects of long-wave and short-wave radiation and gravity wave drag.

3.5.1 Processes controlling temperature

The model processes impacting temperature which we discuss here are the same as those processes identified in affecting UTLS temperature identified in Section 1.2: radiation, latent heating from cloud and convection schemes, and transport through both the model dynamics and vertical diffusion. Before examining the ways in which the influence of the process tendencies on temperature differs between the CTRL and QMOD forecasts, first consider Figure 3.6 (a) which illustrates the mean 12-hourly impact of each of the partitioned processes on temperature. For the CTRL forecast it is clear to see the expected signals of long-wave cooling throughout the depth of the UTLS (which is stronger in the moister upper troposphere), warming through the absorption of short-wave radiation following a similar pattern, and latent heating from the cloud and convection schemes below the tropopause. The model dynamics has a mean warming effect at all tropopause-relative altitudes. Only in the troposphere are increments to temperature by cloud and convection parameterisations of comparable magnitude to radiation, dynamics and vertical diffusion, and the mean impact of gravity wave drag is so small as to not be visible in this figure.

A conclusion drawn from Figures 3.3 and 3.4 is that the lowermost stratosphere cold bias develops more slowly in QMOD than CTRL forecasts. This is also evident from Figure 3.6 (b), with the total mean QMOD minus CTRL difference in 12-hourly temperature change above the tropopause being positive. It is also evident that this difference is primarily due to a reduction of long-wave cooling for QMOD in the lowermost stratosphere, which is expected as a result of the reduction of water vapour in this region. Also expected as a result of the removal of the moist bias from the lowermost stratosphere (and shown in

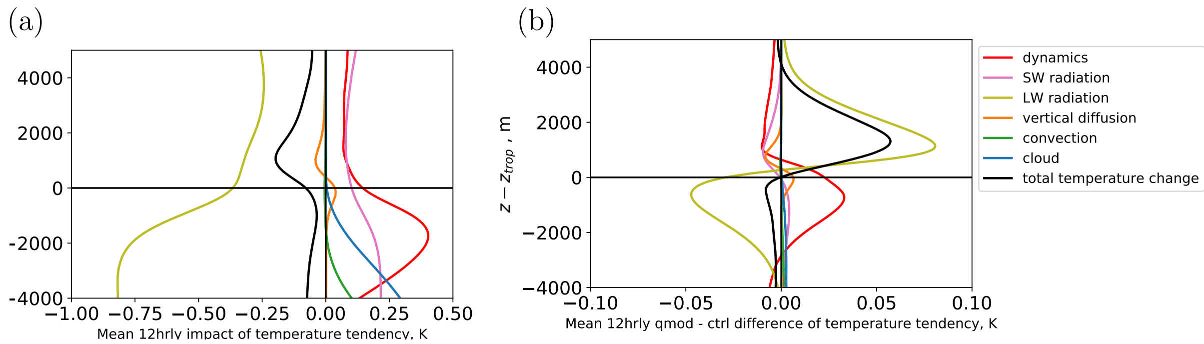

Figure 3.6: Tropopause-relative mean of (a) the Eulerian mean temperature tendencies in a 12-hour period for CTRL and (b) the QMOD minus CTRL difference between these tendencies, where a mean is taken over consecutive 12-hour periods between 24 and 120 hours into the forecast, for model resolution TCo399.

Figure 3.6 (b)) is an increase in long-wave cooling below the tropopause in QMOD. Despite this expected increase in long-wave cooling, there is only a small increase in total upper-tropospheric cooling in QMOD as compared to CTRL. This small increase is because much of the additional tropospheric long-wave cooling is balanced by additional warming from the model dynamics. This additional upper-tropospheric dynamical warming in QMOD appears (through comparison to Figure 3.6 (a)) to be a slight amplification of the dynamic warming balancing the long-wave cooling in the upper troposphere in CTRL, before latent heating takes on a more dominant role lower down. Differences in convection and cloud between QMOD and CTRL are negligible in comparison to those in long-wave radiation. The difference in vertical diffusion temperature tendencies acts to smooth out the temperature dipole introduced by the differences in radiation, as one would expect, and is also of a much smaller magnitude than the difference in the long-wave radiation tendencies.

3.5.2 Processes controlling humidity

In the CTRL forecasts the mean UTLS humidity is approximately in an equilibrium state. This can be seen from Figure 3.2 (a), which shows that the only time variation in the CTRL simulation mean vertical humidity profile is a slight moistening in the lowermost stratosphere (which we will address later). As for temperature the specific humidity tendencies are of the processes identified in Section 1.2. Here too transport is separated into the resolved advection resulting from the solution of the model dynamics, and the parameterised part from the vertical diffusion scheme. The breakdown of mean 12-hourly increments to specific humidity applied by the partitioned processes in Figure 3.7 (a) (and zoomed in upon in (c)) shows that the reduction in atmospheric water vapour due to condensation to cloud is largely balanced by the model dynamics. This smaller residual humidity tendency from the sum of the cloud and dynamics is in turn balanced by vertical diffusion. Hence, the specific humidity is in an approximate equilibrium state between cloud, dynamics and vertical diffusion in the UTLS.

In the QMOD forecasts we unbalance this equilibrium with the initial removal of the lowermost-stratospheric analysis moist bias. To see the impact of this removal we consider the differences in the twelve-hourly development of the humidity field between QMOD and CTRL as shown in Figure 3.7 (b). The mean QMOD-CTRL difference in the mean Eulerian fractional change in humidity over twelve hours, which is bringing the QMOD humidity field back towards that of the CTRL, is bimodal. The dry layer is being eroded from both the top and bottom over the course of the forecast as can be seen by the total fractional humidity change (black line). At the top of the dried region, around 2–

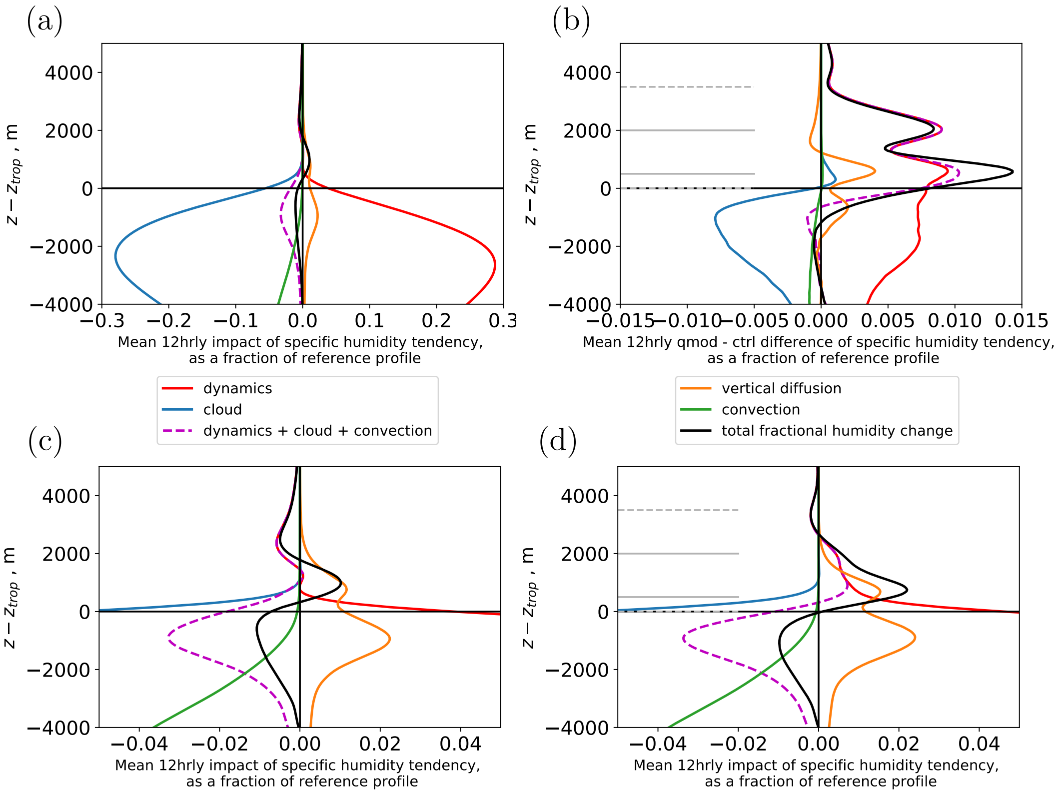

Figure 3.7: (a) Mean CTRL specific humidity tendencies and (b) the QMOD minus CTRL difference between these tendencies, as for Figure 3.6 (a) and (b), where here tendencies are as a fraction of the local humidity profile. (c) shows the same lines as (a), but with a narrower x-axis. (d) is the same as (c) but using data from QMOD rather than CTRL. Grey horizontal lines are as in Figure 3.1 to indicate the region in which the humidity was modified in QMOD.

3 km above the tropopause, the additional relative moistening can be attributed to the dynamics. It can be seen in Figure 3.7 (c) that in the CTRL this region is dominated by subsidence drying. However, for QMOD in Figure 3.7 (d), as we have dried this region, the subsidence within the lower stratosphere no longer results in drying. At some levels around 2 km above the tropopause this goes as far as to reverse, giving a moistening.

Also shown by Figure 3.7 (c) is that below the region of subsidence drying the tropopause-relative Eulerian trend is for the model to moisten the lowermost stratosphere through vertical diffusion for CTRL. Notably, in this tropopause-relative altitude framework there is no apparent sink locally balancing this source of moisture from vertical diffusion in the lowermost stratosphere. It can be seen in Figure 3.2 (c) that CTRL becomes moister than the analysis over time in the lowermost stratosphere. Therefore although the analysis is moist biased, it is still better at representing tropopause sharpness and lowermost stratospheric dryness than the model equilibrium state. This can also be found from considering that analysis increments dry the lowest $\sim$1.5 km of troposphere (not shown), although to a lesser extent than the moistening from vertical diffusion, and to an insufficient extent to prevent the presence of the moisture bias. For QMOD, due to the correction of the moist bias in the lowermost stratosphere, the tropopause-relative Eulerian trend for moistening is increased. Figure 3.7 (b) shows that at the bottom of the dried region the additional relative moistening can be attributed to a combination of vertical diffusion and the dynamics. This moistening from the dynamics is likely to be local transport of moisture across the sharp gradient around the tropopause, with some aspect of the contribution from the model dynamical core being due to numerical mixing.

An additional large difference between CTRL and QMOD shown by Figure 3.7 (b) is in the change to humidity by condensation to cloud and the model dynamics in the upper troposphere. As in Figure 3.7 (a) the terms continue to cancel, but there is more drying due to condensation to cloud in the upper troposphere in the QMOD forecasts than in the CTRL forecasts. The upper troposphere is cooler in QMOD than CTRL, and hence has a lower saturation vapour pressure, which is a likely reason for this increased quantity of upper-tropospheric cloud development. This increased removal of water vapour through condensation to cloud is balanced by an increased moistening by the model dynamics.

As well as being associated with a minimum in temperature and a sharp change in static stability, in the vicinity of the tropopause there is also a sharp change in the vertical gradient of specific humidity, accompanied by a local humidity minimum called the hygropause (which we further amplify when initialising the QMOD forecasts). The net moistening of the lowermost stratosphere we see for CTRL in Figure 3.7 (c) is accompanied by a slight drying above, and a larger drying below. Necessarily (as the gradient of the sum of two fields is equivalent to the sum of the gradients), the humidity gradient across the tropopause is reduced between the local maxima of drying below and moistening above. Additionally, moistening in the lowermost stratosphere and drying above lead to an increase in the altitude of the humidity minimum. This is consistent with the result from Chapter 2 for temperature, that over the course of five-day forecasts the tropopause is smoothed, and the altitude of the tropopause systematically increases. When we artificially sharpen this moisture gradient in QMOD, it is smoothed even faster by the model over the course of the forecast. That the CTRL forecasts moisten in the lowermost stratosphere can also be seen earlier from Figure 3.2, compared to the analysis.

It is important to note that this smoothing of the humidity field is a very local effect. The use of Eulerian tendencies in the vicinity of the tropopause which is a very dynamic surface means that when using a different vertical coordinate (one which is not tropopause-relative), or using much longer timescales, the signal of vertical diffusion moistening the lowermost stratosphere is not as apparent (not shown). This lack of signal is because the tendency for moistening is smoothed in the vertical with the mean drying above and particularly below: remembering that humidity decreases logarithmically with altitude and that Figure 3.7 shows tendencies as a fraction of the background humidity field, the local signal of drying in the upper troposphere is an order of magnitude larger than the local moistening above.

3.6 Effects of horizontal resolution on humidity balance

To investigate further the possibility that excessive horizontal cross-tropopause mixing results from an insufficient horizontal spectral resolution, the set of 30 five-day forecasts have been run at the three resolutions TCo199, TCo399 and TCo1279. The main conclusion from examination of these forecasts is that differences between the three considered horizontal resolutions are fairly small in proportion to the fields themselves.

