The Representation of the Extratropical Tropopause in the ECMWF Model Evaluated by Lidar and Radiosonde Observations

Konstantin Krüger

München 2024

The Representation of the Extratropical Tropopause in the ECMWF Model Evaluated by Lidar and Radiosonde Observations

Konstantin Krüger

Dissertation
an der Fakultät der Physik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Konstantin Krüger
aus Detmold

München, den 29. August 2024

Erstgutachter: Prof. Dr. George C. Craig
Zweitgutachter: Prof. Dr. Martin Weissmann
Tag der mündlichen Prüfung: 22. November 2024

Parts of this thesis are included in:

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

Krüger, K., Schäfler A., Weissmann, M., and Craig, C. G., Influence of radiosonde observations on the sharpness and altitude of the midlatitude tropopause in the ECMWF IFS, 2024, *Weather and Climate Dynamics*, 5, 491–509, https://doi.org/10.5194/wcd-5-491-2024.

Contents

Abstract

Accurate initial conditions in numerical weather prediction (NWP) models are a fundamental prerequisite for reliable weather forecasts. A key to further improving weather forecasts is to remove systematic errors (i.e., biases) in the initial conditions, which requires a robust identification of these biases and their causes in NWP models. This work focuses on biases at the extratropical tropopause and its surrounding layer, the Upper Troposphere and Lower Stratosphere (UTLS). This region is characterised by sharp vertical gradients in temperature, wind and trace species such as water vapor. These gradients determine the distribution of potential vorticity whose distribution at the tropopause act as a waveguide for Rossby-waves and in turn, impact the weather evolution in the midlatitudes. Meteorological profile observations with high precision and vertical resolution are required that sufficiently capture the sharp temperature, wind and humidity gradients at the tropopause. However, the global observing network largely lacks such observations. In this thesis, campaign observations that fulfil these requirements are used to investigate systematic biases in initial conditions in the leading global NWP model, the Integrated Forecast System (IFS) of the European Centre for Medium-Range Weather Forecasts (ECMWF), from a new perspective.

In the first part, highly-resolved and accurate airborne water vapor profile observations measured with a differential absorption lidar (DIAL) are used to study the lower-stratospheric moist bias in the ECMWF's global reanalysis ERA5. 41 flights (six campaigns) provide 33,000 water vapor profiles over more than 200,000 km flight distance with high data coverage across the UTLS covering different seasons and synoptic situations representative of midlatitude weather. This unique dataset is suitable to address open questions regarding the LS moist bias in ERA5, in particular its vertical structure, its seasonal and synoptic variability and potential processes leading to this bias. The statistical comparison of the DIAL with the ERA5 humidity data shows a strong LS moist bias in the model with a maximum of 55 % at 1.3 km altitude above the tropopause. In addition, the analysis reveals the evidence for a decrease of this bias towards 4 km altitude in the LS. The LS moist bias is also found to be stronger in summer and covers a deeper layer at low tropopause heights associated with trough situations. Another novelty is that, for the first time, a connection of LS moist bias to mixing processes is investigated. Collocated ozone and water vapor DIAL data are used to classify the UTLS into tropospheric, stratospheric, and mixed air. The moist bias turned out to be especially strong in the mixing layer while it is weak in the tropospheric and stratospheric air. This highlights that overestimated transport processes across the tropopause in the model play a crucial role for the formation of the bias. One particular mixing process that transports large amounts of water vapor into the LS is overshooting convection. To test the sensitivity of

the LS moist bias to the representation of convection in the IFS, DIAL data of a research flight with observed high convective activity are used to evaluate model output of two IFS versions with different convection parameterization. It is found that the IFS version with an overshooting convection limitation exhibits a reduced LS moist bias.

In the second part, data assimilation output is used to evaluate UTLS temperature and wind biases in the IFS and their impact on the tropopause representation in this model. Ideally, data assimilation should act to reduce biases in the background forecast and improve the analysis, however, the influence of data assimilation on the sharp vertical gradients near tropopause was an open question. This study investigates the influence of assimilated radiosondes on tropopause sharpness and altitude based on 9,729 profiles observed over North America, the Atlantic, and Europe in a one-month autumn period during the North Atlantic Waveguide Downstream Impact Experiment (NAWDEX) campaign. The observed temperature and wind profiles are compared with their model equivalents, i.e. the background and the analysis at the position and time of the observations (observation-space). The background overestimates (underestimates) temperature at the tropopause (in the LS) leading to a smoother change in static stability (N^2), and thus to a less sharp tropopause. Additionally, the background underestimates wind speed across the entire UTLS, especially near the UT wind maximum. Data assimilation improves temperature and wind in the analysis and acts to sharpen temperature and wind gradients near the tropopause. This influence on temperature (wind) is stronger for sharper tropopauses (strong wind situations) which are also characterised by high background errors. In addition, a positive influence on the tropopause altitude is found. 500 non-operational NAWDEX radiosondes are used in a special observing system experiment (OSE) that comprises two independent forecast runs, one with and one without assimilation of the 500 NAWDEX radiosondes. The OSE evaluation provides evidence that the tropopause sharpening and the improved wind profile can be mainly attributed to the assimilated radiosondes. An additional evaluation in model-space shows a similar average influence of data assimilation on temperature and wind beyond the observation location. Furthermore, a stronger influence on temperature and wind is indicated in regions associated with enhanced diabatic activity, such as the warm conveyor belt outflow.

This dissertation demonstrates how independent and assimilated campaign data can be effectively used to approach model biases at the tropopause from a novel perspective supporting progressive development of current NWP models. Findings on the LS moist bias in the IFS, obtained with the unique DIAL data, may contribute to reduce this error (and the associated temperature bias) in future IFS versions for instance by adjusting the representation of mixing processes, such as overshooting convection. The tropopause-based evaluation of data assimilation output not only provides new insights into the influence of assimilated radiosondes on temperature and wind at the tropopause in the IFS, but also paves the way for the application of these diagnostics to study observation impact and biases in other atmospheric phenomena relevant to NWP in the future.

Zusammenfassung

Genaue Anfangsbedingungen in numerischen Wettervorhersagemodellen sind eine grundlegende Voraussetzung für zuverlässige Wetterprognosen. Ein Schlüssel zur weiteren Verbesserung der Wettervorhersage ist die Beseitigung systematischer Fehler (engl. „biases“) in den Anfangsbedingungen, was eine präzise Identifikation dieser Fehler und ihrer Ursachen in den Modellen erfordert. Die vorliegende Dissertation konzentriert sich auf die Identifikation von Biases an der extratropischen Tropopause und in den sie umgebenden Schichten, der oberen Troposphäre und der unteren Stratosphäre (engl. „UTLS“). Die UTLS ist durch starke vertikale Gradienten in Temperatur, Wind und Wasserdampf geprägt. Diese Gradienten bilden Potenzielle Vortizität, deren Verteilung an der Tropopause als Wellenleiter der Rossby-Wellen wirkt, und somit die Wetterentwicklung in den mittleren Breiten bestimmt. Um die scharfen Temperatur-, Wind- und Wasserdampfgradienten der UTLS ausreichend zu erfassen, sind meteorologische Profilmessungen mit hoher Genauigkeit und vertikaler Auflösung nötig, die dem globalen Beobachtungssystem jedoch fehlen. In dieser Arbeit werden hochaufgelöste, präzise Kampagnenmessungen verwertet, um Biases und ihre Ursachen im weltweit führenden Modell dem Integrated Forecasting System (IFS) des European Centre for Medium-Range Weather Forecast (ECMWF) aus einer neuen Perspektive zu evaluieren.

Im ersten Teil werden Wasserdampfprofile eines flugzeuggetragenen differentiellen Absorptionslidars (DIAL) genutzt, um den vermuteten Feuchte-Bias der LS in der globalen Reanalyse ERA5 des ECMWF unabhängig zu bestimmen. Während 41 Flügen (6 Kampagnen) wurden 33.000 DIAL-Profile mit hoher Datenanzahl in der UTLS über eine Flugdistanz von mehr als 200.000 km gemessen. Die Messungen decken zudem verschiedene Jahreszeiten und verschiedene synoptische Situationen ab, die das typische Wetter der mittleren Breiten widerspiegeln. Dieser einzigartige Datensatz ermöglicht es, bisher unbeantwortete Fragen bezüglich des LS-Feuchtebias im ECMWF zu adressieren, die seine vertikale Struktur, seine saisonale/synoptische Variabilität, sowie seine Ursachen im Modell betreffen. Die statistische Auswertung zeigt einen ausgeprägten LS-Feuchtebias in ERA5, mit einem Maximum von 55 % in einer Höhe von 1.3 km oberhalb der Tropopause. Darüber zeigt die Analyse einen robusten Beweis einer stetigen Abnahme des Bias bis 4 km oberhalb der Tropopause. Der LS-Feuchtebias ist im Sommer stärker und nimmt eine tiefere Schicht in Situationen niedriger Tropopausenhöhe ein, die für Troglagen typisch sind. Ein weiterer neuer Aspekt dieser Arbeit ist, dass der LS-Feuchtebias direkt im Zusammenhang mit Mischungsprozessen untersucht wird. Dafür werden simultane Ozon- und Wasserdampf-DIAL-Daten verwendet, um die UTLS in troposphärische, stratosphärische und gemischte Luft zu klassifizieren. Es zeigt sich, dass der LS-Feuchtebias in der Mischungsschicht besonders stark ausgeprägt ist, während er in den troposphä-

risch und stratosphärischen Luftmassen gering ist. Dies ist ein bedeutender Hinweis, dass überschätzte Transport-/Mischungsprozesse über die Tropopause zu Entstehung des LS-Feuchtebias im Modell beitragen. Ein spezifischer Mischungsprozess, verantwortlich für den Transport großer Mengen an Wasserdampf in die LS, ist „Overshooting“ Konvektion. Um die Sensitivität des LS-Feuchtebias auf die Repräsentation von Konvektion im Modell zu testen, werden DIAL-Daten für einen Forschungsfluges mit starker Konvektion mit zwei IFS-Versionen mit unterschiedlicher Konvektions-parameterisierung verglichen. Die IFS-Version mit einer Limitierung von „Overshooting“ Konvektion weist einen geringeren LS-Feuchtebias auf.

Im zweiten Teil werden Datenassimilationsdiagnostiken verwendet, um Temperatur- und Wind-Biases in der UTLS und ihre Konsequenzen für die Tropopause im Modell zu untersuchen. Beobachtungen sollten im Assimilationsprozess dazu beitragen, Fehler in der Hintergrundvorhersage (engl. Background) zu reduzieren und die Analyse zu verbessern. Der Einfluss von Datenassimilation auf die scharfen vertikalen Gradienten an der Tropopause ist jedoch ein kontrovers diskutiertes Thema. Die vorliegende Arbeit analysiert den Einfluss assimilierter Radiosonden auf die Schärfe der Tropopause im IFS mithilfe von 9729 Profilen, die im Rahmen der Messkampagne "North Atlantic Waveguide Downstream Impact Experiment"(NAWDEX) gemessen. Temperatur- und Windprofile der Radiosonden werden mit ihren Modelläquivalenten verglichen, d.h. dem Background und der Analyse an der Position und zum Zeitpunkt der Beobachtung. Der Background überschätzt (unterschätzt) die Temperatur an der Tropopause (in der LS), was zu einer weniger abrupten Änderung der statischen Stabilität (N^2) und damit zu einer weniger scharfen Tropopause führt. Darüber hinaus unterschätzt der Background insbesondere die Windgeschwindigkeit in der UT. Ein bedeutendes Ergebnis ist, dass Datenassimilation die Temperatur- und Windprofile der Analyse verbessert, und somit deren vertikale Gradienten an der Tropopause schärft. Ein stärkerer Einfluss auf Temperatur (Wind) zeigt sich in Fällen schärferer Tropopausen (Jetstream), welche durch hohe Hintergrundfehler gekennzeichnet sind. Des Weiteren wird ein Öbserving System Experiment"(OSE) mit 500 nicht-operationellen NAWDEX-Radiosonden durchgeführt. Dieses umfasst zwei unabhängige Modellläufe, einen mit und einen ohne Assimilation der 500 NAWDEX-Radiosonden. Das OSE zeigt, dass die Schärfung der Tropopause und das verbesserte Windprofil auf die Assimilation der Radiosonden zurückzuführen ist. Eine zusätzliche Auswertung in „model-space“ zeigt einen ähnlichen, positiven Einfluss der Datenassimilation auf Temperatur und Wind an der Tropopause auch über den Ort der Radiosonden hinaus. Insbesondere wird ein großer Einfluss in Regionen mit potentiell hoher diabatischer Aktivität gefunden.

Diese Dissertation zeigt, wie unabhängige und assimilierte Kampagnendaten genutzt werden können, um Modellfehler an der Tropopause und ihre Ursachen effektiv zu analysieren, und somit die Weiterentwicklung moderner Modelle voranzutreiben. Die Erkenntnisse über den LS-Feuchte-Bias im IFS, können dazu beitragen, diesen Fehler (und den damit verbundenen Temperaturfehler) in zukünftigen IFS-Versionen zu reduzieren, z.B. durch Anpassung von Mischungsprozessen. Die Tropopausen-basierte Auswertung Datenassimilations-output liefert nicht nur neue Erkenntnisse über den Einfluss von assimilierten Radiosonden auf die Tropopause im Modell, sondern ebnet auch den Weg für zukünftige Untersuchungen systematischer Modellfehler in anderen atmosphärische Phänomenen, die für die Wettervorhersage von Bedeutung sind.

Chapter 1

Introduction

Weather is ubiquitous. Over 8 billion people on Earth are inevitably affected by the weather across all aspects of life, and thus reliable forecasts are universally relevant. Accurate forecasts not only contribute greatly to financial prosperity in a number of sectors such as energy, agriculture and transport, but are also critical for protecting lives. A good example of this is the forecast of the tropical cyclone SANDY (October 2012), whose intensity and track was predicted several days in advance (Magnusson et al., 2013), enabling authorities to avert greater damage and evacuate people from threatened areas.

"Weather" refers to the physical state of the Earth's atmosphere, fundamentally defined by its key variables pressure, density, temperature, composition and motion. Forecasting the weather is essentially determining how these variables will change over time. In the early 20th century, Abbe and Bjerknes recognized that a set of differential equations - the Navier-Stokes-Equations, the 1st law of thermodynamics, and the mass continuity equation, can be used to describe changes of the atmosphere's physical state. Based on the knowledge of the initial state of the atmosphere, e.g. through meteorological observations of the key variables, its future state can be predicted by numerically solving these equations (Abbe, 1901, Bjerknes, 1904). This is the fundamental approach of modern numerical weather prediction (NWP). The first pioneering NWP attempts at solving these equations to derive 6-72 hour pressure tendencies proved the concept in principle (Richardson, 1922, Bolin, 1955), 1955, but the forecast accuracy was poor. Over the recent six decades, forecast skill of short-, mid- and long-term predictions have improved drastically (Figure 1.1) (Stern and Davidson, 2015, Bauer et al., 2015) with a gain in forecast lead time of about one day per decade in forecast skill. This ongoing progress, termed "the quiet revolution of NWP" (Bauer et al., 2015), is driven by major model developments such as improved parameterizations, the introduction of ensemble methods, advanced data assimilation systems, and by the enhanced global observing system. The "NWP's revolution" is also supported by the increased computing power available, which enables a higher spatial and temporal grid resolution of current NWP models.

This progressive trend, however, will probably not continue forever, as various studies consistently expect an intrinsic limit of NWP forecasts between two and three weeks (Lorenz, 1969, Selz, 2019, Selz et al., 2022) – however, this intrinsic limit has not yet been exhausted by modern NWP models. This is related to the presence of model deficiencies (e.g., erroneous parameterizations, simplified dynamics) and of systematic errors in the

initial conditions of NWP models, which grow rapidly in the prediction across scales and times, leading to loss of predictability (e.g., Selz and Craig, 2015, Craig et al., 2021, and references therein). Due to the highly non-linear and chaotic nature of the atmosphere (its underlying equations, respectively) (Lorenz, 1963), the forecast is highly sensitive to the accuracy of the initial state. Precise initial conditions in NWP systems are therefore critical for reliable weather forecasts. Selz et al. (2022) demonstrate that a potential reduction of initial condition errors in a state-of-the-art NWP model would lead to a gain in predictability and thus to an approach to the intrinsic limit. Identifying systematic initial condition errors and their underlying sources in models is a key prerequisite to support the ongoing progress of NWP (Bauer et al., 2015), which is the overarching motivation of this work.

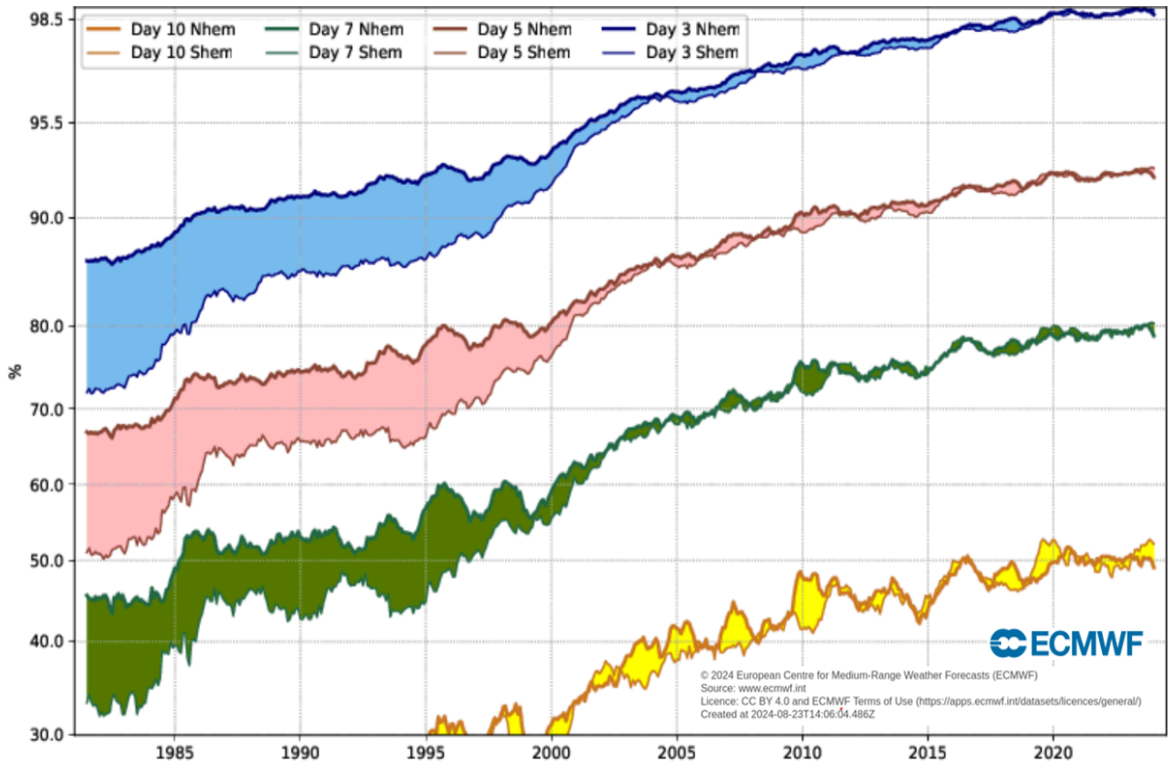

Figure 1.1: A measure of forecast skill at three-, five-, seven- and ten-day ranges, for the mid-latitude northern and the southern hemisphere. Forecast skill is the correlation between the forecasts and the verifying analysis of the height of the 500-hPa level, expressed as the anomaly with respect to the climatological height. Courtesy of the ECMWF.