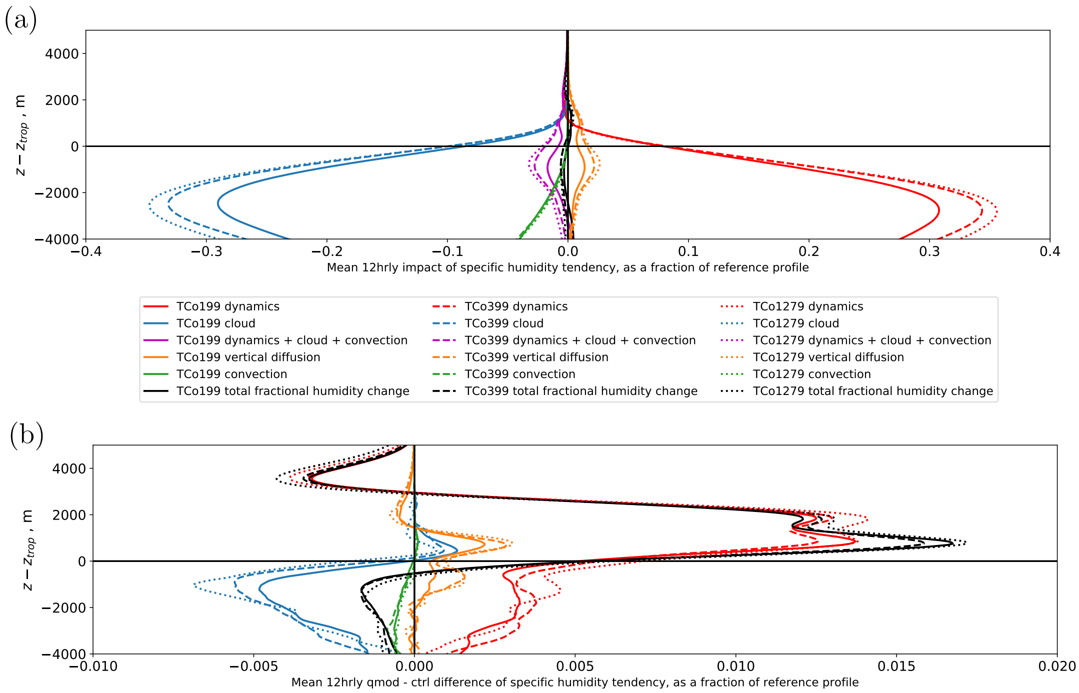

Figure 3.8: (a) Mean CTRL specific humidity tendencies and (b) the QMOD minus CTRL difference between these tendencies, in exactly the same way as Figures 3.7 (a) and (b). Here solid lines use data with forecasts at horizontal resolution TCo199, dashed lines TCo399 and dotted lines TCo1279. The difference is between humidity tendencies at 48 hours and 24 hours into forecast, and halved such that the scale is comparable to the 12-hourly differences in Figure 3.7; this rescaling is required because the data are only available daily for the highest resolution forecast. Note that the aspect ratio of these figures is different to that of previous figures to make the (small) differences visible.

It is shown by Figure 3.8 (a) that the total magnitudes of net condensation to cloud, addition of water vapour by dynamics, vertical diffusion, and net reduction of upper-tropospheric humidity by the difference between dynamics and cloud, all increase with increasing horizontal resolution. We note here that explicit diffusion does not depend on resolution. However, the total mean change to the humidity profile is nearly zero because these changes to tendency fields with resolution are in proportion to one another such that the net total change to the humidity field in the CTRL is approximately zero, maintaining the equilibrium state.

It was learned from Figure 3.2 that the TCo1279 QMOD forecast returned towards analysis faster than in the lower resolution forecasts. Figure 3.8 (b) illustrates that this slight difference is due to both an increase in the remoistening of the top of the dried layer by subsidence, and increased vertical diffusion across the tropopause. This difference is despite the vertical level spacing being the same across all three horizontal resolutions.

3.7 Conclusions

Motivated by the presence of cold biases in IFS medium-range forecasts of the lowermost stratosphere linked to moist bias in the same region, we have endeavoured to initialise forecasts with an initial humidity field that is as realistic as possible (QMOD) so as to compare the temperature and humidity forecasts with those initialised from the unmodified analysis (CTRL). As humidity measurements are not used to constrain the analysis state above the tropopause these humidity values are instead dependent on the model configuration, leading to a moist-biased “model climatology”. QMOD has on average around 40% of the water vapour removed in the lowermost stratosphere to bring it closer to observations at initial time. The first two aims were to determine the differences in the mean vertical structures of specific humidity and temperature in the Upper Troposphere Lower Stratosphere (UTLS) between CTRL and QMOD over 5- and 15-day forecasts as compared to one another and the analysis. The third aim was to attribute these temperature and humidity biases to model processes, and the fourth was to determine any sensitivity of these results to horizontal model resolution.

Addressing the first two aims, following the removal of humidity from the lowermost stratosphere in QMOD the humidity in the corrected region returns to analysis values with a half-life of 8–9 days. Also during this time the development of the lowermost stratosphere cold bias is approximately halved in QMOD as compared to CTRL. Additionally, the systematic bias of the tropopause being at too high an altitude in CTRL is reduced for QMOD.

Addressing the third aim of attribution, the reduction in the lowermost stratosphere cooling is almost entirely a result of a reduction in long-wave radiation when the lowermost stratosphere has less water vapour. There is little concurrent increase in cooling in the upper troposphere in QMOD because the increase in below-tropopause radiative cooling is largely balanced by the model dynamics. Moisture is reintroduced to the layer from which it was removed both at the top and bottom by the model dynamics, and additionally at the bottom by vertical diffusion. In the CTRL forecasts, over 12 hours in a tropopause relative frame of reference, vertical diffusion acts to moisten the lowest km of the stratosphere, and there is no sink of moisture to balance this. There is very little difference made to these results when comparing forecasts run at varying horizontal resolutions (fourth aim).

One key novel conclusion from this study is that changes to the horizontal resolution make very little difference to the overall rate of recovery of the moist bias. There is only a slight tendency for the forecasts run at the highest resolution to return towards the model climatology faster than the other two. If the bias were as a result of an overly-diffusive dynamical representation of horizontal cross-tropopause transport, we would expect a more rapid growth of the moist bias at the lowest horizontal resolutions. That we do not find this therefore indicates that this is not the dominant cause of the extratropical lowermost-stratospheric moist bias. This result seems contrary to the results of Stenke *et al.* (2008), who show that horizontal diffusion in the ECHAM4 model is the main contributor to their wet bias by comparing a new Lagrangian advection scheme to the existing semi-Lagrangian one. This difference may be due to the comparatively lower spectral resolution of T30 and 39 vertical levels used in their experiment. Furthermore, the relative remoistening of the lowermost stratosphere in QMOD compared to CTRL is not substantially achieved through changes to the treatment of water in the cloud microphysics scheme. This independence of cloud processes from the lowermost stratospheric humidity field, in conjunction with the model tropopause being if anything too cold (Polichtchouk *et al.* 2019a), suggests that local cloud processes have little control on the extratropical lowermost stratosphere moist bias. In addition to identifying that horizontal resolution and cloud processes are not important to the lowermost stratosphere moist bias we have confirmed and quantified the relative importance of the contributions from vertical diffusion and the dynamics, and the magnitude of their response to the imposition of a more realistic lower stratospheric humidity field.

The result that the bias seems to be largely controlled by vertical diffusion and the dynamics is consistent with the conclusions of Krüger *et al.* (2022) that the moist bias is likely to be a result of the model representation of mixing and can be used to direct future investigation. The dynamical aspect would either be as a result of a moist bias in the source region, a misrepresentation of the resolved transport, or excessive implicit diffusion from the dynamics in the vertical. Due to the short timescales of the moist bias reintroduction, and the isolation of the bias to only the lowest 2 km of the extratropical stratosphere, the key regions are most likely to be upwards mixing from the extratropical upper troposphere or pseudo-horizontal isentropic transport from the subtropical upper troposphere. However, as we find no dependence on model horizontal resolution it cannot be attributed to insufficiently resolved horizontal motion. The accuracy of resolved transport in forecasts can be assessed by comparing to observations or analysis products. The isentropic transport from the subtropical upper troposphere into the extratropical lowermost stratosphere is observed to occur often in thin layers (Vaughan and Timmis 1998, Bradshaw *et al.* 2002). A consequence is that limited vertical resolution in the model will mix the higher humidity from these layers too rapidly.

Furthermore, we find that the cold bias development in QMOD forecasts is reduced only to half of that in CTRL. It is noted that though there will be some influence from the remoistening of the lowermost stratosphere, this effect is small. Therefore to fully prevent the development of this temperature bias it may be insufficient to solely resolve the analysis moist bias and the tendency of the model to reintroduce this if removed. Experiments by R. Hogan (pers. comm.) have shown that improvements to the gas optics parameterisations can also lead to improvements in extratropical lowermost stratospheric temperatures. Nevertheless, it is clear from the results of this Chapter that improvement to the cold biased temperature in the extratropical lowermost stratosphere in forecast models is possible through correction of the representation of water vapour. It is demonstrated that it is insufficient to solely correct the specific humidity in the initial conditions as it is re-moistened, and rather it is most important to focus on improving vertical diffusion and advection in order to reduce the model moist bias.

Chapter 4

Impacts of lowermost stratospheric humidity biases on extended-range and seasonal forecast skill

Abstract

While skillful forecasts are possible at lead times of one to two weeks, a challenge is to increase the usefulness of the information that can be gained from longer-range forecasts. Because the simulated atmosphere is a chaotic system, relatively small misrepresentations can grow into large forecast errors given enough time. In this study we outline a mechanism by which sub-seasonal forecast skill can be increased and seasonal forecast bias reduced by improving the model representation of the lowermost stratosphere. By performing experiments with a realistic, if simplistic, correction applied to humidity in the lowermost stratosphere, as seen by the radiative transfer scheme, we provide quantitative evidence of systematic forecast improvements in temperature and winds at all levels. It is shown that by preventing the development of a common extratropical lowermost stratospheric cold bias forecast skill is significantly improved in week three of forecasts, particularly over Europe and the North Atlantic. While there are also many improvements also to seasonal forecasts, particularly in Northern hemisphere winter (December-February), these are not globally uniform. It is hoped that knowledge of the impacts of this moist bias which exists in current models (as opposed to general knowledge of the impacts of water vapour) will motivate further research to reduce this bias.

4.1 Introduction

In simulations of the atmosphere using numerical weather prediction (NWP) and climate models, the temperature in the lowermost stratosphere in the extratropics becomes colder than is observed (Polichtchouk *et al.* 2019a). A large portion of this can be attributed to a collocated moisture bias (Chapter 2). Evidence suggests that this moist bias is due to a misrepresentation of advective transport and mixing, however increasing vertical resolution is computationally expensive and difficult, as parameterisations can require tuning to the vertical resolution. Furthermore, unlike the surface boundary layer the tropopause layer is mobile and so it would be challenging to try to increase resolution in just this region. The strength and locations of the subtropical jets and the Hadley circulation have been shown to be sensitive to lower stratospheric temperature (Tandon *et al.* (2011); Boljka and Birner (2022)), misrepresentations of which can lead to forecast errors in values throughout the atmosphere. To motivate the effort and justify the cost of correcting the lowermost stratospheric biases shown to exist in Chapter 2 and Chapter 3, it is important to have information on the quantitative benefits which can be attained. Hence, the purpose of this chapter is to provide guidance as to these benefits.

The dominant effect of the extratropical lowermost stratospheric specific humidity bias is to cause excessive cooling due to the excessive emission of long-wave radiation (Shine and Myhre 2020). In previous Chapters this excessive cooling is demonstrated to increase the mean tropopause altitude and smooth cross-tropopause gradients, reducing tropopause sharpness. Harvey *et al.* (2016) show that smoothing PV gradients reduce Rossby wave phase speed, due to the reduction of advection dominating over the reduction of the ability of waves to propagate upstream. However, an extratropical cold bias will also lead to an increased meridional temperature difference, which would increase the thermal wind. The accurate representation of planetary-scale wave propagation along the gradients at the tropopause are important for forecasts of fields throughout the depth of the troposphere. One mechanism for this is the interactions of upper- and lower-level waves during baroclinic extratropical cyclone development (Hoskins *et al.* 1985). If waves on the tropopause are systematically moving eastwards unrealistically quickly in forecasts, then there will be a systematic error in the phase shift between upper- and lower-level waves.

Here we use a correction to the water vapour seen by the radiation scheme carefully calculated from observations to compare skill and error between corrected forecasts (QMOD) and control forecasts (CTRL). Whereas many previous studies have focused on multi-annual timescales, because it is shown in Chapter 3 that the cold bias develops on timescales of weeks, in this study we consider rather the first few weeks and months of the forecast. In this way we can identify achievable improvements in sub-seasonal to seasonal forecasts. The observations the humidity correction are based on were made during a field campaign investigating cyclone development over the North Atlantic in September and October (Schäfler *et al.* 2018), and so we look with particular interest at this and other storm track regions. In Section 4.2 the method used to correct the humidity field for the QMOD forecasts is detailed, as are the skill metrics used for the evaluations. In order to first provide the context of the direct consequences of the applied humidity correction, the effect of the correction on temperature and zonal wind at tropopause level is shown in Section 4.3. This is followed by investigation of impacts on a broader range of fields. We analyse systematic effects of the correction on forecast skill in the medium-range in Section 4.4, and on mean error in seasonal forecasts in Section 4.5. This is concluded by summary and discussion in Section 4.6.

4.2 Methods

4.2.1 Correction of humidity bias

To evaluate the difference in forecast evolution resulting from the development of an extratropical lowermost stratospheric cold bias consequent from a collocated moist bias we run one set of forecasts in which the cold bias is allowed to develop as normal (CTRL) and a second set in which the humidity bias is corrected in the input to the radiation scheme (but not the advected humidity field itself) to inhibit the development of the cold bias (QMOD) in an attempt to keep values close to the analysis. The humidity correction used is identical to that shown in Figure 3.1. However, it is shown in Chapter 3 that if the humidity correction is applied at the initial time step and the forecast is subsequently allowed to run freely, the moist bias is reintroduced with a half-life of around 8 days. Therefore we instead do not modify the specific humidity field directly, but rather reduce the humidity that the radiation scheme “sees” for our QMOD forecasts. In doing so we assume that the background humidity is in equilibrium with the initial unbiased temperature, and hence the prognostic QMOD specific humidity field will maintain a constant moist bias as in CTRL; hence, applying a constant humidity modification is suitable throughout the forecast duration. That CTRL (and equivalently operational forecasts) maintain an approximately constant moist bias during forecasts is illustrated in both Chapters 2 and 3. While it will be the case that any differences in temperature between QMOD and CTRL resulting from the perturbation applied to QMOD will impact the equilibrium lowermost stratospheric humidity, the humidity field in the QMOD of Chapter 3 when directly modified still returns towards roughly the same equilibrium state as CTRL (as shown by Figure 3.2) despite the warmer lowermost stratosphere illustrated in Figures 3.3 and 3.4.

It is worth noting that the uniform application of the humidity correction as in Chapter 3 is not entirely without issue. The humidity correction is constructed using only data from Northern Hemisphere (NH) Autumn. However Dyroff *et al.* (2015) identify that the moist bias is larger than this in NH summer. We therefore expect that analysis of simulation data for NH June, July, August (JJA) may underestimate any differences from CTRL that could result from a more accurate humidity correction in QMOD. Furthermore we are also assuming that the dominant impact of the moist bias is via the radiation scheme in all seasons in the same way that it has been demonstrated to be for NH Autumn in Chapter 3. The method used in this Chapter will also miss any impacts through other mechanisms that the correction of this large summer moist bias would result in.

4.2.2 Extended and seasonal range ensemble suites

Both seasonal and extended range forecasts are run as ten member ensembles using the European Centre for Medium-Range Weather Forecasting (ECMWF) Integrated Forecasting System (IFS) version 47r1 with horizontal resolution TCo199 and 137 model levels in the vertical. For the extended-range experiment, 324 forecasts are initialised monthly from 1989 to 2016, and run for 45 days. For the seasonal experiment, 70 7 month forecasts are initialised on the 1st May and the 1st November from 1981 to 2016.

4.2.3 Forecast skill

The measure which will be used in later sections to quantify forecast skill is the Continuous Ranked Probability Skill Score (CRPSS). The continuous ranked probability score (CRPS) is defined in Mylne *et al.* (2021) as being "a quadratic measure of the difference between the forecast cumulative distribution function (CDF) and the empirical CDF of the observation," where the forecast cumulative distribution is over the ensemble and the empirical CDF is a Heaviside function. CRPS is calculated as

$$CRPS = \int_{-\infty}^{+\infty} [F(x) - H(x - x_0)]^2 dx,$$

where $H(x)$ is the Heaviside function, x_0 is the verifying value, and $F(x)$ is the distribution of the ensemble forecast. The skill score is then calculated from this as

$$CRPSS = 1 - \frac{CRPS}{CRPS_{ref}},$$

where $CRPS_{ref}$ is a reference score calculated from climatology. A CRPSS of one therefore indicates an ensemble in which all members perfectly replicate the verifying value, and a CRPSS of zero represents a performance no better than a climatological forecast.

4.3 Tropopause-level biases

The imposed forcing in this experiment directly affects the temperature (T) immediately above the tropopause level, so first we evaluate the success of the bias reduction for this and the closely related zonal wind (U) on a seasonal timescale. Seasonal forecasts are initialised in November and May, and shown are the December-January-February (DJF) and JJA means respectively.

From Figure 4.1 (a-c) it is clear that the NH winter extratropical lowermost stratospheric cold bias is dramatically reduced. The temperature bias is reduced from nearly $-4K$ in CTRL to less than half of this in QMOD. In NH summer (Figure 4.1 (g-i)) the cold bias is also reduced by around $4K$ from $\approx -6K$ to $\approx -2K$. As the moisture bias correction was based on autumn data and it is known that the biases are larger in summer, it being insufficient to fully remove the bias here is to be expected. Furthermore, while it is clear that the temperature is improved in southern hemisphere midlatitudes, the correction seems to be inappropriate in southern polar regions, creating a slight warm bias. Nevertheless, the overall picture shows a representation of global zonal mean temperature closer to that in the ECMWF Reanalysis v5 (ERA5) product in QMOD than in unmodified forecasts.

The notable biases in zonal mean zonal wind of CTRL compared to ERA5 as shown in Figure 4.1 (d) and (j) are that the subtropical jets are too strong by up to $10ms^{-1}$, too poleward, and at too low a pressure. The main consequence of the correction applied for

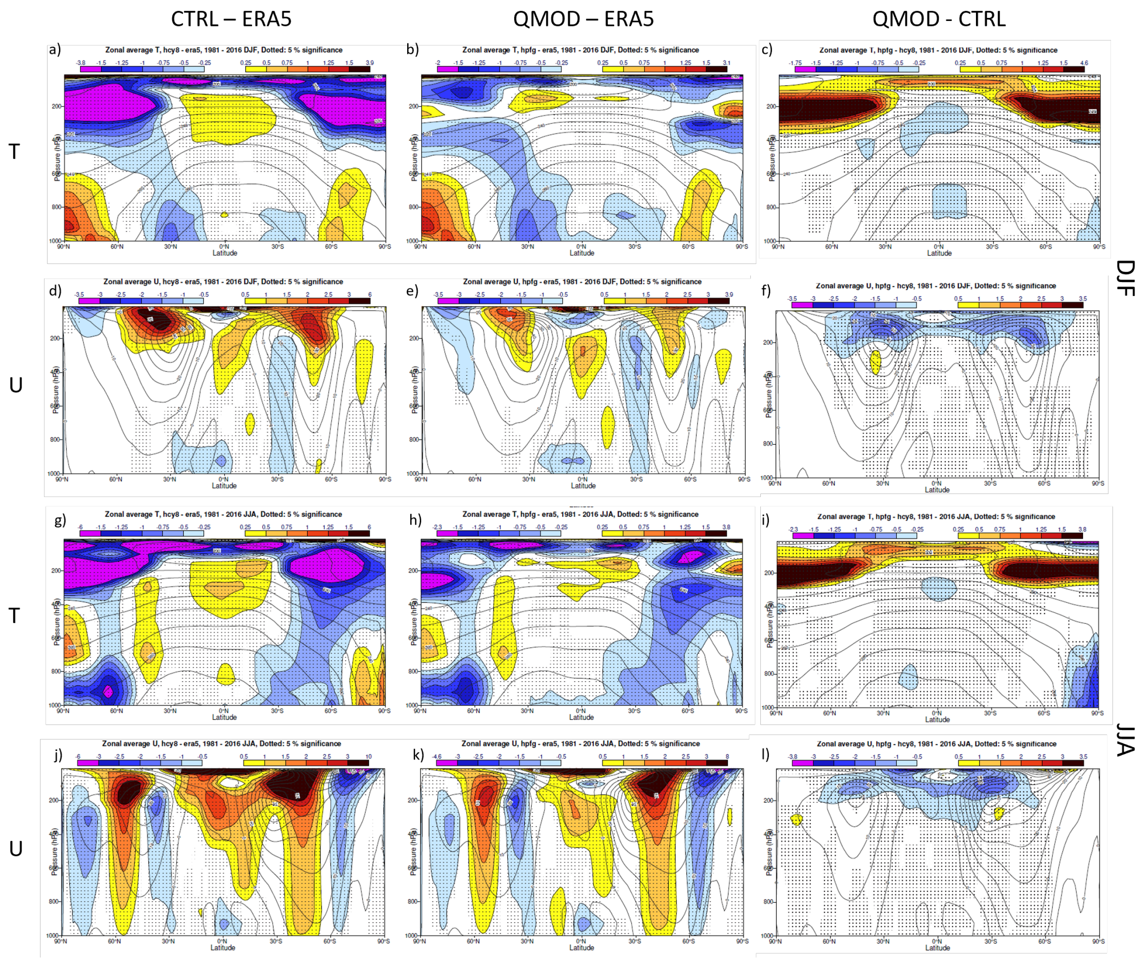

Figure 4.1: Zonal mean (a-c,g-i) temperature and (d-f,j-l) zonal wind differences between (a,d,g,j) CTRL minus ERA5, (b,e,h,k) QMOD minus ERA5, and (c,f,i,l) QMOD minus CTRL for (a-f) DJF and (g-l) JJA. Dotted regions indicate significance at the 5% level. Black contours show ERA5 (a-c,g-i) potential temperature, and (d-f,j-l) zonal wind. Note that the north pole is to the left of the figures, and the south pole to the right.

the QMOD forecasts on the zonal winds is a reduction in zonal velocity above the jets. This may be driven by a reduction in the cross-tropopause temperature gradient. It can also be seen from Figure 4.1 (f) and (l) that the maximum in wind reduction is slightly poleward of the jet maximum, particularly in the summer hemisphere. Hence, the correction acts to improve the bias aspects identified at the start of this paragraph, reducing jet speed and moving it equatorward and towards the surface. As with temperature, the correction applied to QMOD does much better at making forecasts better resemble the reanalysis in DJF than JJA.

4.4 Extended-range skill

With the impacts of the forcing established, we now consider the consequent changes in extended-range forecast skill, beginning with domain averages and then considering spatial distributions.

4.4.1 Domain averages

Regional mean changes to CRPSS are shown in Figure 4.2, and the shorthand variable names used in the leftmost column are explained in the subsequent Table 4.1. Across the five considered domains there is very little impact on forecast skill in weeks one and two. It is shown in Chapter 3 that the cold bias develops in CTRL at a rate of only 0.2 $Kday^{-1}$, therefore it takes several weeks for the difference between QMOD and CTRL to become sufficiently large to affect differences in the development of other fields. The largest and most significant differences are at week three, particularly in the northern hemisphere, over the North Atlantic and Europe.

These differences are present throughout the depth of the atmosphere—not only at the altitude of the perturbation—as a consequence of the improved representation of radiation and upper-level circulation. Probable mechanisms for this are explored in Section 4.6. However, at week four there are very few significant differences in forecast skill. Forecast skill for CTRL is close to zero in this fourth week and similar is true for QMOD, which suggests that this is the limit to which additional forecast skill can be extended through this improvement.

4.4.2 Spatial distributions

Following the knowledge that the humidity correction leads to mean improvement in forecast skill at week 3 (days 19–25), we now look in more detail at some of the pertinent fields highlighted by the scorecard at this lead-time.

Pressure levels 200 hPa and 50 hPa are shown in Figure 4.3 to reflect either side of the region of simulation perturbation, and as spatial patterns of changes in CPRSS for zonal and meridional winds at 500 hPa and 850 hPa are very similar to those at 200 hPa.

Figure 4.2: Difference in Continuous Ranked Probability Skill Score between QMOD and CTRL forecasts for variables described in Table 4.1 over seven-day intervals starting at days 5, 12, 19 and 26 into forecasts, for the five regions of the Northern Hemisphere (NHEM), Tropics, Southern Hemisphere (SHEM), North Atlantic (NATL) and Europe. Blue triangles indicate an increase in CRPSS hence an improvement in skill, and red triangles a decrease. Darker colours indicate significance at the 95% level, and triangle size indicates the magnitude of the CRPSS change.

Consistent with Figure 4.2, the picture for temperature and wind speed is one of an improvement in QMOD over CTRL, most notably over the northern hemisphere. This is far from being uniform however, with reduction in skill apparent over the Southern Atlantic and Asia at 200 hPa, for example. Nevertheless, the patterns of regions of significant increase in CPRSS at 200 hPa and 50 hPa suggest that QMOD may have a better representation of the tropospheric jets and Arctic polar stratospheric vortex, respectively. The improvements are greatest in v at 50 hPa over the Euro-Atlantic sector.