The scientific objective of this dissertation is to evaluate systematic model errors in the initial conditions, specifically at the mid-latitude tropopause, a region with particular importance for mid-latitude weather forecasting. To provide more context to the aims of this work, Sect. 1.1 introduces the basic characteristics of the extratropical tropopause and their relevance for NWP. Sect. 1.2 presents the necessary knowledge on existing initial condition errors near tropopause in current NWP systems and enlightens the need for data assimilation methods and independent campaign observations to address open aspects on these errors. The key scientific research questions are then stated in Sect. 1.3.

1.1 The mid-latitude tropopause

In the period before the 20th century, scientists believed that the temperature in the Earth's atmosphere steadily decreases with increasing altitude. The invention of manned balloon systems around the turn of the 20th century, allowed temperature measurements to be conducted at high altitudes ($>10\,\mathrm{km}$) (e.g., Assmann, 1902). A group of researchers around the German meteorologist Richard Assmann made an unexpected discovery evaluating balloon-based temperature measurements in Berlin in 1901. They found that the air temperature reaches a minimum at a certain altitude (which varies with time, see Figure 1.2). Above this minimum, temperature remained constant or even increased with altitude, which they termed a "warm air current".

Based on this finding on the vertical temperature profile, the Earth's atmosphere can be divided into different layers, that differ in terms of thermal stratification, and boundaries between two layers are termed "pauses". The lowermost layer of the atmosphere, in which the temperature decreases with altitude, is the troposphere. The layer above it, in which the temperature is constant or increasing with altitude, is the stratosphere. The tropopause is the boundary between the troposphere and the stratosphere.

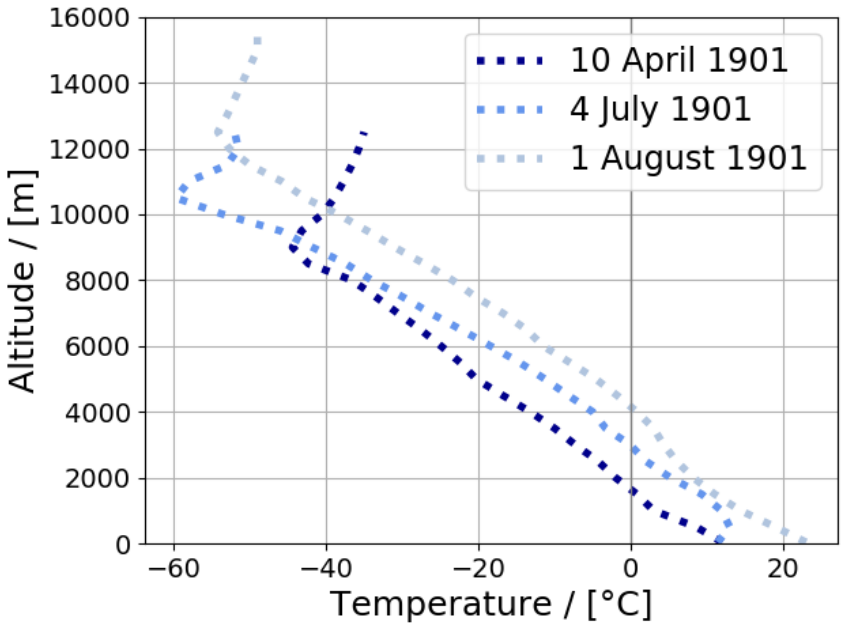

Figure 1.2: Temperature profiles of three balloon ascents (1901, Berlin) as constructed from Assmann's historic balloon ascents (Assmann, 1902).

Today, the tropopause is recognized as an elementary feature of the atmosphere. The physical conditions at the extratropical tropopause and in the surrounding region ($\pm 5\,\mathrm{km}$), the upper troposphere and lower stratosphere (UTLS) are known to affect the atmospheric general circulation (e.g., position and intensity of jet streams) and in turn mid-latitude weather (Harvey et al., 2016, Haualand and Spengler, 2021, Boljka and Birner, 2022). In the following, the typical physical conditions that characterise the tropopause, their relation to atmospheric dynamics, and their role for NWP are elucidated.

Climatological distributions of the key variables across the tropopause
Birner (2006) illustrates the average vertical profiles of key variables in tropopause-based coordinates as observed by radiosondes in the mid-latitudes (see Figure 1.3). In the UT, temperature decreases with altitude with a typical lapse rate of approximately 6-10 K per kilometre up to the tropopause, which is characterised by a distinct temperature minimum. Above the tropopause, a temperature inversion (variable in thickness, but on average roughly 1-2 km deep) is followed by a nearly isothermal temperature profile in the LS. The reversal of the vertical temperature gradient across the tropopause leads to a rapid change of the stratification from the UT to the LS, which is reflected in the profile of the buoyancy frequency squared (N^2; a measure of static stability). From a typically low mean value in the UT (about $1 \times 10^{-4}\, s^{-2}$), N^2 abruptly maximises in a distinct peak (usually $> 5 \times 10^{-4}\, s^{-2}$) occurring at the tropopause altitude. Within the temperature inversion, N^2 decreases to typical stratospheric values of about $4 \times 10^{-4}\, s^{-2}$; above N^2 remains approximately constant in the isothermal LS. The region of enhanced N^2 just above the tropopause is known as the Tropopause Inversion Layer (TIL) (Birner, 2006). The high static stability within the TIL restricts vertical motion and thus acts as a barrier to mass exchange between the UT and the LS (Gettelman et al., 2011). This leads to sharp vertical distributions of trace species across the tropopause, with the highest concentrations occurring in the associated source layers (Pan et al., 2007).

Consequently, water vapour concentrations are large (> 0.1-$1\,\mathrm{g\,kg^{-1}}$) in the UT and decrease sharply with altitude towards low concentrations ($< 10^{-2}\,\mathrm{g\ kg^{-1}}$) in the LS, with the strongest gradients occurring just above the tropopause within the TIL (for example, Kaufmann et al., 2018, Bland et al., 2021). Ozone concentrations, on the contrary, increase with altitude, with low values in the UT and increasing concentrations in the LS (Zahn et al., 2014, Ferreira et al., 2016). Water vapour and ozone are greenhouse gases, and their described distributions have important implications for radiative transfer and the local radiation budget: The rapid decrease of water vapour above the tropopause allows long-wave radiation to be effectively emitted into space, leading to strong cooling near the tropopause while short-wave heating of ozone contributes to an increasing warming with altitude in the stratosphere. This radiative influence of water vapour and ozone on the UTLS temperature profile is a major factor for maintaining the TIL (Randel et al., 2007, Kunz et al., 2009, Ferreira et al., 2016), and thus modifies the sharpness of the tropopause.

The vertical distribution of wind speed and wind shear in the extratropical UTLS is highly variable, which is due to the frequent changes in position and intensity of the jet streams (e.g., Schäfler et al., 2020). On average, wind speed in the troposphere increases linearly with altitude, up to a wind maximum occurring at about 1 km below the tropopause that is related to the jet streams (see Figure 1.1). Above this maximum and in the LS, wind speed decreases constantly with altitude. Corresponding to this wind distribution, vertical shear of the horizontal wind is positive below and negative above the UT wind maximum, with a 1 km-thick layer of enhanced wind shear just above the tropopause (Birner, 2006, Schäfler et al., 2020) also referred to as the tropopause shear layer (TSL) (Kaluza et al., 2021). The increased shear in the TSL is associated with enhanced turbulence (Schäfler et al., 2023) and may impact troposphere-stratosphere exchange processes (Stohl et al., 2003).

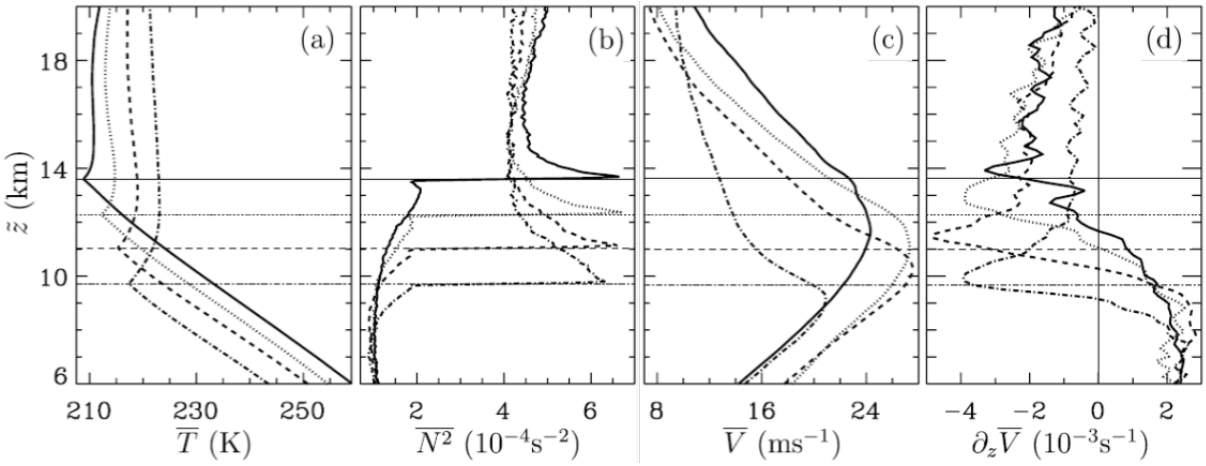

Figure 1.3: Annual tropopause-based profiles of (a) temperature, (b) N^2, (c) horizontal wind, and (d) vertical shear of the horizontal wind of 4 mid-latitude stations in the US: Miramar (33°N, 117°W, solid); Reno (40°N, 120°W, dotted); Quillayute (48°N, 125°W, dashed); and Yakutat (60°N, 140°W, dash-dotted). Horizontal lines denote the average tropopause altitude. Figure adopted from Birner (2006).

Birner's radiosonde climatology (Figure 1.3) and Assmann's observations (Figure 1.2) show the altitude of the extratropical tropopause to be highly variable (i.e., with latitude and time) which is associated with the seasonal and synoptical patterns that characterise the nature of mid-latitude weather. Figure 1.4 presents a cross section of the Northern Hemisphere at an example winter day in 2006 in which the UTLS is highlighted (purple shading) as a layer extending polewards of the subtropical jets spanning an altitude range from about 6 to 20 km (Gettelman et al., 2011). The UTLS is a transitional zone interconnecting the free troposphere and the stably stratified stratosphere, and any exchange processes between the UT and the LS involves the tropopause (Gettelman et al., 2011). The tropopause (red dashes in Figure 1.4) is embedded within the UTLS and its altitude decreases from about >15 km equatorwards of the subtropical jet towards altitudes <10 km in the mid- and high latitudes. However, this latitudinal decrease is not continuous, as the tropopause rapidly drops in altitude across the jet streams.

The transition from the UT to the LS (that fundamentally differ in thermal, dynamical and chemical properties) is characterised by sharp vertical gradients of key variables that culminate near the tropopause (Pan et al., 2004, Birner, 2006, Zahn et al., 2014). For these variables, several definitions of the tropopause have been established over the years relying on the different thermal, dynamical and chemical properties of the UTLS. The thermal definition is based on the temperature gradient (WMO, 1957) and a widely used reference to capture the sharp vertical distributions of key atmospheric quantities (e.g., Birner et al., 2002, Hoinka, 1999, Randel et al., 2007). Another stability-based definition, recently developed by Tinney et al. (2022), uses the potential temperature gradient to determine the tropopause, but this definition largely agrees with the thermal tropopause. Chemists use concentrations of trace species to define a chemical tropopause, e.g., by using a threshold value of ozone concentration (e.g., WMO, 1986, Hoor et al., 2002, Zahn et al., 2004). In addition, a dynamical tropopause definition has established in the field of atmospheric dynamics in recent years (Hoerling et al., 1991). Figure 1.4 shows that

the dynamic tropopause is at a similar height to the thermal tropopause over almost the entire section, with the largest differences occurring near the jet streams. The dynamic tropopause is based on the potential vorticity (PV), a quantity, that takes both, stability and wind gradients into account, providing a direct explanation for the deviations of the dynamic tropopause from the thermal tropopause near the jet streams. The PV concept and its role for the dynamics in the mid-latitudes are explained in the following paragraph.

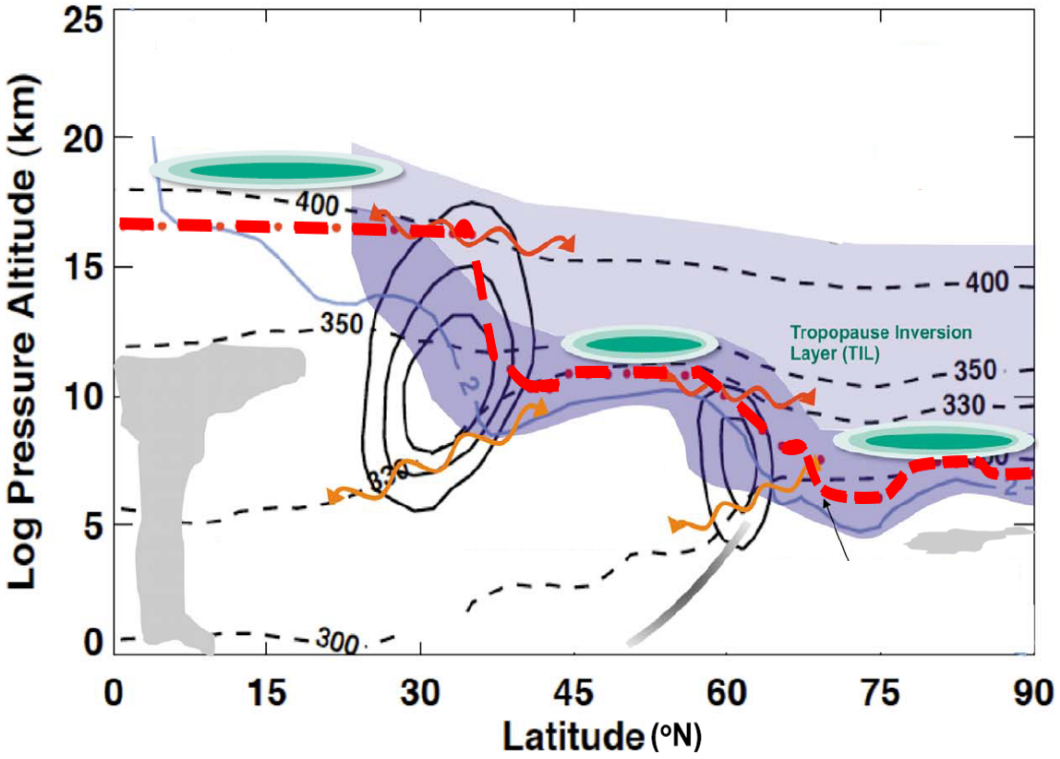

Figure 1.4: Schematic snapshot along a Northern Hemisphere section on 15 February 2006. Wind (black contours, 10 m s^{-1} interval), potential temperature surfaces (black dashes), thermal tropopause (thick red dashes), and dynamical tropopause (blue solid line). Illustrated are the UTLS (dark and light purple shading), clouds and fronts (gray shading), static stability contours (green shading), quasi-isentropic exchange (red wavy arrows), cross-isentropic exchange (orange wavy arrows). Figure adapted from Gettelman et al. (2011).

The role of the tropopause for the mid-latitude dynamics

The outlined sharp cross-tropopause gradients of temperature and wind determine the distribution of PV, a vital quantity in atmospheric dynamics combining kinematic and thermodynamic properties of air. PV is conserved for adiabatic and friction-less motions which allows to understand the large-scale flow of the atmosphere, where an air parcel moves along isentropes (i.e., isolines of potential temperature) (Hoskins et al., 1985). Based on Ertel's definition (Ertel, 1942), PV is composed of a vorticity term (I) and a stability term (II) (see Equation 1.1):

$$PV = \frac{1}{\rho} \underbrace{(\vec{\nabla} \times \vec{\mathbf{v}} + 2\vec{\Omega})}_{\text{(I)}} \cdot \underbrace{(\vec{\nabla}\theta)}_{\text{(II)}} \tag{1.1}$$

with the air density ρ, the wind vector $\vec{\mathbf{v}}$, the angular velocity of the rotating Earth $\vec{\mathbf{\Omega}}$, the potential temperature θ and the Nabla operator $\vec{\mathbf{\nabla}}$. For synoptic-scale motions, the third term of the scalar product of Ertel's PV is dominant. Using the hydrostatic approximation, the isentropic PV can be formulated as

$$PV_\theta \approx -g(\zeta_\theta + f) \cdot \frac{\partial \theta}{\partial p}, \tag{1.2}$$

where PV_θ is a product of absolute vorticity and vertical stability which contain the vertical gradients of wind, and (potential) temperature. The absolute vorticity comprises the latitude-dependent planetary vorticity (with the Coriolis parameter $f = 2\vec{\mathbf{\Omega}}\sin(\phi)$, where ϕ is the latitude) and the vertical component of the relative vorticity $\zeta_\theta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}$. PV is given in "PV units" (PVU) with 1 PVU being equivalent to $10^{-6}\,\mathrm{K\,m^2\,kg^{-1}\,s^{-1}}$.

Hoskins et al. (1985) visualised the PV on isentropic maps and show that the PV distribution in the extratropics is related to the occurrence of large-scale dynamic systems (i.e., ridges and troughs). Furthermore, PV varies with latitude and altitude. Typically, PV is low in the troposphere (PVU < 1), and high (> 2 PVU) in the stratosphere which is caused by the differing stratification. Based on this transition, PV thresholds between 1.5 and 3.5 PVU are used to define the dynamical tropopause (Hoerling et al., 1991, Gettelman et al., 2011). The meridional distribution of the dynamical tropopause (Figure 1.4) shows an abrupt increase of the tropopause across the subtropical and mid-latitude jet streams. The associated sharp meridional contrasts of the isentropic PV (illustrated in Figure 1.6) across the tropopause act as a waveguide for the large-scale, meandering waves in the extratropical westerly flow that are known as Rossby waves (Schwierz et al., 2004, Martius et al., 2010). The Rossby waves play a key role in the development of the synoptic scale features that characterise the nature of mid-latitude weather (Hoskins et al., 1985). Using model experiments, Boljka and Birner (2022) demonstrate that sharper PV gradients strongly influence on atmospheric features (e.g., the jet stream). Hence, a realistic representation of the sharp PV gradients and in turn, of the temperature and wind gradients at the tropopause in NWP models, is crucial for accurate weather and climate forecasts (e.g., Harvey et al., 2018).

Although PV is conserved under the assumption of adiabatic and frictionless flows, it is altered by non-conservative, diabatic processes. As diabatic processes are sources and sinks of PV, they essentially influence the evolution and propagation of synoptic weather patterns (Wernli and Gray, 2023). In turn, their accurate parameterization is critical to correctly represent the PV distributions. The PV gradients in NWP models are continuously modified by both, adiabatic and diabatic processes (Bukenberger et al., 2023), and rapidly decay within short (< 24 h) forecast lead times (Gray et al., 2014). This is attributed to a smoothing effect of the advection scheme that dominates sharpening effects of parameterized diabatic processes such as radiative cooling by water vapour, microphysics, and turbulent mixing (Chagnon et al., 2013, Saffin et al., 2017). The weakening PV gradients are likely associated with background forecast errors of temperature, humidity and wind at the tropopause. Due to the importance of the PV distribution at the tropopause for the dynamics in the mid-latitudes, a robust identification of systematic temperature, water vapour (through its radiative influence on temperature) and wind errors in NWP models is needed (e.g., Lavers et al., 2023). This would support model development and thus pave the way for improved forecast quality in the extratropics.

1.2 Initial condition errors in the ECMWF IFS

In this dissertation, the focus is on the evaluation of systematic initial condition errors (termed "biases") in temperature, humidity and wind across the extratropical tropopause in the European Centre for Medium-Range Weather Forecast's (ECMWF) Integrated Forecasting System (IFS) that is among the most advanced NWP models on Earth (e.g., Haiden et al., 2023). There are two main approaches to investigate initial condition biases in models: The application of data assimilation diagnostics and the use of (model) independent campaign observations. Before these approaches are detailed out in Sect. 1.2.3 and Sect. 1.2.4, first the term "initial conditions", and the fundamental challenges in their generation are explained in Sect. 1.2.1. Subsequently, a brief review on known initial condition biases near tropopause in the ECMWF IFS is given (Sect. 1.2.2).

1.2.1 Challenges in the definition of the initial conditions

NWP is an initial value problem. To calculate a weather forecast, an NWP model requires information of the atmosphere's initial state (i.e., meteorological variables on a model grid) that serves as starting point to calculate the dynamical development of its future state. The mathematical process that generates the initial conditions is called "data assimilation". A data assimilation scheme combines observations with a previous short-term forecast (referred to as "background") to produce the "best estimate" of the atmosphere's initial conditions (referred to as "analysis"). Both, background and observations are imperfect, which can lead to an erroneous analysis (e.g., Laloyaux et al., 2020).

Errors in the background forecast can have a variety of causes. Firstly, the previous analysis may contain errors already, which are then propagated into the short-term forecast. Further background forecast errors are related to imperfections in the model itself such as approximations in the numerics, or parameterizations. Figure 1.5a provides an overview of important parameterizations of physical processes in the atmosphere that are too small-scale to be explicitly resolved by the model grid, such as convective processes, radiative processes, chemistry and cloud microphysics. NWP models attempt to represent these sub-grid scale processes by means of simplified assumptions which may not sufficiently capture the full complexity of real processes. Hence, parameterization schemes are important sources of background errors, which in turn also propagate into initial condition errors.

Another critical source of initial condition errors are deficiencies in the global observation network (Figure 2.7b). Meteorological observations are carried out from conventional observations such as ground stations, buoys, aircraft and radiosondes. These data however, are spatiotemporally inhomogeneously distributed and therefore cannot fully reflect the real state of the atmosphere at a certain point in time (Laloyaux et al., 2020). The vast majority of the data are non-conventional observations such as active and passive remote sensing satellite instruments. They provide a fairly uniform global coverage of data available for data assimilation, but are limited in horizontal and vertical resolution and accuracy. Observations in general have measurement uncertainties, a different spatial resolution than the NWP system and may not measure the prognostic variables directly (e.g., satellite radiances), leading to representativeness errors when assimilated by the models (e.g., Janjić et al., 2018) which in turn affects the formulation of the initial states.

It should be noted that current weather and climate prediction models suffer greatly from the sparse global observing system, which provides only a limited number of conventional observations with high vertical resolution, although required for the initialization of the tropopause in the data assimilation process (Haualand and Spengler, 2021). Even due to their absence, the tropopause region is subject to increased errors in the initial conditions, but also in forecasts (Hakim, 2005).

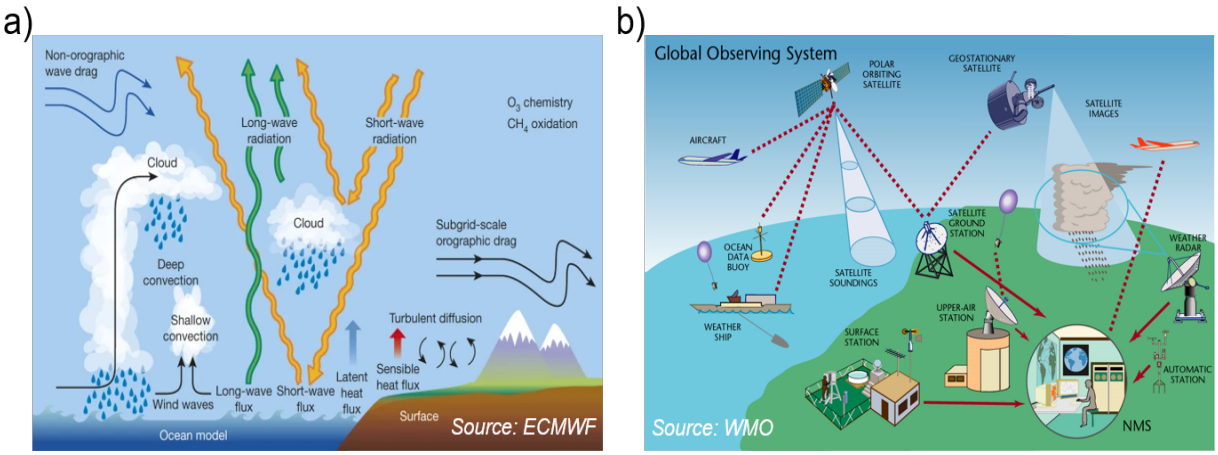

Figure 1.5: a) Atmospheric (sub-grid) scale processes that are parameterized in NWP models (adopted from Bauer et al. (2015)) and b) the global observing network.

.

1.2.2 Systematic errors near tropopause in the ECMWF IFS

The physical conditions (i.e. the sharp vertical gradients) at the tropopause are difficult to resolve for both models and observations, making the tropopause an inherently favourable region for systematic errors in the initial conditions. The following is a brief overview of such existing knowledge, but also on open aspects on temperature, humidity and wind biases in current NWP models that are documented in the literature.

Temperature bias

A number of recent studies showed a warm bias (few tenths of K) near the tropopause to be present in ECMWF IFS analyses, short-range forecast and reanalysis data (Bonavita, 2014, Ingleby et al., 2016, Bland et al., 2021, Ingleby, 2017). This warm bias may be related to the limited ability of the IFS to fully resolve the pronounced temperature minimum of the tropopause (Ingleby et al., 2016) or to the assimilation of warm-biased aircraft observations at altitudes associated with the tropopause (Ingleby, 2017). Beyond the tropopause, the IFS is influenced by a cold bias in the LS (Shepherd et al., 2018), which has a similar magnitude to the warm bias and peaks at about 1 km above the tropopause (e.g., Dyroff et al., 2015, Carminati et al., 2019, Bland et al., 2021). There is a strong indication in the literature, that the cold bias in the LS is caused by excessive radiative cooling of water vapour (Stenke et al., 2008, Diamantakis and Flemming, 2014, Shepherd et al., 2018, Bland et al., 2021).

Humidity bias

The existence of a systematic lower-stratospheric moist bias has been known for decades from previous versions of IFS analyses and reanalyses (e.g., Oikonomou and O'Neill, 2006, Kaufmann et al., 2018, Kunz et al., 2014, Woiwode et al., 2020) and other models (Davis et al., 2017). As little is known about the underlying sources, the LS moist bias is still a problem in modern versions of the IFS (Shepherd et al., 2018), and consequently has attracted increasing interest in recent years. The following briefly summarizes what is known about this bias:

Bland et al. (2021) used radiosonde observations of a two-month period in autumn to address a potential humidity bias in operational analyses of the IFS and the Met Office's Unified Model (METUM). They confirmed the magnitude of this moist bias (of about 70 % in the LS) documented in other studies (Dyroff et al., 2015, Kaufmann et al., 2018). The vertical structure of the moist bias was characterised by a small positive bias in the UT below the thermal tropopause, followed by a vertical increase in the LS to a maximum at 1-2 km above the tropopause (Bland et al., 2021).

However, the vertical thickness of the bias, in particular its vertical structure above its presumed maximum is under debate. A recent study presents case studies of vertically increasing bias, constant bias and decreasing moisture bias with height in the LS in IFS analysis and forecast data based on humidity cross sections from an airborne passive imager (Woiwode et al., 2020). In addition, Bland et al. (2021) show an opposing vertical structure of the LS moist bias using two different radiosonde types. While the radiosonde type RS92 type suggests a vertical constant bias in the LS up to 5 km above the tropopause, the type RS41 indicates a decreasing bias above its maximum. It should be stressed that due to the high uncertainty of the humidity sensors of both radiosonde types, Bland et al. (2021) restricted their analysis to 2 km above the tropopause.

In the literature, there are only assumptions about the potential origin and processes shaping the LS moist bias in NWP systems: one hypothesis is that it is caused by misrepresented dynamical transport and mixing processes (Kunz et al., 2014, Shepherd et al., 2018) which would be supported by the seasonal nature of the bias (Dyroff et al., 2015). Another suggested bias source is that insufficient model resolution, and numerical diffusion in the semi-Lagrangian advection scheme used in the IFS, lead to an excessive transport of moisture from the UT into the LS (Stenke et al., 2008, Kunz et al., 2014, Dyroff et al., 2015, Shepherd et al., 2018). However, a LS moist bias of similar order is also indicated for "Eulerian"-formulated models (e.g., Davis et al., 2017). Moreover, Woiwode et al. (2020) and Shepherd et al. (2018) showed the bias to be already present in the initial conditions and demonstrated a low response of the moist bias to variable vertical or temporal model resolutions.

Wind bias

Schäfler et al. (2020) used Doppler-Wind-Lidar (DWL) (Weissmann et al., 2005, Witschas et al., 2017) profile data in the midlatitudes to investigate wind errors in analyses of the ECMWF IFS and the METUM. Along a time series, they show large wind errors of more than 10 $\mathrm{m\,s^{-1}}$ in both models, with the strongest errors occurring near the tropopause and at situations with high observed wind speeds, suggesting a connection with the passage of jet streams. In particular, they found a systematic underestimation (i.e., a slow bias)

of wind speed maximum of the jet stream and in the lowermost LS. The slow wind bias related to the jet stream was confirmed by Lavers et al. (2023) who detected a vertically increasing slow wind bias ($> 0.5\,\mathrm{m\,s^{-1}}$) in the IFS analysis in the UT at observed wind speeds larger than $50\,\mathrm{m\,s^{-1}}$. The systematic wind speed underestimation in models is in line with Birner et al. (2002) who found notably underestimated UT wind maxima and vertical wind shear already in the ERA–15 reanalysis. Quantitative studies revealed an underestimation of the magnitude of vertical wind shear at the tropopause by a factor of 2–5 in the IFS (Houchi et al., 2010, Schäfler et al., 2020). Lavers et al. (2023) speculate that the slow wind bias may be related to either numerical diffusion processes or an overactive turbulent mixing parameterization scheme, which could lead to excessive vertical transport in regions with increased wind shear and thus possibly foster the LS moist bias in the IFS.

The outlined model biases in temperature, humidity and wind are also expected to influence the representation of the tropopause and their sharp gradients in models (e.g., Birner et al., 2002, Jung et al., 2003). As described at the beginning of Sect. 1.2, data assimilation methods and meteorological observations are key elements for the generation of initial conditions in NWP models. To identify biases in the initial conditions, there are two related approaches: I) the application of data assimilation diagnostics, and II) a direct comparison of independent observations with model states. Since these two approaches are pursued in this dissertation, they are explained in Sect. 1.2.3 and Sect. 1.2.4.

1.2.3 Data assimilation diagnostics

Data assimilation provides a means of correcting for model errors by optimally blending observation and background forecast data. If their error characteristics are properly known, a data assimilation system should reduce errors in the analysis. The evaluation of data assimilation output is a primary method for determining such bias characteristics (e.g., Bauer et al., 2015, Laloyaux et al., 2022). A well-established approach for identifying model biases is the direct comparison of observations, background and analysis in the so-called "observation-minus-background (O-B)" and "observation-minus-analysis" (O-A) statistics (Weissmann et al., 2012, Martin et al., 2023, Schäfler et al., 2024).

Another useful approach to study the impact of particular observations or at a specific regions of interest on the analysis are Observing System Experiments (OSEs) (Bouttier and Kelly, 2001, Schindler et al., 2020). A basic OSE comprises two independent model runs, one control (CTR) and one denial (DEN) run. The "CTR run" uses the operational configuration of the model, in which all available observations are assimilated. In the DEN run certain subsets of the observation network (e.g. individual platforms or areas) are excluded from data assimilation. By comparing (O-B/O-A) statistics for CTR and DEN the benefit of the denied observations can be estimated. Although OSEs are an extremely valuable technique for estimating the impact of observations, but require high computational costs.

Both methods, OSEs and also (O-B/O-A) statistics are suitable for diagnosing model biases and allow to study whether assimilated observations act to reduce biases. Numerous studies based on such methods show that assimilation of observations reduces errors in short-range forecast in the UTLS, reflected by an improved representation of temperature

(e.g., Radnoti et al., 2010, Bonavita, 2014) and wind (e.g., Weissmann and Cardinali, 2007, Weissmann et al., 2012, Lavers et al., 2023, Martin et al., 2023, Borne et al., 2023) profiles in the analysis. Despite this positive influence in a statistical sense, it is not clear how the assimilation of observations affects the sharp gradients at the tropopause. Bannister et al. (2020) demonstrate that data assimilation can act to smooth sharp vertical temperature and humidity gradients in the planetary boundary layer, where gradients are roughly comparable in strength to those near the tropopause.

There are only two dedicated studies that focus on the influence assimilated observations on the representation of the mid-latitude tropopause, specifically on tropopause sharpness. Birner et al. (2006) investigate the role of assimilated satellite data on tropopauses conducting an OSE in the Canadian Middle Atmosphere Model (CMAM). A decrease of N^2_{max} in the experiment with assimilated satellite data compared to a model run without satellite data suggested that satellite data assimilation smears out the gradients near the tropopause. A more recent study by Pilch Kedzierski et al. (2016) analysed the data assimilation influence on the tropopause sharpness comparing background and analysis at the positions of GPS radio occultation (GPS–RO) observations in the ECMWF's ERA–Interim reanalysis and IFS analysis. The detected increase of N^2 in an about 1 km-thick layer just above the tropopause and a decrease of N^2 above and below this layer is corresponding to a tropopause sharpening, which was attributed to the assimilation of GPS–RO observations. Notably, both studies (Birner et al., 2006, Pilch Kedzierski et al., 2016) not only show a different data assimilation influence on the tropopause sharpness, but also differ in essential aspects, which makes them rather difficult to compare. In particular, the studies differ in terms of the applied methods to diagnose the influence, the used observation type, the spatial resolution of observations and model, and different generations of data assimilation schemes. Moreover, it has to be noted that the IFS has also undergone significant further developments in the recent years. These relate, for example, to increased model resolution and the introduction of a hybrid data assimilation system (Bonavita et al., 2016) (more information in Sect. 2.2). It remains open how assimilated observations in the current configuration of the ECMWF IFS would influence the tropopause.

1.2.4 Independent campaign data

Through the data assimilation process, operational meteorological observations are directly involved in the formulation of the optimal initial conditions. Field campaigns are the only source providing observational data independent of the model states making them highly suitable to characterise signature of biases in the initial conditions. Campaign data also offers the opportunity to analyse specific atmospheric processes and their representation in models in specific regions.

In this work, a high number of observations from various measurement campaigns are evaluated. In the following the research campaign providing the key observational data sets is introduced. The North Atlantic Waveguide and Downstream Impact EXperiment (NAWDEX) was an international field campaign that took place between 17 September and 22 October 2016. NAWDEX aimed to better understand the influence of diabatic processes on jet stream disturbances, their development, and their role on high-impact

weather downstream (Schäfler et al., 2018). Weather systems with strong diabatic activity are suspected to be a potential source of strong forecast errors (e.g., Harr et al., 1992, Wernli and Gray, 2023). NAWDEX targeted regions of high diabatic activity along the Northern Atlantic's mid-latitude storm track, such as the inflow, ascent and outflow of Warm Conveyor Belts (WCBs) (Figure 1.6). To address the influence of diabatic processes, accurate and highly resolved observations of temperature, moisture, wind and cloud properties from the surface to the lower stratosphere are sampled. As the data coverage of conventional observations alone in the Northern Atlantic region is sparse (e.g., Ingleby, 2017), 4 research aircraft and about 40 ground-based stations were coordinated in 13 intensive observing periods to intensively collect observational data across Northern Atlantic's mid-latitude storm track.

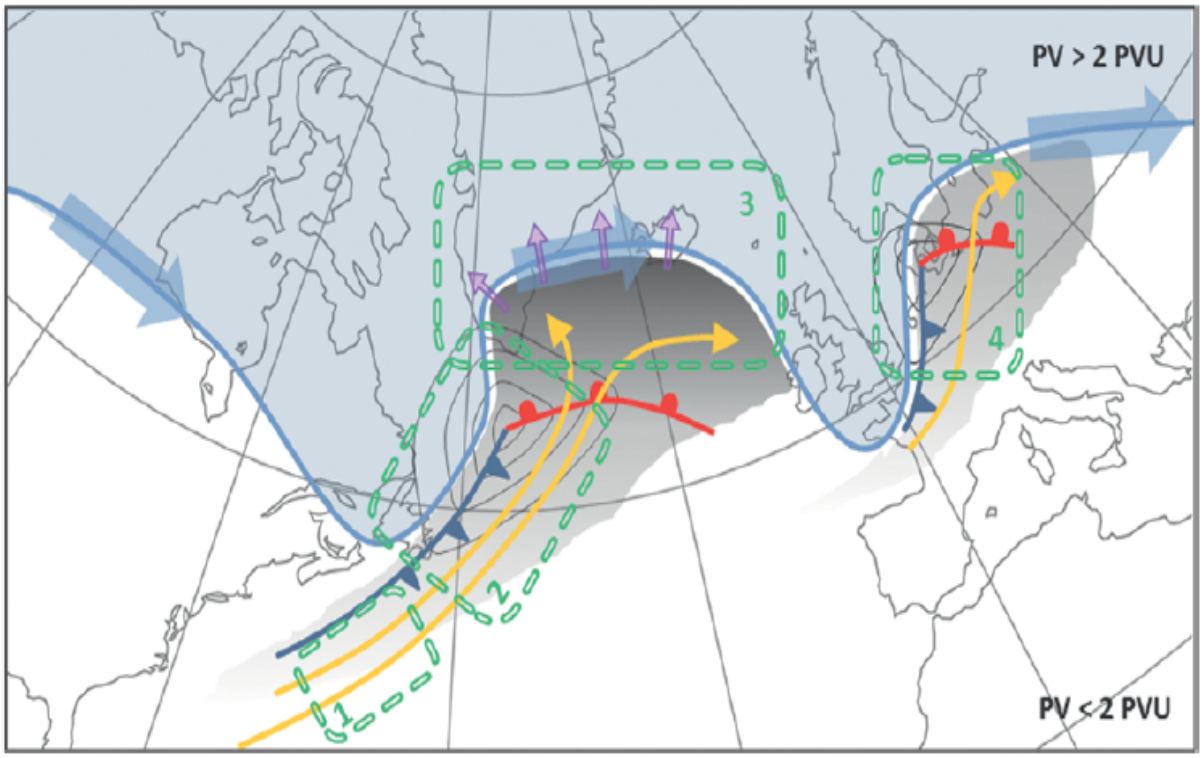

Figure 1.6: An idealized weather situation during NAWDEX. The blue line marks the waveguide with a strong isentropic PV gradient separating LS (blue shading) from UT air (white shading). The jet stream (blue arrows) follows the waveguide. The surface low (gray lines indicate sea level pressure, dark blue/red lines its cold/warm front) develops below the leading edge of upper-level positive PV anomalies. Gray-shaded areas illustrate clouds related to ascending WCBs (yellow arrows). Purple arrows mark divergent outflow at the tropopause. The green boxes show the key target regions: inflow (1), ascent (2), outflow (3) of WCBs, and a region of expected downstream impact (4). Figure adopted from Schäfler et al. (2018).

The research aircraft were equipped with several types of advanced passive and active remote sensing and in-situ instruments. This payload ensures excellent data coverage from the boundary layer to the lower stratosphere (i.e. altitude of the aircraft) as well as inside and outside the clouds. In addition, a high number of dropsondes (289) were released during the flights. Complementing the aircraft data, ground-based stations from Canada,

Iceland and Europe contributed with a higher frequency (about 589) of operational radiosonde releases. Both, the radiosonde and the dropsonde data were sent directly to the weather centres and used for operational weather forecasting. More details on the instrumentation, research targets and a synoptic overview are given in Schäfler et al. (2018).