While the pattern in ΔCRPSS for winds is similar throughout the troposphere to that at 200 hPa, this is not the case for temperature. This can be seen in Figure 4.4 where surface

shorthand	variable
tprate	total rainfall
2t	2m temperature
msl	mean sea level pressure
u#	zonal wind at #hPa
v#	meridional wind at #hPa
t#	temperature at #hPa
strf200	streamfunction at 200 hPa
vp200	velocity potential at 200 hPa
rws200	rossby wave source at 200 hPa
sst	sea surface temperature
ci	sea ice concentration

Table 4.1: Descriptions of the shorthand variable names from Figure 4.2.

temperature is shown along with geopotential height at 500 hPa. There is significant improvement in skill for Z500 in the storm track region over the North Atlantic, and also over Greenland and Eastern Europe. At the same time there is a significant decrease in skill in the tropics, although skill at three weeks is typically higher for the CTRL in the tropics than in the midlatitudes. There is no such clear pattern in two-metre temperature, but nevertheless the signal of improvement over the North Atlantic is still apparent.

Now we look further at the timeline of the development of this difference in the representation of the North Atlantic Oscillation (NAO). Here we use the NAO Z500 index - the principal component of the Empirical Orthogonal Function decomposition of the field of geopotential height at 500 hPa. Unlike for the bias and ΔCRPSS fields shown in Figures 4.1–4.4, differences in the NAO Z500 index correlation and root mean square error (RMSE) between QMOD and CTRL are not significant at many forecast lead times. Nevertheless, it is worth noting that although the correlation between CTRL and the analysis drops off at day 21 (Figure 4.5 (a)), there is no change in model resolution or configuration at this time to precipitate this. The lowermost stratospheric temperature improvement in QMOD also increases the ensemble spread at three weeks and slightly decreases RMSE, bringing these closer together, as shown in Figure 4.5 (b).

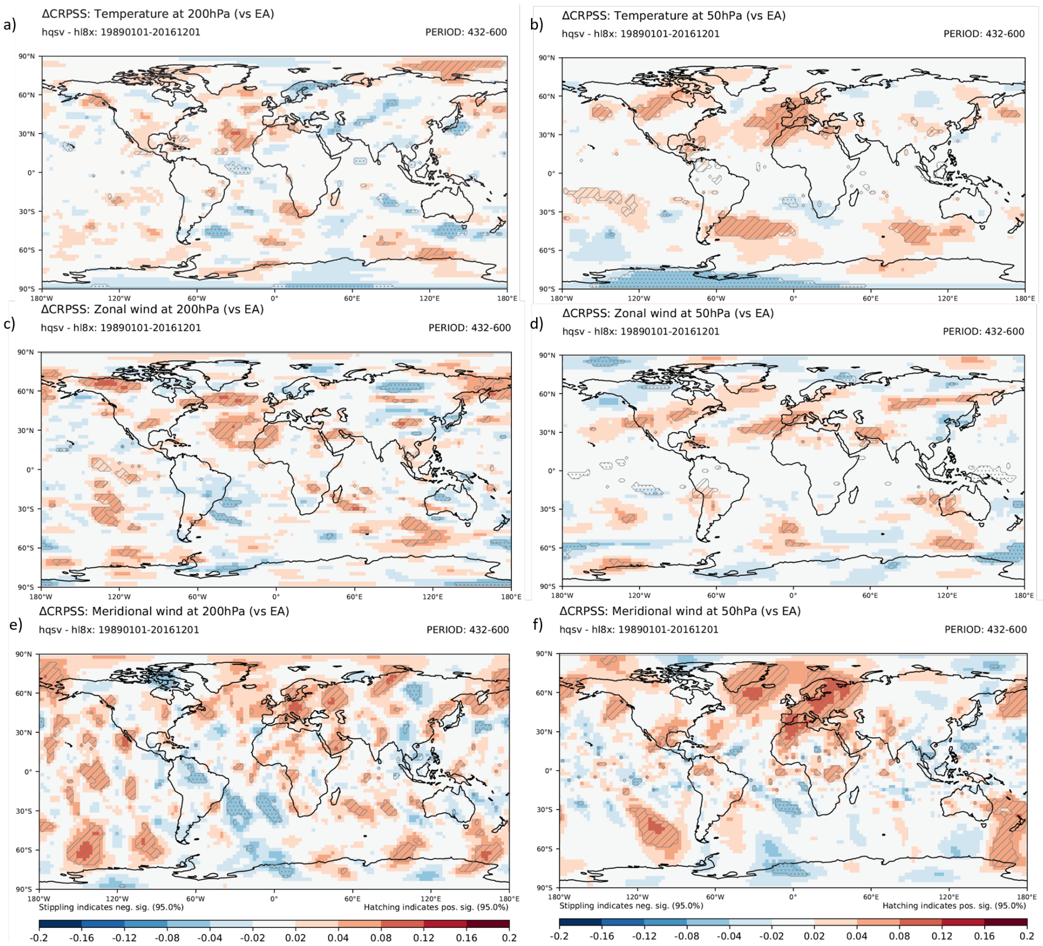

Figure 4.3: QMOD minus CTRL Difference in CRPSS. The opposite to Figure 4.2, red colouring indicates an improvement in skill of QMOD over CTRL and blue a degradation. Hatched (positive) and stippled (negative) regions indicate significance at the 5% level. (a, c, e) show pressure level 200 hPa and (b, d, f) 50 hPa. The fields concerned are (a, b) temperature, (c, d) zonal wind and (e, f) meridional wind.

4.5 Systematic seasonal effects

We move on now from sub-seasonal to seasonal systematic forecast improvements. Here we consider mean error compared to ERA5 of 2–4 month forecasts. In JJA equatorial zonal winds at 200 hPa are positively biased in CTRL as compared to ERA5, in that they are insufficiently easterly. This is improved in QMOD, as can be seen by comparison of Figures 4.1 (j) and (k). This improvement is more-or-less zonally uniform in structure. Another significant reduction in error of QMOD over CTRL is in polar Z500 in both hemispheres (not shown). Differences in DJF, however, tend to be more localised, which we examine in more detail below.

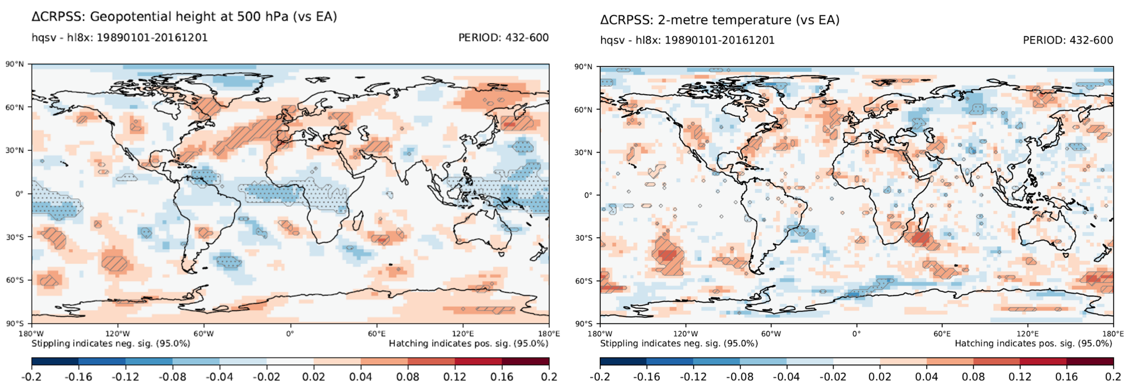

Figure 4.4: As for Figure 4.3, but for (a) Geopotential height at 500 hPa (Z500), and (b) 2-metre temperature.

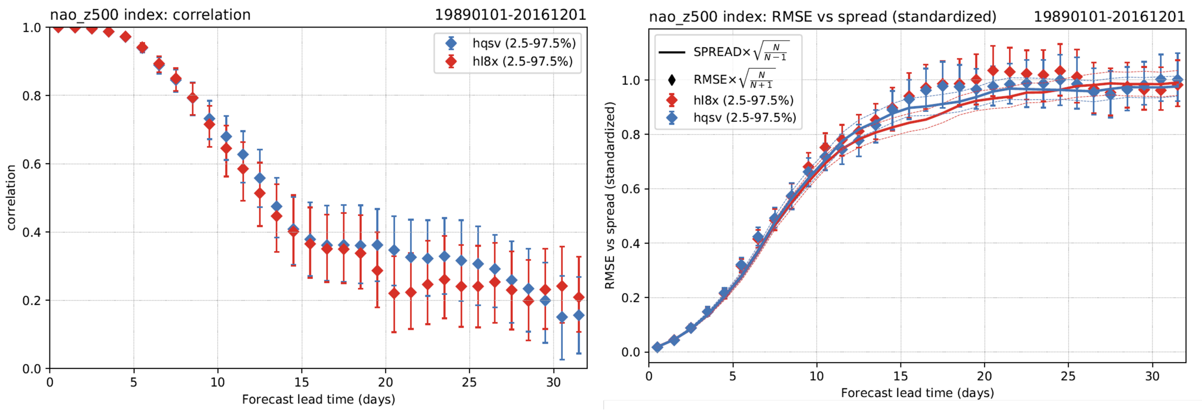

Figure 4.5: (a) Correlation of NAO Z500 index with analysis for CTRL (red) and QMOD (blue) with 95% confidence intervals. (b) RMSE of NAO index with confidence intervals (shapes), and ensemble spread (lines).

4.5.1 Northern Hemisphere Winter

In addition to showing that equatorial zonal winds at 200 hPa are positively biased, Figure 4.1 (d) and (j) also show that the zonal-mean subtropical jets are also excessively strong. This is apparent also from Figure 4.6 (a). Figure 4.6 (b) illustrates how the reduction in zonal wind bias is focused in the jet regions, particularly over ocean basins.

Similarly, while Figure 4.5 shows that QMOD better represents the NAO Z500 index than CTRL at a timescale of up to 30 days, Figures 4.6 (c) and (d) show that CTRL is under-representing the climatological ridge over Western Europe in months 2-4, and that this is improved in QMOD. The difference between seasonal forecasts and ERA5 over the eastern North Pacific is larger in QMOD than CTRL however, in what appears to be

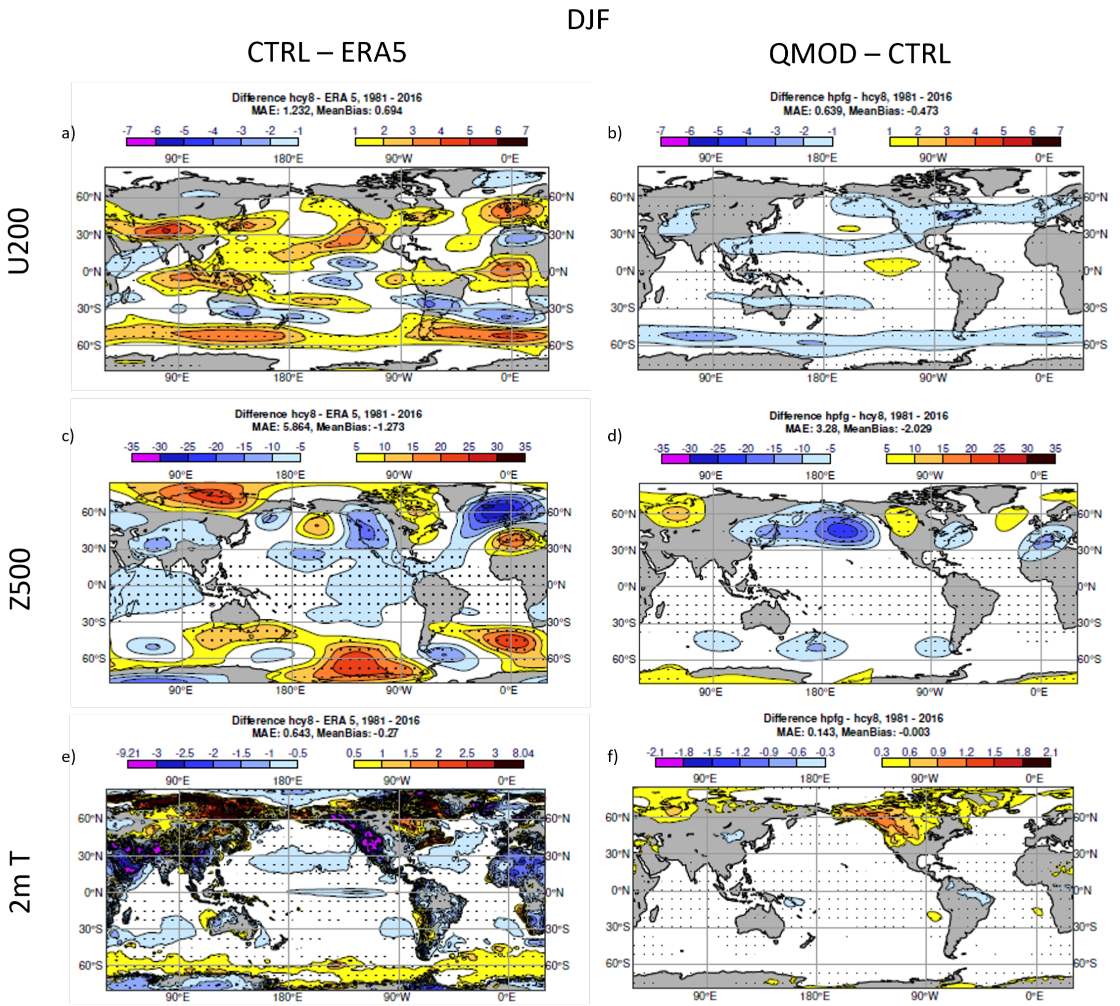

Figure 4.6: Mean difference in (a,b) 200 hPa zonal wind, (c,d) 500 hPa geopotential and (e,f) 2 metre temperature between (a,c,e) CTRL minus ERA5 and (b,d,f) QMOD minus CTRL. As before, stippling indicates significance.

an excessively western placement of the climatological trough over North America. This too while it is shown by Figure 4.6 (f) that North America is the location of the largest difference in DJF surface temperature between QMOD and CTRL.