Overall, the vast variety of the observation types operated during the NAWDEX field campaign provide a unique and comprehensive data set comprising accurate measurements of wind temperature, wind and water vapour profiles as well as cloud and aerosol properties with outstanding coverage of the UTLS. The active water vapour lidar and radiosonde measurements, which are used in this dissertation, are characterised by high vertical resolution and accuracy, allowing to capture the sharp vertical gradients at the tropopause, and thus to validate model biases in this region accordingly. It has to be emphasized that the majority of the NAWDEX is sampled in remote regions (i.e., North Atlantic) where NWP models lack precise and vertically highly resolved conventional observations. In addition, weather situations were probed with large expected forecast errors related to the strong diabatic activity (Wernli and Gray, 2023). The NAWDEX radiosonde observations are also used by the ECMWF to conduct an OSE, which allows to study the particular influence of the assimilated observations on the tropopause.

1.3 Research questions and thesis outline

The scientific objective of this dissertation is the localisation and quantification of temperature, wind and humidity biases across the tropopause in the initial conditions of the leading global NWP system ECMWF IFS (Haiden et al., 2023). Such biases could deteriorate the representation of the tropopause and its sharpness in the NWP models (Jung et al., 2003), which is expected to affect the quality of NWP forecasts. Various unique observational data sets from field campaigns comprising radiosonde and water vapour lidar data allow model biases from different perspectives: 1) through independent validation, and 2) by the application of data assimilation diagnostics. This combined approach is pursued to address two major unanswered research questions (RQs) on biases at the tropopause.

1. Evaluation of the lower-stratospheric moist bias in the IFS

The lower-stratospheric moist bias remains a prominent systematic error in various NWP models unresolved for decades. As pointed out in Sect. 1.2.2, there are currently open aspects of this bias, in particular concerning its unknown vertical structure especially at high altitudes in the LS, and the possible processes causing this bias. To approach these aspects, accordingly, three research questions are stated in the first part of this thesis:

- **RQ1.1 What is the vertical structure of the moisture bias in the LS?**
- **RQ1.2 Is the bias variable for different seasons and synoptic situations?**
- **RQ1.3 Is the moist bias related to mixing processes?**

To answer these research questions, a unique multi-campaign data set of airborne water vapour lidar observations is used to evaluate UTLS humidity errors in the ECMWF's most recent reanalysis ERA5. This water vapour lidar data set provides a high number of accurate and highly resolved profile observations, measured in data sparse regions of the mid-latitudes. This data set is furthermore not assimilated by the ECMWF allowing an independent validation of potential humidity biases and their potential causes. There was -up to this dissertation- no observation capability available providing highly resolved profiles humidity observations to reliably address the vertical structure of the bias (RQ1.1). The water vapour lidar observations cover three seasons and a broad range of synoptic situations representative of the mid-latitude weather, and thus allow to evaluate the seasonal and synoptical variability of the bias (RQ1.2). In addition, collocated water vapour and ozone lidar observations, available for one campaign, are utilized to explore a specific hypothesized cause of the moist bias (mixing processes). For this purpose, the UTLS air masses classified by the lidar observations are correlated with the moist bias (RQ1.3).

2. Influence of assimilated radiosondes on the mid-latitude tropopause

The potential influence of assimilated observations on the tropopause is currently under debate (Sect. 1.2.3). The two available studies (Birner et al., 2006, Pilch Kedzierski et al., 2016) present different conclusions as to whether assimilated observations sharpen or smoothen the vertical temperature and wind gradients near the tropopause. To address this existing gap of knowledge, three RQ's are raised in the second part of the thesis:

- **RQ2.1 How are the vertical temperature and wind gradients at the tropopause represented in the IFS background and what is the influence of data assimilation on the analysis?**
- **RQ2.2 Can the diagnosed influence be attributed to the assimilation of radiosondes?**
- **RQ2.3 What is the influence of data assimilation on the tropopause beyond the observation location?**

In this analysis, the influence of assimilated radiosondes on the tropopause is investigated in the ECMWF IFS which operates the most advanced data assimilation system among all NWP models. The evaluation is based on a one-month radiosonde data set comprising operational and non-operational radiosondes observed during the NAWDEX field campaign. These observations provide accurate temperature and wind profiles resolving the sharp vertical temperature and wind gradients at the tropopause, and are assimilated by the IFS. Although radiosondes represent only a small part of the global observing system, they have a particularly high influence per assimilated observation on the forecast quality (Pauley and Ingleby, 2022). Due to their high vertical resolution and precision, they could potentially improve temperature and wind gradients at the tropopause during the assimilation. Data assimilation diagnostics (O-B, O-A and increments statistics) are applied to study the representation of the tropopause sharpness in the background and the influence on the analysis (RQ2.1). Furthermore, an OSE was set up in the IFS that consists of two separate IFS run, one with and one without NAWDEX radiosonde observations. The examination of this experiment will uncover to what extent the diagnosed influence on tropopause sharpness can be attributed to radiosonde data assimilation (RQ2.2). RQ2.1 and RQ2.2 adress the data assimilation influence at the position of the radiosondes. To characterise the influence of data assimilation on the tropopause beyond the location of the radiosondes, a case study in model-space is presented in RQ2.3.

Structure of this thesis

Sect. 2 provides the theoretical background to support a better understanding of the findings of this work. This includes an explanation of the atmospheric lidar technique in Sect. 2.1 and a brief introduction of data assimilation theory and the data assimilation system of the ECMWF IFS Sect. 2.2. Sect. 3 contains results on the first research topic of this dissertation addressing the evaluation of the lower-stratospheric moist bias in the IFS. Results on the second research topic, the influence of radiosonde data assimilation on the tropopause are presented in Sect. 4. The main conclusions of this work are summarized in Sect. 5 which also provides an outlook of further research in this field.

Chapter 2

Theoretical background

This chapter aims to provide an overview of the theoretical principles (of lidar measurements and data assimilation) that are applied to answer the RQs. Sect. 2.1 provides the theoretical basis of the atmospheric lidar technique and presents the differential absorption instrument providing the water vapour lidar data set. Sect. 2.2 introduces the fundamental concept of variational data assimilation and its implementation in the ECMWF IFS, of which the data assimilation output is evaluated in this study.

2.1 Atmospheric lidar technique

Already in 1916, Albert Einstein described the principle of stimulated emission of radiation from molecules as a result of absorption of light (Einstein, 1916) and thus provided the theoretical framework for the modern laser (Light Amplification by Stimulated Emission of Radiation) technology. It took about 45 years to realize the first operational laser device (Maiman et al., 1960). This technological milestone was the starting point for the development of so-called lidar (LIght Detection And Ranging) systems in the 1960s. Lidars systems are active remote sensing instruments that have become a key technology in atmospheric sciences. The lidar concept is based on interaction of light (usually in the spectrum from ultraviolet (UV) to infrared (IR)) with atmospheric constituents e.g., atoms, molecules as well as cloud and aerosol particles. A lidar emits monochromatic light along a narrow beam into the atmosphere, and the share of light scattered back to the instrument, is used to derive physical parameters of the atmosphere, such as temperature, concentrations of water vapour and other trace gases, wind velocity as well as properties of clouds and aerosols. Nowadays, lidar instruments are operated from ground-based stations, aircraft, ships, and from space providing highly valuable information of the Earth's atmosphere. For example, data products from lidars can be applied to study specific processes in the atmosphere and also their representation in NWP models. Lidar data can also be directly assimilated by NWP models and thus contribute to a precise prediction, as the successful Aeolus mission recently demonstrated (e.g., Martin et al., 2023). In this chapter, the fundamental principles of an atmospheric lidar system are introduced Sect. 2.1.1. This study uses water vapour data from a differential absorption lidar (DIAL). The theoretical basics of the DIAL technique and its instrumental realisation are explained in Sect. 2.1.2 and Sect. 2.1.3, respectively.

2.1.1 Basic principle and the lidar equation

A basic atmospheric lidar system consists of three main components: a transmitter, a receiver (or telescope), and a data acquisition system. An example space-borne lidar is illustrated in Figure 2.1. The transmitter contains the laser system, which generates and emits short, monochromatic light pulses along a narrow beam into the atmosphere. Along its path in the atmosphere the light pulse is scattered by the molecules and particles in all directions. A receiving telescope detects the proportion of the light that is backscattered to the lidar system.

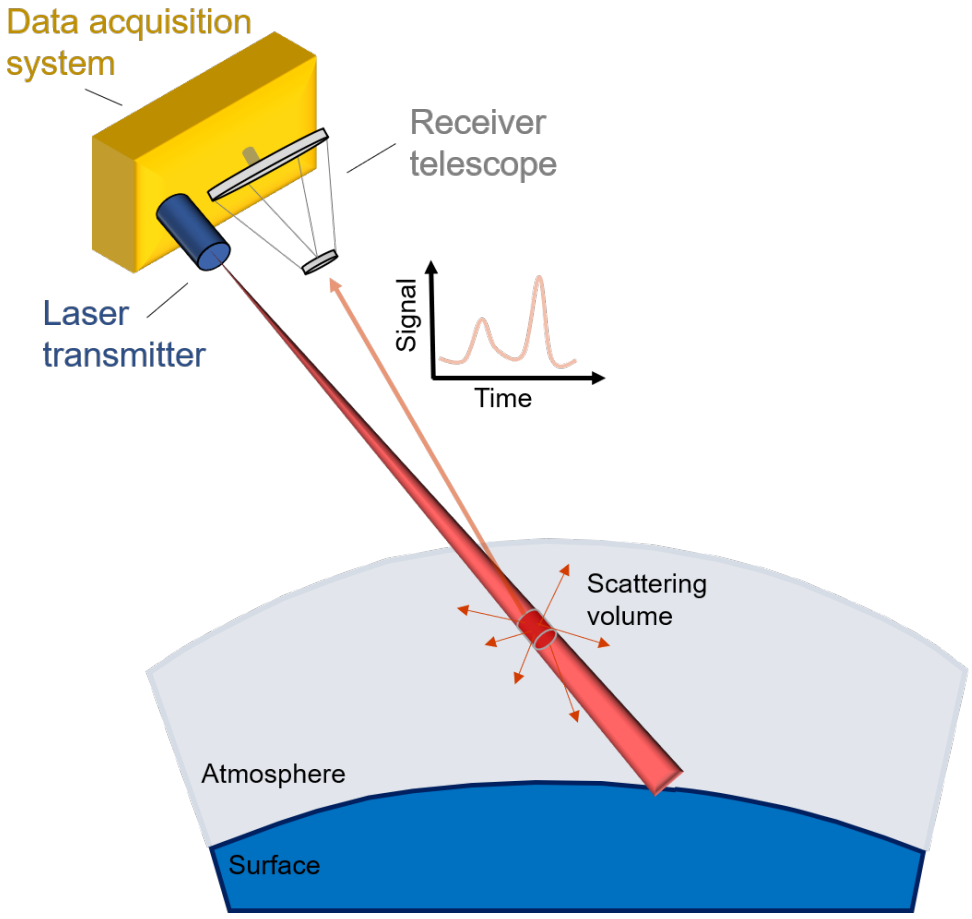

Figure 2.1: An example spaceborne lidar system.

From the time difference between the pulse emission at time t_0 and the subsequent detection of the backscattered light at time t, the distance R at which the scattering process occurred can be derived through the relation $R = \frac{c(t-t_0)}{2}$ with c being the vacuum light velocity. The data acquisition system counts the number of photons in certain time intervals τ from which the corresponding range bin length can be derived via $\Delta R = \frac{c\tau}{2}$. It has to be noted, that the time interval is usually larger than the pulse length τ_{pulse}.

The power of the backscattered light received by the telescope in a one-wavelength setup is described by the lidar equation (Equation 2.1). The following description and notation of the lidar equation and its components is based on Gimmestad and Roberts (2023):

$$P(R) = \underbrace{P_0\, k_{T,R}\, G(R)\, \frac{A}{R^2}\, \frac{c\,\tau_{pulse}}{2}}_{\text{(I)}}\ \underbrace{\beta(R)\, \exp\left[-2\int_0^R \alpha(r)\, dr\right]}_{\text{(II)}} \tag{2.1}$$

The parameters of the lidar equation that influence the received Power P(R) at the telescope are grouped into (I) instrument-related parameters and (II) parameters related to optical atmospheric processes. Instrument parameters include the average power of the emitted laser pulse (P_0), which is the pulse energy (E_{pulse}) divided by the pulse width (τ_{pulse}), the optical efficiencies of the transmitter and receiver ($k_{T,R}$), the geometrical function $G(R)$, the receiver telescope area A and the quadratic attenuation of signal intensity with distance R^2. Each of these instrument parameters needs to be determined for a particular lidar system. Term (II) does not include any instrument constants, but parameters that describe the loss of signal power due to scattering and extinction processes in the measured air volume. First, the proportion of light that is backscattered to the lidar system are described by the total backscatter coefficient $\beta(R)$. Molecular scattering (i.e., scattering by atoms and molecules) and particle scattering contribute to the total backscatter. Note that the Lidar equation makes the assumption that the laser pulse has received only one scattering event. Second, the total power loss of the beam related to absorption and scattering by molecules and particles is expressed by the term $\exp\left[-2\int_0^R \alpha(r)dr\right]$ (also known as Beer-Lambert's Law) which expresses the two-way transmittance from the lidar to the scattering air volume. The integral in the Beer-Lambert's law $\int_0^R \alpha(r)dr$ represents the optical thickness (or optical depth) with the extinction coefficient $\alpha(r)$. The transmission of light trough the atmosphere depends on molecular and particle scattering and absorption (see Equation 2.2):

$$\alpha(r) = \alpha^{\text{mol.}}_{\text{scat.}}(r) + \alpha^{\text{mol.}}_{\text{abs.}}(r) + \alpha^{\text{part.}}_{\text{scat.}}(r) + \alpha^{\text{part.}}_{\text{abs.}}(r). \tag{2.2}$$

In the electromagnetic spectrum relevant for lidar systems (UV, visible, IR), extinction by aerosol and cloud particles as well as by molecular scattering is only relatively weakly dependent on the wavelength. In contrast, molecular absorption is highly wavelength-dependent in the visible and IR spectrum, and this absorption is characterised by fine spectral absorption lines (e.g., Gimmestad and Roberts, 2023). These absorption lines are individual for each atmospheric gas and can be precisely determined in a laboratory. An example for the absorption cross section of water vapour in the near IR spectrum is given in Figure 2.2. It shows that water vapour absorption cross section is highly variable by several orders of magnitude within a small spectral band between 935 and 936 nm. The figure also illustrates that cross-sections vary not only with wavelength, but also with concentration, and temperature.

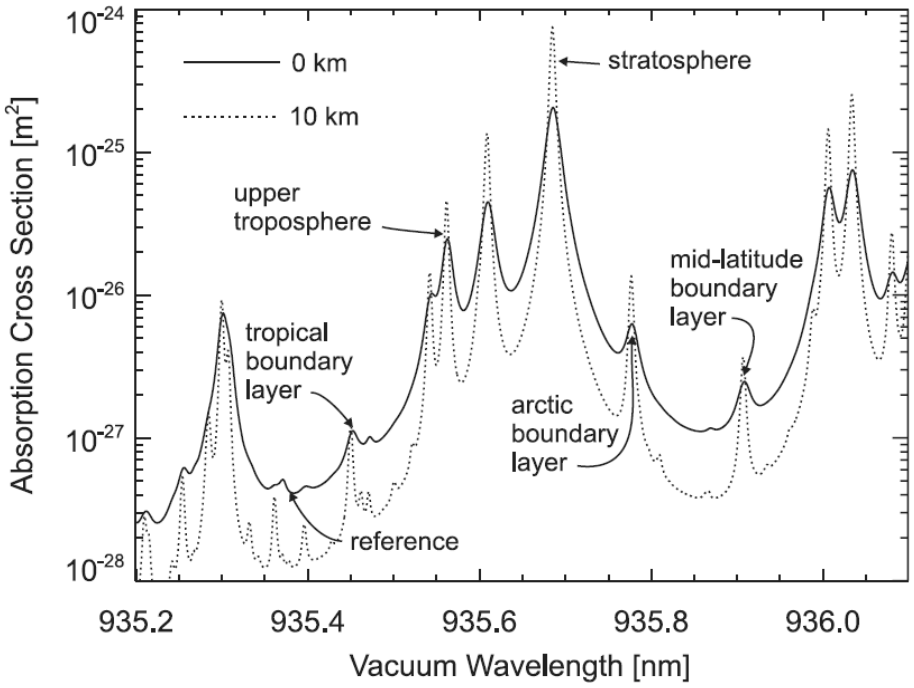

Figure 2.2: Water vapour absorption cross section from the HIRTRAN (HIgh-Resolution TRANsmission molecular absorption) data base (Rothman et al., 2009) given for sea level conditions (solid) and at 10 km altitude (dashed). Adopted from (Wirth et al., 2009).

2.1.2 Differential absorption lidar

The DIAL (Differential Absorption lidar) technique is based on the concept of the differential absorption by atmospheric trace species for differing wavelengths. A DIAL principally operates at least two wavelengths (one online and one offline channel). On- and offline wavelengths are chosen spectrally so narrow that their scattering and extinction properties are almost identical with regard to aerosols, cloud particles and oxygen/nitrogen. The selected on- and offline wavelengths should only be sensitive to the absorption of the trace gas to be measured. The online wavelength is strongly absorbed by water vapour (resulting in a weaker signal) and the offline wavelength is barely absorbed (stronger signal) (see graphs illustrated in Figure 2.3). From the measured differential absorption, a DIAL allows the concentration of a trace gas of interest (e.g., H_2O, O_3, CO, CH_4, N_2O) to be determined.

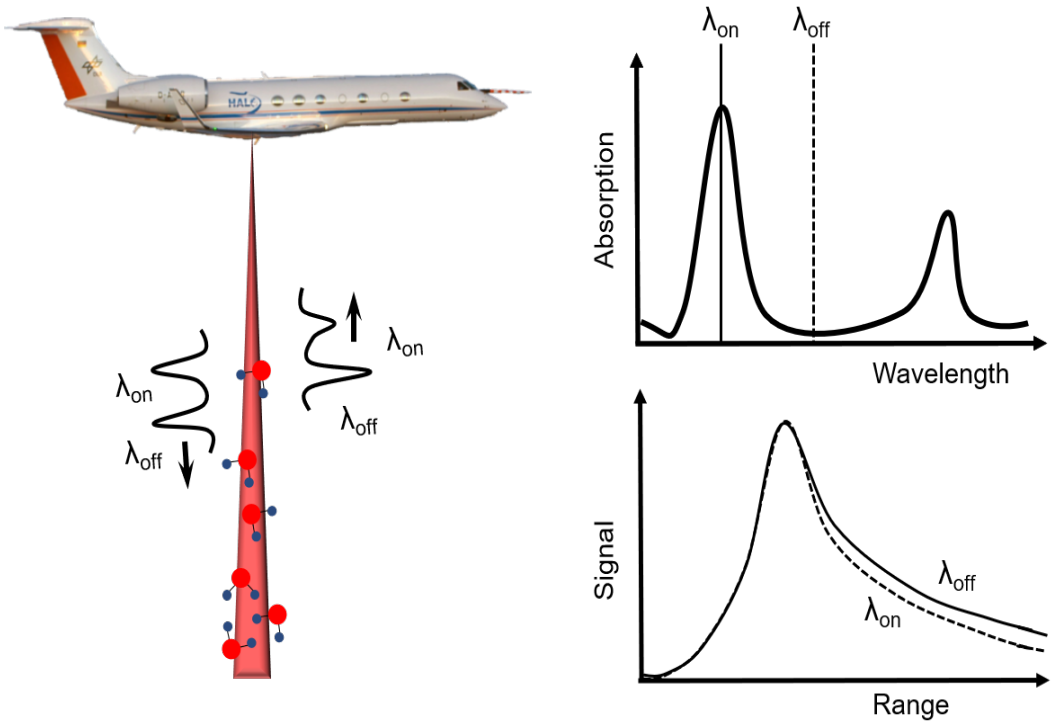

Figure 2.3: Example airborne water vapour DIAL emitting on- and offline wavelength. The differential absorption for on- and offline wavelength and its response to the received signal is schematically illustrated on the right.

A separate lidar equation is formulated for the online wavelength(λ_{on}) and the offline wavelength (λ_{off}) (see Equation 2.4 and Equation 2.5). For the sake of simplicity, the instrument parameters (term (I); Equation 2.1) are replaced by a wavelength-dependent constant C_λ. In addition, for the two spectral closely selected on- and offline wavelengths, the extinction term (Equation 2.2) simplifies to

$$\alpha(r) \approx \alpha_{\text{abs.}}^{\text{mol.}}(r). \tag{2.3}$$

$$P_{\lambda_{on}}(R) = C_{\lambda_{on}}\,\beta_{\lambda_{on}}(R)\,\exp\left[-2\int_0^R \alpha_{\lambda_{on}}(r)\,dr\right]. \tag{2.4}$$

$$P_{\lambda_{off}}(R) = C_{\lambda_{off}}\,\beta_{\lambda_{off}}(R)\,\exp\left[-2\int_0^R \alpha_{\lambda_{off}}(r)\,dr\right]. \tag{2.5}$$

$P_{\lambda_{off}}(R)$ and $P_{\lambda_{on}}(R)$ are the measured quantities, the ratio of which is written as follows

$$\frac{P_{\lambda,\mathrm{on}}(R)}{P_{\lambda,\mathrm{off}}(R)} = \frac{C_{\lambda_{on}}}{C_{\lambda_{off}}}\frac{\beta_{\lambda_{on}}}{\beta_{\lambda_{off}}}\exp\left[-2\int_0^R (\alpha_{\lambda_{on}}(r) - \alpha_{\lambda_{off}}(r))\,dr\right]. \tag{2.6}$$

The instrument constants, and the backscatter coefficients assuming $\beta_{\lambda_{on}}(R) \widehat{=} \beta_{\lambda_{off}}(R)$, cancel out of Equation 2.6. Vanishing the integral, followed by applying the range derivative yields:

$$\alpha_{\lambda_{on}}(R) - \alpha_{\lambda_{off}}(R) = \frac{1}{2}\left[\frac{d}{dR}\ln\left(\frac{P_{\lambda,\mathrm{off}}(R)}{P_{\lambda,\mathrm{on}}(R)}\right)\right]. \tag{2.7}$$

The extinction coefficient is determined by the product of the number density or number concentration N(R) of the absorbing molecules and the wavelength-dependent cross section (σ) of the absorbing species which can be precisely determined in laboratory measurements or taken from the HIRTRAN database (Rothman et al., 2005). Hence, the left hand side of Equation 2.7 can be rewritten as:

$$\alpha_{\lambda_{on}}(R) - \alpha_{\lambda_{off}}(R) = \Delta\alpha = N(R)\,\Delta\sigma. \tag{2.8}$$

Inserting Equation 2.8 in Equation 2.7 yields the DIAL equation:

$$\begin{aligned} N(r) &= \frac{1}{2\Delta\sigma}\frac{d}{dR}\ln\left(\frac{P_{\mathrm{off}}(R)}{P_{\mathrm{on}}(R)}\right) \\ &= \frac{1}{2\Delta\sigma\Delta R}\ln\left(\frac{P_{\mathrm{off}}(R+\Delta R)}{P_{\mathrm{off}}(R)}\frac{P_{\mathrm{on}}(R)}{P_{\mathrm{on}}(R+\Delta R)}\right) \end{aligned} \tag{2.9}$$

From the DIAL equation, the number concentration can be directly determined by the received power, measured for on- and online wavelengths at a given range, and with respect to the difference of the absorption cross section of the chosen wavelengths. The DIAL equation does not contain any system constants which makes the DIAL a so-called self-calibrating technique. In this analysis however, the measured water vapour number concentration is not the desired quantity. Instead, it is converted within a few steps into mixing ratio and specific humidity using the ideal gas law with a given temperature and pressure profile. The fundamental equations utilized for the following steps can be found in (e.g., Malberg, 2007).

Step 1: Conversion of the water vapour number concentration to the mass concentration of water vapour (m_{H_2O}) in $\mathrm{kg\,m^{-3}}$:

$$m_{H_2O} = \frac{N \times M_{\mathrm{H_2O}}}{N_A} \tag{2.10}$$

where N_A is Avogadro's number, approximately $6.022 \times 10^{23}\,\mathrm{mol}^{-1}$ particles per mole and M_{H_2O} is the molar mass of water vapour ($18.015\,\mathrm{g\,mol}^{-1}$).

Step 2: From the ideal gas law the mass of dry air (m_d) in kg m^{-3} can be derived:

$$m_d = \frac{p}{R_d T}. \tag{2.11}$$

In this equation T is the air temperature, p the pressure and R_d the specific gas constant of dry air (about 287 $\frac{J}{kgK}$).

Step 3: The humidity mixing ratio (ω) is obtained via the ratio of the mass of water vapour and mass of dry air :

$$\omega = \frac{m_{H_2O}}{m_d}. \tag{2.12}$$

Step 4: Specific humidity (q) is related to the mixing ratio via:

$$q = \frac{\omega}{1+\omega}. \tag{2.13}$$

2.1.3 The water vapour DIAL WALES

The WAter vapour Lidar Experiment in Space (WALES) is a lidar instrument that was conceptualized to prepare a space-based lidar mission with the goal to provide accurate and vertically highly-resolved water vapour profiles with global coverage which are highly demanded by NWP models. In the 2000s, WALES was developed as an airborne demonstrator at the German Aerospace Center (DLR) (Wirth et al., 2009) and has been successfully operated on board the German research aircraft HALO (High Altitude LOng Range; (Krautstrunk and Giez, 2012)) and on the DLR Falcon aircraft during in total 18 campaigns starting in 2007.

WALES unites three different lidar techniques, a DIAL, a depolarization lidar and a high spectral resolution (HSRL) lidar. The DIAL module operates four wavelengths (three online, and one offline) targeting the strong water vapour absorption band in the near IR spectrum around 935-936 nm (Figure 2.2). The online channels are sensitive to different trace gas concentrations and in turn to different altitude levels. The exact wavelengths of the online-channels are selected such, that they are sensitive to the typical water vapour concentrations of the boundary layer, the upper troposphere and the lower stratosphere, allowing precise water vapour profiles to be measured from the surface to flight altitude (up to approx. 15 km). WALES furthermore operates two aerosol-sensitive channels to measure aerosol extinction with the HSRL technique at 532 nm (e.g., Esselborn et al., 2008) and depolarization (at 1064 nm and also 532 nm) enabling to study aerosol characterization (Groß et al., 2013, 2015). In an experimental configuration, implemented at different research campaigns, two of the NIR channels were replaced by ozone-sensitive channels in the UV spectrum (305-315 nm), so that ozone and water vapour can be measured concurrently (Fix et al., 2019). In the following, the operational setup of WALES (only H_2O channels) is briefly described with assistance of the schematic illustration of the

Table 2.1: Transmitter and system parameters of the WALES instrument.

Transmitter Parameters	**Values**	**Unit**
Pulse repetition per laser	100	Hz
Pump laser pulse energy at 1064 nm	400	mJ
Pump laser pulse energy at 532 nm	220	mJ
Pulse length at 935 nm	7.6	ns
WALES system output at 1064 nm	120	mJ
WALES system output at 532 nm	75	mJ
WALES system output at 935 nm	45	mJ
Spectral purity at 935 nm	>99.9	%
Beam divergence	1	mrad
System Parameters	**Values**	
Weight	450	kg
Total power consumption	2000	W
Power consumption of the laser system	1500	W
Telescope diameter	48	cm
Beam diameter on ground	~10	m
Effective vertical resolution of the humidity profiles	300	m

transmitter Figure 2.4, and the system design Figure 2.5. Table 2.1 provides an overview of the key system and transmitter parameters.

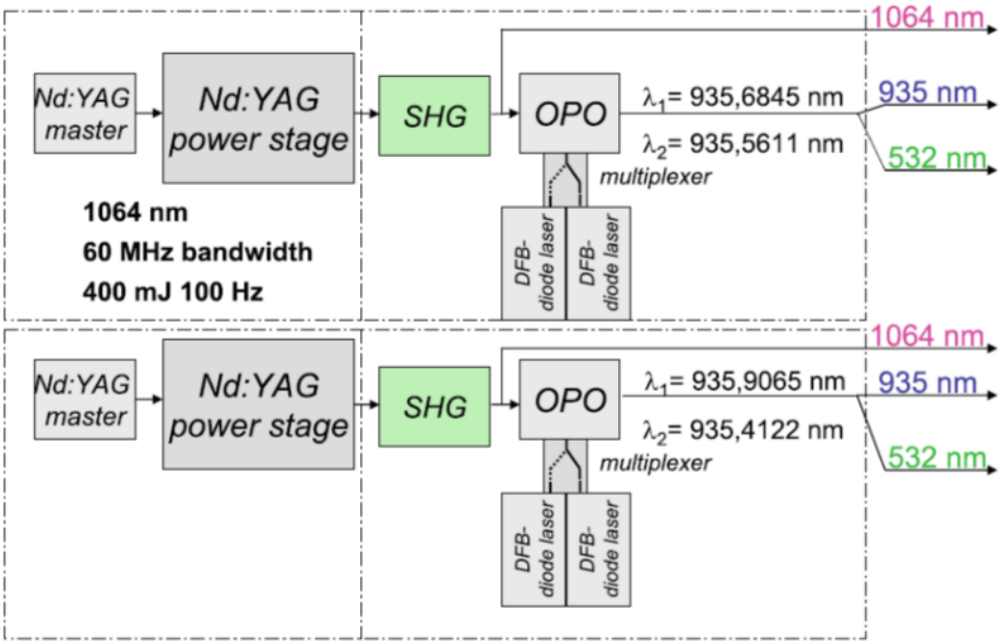

Figure 2.4: WALES transmitter setup, adopted from (Wirth et al., 2009).

WALES transmitter setup

The WALES transmitter operates two identical laser systems (each generating two NIR pulses) providing temporarily shifted output with high repetition rates of 200 Hz. Each of the laser systems uses a Neodymium-doped Yttrium Aluminum Garnet (Nd:YAG) laser in a master oscillator power amplifier configuration to produce the sufficient beam quality and power. The master laser is a Q-switched monolithic ring laser that generates pulses at 1064 nm which is then amplified in the Nd:YAG power stage to 400 mJ. Part of this beam is then frequency-doubled (532 nm; 200 mJ pulse energy) using a second harmonic generator (SHG). An optical parametric oscillator (OPO), seeded by four distributed feedback laser diodes, generates the required DIAL (935 nm) channels. The output power of each laser pulse at 935 nm is about 45 mJ (corresponds to about 2.12×10^{17} photons) while the received power at the detector is about 200 fJ (roughly equivalent to 10^6 photons). This means than only one photon per about 212,000,000,000 photons emitted by WALES is scattered back to the instrument providing information on water vapour in the atmosphere.

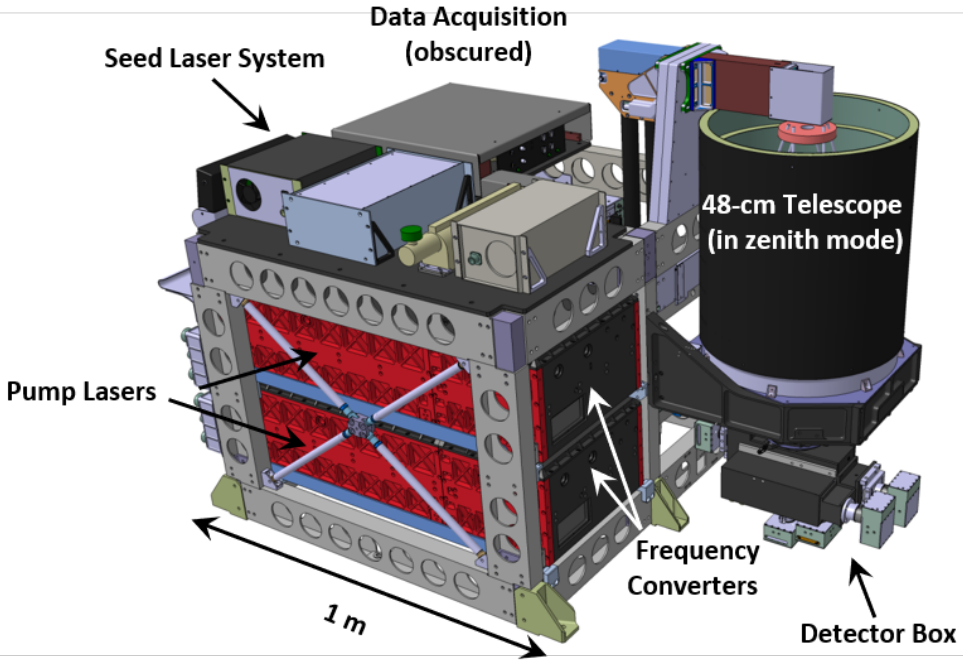

Figure 2.5: Mechanical design of the WALES instrument on-board HALO adopted from Wirth et al. (2009). Note that the telescope in the sketch is aligned upwards, although it was oriented downwards during all campaigns considered in this work (see Sect. 3.1.1).

Characteristics and errors of the retrieved water vapour profiles

In all campaigns considered in this work WALES was operated in a near-nadir mode (2-3 % from zenith), despite an upward-looking set up was also successfully tested in one campaign in 2015 (see Figure 2.5). In optically thick clouds the extinction by cloud particles is too strong so no water vapour information can be retrieved within or below the cloud. Hence, the availability of WALES humidity profiles is restricted to cloud-free or to

optically thin clouds such as Cirrus ice clouds in the UT (Groß et al., 2014, Dekoutsidis et al., 2023, Jurkat-Witschas et al., 2024). The derived water vapour profiles have an effective vertical resolution of about 300 m, which corresponds to the full width of half maximum of the weighting functions used in the retrieval. The horizontal resolution is determined by the aircraft speed and the selected 12 s time window in which the measurements are averaged. For an estimated aircraft speed of 220 m s^{-1} the horizontal resolution would be approximately 2.6 km. Different error sources contribute to the accuracy (systematic error) and the precision (statistical error) of the retrieved water vapour profiles and these are assessed in detail by Kiemle et al. (2008). Statistical (or random) errors are uncorrelated for each measurement, and depend on the individual shape of the water vapour profile. The statistical error can be significantly reduced by vertically or horizontally averaging a large number of data. It should be noted that the retrieval provides an estimate of the statistical error, which makes it possible, for example, to filter out profiles with high noise (as applied in Sect. 3.1.1). The estimated average statistical error due to photon noise at the detector is expected to be within 5 % (Kiemle et al., 2008). Systematic errors (instrument bias) can not be reduced by averaging and lead to a shift between measurement and true parameter. The moist important systematic error sources for WALES water vapour profiles are uncertainties in the spectral data bases, spectral impurities of the laser source, and Rayleigh–Doppler broadening (e.g., Bösenberg, 1998). In addition, due to the non-linearity of the DIAL equations, noisy data can also affect the instrument accuracy. Kiemle et al. (2008) estimated the combined measurement uncertainty of the WALES humidity product in the UTLS to be within ± 10 %. The high reliability of WALES was shown in various intercomparisons, for example, with Lyman-α in-situ hygrometers (Kiemle et al., 2008), similar airborne and ground-based DIAL instruments (Bhawar et al., 2011), and radiosondes with a frost point hygrometer (Trickl et al., 2016).

2.2 Data assimilation theory

The results presented in Sect. 4 are based on a statistical analysis of data assimilation output from the ECMWF IFS. The objective of Sect. 2.2 is to provide the necessary context for a better understanding of these results. First, the theoretical principles of data assimilation are introduced below. This is followed by a brief outline of the workflow of the IFS' operational data assimilation system.

As stated in Sect. 1.1 the quality of a weather forecast is highly sensitive to the accuracy of initial conditions, known as the "initial value problem" of NWP. Data assimilation is a mathematical technique that tackles this initial value problem and aims to determine the best estimate of the atmosphere's state at the model initial time (i.e, termed analysis). From this analysis, the numerical weather forecast is calculated. To derive the analysis, a data assimilation system attempts to optimally combine information of a previous short-range forecast (called background or "first-guess") with latest available meteorological observations.

2.2.1 From Bayes' theorem to variational data assimilation

Most popular approaches to solve this data assimilation problem rely on the Bayesian probabilistic perspective (Bannister, 2017). Bayes' theorem states, that the posterior probability of the occurrence of an event X (under the condition that an event Y has previously occurred) can be calculated from the information of the prior probability of occurrence of X and the likelihood of occurrence of Y (under the condition that X has occurred) (Lorenc, 1986). By defining event X = $\mathbf{x}$ as the atmospheric state and event Y = $\mathbf{y}$ as the observed values this concept is applicable to NWP. This section provides an overview of how Bayes' theoretical framework is applied to the data assimilation problem.

In general there are three strategies: (I) Variational data assimilation schemes, (II) ensemble data assimilation and (III) non-Gaussian Monte-Carlo methods. In the subsequent, we focus on the Variational (VAR) approach which is the core assimilation scheme implemented in the ECMWF IFS. Variational assimilation algorithms in its principle derive the best analysis estimate by an iterative minimization of a cost function that contains information about the prior state (background), a set of observations, and their corresponding assumed uncertainties. The following description is primarily based on the work by Lorenc (1986), Ingleby and Lorenc (1993), Bannister (2008, 2017) and Evensen et al. (2022). An overview of the variable and operator notation used in this section is provided in Table 2.2.

The Bayes' equation (Equation 2.14) in its general form is used as a starting point. Here, $p(\mathbf{x})$ serves as the prior probability distribution function (PDF) of the NWP system's state $\mathbf{x} \in \mathbb{R}^n$ comprising all information before assimilation of the latest observations $\mathbf{y} \in \mathbb{R}^m$ and $p(\mathbf{y}|\mathbf{x})$ represents the likelihood of the observations given state $\mathbf{x}$ (e.g., Lorenc, 1986).

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})\, p(\mathbf{x}) \tag{2.14}$$

Variable	Description
$\mathbf{B}$	Background error covariance matrix
$\lvert\mathbf{B}\rvert$	Determinant of $\mathbf{B}$
H	Nonlinear observation operator matrix
$\mathbf{H}$	Linear observation operator matrix
$\mathbf{M}$	Nonlinear model operator
$\mathbf{R}$	Observation error covariance matrix
$\lvert\mathbf{R}\rvert$	Determinant of $\mathbf{R}$
$J(\mathbf{x})$	Cost function
$p(\mathbf{x}\lvert\mathbf{y})$	Posterior distribution
$p(\mathbf{x})$	Prior distribution
$p(\mathbf{y}\lvert\mathbf{x})$	Observations likelihood
$\mathbf{x}_t$	True atmospheric state
$\mathbf{x}_a$	Analysis state
$\mathbf{x}_b$	Background state
$\mathbf{y}_o$	Observation vector

Table 2.2: Overview of variable and operator definitions.

The Bayesian equation (Equation 2.14) can be analytically solved by making assumptions on the contained PDFs. Commonly Gaussian shaped distributions for both, the observations likelihood $p(\mathbf{y}|\mathbf{x})$ and the prior PDF $p(\mathbf{x})$ are assumed (e.g., Ingleby and Lorenc, 1993, Evensen et al., 2022).