4.6 Conclusions

Based on the understanding of temperature and humidity biases from previous Chapters, in this Chapter we have explored the impact that the presence of such biases have on the rest of the atmosphere at longer time scales. By removing the moist bias from the specific humidity field that the radiation scheme uses to calculate temperature increments of every time-step of extended- and seasonal- range forecasts it is shown that both the temperature and horizontal winds are improved at all levels. The largest improvements are at week 3 of forecasts and over the North Atlantic and Europe, particularly in the storm-track region. Although this method is unable to significantly improve skill at week 4 and beyond, reduction in mean biases can be seen out to at least 4 months.

A possible reason that many of the largest improvements to skill and bias are over the North Atlantic and Europe would be that one of the key large-scale processes sensitive to improvements to tropopause level temperatures is the development of extratropical cyclones. A physical mechanism for this is proposed in Section 4.1, in which the forecast temperature bias leads to biases in the structure of the tropopause layer and hence to the propagation of waves along it. By improving the temperature in the experiments in this Chapter, it is possible the improvements in the skill in the forecasts of the global circulation are due to the systematic improvement of the representation of the relative positions of upper-level waves which are important for the baroclinic development of cyclones in these regions.

Chapter 5

Conclusions

In this thesis extratropical lowermost stratospheric biases in temperature and humidity have been analysed to determine the detail of their structure, the reason for their presence in the model, and any detrimental impact they may have on forecast quality. With an overall motivation to make progress towards forecast model improvement, in the introduction we introduced four aims to be addressed in this thesis. Key conclusions of work pertaining to each of these aims are summarised below.

1. **Characterise the spatial and temporal structures of model biases in extratropical UTLS temperature, humidity and tropopause altitude**

   The vertical structures of the biases are determined through comparison to in-situ radiosonde observations in a tropopause-altitude-relative frame of reference in Chapter 2. It is shown that the model extratropical lowermost stratosphere specific humidity is moist biased by a factor of ≈ 2 in the analysis, and becomes cooler than observations at a rate of $\approx 0.2\ Kday^{-1}$. Both the humidity and temperature biases are shown to be approximately uniform throughout the extratropical lowermost stratosphere, although there is some discernible dependence on tropopause altitude with the biases above tropopause troughs extending deeper into the stratosphere. Additionally, in operational forecasts the mean tropopause altitude increases by $\approx 100mday^{-1}$.

2. **Quantify the relationship between these biases**

   When simulations are performed with the moist bias removed from the extratropical lowermost stratosphere, lowermost stratospheric temperatures remain closer to the analysis and do not develop a cold bias as quickly as unperturbed simulations. Running such simulations in an idealised single column model yields also the result of additional cooling in the upper troposphere following a reduction in lowermost stratosphere water vapour, but such an effect is not present in full atmospheric models (Chapter 2). Considering contributions from individual parametrized physical processes it is shown that the majority of these temperature differences are driven by changes in long-wave radiation, as would be expected as a consequence of the radiative properties of water vapour at this location in the atmosphere. Although the radiative effects of an excessively moist lowermost stratosphere would be to drive cooling above and warming below, these tendencies too show that the more dynamically active troposphere counters the development of a temperature bias here (Chapter 3).

From these results it is concluded that the dominant cause of the cold bias which develops in forecasts in the extratropical lowermost stratosphere is the collocated moist bias. It is acknowledged that this is not necessarily the sole cause, as some cold bias develops despite the humidity correction and other perturbations such as changes to gas optics have been shown too to improve this cold bias. A schematic is also provided in Chapter 2 (Figure 2.11) to illustrate the mechanism by which the heating/cooling dipole resulting from the moist bias causes a systematic increase in tropopause altitude in forecasts.

3. **Determine the reasons for the presence of these biases in models**

   Investigating the relationship between the biases has led to the conclusions that the tropopause altitude bias is as a result of the temperature bias, and the temperature bias is as a result of the water vapour bias. The question then remains as to what is the reason for the moist bias in models. The bias is shown to be present in the initial conditions, and change very little over the course of forecasts. This leads naturally to the question of whether the problem is merely with the initial state, or if there is also a problem with the forecast model itself. A correction of the initial state is, unfortunately, insufficient to correct forecasts. The moisture bias is reintroduced to the lowermost stratosphere with a half-life of 8–9 days. Partitioned physical process tendencies reveal that vertical diffusion and the model dynamics are primarily responsible, with cloud microphysics making little difference. Despite the importance of diffusion, transport and mixing, the re-moistening has negligible dependence on horizontal resolution, suggesting that there may be processes poorly resolved in the vertical which will ultimately require addressing.

4. **Quantify the impacts that these biases have on forecast skill, and hence the improvements that could be achieved by removing them**

   As the primary impact of the moist bias on global circulation is on collocated temperature through excessive emission of long-wave radiation, the improvements that could be made to forecasts through removing the identified biases can be investigated by correcting the moisture field that the radiation scheme sees. In this way it is shown that such a bias correction leads to significant increases in skill at lead times of around 3 weeks particularly over the North Atlantic and Europe, and a reduction in seasonal average error globally. As well as the expected improvements to near-tropopause temperature and zonal winds, there is additional skill from 50hPa to the surface in temperature, zonal winds, geopotential height and surface pressure in addition to smaller improvements to rainfall, sea surface temperatures and sea ice concentrations. The proposed mechanism for this is via an improvement of the representation of Rossby wave propagation along the tropopause then systematically improving the forecasting of extratropical cyclones.

5.0.1 Implications

New knowledge of both the structure of extratropical lowermost stratospheric temperature and humidity biases and their causes within the model is presented, in addition to quantifying the improvements to forecast skill that can be achieved by utilising this knowledge to remove such biases from forecast models. Although forecasts longer than four months are not considered here, the same mechanisms will apply on multiannual timescales.

It is shown here that the extratropical lowermost stratosphere cold bias that is apparent in seasonal forecasts is relevant too on timescales of weeks, and that both the structure and rate of growth are consistent with the additional long-wave cooling from the moist bias. The moist bias is present in the analysis, but when removed from the analysis is reintroduced also on a timescale of weeks. As the most notable impacts on forecast quality are at lead times of three weeks or greater, an improvement to the data assimilation system alone would be insufficient to provide much in the way of forecast improvement.

For operational centres to significantly improve forecasts it is required that they both improve the representation of water vapour in the analysis and improve the representation of vertical diffusion and vertical advection within the model. Suggested approaches for this are presented below in Section 5.1.2.

5.1 Future Work

5.1.1 Limitations of present study

It is acknowledged that the analysis presented in this study has some limitations. Although a motivation for this work is that the climatological cold bias is present in many models the biases in operational forecasts are only characterised in the MetUM and IFS in Chapter 2, and process tendencies are only output from the IFS in Chapter 3. Despite many other models having a cold bias, it has not been shown that they all too have a moist bias, or that it is necessarily the same processes being misrepresented in all cases.

Furthermore, in addition to only considering two models due to the opportunity presented by the recently performed NAWDEX field campaign Chapter 2 only considers observations from September and October over the North Atlantic. While Chapter 3 includes the remainder of the extratropical northern hemisphere in statistics, hindcasts are still only performed for northern hemisphere autumn. This will have some impact on the results in Chapter 4, as it is known from other studies such as Dyroff *et al.* (2015) that the moisture bias is larger in the Summer. Hegglin *et al.* (2009) deduce that the extratropical lowermost stratosphere is "flushed" with air from the tropics during the summer, and it is identified by Birner (2010) that air in the extratropical lowermost stratosphere is youngest in autumn and oldest in spring. This may imply that a correction to water vapour in the analysis in Spring would not be reintroduced with the same timescale as that found in autumn, or that different processes would be more important.

In addition to performing experiments during different seasons, it may be valuable to analyse bias structures in regions other than the North Atlantic. In Chapter 4 a bias correction is used based on North Atlantic observations, and results show the greatest improvements over the North Atlantic and Europe. There may exist regional variation in the character of these humidity and temperature biases.

5.1.2 Development of present study

It is identified that the source of the model error and hence the key to its resolution is likely in the representation of vertical diffusion, transport and mixing. Inclusion of sufficient turbulent mixing in the model is important, as otherwise the jets get too strong (Richard Forbes, pers. comm.). Parametrized turbulent mixing in the form of vertical diffusion is therefore important for the prevention of detrimental forecast impacts, but there is uncertainty in assumptions made in the parametrization. However, it is shown above in Chapter 4 that the moist bias and hence the cold bias also cause the jets to be too strong. One would therefore hypothesise that a reduction in the parametrized vertical diffusion would lead to a direct increase in jet strength, but also a decrease in lower stratospheric water vapour and therefore a coincident decrease in jet strength. The question raised from this is therefore: would a reduction in vertical mixing in conjunction with a reduction in initial LMS humidity result in a better representation of both temperature and jet velocity?

Another question left unanswered is that of where precisely the mixing processes are being misrepresented. Chapter 3 focused on whole-hemisphere means and found a half-life of humidity re-introduction of 8–9 days, but it is not obvious from this alone whether the remoistening processes are uniform or localised. To begin to answer this question an estimate of a field for remoistening can be created. First it is required to assume that the QMOD forecast state at some sufficiently short time into the forecast approximately represents the “truth” plus some remoistening bias. To then isolate the remoistening field one can subtract from this some approximation to a “truth” state. In this case the CTRL forecast initialised at the same time as the QMOD forecast is used, with the humidity modification applied to the output. Therefore the position of the PV contours between the two compared fields closely aligns given the very similar development of QMOD and CTRL at short lead-times. Let F be the forecast operator and g be the humidity modification defined by $g = 1 - f(p, z)$ where $f(p, z)$ is as defined in equation 3.1. Then QMOD at time zero is $q_{mod}(0) = q_{an} \cdot g$, CTRL at time zero is $q_{ctrl}(0) = q_{an}$, and we are assuming that the remoistening field q' is equal to

$$q'(t) = q_{mod}(t) - q_{ctrl}(t) \cdot g = F_t(q_{an} \cdot g) - F_t(q_{an}) \cdot g \tag{5.1}$$

where q_{an} is the analysis field as before. A preliminary look at this experiment as illustrated in Figure 5.1 suggests that the remoistening processes are localised, with air masses going from unbiased at day 0 to having roughly the same bias as CTRL at day 1, above and

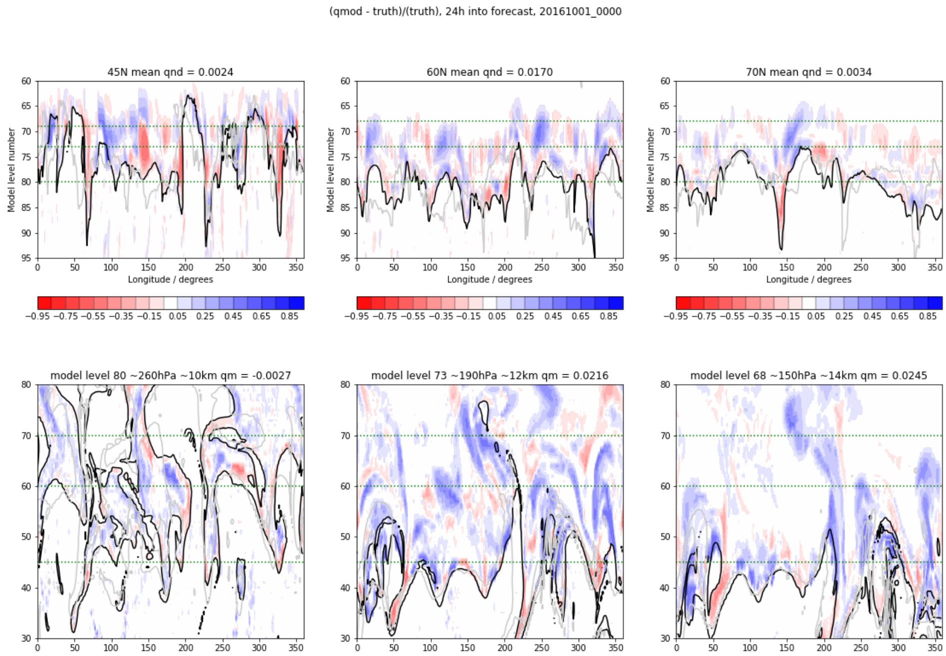

Figure 5.1: Specific humidity normalised difference (as defined in Section 2.3.5) between QMOD (from Chapter 3) at 1 day and *CTRL at 1 day, modified at this later time by the humidity correction from Section 3.3.1.* Equivalent to $q'(1)$ from equation 5.1 above. Black lines show 3.5 PVU from CTRL at 1 day, and grey lines the same from day 0. Panels (a, b, c) are vertical cross sections through latitudes 40N, 60N and 70N respectively. (d, e, f) are horizontal cross sections at model levels 80, 73 and 68 corresponding to altitudes of roughly 10, 12 and 14km respectively.

on the western flanks of ridges. An advantage of considering the same time interval as a field campaign is that this allows the comparison of such snapshots to known case studies: the feature furthest to the right of each panel in Figure 5.1 is seen to be the wave pattern associated with the "Stalactite cyclone" from early October 2016. An alternative approach to further diagnose the origin of the moist bias would be to use source-region tracers, such as the water vapour tagging method utilised in Sodemann *et al.* (2009).