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{(2\pi)^{\frac{m}{2}}|\mathbf{R}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{y}-H(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y}-H(\mathbf{x}))\right\} \tag{2.15}$$

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}|\mathbf{B}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b)\right\} \tag{2.16}$$

The goal is to find the most probable solution of the analysis state $\mathbf{x}_a$ that corresponds to the maximum likelihood of the posterior distribution

$$\mathbf{x}_a = \arg\max_x p(\mathbf{x}|\mathbf{y}). \tag{2.17}$$

To tackle this problem, a cost function $J(\mathbf{x})$ is defined as

$$\begin{aligned} J(\mathbf{x}) &= -\log\left(p(\mathbf{x}|\mathbf{y})\right) + \mathrm{C} \\ &= -\log\left(p(\mathbf{y}|\mathbf{x})\right) - \log\left(p(\mathbf{x})\right) + \mathrm{C}. \end{aligned} \tag{2.18}$$

Inserting the observation likelihood (Equation 2.15) and the prior state (Equation 2.16) in Equation 2.18 with an appropriate choice of the constant C yields the three dimensional cost function of the data assimilation problem:

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + \frac{1}{2}(\mathbf{y}-H(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y}-H(\mathbf{x})) \tag{2.19}$$

.

The different components/matrices of the costfunction (Equation 2.19) will be explained in Sect. 2.2.2. In order to explicitly minimise the cost function of the 3D-Var the nonlinear observation operator H is assumed to be approximately linear which is done by a first order Taylor expansion around $\mathbf{x}_b$:

$$H(\mathbf{x}) \approx H(\mathbf{x}_b) + \mathbf{H}(x - x_b) \tag{2.20}$$

with $\mathbf{H}$ being the Jacobian matrix of $H(\mathbf{x})$. The term $[\mathbf{y} - H(\mathbf{x})]$ in the 3D-VAR cost function can then be rewritten as:

$$\mathbf{y} - H(\mathbf{x}) = \mathbf{y} - H[\mathbf{x}_b + (\mathbf{x} - \mathbf{x}_b)] = [\mathbf{y} - H(\mathbf{x}_b)] - \mathbf{H}(\mathbf{x} - \mathbf{x}_b) \tag{2.21}$$

The model state with the maximum probability corresponds to the one that minimizes this cost function which is expressed by the gradient of $J(\mathbf{x})$. Differentiation of Equation 2.19 yields:

$$\nabla J(\mathbf{x}) = \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + \mathbf{H}^T\mathbf{R}^{-1}[\mathbf{y} - H(\mathbf{x})] \tag{2.22}$$

.

For $\mathbf{x} = \mathbf{x_a}$ the solution of the 3D-Var analysis state can be obtained by equating its gradient to zero ($\nabla J(\mathbf{x_a}) \stackrel{!}{=} 0$):

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{B}\mathbf{H}^T \left(\mathbf{H}^T\mathbf{B}^{-1}\mathbf{H} + \mathbf{R}\right)^{-1} (\mathbf{y} - H(\mathbf{x}_b)). \tag{2.23}$$

As can be seen from Equation 2.23, the desired solution, the analysis $\mathbf{x}_a$, consists of two components. Firstly, the background $\mathbf{x}_a$ and secondly, the distance between the observation and the background at the location of the observation $\mathbf{y} - H(\mathbf{x}_b)$ (also known as "innovation vector") weighted its optimal weight $\mathbf{B}\mathbf{H}^T \left(\mathbf{H}^T\mathbf{B}^{-1}\mathbf{H} + \mathbf{R}\right)^{-1}$, known as the Kalman gain. However, since the matrices are too large (for example $\mathbf{B}$ has 10^9x10^9 elements) to perform the multiplication explicitly, the cost function can practically only be minimised using iterative methods. The choice of iterative methods depends on the properties of the cost function itself. Different iterative methods are detailed out in Evensen et al. (2022).

Hitherto, the derivation of the optimal analysis state has been limited to the 3-dimensional scenario (3D-VAR), in which the observations assimilated over a time window are compared with the model states at a specific point in time (usually the analysis times 00, 06, 12, and 18 UTC). As weather systems move and evolve with time, the temporal mismatch between observation time and analysis time may induce errors. To reduce such errors, four-dimensional variation assimilation (4D-VAR) matches the model background with observations at the same time allowing optimization of a model state over a time window rather than at a fixed point in time (as indicated in Figure 2.6). In November 1997, the 4D-VAR data assimilation scheme was implemented which is an expansion of 3D-VAR in the time dimension (see Equation 2.24). The 4D-VAR algorithm seeks the optimal alignment between the model trajectory and observations to find the initial state $\mathbf{x}_0$, considering the spatial and temporal distribution of observations during the minimization process. In an effective data assimilation system, the resulting analysis typically lies

between background and observations as depicted in Figure 2.6. There are several diagnostics to measure the impact of observations in a data assimilation system. A method is to compare observations, background and analysis states at the position and time of each assimilated observation (marked by arrows in Figure 2.6) which is the approach pursued in this dissertation. The relevant metrics used for the evaluation will be introduced in Sect. 4.1.2.

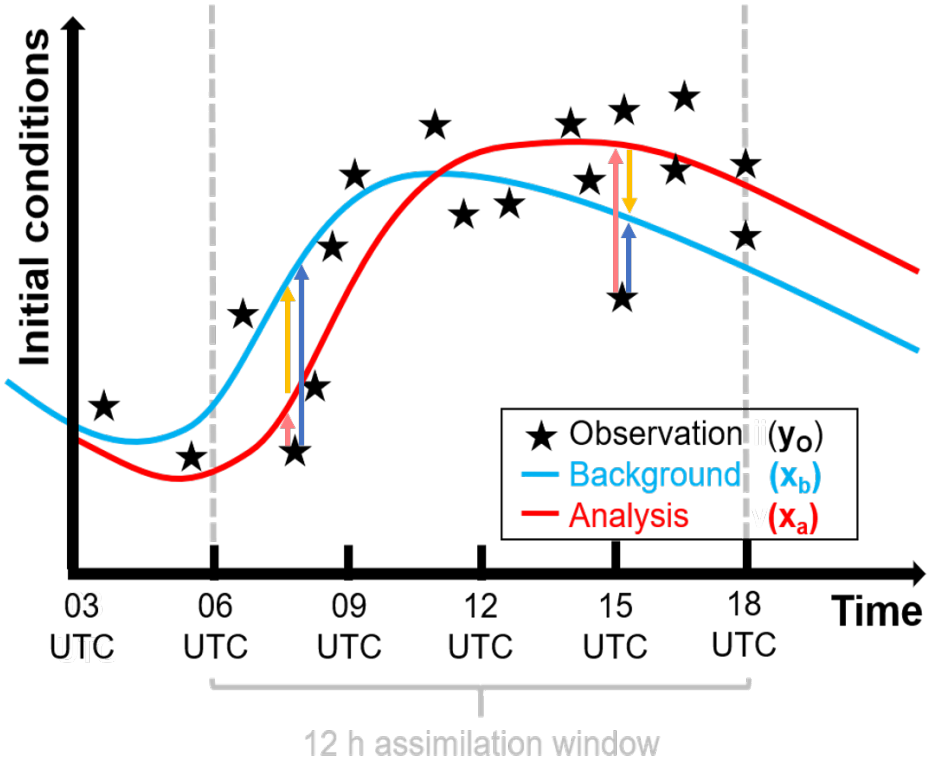

Figure 2.6: Schematic illustration of the principle of the four-dimensional data assimilation system. The analysis is determined by correcting the background (blue line) by fitting with assimilated observations (black stars) over a selected time window (i.e., the assimilation window; grey interval). Arrows mark the key data assimilation metrics, i.e., increments (orange), innovations (dark blue) and residuals (light red), that are examined in Sect. 4. Rebuild from: https://confluence.ecmwf.int/pages/viewpage.action?pageId=288339161, last access: 07 May 2024.

The standard formulation of the 4D-VAR cost reads as

$$J(\mathbf{x_0}) = \frac{1}{2}\underbrace{(\mathbf{x_0}-\mathbf{x}_b)^T\mathbf{B}_0^{-1}(\mathbf{x_0}-\mathbf{x}_b)}_{J_b} + \frac{1}{2}\sum_{i=0}^{n}\underbrace{(\mathbf{y}_i - H_\mathbf{i}(\mathbf{x}_i))^T\mathbf{R}_i^{-1}(\mathbf{y}_i - H_\mathbf{i}(\mathbf{x}_i))}_{J_o} \tag{2.24}$$

where i is the time index, where $H_\mathbf{i}$ and $\mathbf{y}_i$ are time-dependent. The temporal evolution of the model state $\mathbf{x}_i$ from t_i to t_{i+1} is performed by a nonlinear NWP model operator M

$$\mathbf{x}(t_i) = M[\mathbf{x_{i,0}}(t_i)], \tag{2.25}$$

where M being the output from integrating the forecast model from to from t_0 to t_i, with initial conditions $\mathbf{x}(t_0)$. Assuming

$$\mathbf{x}_i = M_i(\mathbf{x_0}), \tag{2.26}$$

$M_i[\mathbf{x_0}]$ depends on $\mathbf{x_0}$ only, so all $\mathbf{x}_i$ can be calculated from the initial state using the model. However, this is subject to the constraint that the model is perfect by assuming the model integration is free of errors (Laloyaux et al., 2020). The perfect model assumption is termed "strong constrained 4D-VAR" (e.g., Evensen et al., 2022).

Inserting Equation 2.26 into the 4D-VAR cost function yields

$$J(\mathbf{x_0}) = \frac{1}{2}(\mathbf{x_0} - \mathbf{x}_b)^T \mathbf{B}_0^{-1}(\mathbf{x_0} - \mathbf{x}_b) + \frac{1}{2}\sum_{i=0}^{n}(H_\mathbf{i} M_\mathbf{i}(\mathbf{x}_0) - \mathbf{y}_i)^T \mathbf{R}_i^{-1}(H_\mathbf{i} M_\mathbf{i}(\mathbf{x}_0) - \mathbf{y}_i). \tag{2.27}$$

To obtain the best analysis state estimate the gradient of Equation 2.27 becomes

$$\nabla J(x_0) = \mathbf{B}^{-1}(\mathbf{x_0} - \mathbf{x}_b) - \sum_{i=1}^{n} \mathbf{M}_{i,0}^T \mathbf{H}_i^T \mathbf{R}_i^{-1}(H_\mathbf{i}(\mathbf{x}_i) - \mathbf{y}_i), \tag{2.28}$$

where $\mathbf{H}_i^T$ is the adjoint of the linearised observation operator and $\mathbf{M}_{i,0}^T$ the adjoint of the linearized forecast model. It should be mentioned that the current ECMWF data assimilation system allows for model errors (weak-constraint 4D-VAR). To do this, another term with a model error covariance matrix Q is added to the cost function. As this contains more control variables, it complicates the solution of the data assimilation problem. More context on the weak-constraint 4D-Var formulation in the ECMWF are provided by (Laloyaux et al., 2020, , and references therein). For the detailed sequence of the 4D-VAR solution algorithm in the IFS, the interested reader is referred to (ECMWF, 2016a).

2.2.2 Error treatment in data assimilation

As illustrated in Equation 2.24 the general form of the cost function is composed of two terms, of which the first (J_b) measures the difference between the state x and the background, and the second (J_o) the misfit between observations and their model equivalents. Observations, and background are always uncertain in accurately representing the true atmospheric state. To account for these uncertainties, in data assimilation, errors for background (ϵ_b), observations (ϵ_o) and analysis (ϵ_a) are defined in the subsequent:

$$\epsilon_b = x_b - x_t, \quad \epsilon_o = y_o - H(x_t), \quad \epsilon_a = x_a - x_t. \tag{2.29}$$

The true state of the atmosphere is an unknown parameter, so assumptions about the statistical characteristics of the errors have to be made. In data assimilation, errors are typically considered to be normally (Gaussian) distributed with an expected value of zero (no bias). The estimated errors are represented by the so-called error covariance matrices (Equation 2.30)

$$\mathbf{B} = E(\epsilon_b\, \epsilon_b^T)_{n\times n}, \quad \mathbf{R} = E(\epsilon_o\, \epsilon_o^T)_{m\times m}, \quad \mathbf{A} = E(\epsilon_a\, \epsilon_a^T)_{n\times n}, \tag{2.30}$$

of whose the diagonal elements are the variances and the off-diagonal elements represent the covariances. The optimal formulation of the error covariances is the overarching challenge in data assimilation, as they are never measured directly and need to be estimated in a statistical sense for instance by means of O-B and O-A statistics (e.g., Bouttier and Courtier, 2002). The background error covariance matrix **B** and the observation error covariance matrix **R** are elementary actors in the 4D-VAR cost function (also in e.g., Ensemble- (Kalman)-based assimilation schemes). Thus, the following is a brief introduction to their tasks and effects in the data assimilation system.

The role of **B** is to spread the information of the background in space, depending on their estimated error covariances. Due to imperfections in the data assimilation system (for instance errors in reality are not Gaussian, but covariances are formed assuming Gaussianity), B tends to have a smoothing effect on sharp gradient would be expected in general. Another important task of **B** is to ensure dynamical balance (i.e. producing increments consistent with the dynamics) via the cross-correlations between the different model variables. For example, a wind observation that causes an increase in wind also induces a pressure increment in order to remain close to geostrophic balance (relates wind and pressure/mass), which is achieved by the cross-correlations in **B** (Bouttier and Courtier, 2002). The covariances also take care that no non-physically large gradients occur between neighboured grid points. Both the spreading and the cross-correlating effect of **B** are illustrated in Bannister (2008) demonstrating wide-spread wind and potential temperature increments ($>1000\,\mathrm{km}$ in the horizontal, several kilometre in the vertical) based on the assimilation of a single pressure observation in the MetUM system.

The observation error covariance matrix **R** determines the weight an observation is given according to its estimated error. Observation error covariances are often assumed to be zero so that **R** is a symmetric, diagonal matrix. This assumption, however, is not necessarily valid e.g., for remote-sensing measurements where errors can be correlated. **R** is comprised of three error components: The first component is the measurement (or instrument) error which is related to systematic or statistical errors of the measurement device. This error is often well-documented and can be "easily" treated in **R**. Second, errors can arise from the observation operator H that conducts the conversion from model-space to observation-space and this process is usually subject to approximations. The third component is the representativeness error, an error caused due to unresolved scales and processes (Janjić et al., 2018) which refers to the differing spatial and temporal resolution of the model variables and the observations. For example, measurement of a parameter at a single point in space and time, such as a surface pressure observation of a ground-based station, does not reflect the average state of the corresponding surrounding grid box. In addition, interpolation errors related to the observation operator and imperfect quality control contribute to the representativeness error. Successful data assimilation requires all three error components to be sufficiently determined in order to optimally configured **R**.

2.2.3 The data assimilation system in the ECMWF IFS

The ECMWF IFS is one of the most advanced NWP model in the world operates the most-developed data assimilation system. To my knowledge, there is no global NWP centre that processes and assimilates more observations (about 100 million per day) than the ECMWF IFS. Its steadily developed and improved data assimilation system is one of the secrets behind the increasing quality of IFS weather forecasts in the recent decades (Bauer et al., 2015). In its currently operational version (Cy48r1), the IFS runs operational medium-range forecasts (10 days) daily at 00 and 12 UTC. These forecasts are produced in high resolution (HRES, (Malardel et al., 2016)) corresponding to a global average resolution of approx. 9 km in the horizontal and 137 levels in the vertical which exceeds the resolution of other well-established global NWP systems such as the German Icosahedral Nonhydrostatic Model (ICON) or the American Global Forecast System (GFS).

The ECMWF IFS employs a hybrid data assimilation system that combines 4D-VAR and ensemble methods (Isaksen et al., 2010, Bonavita et al., 2016) and is called Ensemble Data Assimilation (EDA). Ensemble methods are useful because they allow to quantify atmospheric predictability associated with uncertainties in the initial conditions. An ensemble consists of a set of independent forecast runs called the "ensemble members" that are generated. Depending on the synoptic situation, these differences may lead to diverging forecasts, due to the sensitivity of highly nonlinear systems with respect to the initial conditions (Lorenz, 1963) (see Sect. 1). Typically, the ensemble members diverge with lead time, and the spread of the members is a solid indicator of the atmospheric predictability.

At the ECMWF, the different members are generated by the ensemble forecast suite (ENS). The ENS adds small perturbations of observations, boundary conditions (e.g. sea surface temperature) and model physics to the operational HRES control run to generate 50 independent ensemble members. Subsequently, for each member, the optimal 4D-VAR analysis is solved over a selected time window (called assimilation window). From the divergence of the member trajectories over time within the short-range forecast, information of uncertainties in model and observations over the assimilation window can be derived. This information is used to add a flow-dependent component to the climatological background error covariance matrix $\mathbf{B}$. The better representation of $\mathbf{B}$ leads to an improved analysis and forecasts (Bonavita et al., 2016).

To setup the operational HRES and ENS runs, observations are assimilated in two major interleaving steps termed Long-Window Data Assimilation (LWDA) on the one hand and Short Cutoff Daily Assimilation (SCDA) on the other hand (see Figure 2.7). The following is an simplified example of how the LWDA and the SCDA work together to provide the best analysis estimate at 12 UTC available for the HRES and ENS. The "00 UTC LWDA" (blue boxes in Figure 2.7) is started 14 hours after nominal model time (i.e., 14 UTC) to produce an analysis at 00 UTC using observations of a 12-hourly assimilation window (21-09 UTC). The LWDA also assimilates observations of a 6-hourly assimilation window (03-09 UTC) to create a 06 UTC analysis. This 06 UTC analysis serves as basis to generate short forecasts (T+03 to T+15) which are used as background for the SCDA (red box in Figure 2.7) and for the subsequent LWDA cycle. The SCDA's (red box) task is to determine the 12 UTC analysis. It is started 4 hours after nominal

model time (i.e., 16 UTC) and uses (I) more recent observations from a 6-hour time window (09-15 UTC), and (II) the short range forecasts (T+03 and T+09) to produce the 12 UTC analysis. From this 12 UTC analysis, the HRES medium-range forecast and the ENS are onset. This procedure is repeated twice a day. The interleaving assimilation time windows of LWDA and SCDA allows propagation of information from cycle to cycle and as much observations as possible to be assimilated to best initialize the analysis.

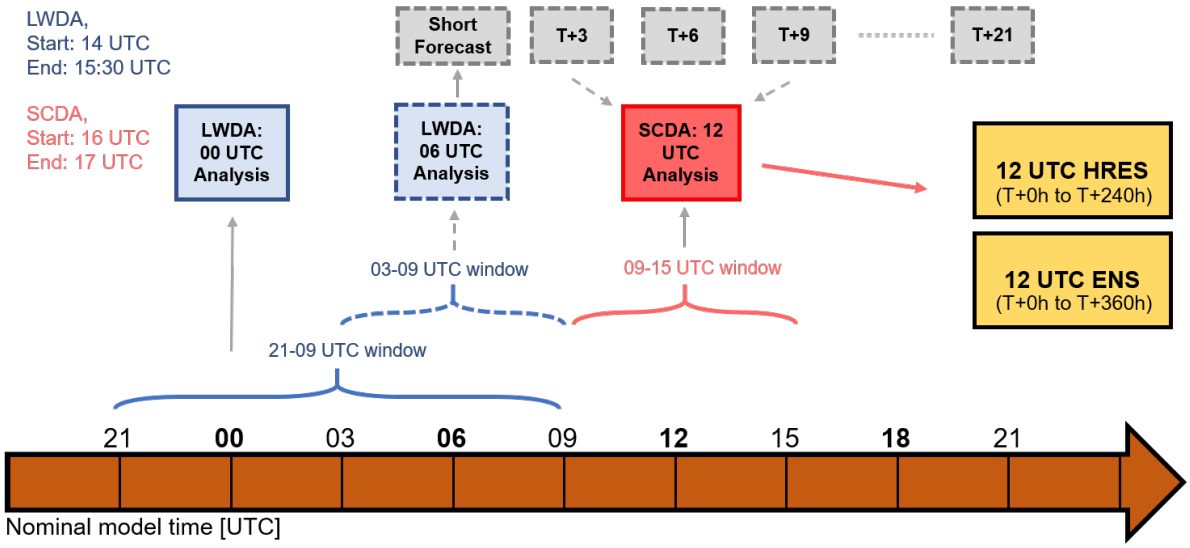

Figure 2.7: Sequence of the ECMWF's operational data assimilation system with its components LWDA (blue boxes), SCDA (red boxes), short forecasts (gray boxes), HRES and ENS forecasts (yellow boxes). Rebuilt based the ECMWF's forecast user guide available at: https://confluence.ecmwf.int/display/FUG/.