A key result from this work is that the model moistening bias is independent of horizontal resolution, suggesting the problem is in vertical resolution. It would therefore be very valuable to complete similar experiments to those in Chapter 3 and Figure 5.1 at higher vertical resolutions.

Further additional work which I did not have time to complete comprised an attempt to directly establish and quantify the physical mechanism linking the correction of the extra-tropical lowermost stratospheric moisture bias to propagation of waves in the tropopause layer, which I propose drives forecast improvement seen in Chapter 4. A rudimentary method for doing so would be to perform a lagged-correlation in longitude and/or time between all-hemisphere fields of tropopause height (calculated from any suitable metric - PV would be the easiest) between QMOD and CTRL forecasts. From this, assuming the wave forms look suitably similar between QMOD and CTRL and the dominant difference is due to difference in propagation speed, the lag with the highest correlation would correlate with the phase speed error. Such assumptions are only likely to be valid at short lead-times, so it is unclear whether the analysis would also be able to be included in such comparisons. To explicitly determine the intermediate step and link this to the work of Harvey *et al.* (2016), one could also calculate the width of the tropopause layer for each of QMOD and CTRL, and attempt to relate any changes in wave propagation directly to the smoothness of the PV front.

Bibliography

AIRS Science Team and Teixeira, J. (2013). AIRS/Aqua L3 Daily Support Product (AIRS-only) 1 degree x 1 degree V006, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC). https://disc.gsfc.nasa.gov/datasets/AIRS3SPD006/summary?keywords=airs%20ozone. Date accessed: 22/10/20. doi: 10.5067/Aqua/AIRS/DATA306.

Bannister, R. N., Chipilski, H. G., and Martinez-Alvarado, O. (2020). Techniques and challenges in the assimilation of atmospheric water observations for numerical weather prediction towards convective scales. *Quarterly Journal of the Royal Meteorological Society*, **146**(726), 1–48. doi: 10.1002/qj.3652.

Birner, T. (2006). Fine-scale structure of the extratropical tropopause region. *Journal of Geophysical Research Atmospheres*, **111**(4). doi: 10.1029/2005JD006301.

Birner, T. (2010). Residual circulation and tropopause structure. *Journal of the Atmospheric Sciences*, **67**(8), 2582–2600. doi: 10.1175/2010JAS3287.1.

Birner, T. and Charlesworth, E. J. (2017). On the relative importance of radiative and dynamical heating for tropical tropopause temperatures. *Journal of Geophysical Research: Atmospheres*, **122**(13), 6782–6797. doi: 10.1002/2016JD026445.

Bland, J., Gray, S., Methven, J., and Forbes, R. (2021). Characterising extratropical near-tropopause analysis humidity biases and their radiative effects on temperature forecasts. *Quarterly Journal of the Royal Meteorological Society*, **147**(741), 3878–3898. doi: 10.1002/QJ.4150.

Boljka, L. and Birner, T. (2022). Potential impact of tropopause sharpness on the structure and strength of the general circulation. *Nature Partner Journals Climate and Atmospheric Science*, **5**(1), 1–9. doi: 10.1038/s41612-022-00319-6.

Bönisch, H., Engel, A., Curtius, J., Birner, T., and Hoor, P. (2009). Quantifying transport into the lowermost stratosphere using simultaneous in-situ measurements of SF6 and CO2. *Atmospheric Chemistry and Physics*, **9**(16), 5905–5919. doi: 10.5194/acp-9-5905-2009.

Bradshaw, N. G., Vaughan, G., Busen, R., Garcelon, S., Jones, R., Gardiner, T., and Hacker, J. (2002). Tracer filamentation generated by small-scale Rossby wave breaking in the lower stratosphere. *Journal of Geophysical Research: Atmospheres*, **107**(D23), ACL 12–1. doi: 10.1029/2002JD002086.

Brioude, J., Cammas, J. P., Cooper, O. R., and Nedelec, P. (2008). Characterization of the composition, structure, and seasonal variation of the mixing layer above the extratropical tropopause as revealed by MOZAIC measurements. *Journal of Geophysical Research: Atmospheres*, **113**(D7), 0–01. doi: 10.1029/2007JD009184.

Butchart, N., Charlton-Perez, A. J., Cionni, I., Hardiman, S. C., Haynes, P. H., Krüger, K., Kushner, P. J., Newman, P. A., Osprey, S. M., Perlwitz, J., Sigmond, M., Wang, L., Akiyoshi, H., Austin, J., Bekki, S., Baumgaertner, A., Braesicke, P., Brühl, C., Chipperfield, M., Dameris, M., Dhomse, S., Eyring, V., Garcia, R., Garny, H., Jöckel, P., Lamarque, J.-F., Marchand, M., Michou, M., Morgenstern, O., Nakamura, T., Pawson, S., Plummer, D., Pyle, J., Rozanov, E., Scinocca, J., Shepherd, T. G., Shibata, K., Smale, D., Teyssèdre, H., Tian, W., Waugh, D., and Yamashita, Y. (2011). Multimodel climate and variability of the stratosphere. *Journal of Geophysical Research: Atmospheres*, **116**(D5). doi: 10.1029/2010JD014995.

Carminati, F., Migliorini, S., Ingleby, B., Bell, W., Lawrence, H., Newman, S., Hocking, J., and Smith, A. (2019). Using reference radiosondes to characterise NWP model uncertainty for improved satellite calibration and validation. *Atmospheric Measurement Techniques*, **12**(1), 83–106. doi: 10.5194/amt-12-83-2019.

Carver, G. (2019). Single column model 43r3 release notes - OpenIFS - ECMWF Confluence Wiki. https://confluence.ecmwf.int/display/OIFS/Single+column+model+43r3+release+notes. Date accessed: 2021-04-26.

Cau, P., Methven, J., and Hoskins, B. (2005). Representation of dry tropical layers and their origins in ERA-40 data. *Journal of Geophysical Research: Atmospheres*, **110**(D6). doi: 10.1029/2004JD004928.

Chagnon, J. M., Gray, S. L., and Methven, J. (2013). Diabatic processes modifying potential vorticity in a North Atlantic cyclone. *Quarterly Journal of the Royal Meteorological*

Society, **139**(674), 1270–1282. doi: 10.1002/qj.2037.

Charlesworth, E., Plöger, F., Birner, T., Baikhadzhaev, R., Abalos, M., Abraham, N. L., Akiyoshi, H., Bekki, S., Dennison, F., Jöckel, P., Keeble, J., Kinnison, D., Morgenstern, O., Plummer, D., Rozanov, E., Strode, S., Zeng, G., Egorova, T., and Riese, M. (2023). Stratospheric water vapor affecting atmospheric circulation. *Nature Communications*, **14**. doi: 10.1038/s41467-023-39559-2.

Choi, Y., Ha, J.-C., and Lim, G.-H. (2015). Investigation of the Effects of Considering Balloon Drift Information on Radiosonde Data Assimilation Using the Four-Dimensional Variational Method. *Weather and Forecasting*, **30**(3), 809–826. doi: 10.1175/WAF-D-14-00161.1.

Davis, S. M., Hegglin, M. I., Fujiwara, M., Dragani, R., Harada, Y., Kobayashi, C., Long, C., Manney, G. L., Nash, E. R., Potter, G. L., Tegtmeier, S., Wang, T., Wargan, K., and Wright, J. S. (2017). Assessment of upper tropospheric and stratospheric water vapor and ozone in reanalyses as part of S-RIP. *Atmos. Chem. Phys*, **17**, 12743–12778. doi: 10.5194/acp-17-12743-2017.

Dee, D. P. and Da Silva, A. M. (2003). The choice of variable for atmospheric moisture analysis. *Monthly Weather Review*, **131**(1), 155–171. doi: 10.1175/1520-0493(2003)131¡0155:TCOVFA¿2.0.CO;2.

Diamantakis, M. and Flemming, J. (2014). Global mass fixer algorithms for conservative tracer transport in the ECMWF model. *Geoscientific Model Development*, **7**, 965–979. doi: 10.5194/gmd-7-965-2014.

Dirksen, R. J., Bodeker, G. E., Thorne, P. W., Merlone, A., Reale, T., Wang, J., Hurst, D. F., Demoz, B. B., Gardiner, T. D., Ingleby, B., Sommer, M., Von Rohden, C., and Leblanc, T. (2019). Progress in managing the transition from the RS92 to the Vaisala RS41 as the operational radiosonde within the GCOS Reference Upper-Air Network. *Geoscientific Instrumentation, Methods and Data Systems.* doi: 10.5194/gi-2019-36.

Dyroff, C., Zahn, A., Christner, E., Forbes, R., Tompkins, A. M., and van Velthoven, P. F. J. (2015). Comparison of ECMWF analysis and forecast humidity data with CARIBIC upper troposphere and lower stratosphere observations. *Quarterly Journal of the Royal Meteorological Society*, **141**(688), 833–844. doi: 10.1002/qj.2400.

ECMWF (2016). *IFS documentation CY41r2.* https://www.ecmwf.int/en/publications/ifs-documentation. Date accessed: 12/10/20. doi: 10.21957/tr5rv27xu.

ECMWF (2017). *IFS Documentation CY43R3, Part IV: Physical processes.* IFS Documentation. ECMWF. https://www.ecmwf.int/en/elibrary/17736-part-iv-physical-processes. Date accessed: 12/10/20. doi: 10.21957/efyk72kl.

ECMWF (2020). IFS Documentation CY47R1 - Part IV: Physical Processes — ECMWF. In *IFS Documentation CY47R1*, chapter 4. ECMWF. https://www.ecmwf.int/en/elibrary/81189-ifs-documentation-cy47r1-part-iv-physical-processes. Date accessed: 18/09/23. doi: 10.21957/cpmkqvhja.

Edwards, D., Anderson, G., Oakley, T., and Gault, P. (2014). Met Office Intercomparison of Vaisala RS92 and RS41 Radiosondes. Technical report. https://www.vaisala.com/sites/default/files/documents/Met Office Intercomparison of Vaisala RS41 and RS92 Radiosondes.pdf. Date accessed: 05/02/24.

Erler, A. R. and Wirth, V. (2011). The Static Stability of the Tropopause Region in Adiabatic Baroclinic Life Cycle Experiments. *Journal of the Atmospheric Sciences*, **68**(6), 1178–1193. doi: 10.1175/2010JAS3694.1.

Feist, D. G., Geer, A. J., Müller, S., and Kämpfer, N. (2007). Middle atmosphere water vapour and dynamical features in aircraft measurements and ECMWF analyses. *Atmospheric Chemistry and Physics*, **7**, 5291–5307. doi: 10.5194/acp-7-5291-2007.

Forster, C. and Wirth, V. (2000). Radiative decay of idealized stratospheric filaments in the troposphere. *Journal of Geophysical Research: Atmospheres*, **105**(D8), 10169–10184. doi: 10.1029/2000JD900052.

Forster, P. M. F. and Shine, K. P. (1999). Stratospheric water vapour changes as a possible contributor to observed stratospheric cooling. *Geophysical Research Letters*, **26**(21), 3309–3312. doi: 10.1029/1999GL010487.

Forster, P. M. F. and Shine, K. P. (2002). Assessing the climate impact of trends in stratospheric water vapor. *Geophysical Research Letters*, **29**(6), 10–1–10–4. doi: 10.1029/2001gl013909.

Fowler, A., Bannister, R., and Eyre, J. (2012). A new floating model level scheme for the assimilation of boundary-layer top inversions: the univariate assimilation of temperature. *Quarterly Journal of the Royal Meteorological Society*, **138**(664), 682–698. doi: 10.1002/qj.955.

Gates, W. L., Boyle, J. S., Covey, C., Dease, C. G., Doutriaux, C. M., Drach, R. S., Fiorino, M., Gleckler, P. J., Hnilo, J. J., Marlais, S. M., Phillips, T. J., Potter, G. L., Santer, B. D., Sperber, K. R., Taylor, K. E., and Williams, D. N. (1999). An Overview of the Results of the Atmospheric Model Intercomparison Project (AMIP I). *Bulletin of the American Meteorological Society*, **80**(1), 29–56. doi: 10.1175/1520-0477(1999)080¡0029:AOOTRO¿2.0.CO;2.

Gettelman, A., Hegglin, M. I., Son, S.-W., Kim, J., Fujiwara, M., Birner, T., Kremser, S., Rex, M., Añel, J. A., Akiyoshi, H., Austin, J., Bekki, S., Braesike, P., Brühl, C., Butchart, N., Chipperfield, M., Dameris, M., Dhomse, S., Garny, H., Hardiman, S. C., Jöckel, P., Kinnison, D. E., Lamarque, J. F., Mancini, E., Marchand, M., Michou, M., Morgenstern, O., Pawson, S., Pitari, G., Plummer, D., Pyle, J. A., Rozanov, E., Scinocca, J., Shepherd, T. G., Shibata, K., Smale, D., Teyssèdre, H., and Tian, W. (2010). Multimodel assessment of the upper troposphere and lower stratosphere: Tropics and global trends. *Journal of Geophysical Research*, **115**(D3), D00M08. doi: 10.1029/2009JD013638.

Gettelman, A., Hoor, P., Pan, L. L., Randel, W. J., Hegglin, M. I., and Birner, T. (2011). The Exrtatropical Upper Troposphere and Lower Stratosphere. *Reviews of Geophysics*, **49**(3). doi: 10.1029/2011RG000355.

Gray, S. L., Dunning, C. M., Methven, J., Masato, G., and Chagnon, J. M. (2014). Systematic model forecast error in Rossby wave structure. *Geophysical Research Letters*, **41**(8), 2979–2987. doi: 10.1002/2014GL059282.

Haiden, T., Janousek, M., Bidlot, J.-R., Ferranti, L., Prates, F., Vitart, F., Bauer, P., and Richardson, D. (2016). Evaluation of ECMWF forecasts, including the 2016 resolution upgrade — ECMWF. Technical report. https://www.ecmwf.int/en/elibrary/16924-evaluation-ecmwf-forecasts-including-2016-resolution-upgrade. Date accessed: 07/05/2020. doi: 10.21957/vltarm3ql.

Hardiman, S. C., Boutle, I. A., Bushell, A. C., Butchart, N., Cullen, M. J. P., Field, P. R., Furtado, K., Manners, J. C., Milton, S. F., Morcrette, C., O'Connor, F. M., Shipway, B. J., Smith, C., Walters, D. N., Willett, M. R., Williams, K. D., Wood, N., Abraham, N. L., Keeble, J., Maycock, A. C., Thuburn, J., and Woodhouse, M. T. (2015). Processes Controlling Tropical Tropopause Temperature and Stratospheric Water Vapor in Climate Models. *Journal of Climate*, **28**(16), 6516–6535. doi: 10.1175/JCLI-D-15-0075.1.

Harvey, B., Methven, J., and Ambaum, M. H. P. (2018). An Adiabatic Mechanism for the Reduction of Jet Meander Amplitude by Potential Vorticity Filamentation. *Journal of the Atmospheric Sciences*, **75**(12), 4091–4106. doi: 10.1175/JAS-D-18-0136.1.

Harvey, B., Methven, J., Sanchez, C., and Schäfler, A. (2020). Diabatic generation of negative potential vorticity and its impact on the North Atlantic jet stream. *Quarterly Journal of the Royal Meteorological Society*, **146**(728), 1477–1497. doi: 10.1002/qj.3747.

Harvey, B. J., Methven, J., and Ambaum, M. H. P. (2016). Rossby wave propagation on potential vorticity fronts with finite width. *Journal of Fluid Mechanics*, **794**, 775–797. doi: 10.1017/jfm.2016.180.

Hegglin, M. I., Boone, C. D., Manney, G. L., and Walker, K. A. (2009). A global view of the extratropical tropopause transition layer from Atmospheric Chemistry Experiment Fourier Transform Spectrometer O3, H2O, and CO. *Journal of Geophysical Research Atmospheres*, **114**(7). doi: 10.1029/2008JD009984.

Holton, J. R., Haynes, P. H., McIntyre, M. E., Douglass, A. R., Rood, R. B., and Pfister, L. (1995). Stratosphere-troposphere exchange. *Reviews of Geophysics*, **33**(4), 403. doi: 10.1029/95RG02097.

Hoskins, B. J. (1991). Towards a PV-θ view of the general circulation. *Tellus B*, **43**(4), 27–35. doi: 10.1034/J.1600-0889.1991.T01-3-00005.X.

Hoskins, B. J. and James, I. N. (2014). *Fluid Dynamics of the Midlatitude Atmosphere*. John Wiley Sons, Ltd, Chichester. doi: 10.1002/9781118526002.

Hoskins, B. J., McIntyre, M. E., and Robertson, A. W. (1985). On the use and significance of isentropic potential vorticity maps. *Quarterly Journal of the Royal Meteorological Society*, **111**(470), 877–946. doi: 10.1002/qj.49711147002.

Ingleby, B. (2017). An assessment of different radiosonde types 2015/2016 — ECMWF. *ECMWF Technical Memoranda*, **807**(80268). https://www.ecmwf.int/en/elibrary/17551-assessment-different-radiosonde-types-2015-2016. Date accessed: 12/11/2018. doi: 10.21957/0nje0wpsa.

Ingleby, B. and Edwards, D. (2015). Changes to radiosonde reports and their processing for numerical weather prediction. *Atmospheric Science Letters*, **16**(1), 44–49. doi: 10.1002/asl2.518.

Ingleby, B., Lorenc, A., Ngan, K., Rawlins, R., and Jackson, D. (2012). Improved variational analyses using a nonlinear humidity control variable: Formulation and trials. https://digital.nmla.metoffice.gov.uk/download/file/IO 498d499b-3f4f-4380-bfd0-c59098ad6bc9. Date accessed: 07/02/24.

Ingleby, B., Pauley, P., Kats, A., Ator, J., Keyser, D., Doerenbecher, A., Fucile, E., Hasegawa, J., Toyoda, E., Kleinert, T., Qu, W., St. James, J., Tennant, W., and Weedon, R. (2016). Progress toward High-Resolution, Real-Time Radiosonde Reports. *Bulletin of the American Meteorological Society*, **97**(11), 2149–2161. doi: 10.1175/BAMS-D-15-00169.1.

Jauhiainen, H., Survo, P., Lehtinen, R., Lentonen, J., and Oyj, V. (2014). Radiosonde RS41 and RS92 key differences and comparison test results in different locations and climates. Technical report, Vaisala, Helsinki. https://docplayer.net/44929074-Radiosonde-rs41-and-rs92-key-differences-and-comparison-test-results-in-different-locations-and-climates.html. Date accessed 07/02/24.

Jensen, E. J., Pan, L. L., Honomichl, S., Diskin, G. S., Krämer, M., Spelten, N., Günther, G., Hurst, D. F., Fujiwara, M., Vömel, H., Selkirk, H. B., Suzuki, J., Schwartz, M. J., and Smith, J. B. (2020). Assessment of Observational Evidence for Direct Convective Hydration of the Lower Stratosphere. *Journal of Geophysical Research: Atmospheres*, **125**(15), e2020JD032793. doi: 10.1029/2020JD032793.

Jensen, M. P., Holdridge, D., Survo, P., Lehtinen, R., Baxter, S., Toto, T., and Johnson, K. L. (2016). Comparison of Vaisala radiosondes RS41 and RS92 at the ARM southern great plains site. *Atmospheric Measurement Techniques*, **9**(7), 3115–3129. doi: https://doi.org/10.5194/amt-9-3115-2016.

Joshi, M. M., Charlton, A. J., and Scaife, A. A. (2006). On the influence of stratospheric water vapor changes on the tropospheric circulation. *Geophysical Research Letters*, **33**(9), L09806. doi: 10.1029/2006GL025983.

Krüger, K., Schäfler, A., Wirth, M., Weissmann, M., and Craig, G. C. (2022). Vertical structure of the lower-stratospheric moist bias in the ERA5 reanalysis and its connection to mixing processes. *Atmospheric Chemistry and Physics*, **22**(23), 15559–15577. doi: 10.5194/ACP-22-15559-2022.

Kunkel, D., Hoor, P., Kaluza, T., Ungermann, J., Kluschat, B., Giez, A., Lachnitt, H. C., Kaufmann, M., and Riese, M. (2019). Evidence of small-scale quasi-isentropic mixing in ridges of extratropical baroclinic waves. *Atmospheric Chemistry and Physics*, **19**(19), 12607–12630. doi: 10.5194/ACP-19-12607-2019.

Kunz, A., Spelten, N., Konopka, P., Müller, R., Forbes, R. M., and Wernli, H. (2014). Comparison of Fast In situ Stratospheric Hygrometer (FISH) measurements of water vapor in the upper troposphere and lower stratosphere (UTLS) with ECMWF (re)analysis data. *Atmospheric Chemistry and Physics*, **14**, 10803–10822. doi: 10.5194/acp-14-10803-2014.

Lambert, A., Read, W., Livesey, N., and Fuller, R. (2020). MLS/Aura Level 3 Daily Binned Water Vapor (H2O) Mixing Ratio on Assorted Grids V004. https://disc.gsfc.nasa.gov/datasets/ML3DBH2O 004/summary. Date accessed 11/12/20. doi: 10.5067/Aura/MLS/DATA/3109.

Laroche, S., Sarrazin, R., Laroche, S., and Sarrazin, R. (2013). Impact of Radiosonde Balloon Drift on Numerical Weather Prediction and Verification. *Weather and Forecasting*,

28(3), 772–782. doi: 10.1175/WAF-D-12-00114.1.

Lawrence, Z. D., Abalos, M., Ayarzagüena, B., Barriopedro, D., Butler, A. H., Calvo, N., de la Cámara, A., Charlton-Perez, A., Domeisen, D. I., Dunn-Sigouin, E., García-Serrano, J., Garfinkel, C. I., Hindley, N. P., Jia, L., Jucker, M., Karpechko, A. Y., Kim, H., Lang, A. L., Lee, S. H., Lin, P., Osman, M., Palmeiro, F. M., Perlwitz, J., Polichtchouk, I., Richter, J. H., Schwartz, C., Son, S. W., Statnaia, I., Taguchi, M., Tyrrell, N. L., Wright, C. J., and Wu, R. W. (2022). Quantifying stratospheric biases and identifying their potential sources in subseasonal forecast systems. *Weather and Climate Dynamics*, **3**(3), 977–1001. doi: 10.5194/WCD-3-977-2022.

Livesey, N. J., Read, W. G., Wagner, P. A., Froidevaux, L., Lambert, A., Manney, G. L., Millán Valle, L. F., Pumphrey, H. C., Santee, M. L., Schwartz, M. J., Wang, S., Fuller, R. A., Jarnot, R. F., Knosp, B. W., Martinez, E., and Lay, R. R. (2020). Version 4.2x Level 2 and 3 data quality and description document. https://mls.jpl.nasa.gov/data/v4-2 data quality document.pdf. Date accessed 07/02/24.

MacPherson, B. (1995). Radiosonde balloon drift — does it matter for data assimilation? *Meteorological Applications*, **2**(4), 301–305. doi: 10.1002/MET.5060020402.

Malardel, S., Wedi, N., Deconinck, W., Diamantakis, M., Kühnlein, C., Mozdzynski, G., Hamrud, M., and Smolarkiewicz, P. (2016). A new grid for the IFS. *Ecmwf Newsletter*, **146**(80059), 23–28. doi: 10.21957/zwdu9u5i.

Maycock, A. C., Shine, K. P., and Joshi, M. M. (2011). The temperature response to stratospheric water vapour changes. *Quarterly Journal of the Royal Meteorological Society*, **137**(657), 1070–1082. doi: 10.1002/qj.822.

Maycock, A. C., Joshi, M. M., Shine, K. P., and Scaife, A. A. (2013). The Circulation Response to Idealized Changes in Stratospheric Water Vapor. *Journal of Climate*, **26**(2), 545–561. doi: 10.1175/JCLI-D-12-00155.1.

Motl, M. (2013). Vaisala RS41 Trial in the Czech Republic. Technical report, Czech Hydrometeorological Institute. https://www.vaisala.com/sites/default/files/documents/VaisalaNews192 Czech%20Trials.pdf. Date accessed 07/02/24.

Müller, S., Hoor, P., Berkes, F., Bozem, H., Klingebiel, M., Reutter, P., Smit, H. G., Wendisch, M., Spichtinger, P., and Borrmann, S. (2015). In situ detection of stratosphere-troposphere exchange of cirrus particles in the midlatitudes. *Geophysical Research Letters*, **42**(3), 949–955. doi: 10.1002/2014GL062556.