Chapter 3

Evaluation of the lower-stratospheric moist bias in the IFS

In this chapter, results on the lower-stratospheric moist bias in the ECMWF model are presented. Although the LS humidity bias has been present in previous versions of the ECMWF for decades, there are still open questions regarding this systematic error. These concern its vertical structure (RQ1.1), its seasonal and synoptic variability (RQ1.2), as well as potential causes (RQ1.3), which are addressed in this paper. For the first time, the bias is also analysed in the ECMWF's most recent reanalysis ERA5. The evaluation is based on unique, independent measurements of the water vapour lidar WALES (Sect. 2.1).

3.1 Data and methodology

The water vapour DIAL observations are compared with ERA5 reanalysis (Sect. 3.2-Sect. 3.4) and IFS analysis (Sect. 3.5) data. The DIAL data and the gridded model output have fundamentally different characteristics (Table 3.1). A few preparatory steps are necessary to compare both data sets adequately which are outlined in Sect. 3.1. The sequence to process the DIAL and IFS model data is shown in Figure 3.1. The DIAL profile data are given as humidity mixing ratio which are first converted to specific humidity. Then, the two filtering steps are carried out to deal with the measurement noise and to remove profiles with uncertain tropopause detection (Sect. 3.1.1, Sect. 3.1.3). The gridded IFS model output is given on model levels of which the geometrical altitude has to be determined. Moreover, the model output is coarser temporal and horizontal resolved

	DIAL measurements	**IFS model data**
Moisture quantity	Mixing ratio in ppm	Specific humidity in kg/kg
Temporal resolution	12 s (24 s for WISE)	Hourly
Horizontal resolution	2-6 km (flight speed-dependent)	31 km (9 km for IFS analysis)
Vertical resolution	About 300 m	300 m near tropopause
Vertical coordinate	Height	Sigma-pressure hybrid level

Table 3.1: Technical characteristics DIAL and IFS model data

than the observations requiring an interpolation towards the observations in space and time (Sect. 3.1.2). Then, thermal tropopause altitudes are calculated using interpolated ERA5 temperature field for each observation profile (Sect. 3.1.3). This allows to perform the statistical evaluation in tropopause-relative coordinates.

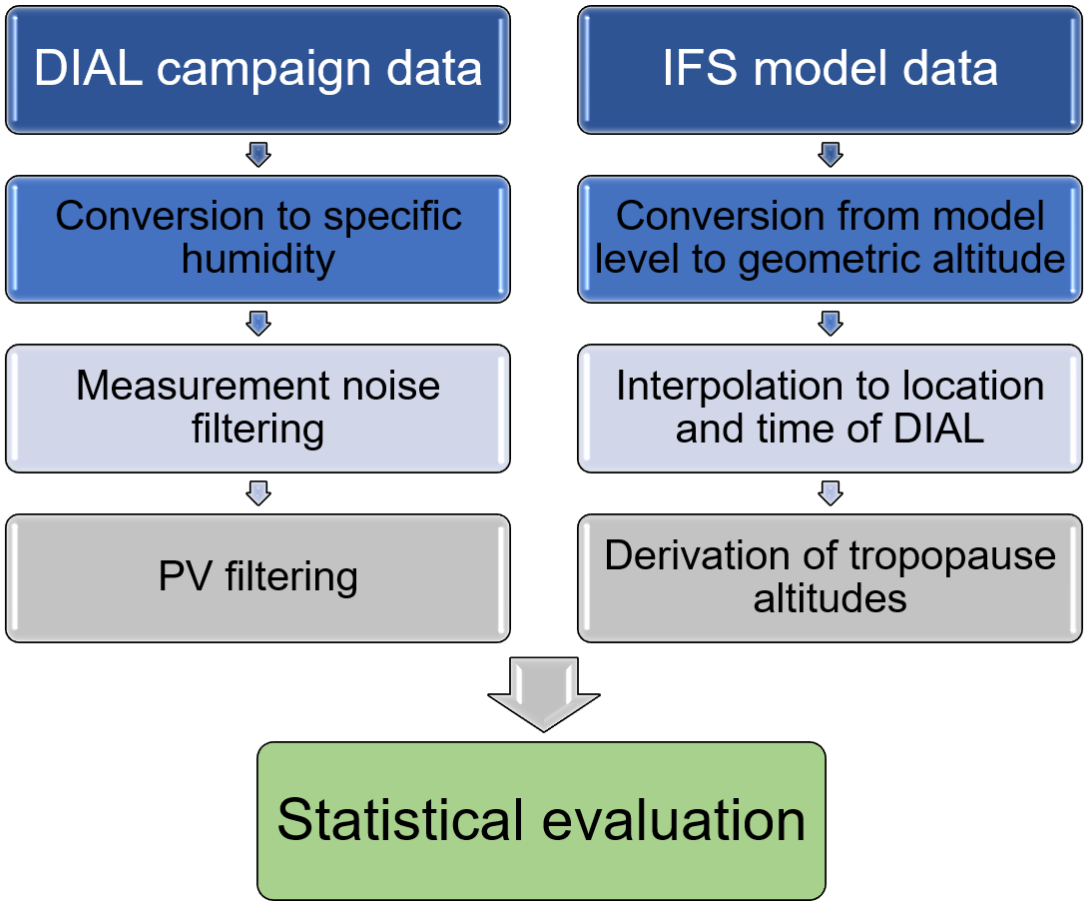

Figure 3.1: Data processing steps for the study of the lower-stratospheric bias in the IFS.

3.1.1 The multi-campaign DIAL data set

In this study, a multi-campaign water vapour lidar data set is used obtained with the airborne DIAL instrument WALES. During six measurement campaigns between 2013 and 2021, 41 research flights were carried out, providing almost 33,000 water vapour profiles in the LS (Table 3.2). The profiles were sampled along the flight track and extend from the surface up to about 14.5 km altitude, corresponding to the maximum flight level of the HALO aircraft (Krautstrunk and Giez, 2012). As the focus of this study is on the mid-latitude UTLS, only flights that provide a substantial amount of data above the tropopause and north of 35°N latitude, are considered. The majority (25) of these flights took place in the northern hemispheric autumn season during the NAWDEX (briefly introduced in Sect. 1.2.4; Schäfler et al., 2018) and the Wave-driven ISentropic Exchange campaign (WISE; Kunkel et al. (2019)). As part of the campaigns ElUcidating the RolE of Cloud-Circulation Coupling in ClimAte (EUREC^{4}A; Stevens et al., 2021), the

Next-generation Aircraft Remote sensing for VALidation studies (NARVAL; Klepp et al., 2014), and NARVAL2 (Stevens et al., 2019) measurements were taken during eight flights in the winter season. In addition, the Cirrus in High-Latitudes (CIRRUS-HL; Jurkat-Witschas et al., 2024) mission provides observations in summer. Figure 3.2 depicts the legs of HALO research flights where DIAL observations are available. In total, HALO covered a flight distance of more than 200,000 km, which is equivalent to five times the Earth's perimeter. Most flights were carried out over the North Atlantic between 48° and 66°N, the North Sea and central to western Europe. Additionally, the subtropics (> 35°N) and the Arctic (> 66°N) were covered by individual flights.

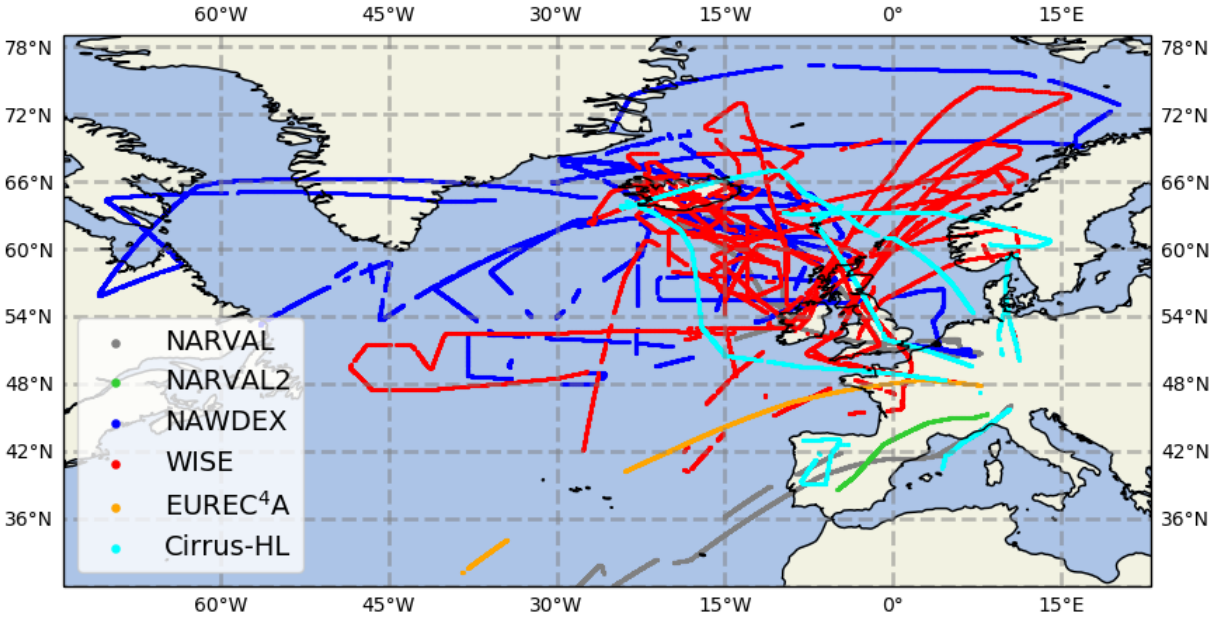

Figure 3.2: Map of HALO flight sections with WALES DIAL water vapour observations during the research campaigns NARVAL, NARVAL2, NAWDEX, WISE, EUREC^{4}A, and CIRRUS-HL.

Table 3.2: Overview of all considered campaigns with DIAL observations. The number of DIAL profiles refers to all profiles that were sampled during 41 research flights.

Campaign	Period	Season	Flights (hours)	Distance	Profiles	Profiles in LS
NARVAL	01/02 2013/14	Winter	7 (41)	31,157	10,973	5,288
NARVAL2	08 2016	Summer	1 (9)	7,729	2,395	485
NAWDEX	09/10 2016	Autumn	11 (75)	55,695	19,139	12,062
WISE	09/10 2017	Autumn	14 (105)	83,041	13,557	9,493
EUREC^{4}A	01/02 2020	Winter	1 (8)	7,011	2,307	1,009
CIRRUS-HL	06/07 2021	Summer	7 (30)	23,675	6,777	4,568
Total	–	–	41 (268)	208,308	55,148	32,905

Pre-processing of the lidar data

As introduced in Sect. 2.1 the retrieved water vapour DIAL profiles are not free of error. Kiemle et al. (2008) give an estimate of their accuracy in the UTLS to be within 10 %. The DIAL data retrieval provides the statistical error (i.e. the measurement noise, q_{noise}) in the observed volume, from which the relative error of the DIAL specific humidity data is calculated via Equation 3.1:

$$\text{Relative error} = \left(\frac{q_{\text{noise}}}{q_{\text{DIAL}}} \right) \times 100\,\% \tag{3.1}$$

The histogram of the relative error for the entire DIAL data set is given in Figure 3.3, and shows a uni-modal shape with a largest density of the relative error located around 6.5 % (median) which is in line with Kiemle et al. (2008). From the maximum, the distribution declines rapidly to both flanks, with the decline to larger errors being faster than that to smaller ones expressing the negative skew of the distribution. For each bin of the relative error the average altitude is shown by the red line in Figure 3.3. Interestingly, lower relative errors ($< 5\,\%$) are located at higher altitudes (about 9 km), while increased errors occur at low altitudes (< 6 km). This indicates a low statistical error of the DIAL observations at high altitudes.

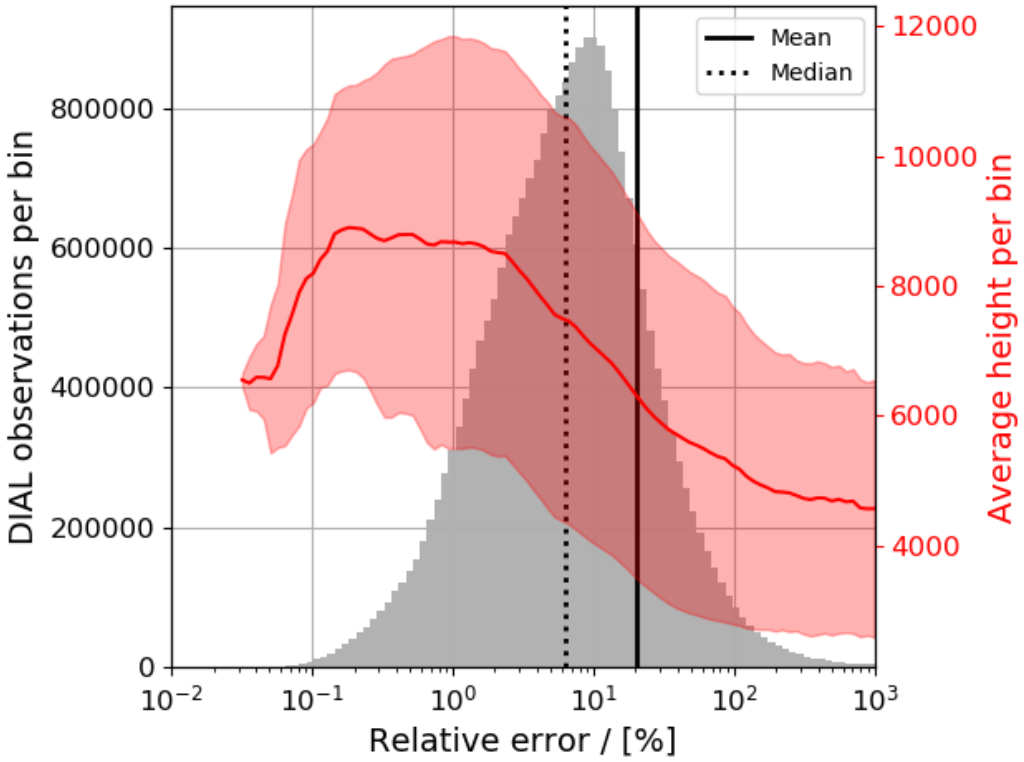

Figure 3.3: Histogram DIAL data counts per bin of the relative measurement error (gray) with a size of 0.01 g kg^{-1} of $\log_{10}(q_{DIAL})$. The mean (median) of the histogram are shown by the black solid (dashed) lines. The red line shows the average height per bin of relative error with its associated standard deviation (red shading).

Figure 3.4 depicts the distribution of the relative error for each research flight, with medians (black lines in the boxes) ranging between 2-20 %. The 95th percentiles shows, that for almost each flight 5 % of the DIAL data possess a relative error of up to 100 %, a value where the noise is as large as the signal (i.e., humidity value). This variability of the relative error in the different research flights can be explained by the different flight altitudes (and air speeds) and the different synoptic situations (different air masses, cloud cover).

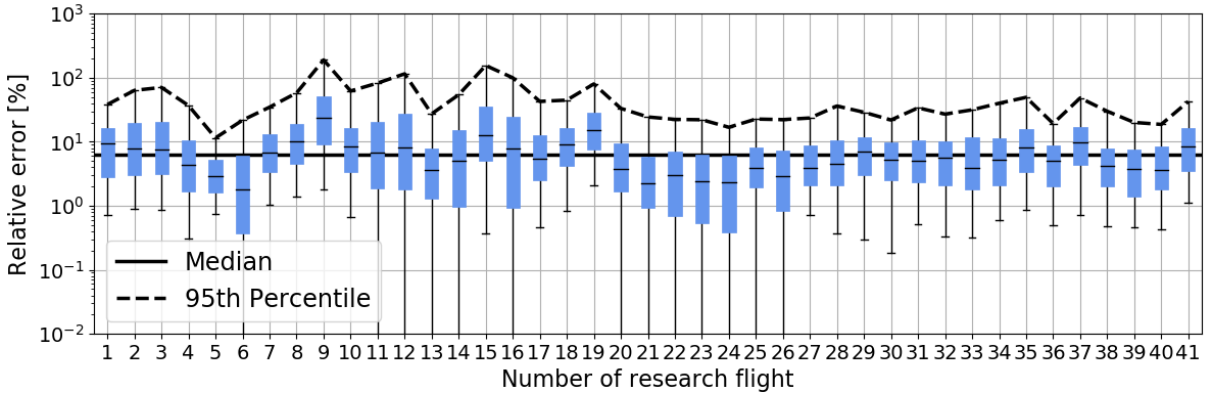

Figure 3.4: The relative error for the individual flights. Boxes define the interquartile (IQR) range, and the whiskers show the 5th/95th percentile. The median for all flights is given by the solid black line and the dashed black line connects the 95th percentiles.

In the following, it is examined to what extent the occurrence of noisy measurements can be reduced in the data set. For this purpose, different relative error thresholds are applied in a sensitivity study to filter noisy data. For each individual flight, (15 %, 5 % (dashed line in Figure 3.4), None) data is filtered out. Figure 3.5 show the distributions of the water vapour DIAL data after applying the different filtering thresholds. In general, all distributions cover specific humidity values of 4 orders of magnitude, ranging from 10^{-3} to $10\,\mathrm{g\,kg^{-1}}$. The bi-modal shape of the histogram is composed of a broad moist part that can be assigned to the troposphere and a fraction of low humidity representing the dry conditions in the stratosphere. This is confirmed by the derived average height per bin (note the blue lines in Figure 3.5) that show a continuous decrease in altitude with increasing specific humidity.

The differently colored histograms (darkblue, blue, lightblue) represent the DIAL data set after filtering by different threshold values of the relative error (No filtering, 95th percentile, 85th percentile). Filtering 5 % of the noisiest data (95th percentile), data coverage is particularly reduced above approx. $0.02\,\mathrm{g\,kg^{-1}}$ and at altitudes below 8 km altitude, while the number of DIAL data associated with the stratosphere (high altitudes) is almost unaffected. On the contrary, if 15 % of the noisiest data are filtered out, the number of DIAL data is reduced substantially at nearly any altitude and over the entire specific humidity spectrum. It is also noteworthy that the average altitude curves diverge at very low specific humidity values ($< 3\times10^{-3}\,\mathrm{g\,kg^{-1}}$), where the unfiltered curve shows a lower altitude (about 2-3 km), compared to the filtered curves. Such low specific humidity at low altitudes (i.e., in the UT) are typical for stratospheric intrusions (Trickl et al.,

2016), where larger measurement errors are expected. However, the small number of data with specific humidity below $3 \times 10^{-3}\,\mathrm{g\,kg^{-1}}$ should also be taken into account here. In conclusion, the 5 % threshold is a useful compromise to effectively reduce "noisy" data without excessively reducing the number of DIAL data (especially in the stratosphere, which is the target of this analysis). The 5 % filtering is therefore applied to the entire data set.