Mylne, K., Chen, J., Erfani, A., Hamill, T., Richardson, D., Vannitsem, S., Wang, Y., and Zhu, Y. (2021). *Guidelines on Ensemble Prediction System Postprocessing*, volume WMO 1254 of *WMO*. Geneva. https://library.wmo.int/viewer/57510?medianame=1254 Guidelines on EPSPP en page=1viewer=pictureo=bookmarkn=0q=. Date accessed: 07/02/24. ISBN 978-92-63-11254-5.

Nash, J., Oakley, T., Vömel, H., and Wei, L. (2010). WMO Intercomparison of High Quality Radiosonde Systems. *WMO Instruments and Observing Methods (IOM)*, **107**(1580).

Noël, S., Weigel, K., Bramstedt, K., Rozanov, A., Weber, M., Bovensmann, H., and Burrows, J. P. (2018). Water vapour and methane coupling in the stratosphere observed using SCIAMACHY solar occultation measurements. *Atmospheric Chemistry and Physics*, **18**(7), 4463–4476. doi: 10.5194/ACP-18-4463-2018.

O'Connor, F. M., Vaughan, G., and Backer, H. D. (1999). Observations of subtropical air in the european mid-latitude lower stratosphere. *Quarterly Journal of the Royal Meteorological Society*, **125**(560), 2965–2986. doi: 10.1002/QJ.49712556009.

Oh, J., Son, S.-W., Williams, K., Walters, D., Kim, J., Willett, M., Earnshaw, P., Bushell, A., Kim, Y., and Kim, J. (2018). Ozone sensitivity of tropical upper-troposphere and stratosphere temperature in the MetOffice Unified Model. *Quarterly Journal of the Royal Meteorological Society*, **144**(715), 2001–2009. doi: 10.1002/qj.3346.

Oikonomou, E. K. and O'Neill, A. (2006). Evaluation of ozone and water vapor fields from the ECMWF reanalysis ERA-40 during 1991–1999 in comparison with UARS satellite and MOZAIC aircraft observations. *Journal of Geophysical Research*, **111**(D14), D14109. doi: 10.1029/2004JD005341.

Pan, L. L., Randel, W. J., Gary, B. L., Mahoney, M. J., and Hintsa, E. J. (2004). Definitions and sharpness of the extratropical tropopause: A trace gas perspective. *Journal of Geophysical Research: Atmospheres*, **109**(D23), 1–11. doi: 10.1029/2004JD004982.

Polichtchouk, I., Bechtold, P., Diamantakis, M., Forbes, R., Hogan, R., Malardel, S., Sandu, I., Stockdale, T., Vana, F., Vitart, F., and Wedi, N. (2019a). Causes and fixes for stratospheric temperature biases in IFS and their impact on predictability. In *Workshop on Stratospheric Prediciability and impact on the troposphere.* https://events.ecmwf.int/event/129/contributions/1110/attachments/286/526/Stratospheric-WS-Polichtchouk.pdf. Date accessed 09/02/24.

Polichtchouk, I., Stockdale, T., Bechtold, P., Diamantakis, M., Malardel, S., Sandu, I., Vána, F., and Wedi, N. (2019b). Control on stratospheric temperature in IFS: resolution and vertical advection. *ECMWF Technical Memoranda*, **847**(81078). doi:

10.21957/cz3t12t7e.

Polichtchouk, I., Bechtold, P., Bonavita, M., Forbes, R., Healy, S., Hogan, R., Laloyaux, P., Rennie, M., Stockdale, T., Wedi, N., Diamantakis, M., Flemming, J., English, S., Isaksen, L., Vána, F., Gisinger, S., and Byrne, N. (2021). Stratospheric modelling and assimilation. *ECMWF Technical Memoranda*, **877**(19875). doi: 10.21957/25hegfoq.

Pope, V. D., Pamment, J. A., Jackson, D. R., and Slingo, A. (2001). The Representation of Water Vapor and Its Dependence on Vertical Resolution in the Hadley Centre Climate Model. *Journal of Climate*, **14**(14). doi: 10.1175/1520-0442(2001)014¡3065:TROWVA¿2.0.CO;2.

Randel, W. J., Wu, F., and Forster, P. (2007). The extratropical tropopause inversion layer: Global observations with GPS data, and a radiative forcing mechanism. *Journal of the Atmospheric Sciences*, **64**(12), 4489–4496. doi: 10.1175/2007JAS2412.1.

Rodwell, M. J. and Palmer, T. N. (2007). Using numerical weather prediction to assess climate models. *Quarterly Journal of the Royal Meteorological Society*, **133**(622), 129–146. doi: 10.1002/QJ.23.

Saffin, L., Gray, S. L., Methven, J., and Williams, K. D. (2017). Processes Maintaining Tropopause Sharpness in Numerical Models. *Journal of Geophysical Research: Atmospheres*, **122**(18), 9611–9627. doi: 10.1002/2017JD026879.

Saffin, L., Methven, J., Bland, J., Harvey, B., and Sanchez, C. (2021). Circulation conservation in the outflow of warm conveyor belts and consequences for Rossby wave evolution. *Quarterly Journal of the Royal Meteorological Society*, **147**(740), 3587–3610. doi: 10.1002/QJ.4143.

Schäfler, A., Craig, G., Wernli, H., Arbogast, P., Doyle, J. D., McTaggart-Cowan, R., Methven, J., Rivière, G., Ament, F., Boettcher, M., Bramberger, M., Cazenave, Q., Cotton, R., Crewell, S., Delanoë, J., Dörnbrack, A., Ehrlich, A., Ewald, F., Fix, A., Grams, C. M., Gray, S. L., Grob, H., Groß, S., Hagen, M., Harvey, B., Hirsch, L., Jacob, M., Kölling, T., Konow, H., Lemmerz, C., Lux, O., Magnusson, L., Mayer, B., Mech, M., Moore, R., Pelon, J., Quinting, J., Rahm, S., Rapp, M., Rautenhaus, M., Reitebuch, O., Reynolds, C. A., Sodemann, H., Spengler, T., Vaughan, G., Wendisch, M., Wirth, M., Witschas, B., Wolf, K., and Zinner, T. (2018). The North Atlantic Waveguide and Downstream Impact Experiment. *Bulletin of the American Meteorological Society*, pages BAMS–D–17–0003.1. doi: 10.1175/BAMS-D-17-0003.1.

Schäfler, A., Harvey, B., Methven, J., Doyle, J. D., Rahm, S., Reitebuch, O., Weiler, F., and Witschas, B. (2020). Observation of jet stream winds during NAWDEX and characterization of systematic meteorological analysis errors. *Monthly Weather Review*, **148**(7), 2889–2907. doi: 10.1175/MWR-D-19-0229.1.

Shepherd, T. G., Polichtchouk, I., Hogan, R. J., and Simmons, A. J. (2018). Report on Stratosphere Task Force. *ECMWF Technical Memoranda*, **824**(18259), 32.

Shine, K. and Myhre, G. (2020). The Spectral Nature of Stratospheric Temperature Adjustment and its Application to Halocarbon Radiative Forcing. *Journal of Advances in Modeling Earth Systems*, **12**(3). doi: 10.1029/2019MS001951.

Smith, C. A., Haigh, J. D., and Toumi, R. (2001). Radiative forcing due to trends in stratospheric water vapour. *Geophysical Research Letters*, **28**(1), 179–182. doi: 10.1029/2000GL011846.

Sodemann, H., Wernli, H., and Schwierz, C. (2009). Sources of water vapour contributing to the Elbe flood in August 2002-A tagging study in a mesoscale model. *Quarterly Journal of the Royal Meteorological Society*, **135**(638), 205–223. doi: 10.1002/qj.374.

Solomon, S., Rosenlof, K. H., Portmann, R. W., Daniel, J. S., Davis, S. M., Sanford, T. J., and Plattner, G.-K. (2010). Contributions of Stratospheric Water Vapor to Decadal Changes in the Rate of Global Warming. *Science*, **327**(5970), 1219–1223. doi: 10.1126/science.1182488.

Sonntag, D. (1990). Important new Values of the Physical Constants of 1986, Vapour Pressure Formulations based on the ITS-90, and Psychrometer Formulae. *Zeitschrift für Meteorologie*, **70**(5), 340–344.

Sonntag, D. (1994). Advancements in the field of hygrometry. *Meteorologie Zeitschrift*, (3), 51–66.

Stenke, A., Grewe, V., and Ponater, M. (2008). Lagrangian transport of water vapor and cloud water in the ECHAM4 GCM and its impact on the cold bias. *Climate Dynamics*, **31**(5), 491–506. doi: 10.1007/s00382-007-0347-5.

Tandon, N. F., Polvani, L. M., and Davis, S. M. (2011). The response of the tropospheric circulation to water vapor-like forcings in the stratosphere. *Journal of Climate*, **24**(21), 5713–5720. doi: 10.1175/JCLI-D-11-00069.1.

Vaisala (2013). Vaisala Radiosonde RS92-SGP. Technical report. https://www.bodc.ac.uk/data/documents/nodb/pdf/RS92SGP-Datasheet-B210358EN-F-LOW.pdf. Date accessed 04/09/19.

Vaisala (2014). Comparison of Vaisala Radiosondes RS41 and RS92. Technical report. https://www.vaisala.com/sites/default/files/documents/RS-Comparison-White-Paper-B211317EN.pdf. Date accessed: 07/02/24.

Vaisala (2017). Vaisala Radiosonde RS41 Measurement Performance. Technical report. https://www.vaisala.com/sites/default/files/documents/White%20paper%20RS41 %20Performance%20B211356EN-A.pdf. Date accessed 07/02/24.

Vaisala (2018). Radiosonde RS41-SG. Technical report. https://docs.vaisala.com/v/u/B211321EN-K/en-US. Date accessed 04/09/19.

Vaisala (2020). Vaisala RS92 Radiosonde Technical Changes. https://www.vaisala.com/en/sounding-data-continuity. Date accessed 22/04/20.

Van Thien, L., Gallus, W. A., Olsen, M. A., and Livesey, N. (2010). Comparison of Aura MLS Water Vapor Measurements with GFS and NAM Analyses in the Upper Troposphere–Lower Stratosphere. *Journal of Atmospheric and Oceanic Technology*, **27**(2), 274–289. doi: 10.1175/2009JTECHA1317.1.

Vaughan, B. G. and Timmis, C. (1998). Transport of near-tropopause air into the lower midlatitude stratosphere. *Quarterly Journal of the Royal Meteorological Society*, **124**(549), 1559–1578. doi: 10.1002/QJ.49712454910.

Walters, D., Boutle, I., Brooks, M., Melvin, T., Stratton, R., Vosper, S., Wells, H., Williams, K., Wood, N., Allen, T., Bushell, A., Copsey, D., Earnshaw, P., Edwards, J., Gross, M., Hardiman, S., Harris, C., Heming, J., Klingaman, N., Levine, R., Manners, J., Martin, G., Milton, S., Mittermaier, M., Morcrette, C., Riddick, T., Roberts, M., Sanchez, C., Selwood, P., Stirling, A., Smith, C., Suri, D., Tennant, W., Luigi Vidale, P., Wilkinson, J., Willett, M., Woolnough, S., and Xavier, P. (2017). The Met Office Unified Model Global Atmosphere 6.0/6.1 and JULES Global Land 6.0/6.1 configurations. *Geoscientific Model Development*, **10**(4), 1487–1520. doi: 10.5194/gmd-10-1487-2017.

Wang, J., Zhang, L., Dai, A., Immler, F., Sommer, M., and Vömel, H. (2013). Radiation Dry Bias Correction of Vaisala RS92 Humidity Data and Its Impacts on Historical Radiosonde Data. *Journal of Atmospheric and Oceanic Technology*, **30**(2), 197–214. doi: 10.1175/JTECH-D-12-00113.1.

WMO (1957). Meteorology - a three dimensional science. *WMO Bulletin*, **6**(4), 134–139.

Woiwode, W., Dörnbrack, A., Polichtchouk, I., Johansson, S., Harvey, B., Höpfner, M., Ungermann, J., and Friedl-Vallon, F. (2020). Technical note: Lowermost-stratospheric moist bias in ECMWF IFS model diagnosed from airborne GLORIA observations during winter/spring 2016. *Atmospheric Chemistry and Physics*, **20**(23), 15379–15387. doi: 10.5194/acp-2020-367.

WOUDC (2020). WOUDC Ozonesonde Dataset. https://geo.woudc.org/def/data/ozone/vertical-ozone-profile/ozonesonde. Date accessed: 22/10/20. doi: 10.14287/10000008.

Wu, Z. and Reichler, T. (2020). Variations in the Frequency of Stratospheric Sudden Warmings in CMIP5 and CMIP6 and Possible Causes. *Journal of Climate*, **33**(23), 10305–10320. doi: 10.1175/JCLI-D-20-0104.1.