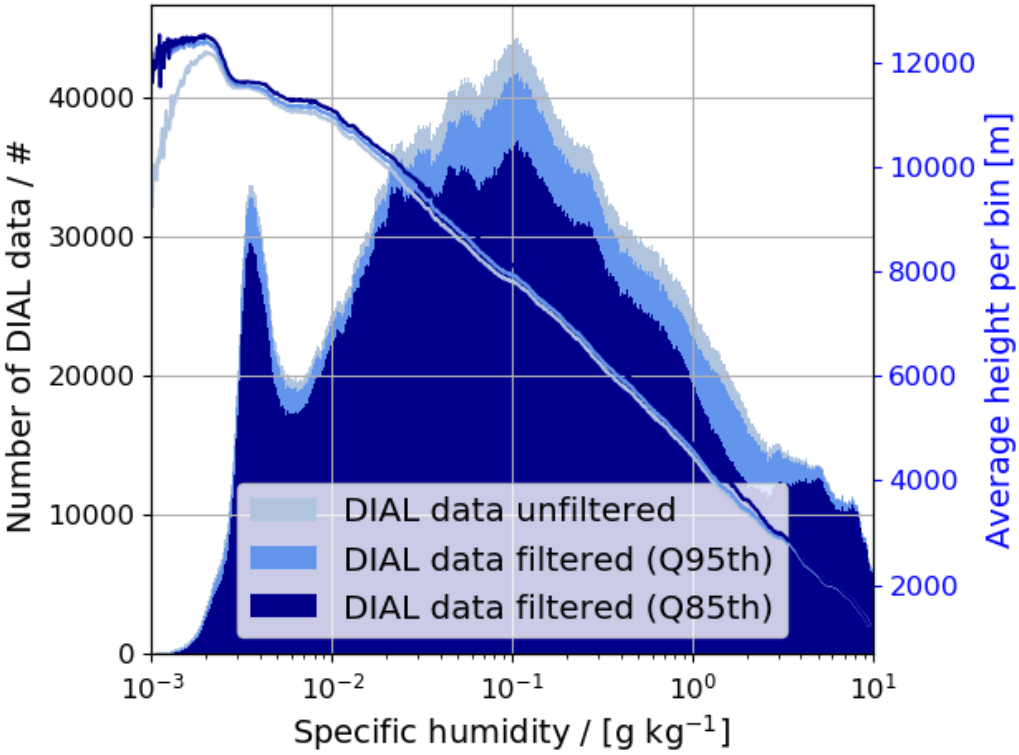

Figure 3.5: Histograms of DIAL data per humidity bin with a size of $0.01\,\mathrm{g\,kg^{-1}}$ of $\log_{10}(\mathrm{q}_{DIAL})$ for different filtering thresholds (see text). For each filtering the average height per bin is illustrated by the solid line (same colouring).

For a better understanding of the DIAL data set used in this evaluation, the geographical distribution of the data and the measured distributions of the specific humidity for the individual campaigns are shown in Figure 3.6. The number of DIAL observations shows a high data availability in the mid-latitudes (particularly between 50°-67°N), which is the region of interest in this study (see Figure 3.6a). This data set that covers humidity observations in a broad spectrum of synoptic situations is considered to be representative of mid-latitude weather. Figure 3.6b shows the histograms of the observed specific humidity data, in which the data are classified in which the data points were classified according to their altitude region (UT, LS, below UTLS). It should be mentioned that this data classification is based on the tropopause height, whose derivation is first described in Sect. 3.1.3. The grey shading clearly shows that the low measured specific humidity values (left part of the spectrum, dark grey) are assigned to the LS, while higher specific humidity values (right part) are assigned to the UT and the troposphere below.

The coloured lines in Figure 3.6b represent the frequency for the different research campaigns. Each campaign exhibits an individual footprint of measured humidity, depending on the season, observation areas, and the flight level selection. For instance, the histograms for NAWDEX and WISE are remarkably similar since both campaigns took place over the North Atlantic in autumn. However, as only two NIR wavelengths were operated to measure water vapour during WISE, fewer measurements are available at high humidity levels (i.e, $> 10^{-1}\,\mathrm{g\,kg^{-1}}$). NARVAL shows a distinctive dry spectrum of measured humidity corresponding to the winter season and less data is available for the LS due to the low altitudes frequently flown during this campaign. The humidity distribution of the CIRRUS-HL summer campaign stands out because a large proportion of high moisture values was observed. Its associated marked peak at $10^{-1}\,\mathrm{g\,kg^{-1}}$ is related to observations in the UT. The NARVAL2 and the EUREC^{4}A campaign provide UTLS measurements only for one flight and, thus, compared to the other field campaigns, provide a small number of observations (Table 3.2).

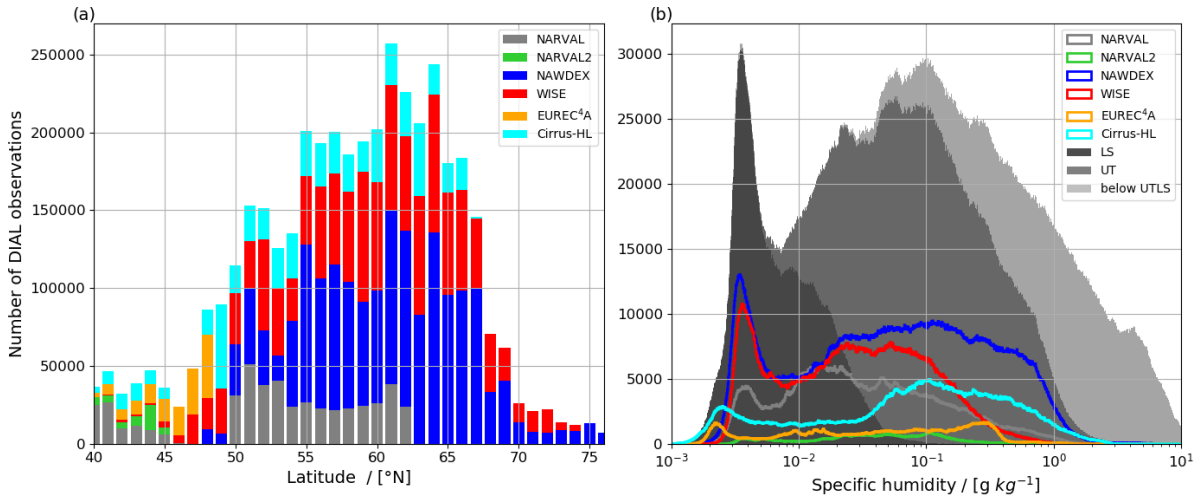

Figure 3.6: (a) Stacked distribution of the number of DIAL observations in 1° latitude bins per individual campaign (coloured bars). (b) Histogram of observations per humidity bin with a size of $0.01\,\mathrm{g\,kg^{-1}}$ of $\log_{10}(\mathrm{q}_{DIAL})$ for individual campaigns in the UTLS (coloured lines). Shading shows frequencies separated for the LS (all data above the thermal tropopause, dark grey shading), the UT (all data between the thermal tropopause and 5 km below, medium grey shading), and the remaining data in the troposphere below UT (light grey shading).

3.1.2 ERA5 reanalysis and IFS analysis data sets

Reanalysis provides the most comprehensive picture of past weather and climate conditions. Atmospheric reanalysis data is generated by combining past "short-range forecasts" with past observations using data assimilation methods (see Sect. 2.2), and so re-construct the best possible estimate of the atmosphere. The most modern reanalyses (such as the ECMWF's ERA5 reanalysis) cover a period from about 1950 to the present day. Reanalyses are generally based on a fixed model cycle, which means that they are based on both a consistent model physics (i.e. parameterizations) and an unchanged imple-

mented data assimilation scheme. Thus, reanalyses provide valid model data sets over climatological periods, making them a solid reference for long-term studies. This is a major advantage, as this study uses campaign observations from a period of 8 years. The ERA5 reanalysis data are the basis for the evaluation of humidity errors in the UTLS (Sect. 3.2-Sect. 3.4). ERA5 is based on the IFS cycle Cy41r2 that was operationally used in 2016. In addition to the ERA5 reanalysis data, the results of newer IFS versions (Cy47r1 and a pre-operational version of Cy47r3, operationally used in the IFS in 2020 and 2021 respectively) are examined in a separate sub-study (Sect. 3.5).

ERA5 is the latest-generation reanalysis of the ECMWF and offers the highest spatial, temporal and vertical resolution of the output data compared to other modern global reanalyses. Atmospheric quantities are provided on a global grid with a horizontal resolution (TL639) of about 31 km and on 137 model levels, ranging from the surface up to 0.01 hPa (about 80 km) in the vertical. The altitude range of the DIAL observations (i.e., from the surface to the LS) is covered by the lowermost 70 model levels. The vertical grid spacing of the model levels ranges from a few metres in the boundary layer to about 300 m at the tropopause level (Schäfler et al., 2020). ERA5 reanalyses are available with a time resolution of 1 h, which is an improvement compared to a 6-hourly resolution of its predecessor ERA-Interim (Dee et al., 2011). Further details about ERA5 are documented in Hersbach et al. (2020).

For this study, ERA5 reanalysis data is retrieved on a 0.36°×0.36° latitude-longitude grid corresponding to a resolution of 31 km. As vertical coordinate, the ERA5 reanalysis output is based on hybrid sigma-pressure model levels, which represent a combination of pressure and terrain-following coordinates. The pressure of the coordinates and the height of each model level is derived from the documentation (ECMWF, 2016c). The pressure of the model level can be trivially calculated from the given surface pressure. From the pressure of each level (k), the layer thickness (Δz) of two neighboring model levels are calculated via Equation 3.2 which is based on the hydrostatic relation. In Equation 3.2, R_{dry} is the gas constant of dry air, T_v is the virtual temperature, g is the gravitational acceleration

$$\Delta z_k = z_{k+1/2} - z_{k-1/2} = -\frac{R_{\mathrm{dry}}(T_v)_k}{g} \ln\left(\frac{p_{k+1/2}}{p_{k-1/2}}\right). \tag{3.2}$$

The height of each model level is obtained by summing up the calculated layer thicknesses following Equation 3.3

$$z_j = z_{surface} + \sum_{j=k+1}^{Nlev} \frac{R_{\mathrm{dry}}(T_v)_j}{g} \ln\left(\frac{p_{j+1/2}}{p_{j-1/2}}\right). \tag{3.3}$$

Compared to ERA5, the IFS analysis data have the same vertical and temporal resolution, but a higher horizontal resolution (about 9 km; 0.125° × 0.125°, compare Figure 3.7). To statistically compare IFS reanalysis, analysis and WALES data, the gridded model data are interpolated in space and time to the observation location. The interpolation sequence applied in this work uses a horizontally bi-linear interpolation, followed by a linear interpolation in the vertical. Finally, a linear interpolation in time of the hourly model profiles towards the observation time is carried out. This sequence of interpolation steps has been applied similarly in other studies (e.g., Schäfler et al., 2010). It should be

noted that the procedure outlined (derivation of the model level height and interpolation) is used identically for both the ERA5 reanalysis and the IFS output data.

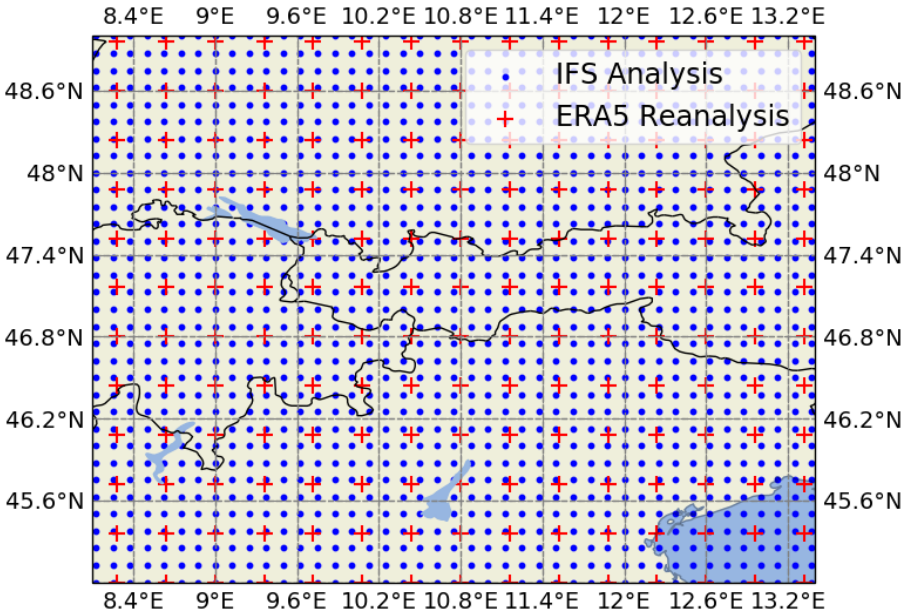

Figure 3.7: Example location of the grid points of the ERA5 reanalysis product ($0.36 \times 0.36°$ resolution, red crosses) and the IFS analysis product ($0.125 \times 0.125°$ resolution, blue circles). Note that the model data is originally spectral and is converted into a grid with the above-mentioned resolution during the data retrieval.

3.1.3 Determination of the thermal tropopause altitudes

Due to the variable altitude of the tropopause (indicated in Figure 1.4), the distribution of water vapour in the UTLS at individual altitudes is also highly variable. Hence, averaging of the humidity profiles in geometrical coordinates strongly blurs the vertical gradients across the tropopause. Therefore, bias statistics are commonly performed in tropopause-relative coordinates, such as in Kunz et al. (2014) and Bland et al. (2021). By definition the thermal tropopause marks the reversal of the vertical temperature gradient and thus the abrupt increase in static stability, which is reflected in the sharp distribution of trace species across the tropopause (Gettelman et al., 2011). The thermal tropopause is used in this analysis as it best reflects the strongest vertical gradients of water vapour (Birner et al., 2002, Pan et al., 2004). From each ERA5 temperature profile interpolated to the 15 m vertical grid of the DIAL observations, the altitude of the thermal tropopause is calculated using the World Meteorological Organization's (WMO) lapse rate-based definition (WMO, 1957). A tropopause is detected as the lowest level at which the vertical temperature gradient Γ drops below $2\,\mathrm{K\,km^{-1}}$ and is only defined if the average lapse rate between this and any other level within a 2 km deep layer remains equal or lower than $2\,\mathrm{K\,km^{-1}}$. The vertical temperature gradient, i.e. the lapse rate, is computed as

$$\Gamma = -\frac{\Delta T}{\Delta z}, \tag{3.4}$$

with ΔT being the temperature difference and Δz the height difference between two neighboured levels.

In this analysis, the detection of the thermal tropopause is directed upwards from 5 km altitude to avoid false detection of the tropopause due to local temperature variations in the lower to middle troposphere. When a tropopause is detected, the (thermal) tropopause-relative coordinates $z_{\text{tropopause-relative altitude}}$ are obtained by simply subtracting the tropopause altitude ($z_{tropopause}$) from the geometric height vector ($z_{geom.}$):

$$z_{\text{tropopause-relative altitude}} = z_{\text{tropopause}} - z_{\text{geom.}} \tag{3.5}$$

There are conditions in which tropopause detection is ambiguous, especially in the vicinity of the jet streams and associated tropopause folds, where double tropopauses can occur (e.g., Shapiro, 1980, Gettelman et al., 2011, Tinney et al., 2022). It is found that in situations of weak vertical temperature gradients near the jet streams, the lapse rate threshold in the WMO definition may lead to vertical jumps of the tropopause altitudes for adjacent profiles. These fluctuations result in a wrong vertical allocation of water vapour in tropopause-relative coordinates. Such incorrectly allocated profiles become visible in the data set due to unusual PV values. PV is typically large (i.e., » 2 PVU) above and small (i.e., « 2 PVU) below the thermal tropopause. To remove such profiles in the overall statistic, a filtering method based on mean potential vorticity (MPV; Shapiro et al. (1999)) is applied, which is the average PV calculated for the 5 km layer above and below the tropopause. MPV <3.5 PVU above and MPV >3.5 PVU below the tropopause are found to be an efficient metric to filter profiles within an erroneously assigned tropopause.

Figure 3.8 shows the vertical distribution of tropopause altitudes for the 32,905 profiles, which lies between 5.5 km and more than 15 km altitude, reflecting the broad spectrum of synoptic situations covered by the data set. The majority of all tropopause levels are found between 10 and 13 km, which represents the typical location of the mid-latitude tropopause with respect to interannual or synoptic variations (e.g., Hoerling et al., 1991).

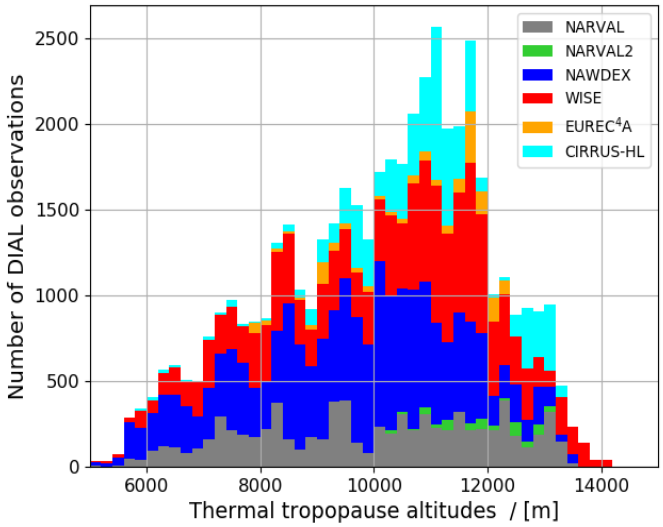

Figure 3.8: Histogram of the number of observations per thermal tropopause altitude bin (1000 m) and per campaign (coloured bars).

3.1.4 Statistical metrics

The selection of a suitable difference metric is crucial for a robust quantification of model humidity errors, and different statistical approaches can be found in the literature (Kunz et al., 2014, Dyroff et al., 2015, Bland et al., 2021). As specific humidity rapidly decreases across the tropopause, absolute humidity differences are not appropriate, and most studies rely on a relative formulation of the error. However, since the simple ratio of model and observation as well as the absolute bias divided by the observed value are statistical asymmetric quantities, a logarithmic formulation with base 2 is applied (see Equation 3.6), that was introduced by Kunz et al. (2014):

$$\text{humidity bias} = \log_2 \left(\frac{Q_{\text{Model}}}{Q_{\text{DIAL}}} \right) \tag{3.6}$$

with Q_{DIAL} being the measured and Q_{Model} being the specific humidity of the ERA5 reanalysis data (IFS data, respectively). This dimensionless definition of the relative bias is symmetrically centred around zero and thus not distorted when averaged. A perfect agreement (humidity bias = 0) between the ERA5 and the DIAL specific humidity is reached if $Q_{Model} = Q_{DIAL}$. A positive humidity bias $\in [0, \infty]$ indicates an overestimation of humidity by the model ($Q_{Model} > Q_{DIAL}$), whereas a negative humidity bias $\in [-\infty, 0]$ implies an underestimation ($Q_{Model} < Q_{DIAL}$). Table 3.3 gives some example bias values for selected moisture observations.

Table 3.3: Example values of specific humidity and the corresponding derived humidity bias.

Q_{Model}	$g\,kg^{-1}$	0.50	0.75	1.00	1.25	1.50	2.00	2.50	3.00
Q_{DIAL}	$g\,kg^{-1}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Humidity bias	Unitless	−1.00	−0.41	0	0.32	0.58	1.00	1.32	1.58
Percentage	%	−50	−25	0	25	50	100	150	200

3.2 Cross sections of water vapour and humidity bias for a selected case

In order to provide an overview of the typical spatial distribution of water vapour in the UTLS in the observation and in ERA5, an example cross-section of water vapour from the research flight on 1 October 2017 during the WISE campaign is first presented in Figure 3.9. The case is selected as it possesses a good data coverage across the UTLS and as it additionally provides ozone observations. HALO flew meridional transects approximately between Iceland and Ireland over the North Atlantic (50°N to 60°N) at 13°W with the aim of conducting observations across the strong, zonally oriented jet stream. Schäfler et al. (2021) provides a detailed insight into the prevailing synoptic situation during this flight. Presenting the isentropic PV distribution, they show how the strong jet stream separates stratospheric (high PV north of it) and tropospheric (low PV south of it) air masses. Due to these air mass contrasts, sharp trace gas gradients were expected, that were the main research target of this particular flight. Further details on the underlying weather situation and the corresponding mission objectives are documented in (Schäfler et al., 2021). During the transect, HALO flew perpendicular to the jet stream and overpassed it several times, allowing excellent coverage with the WALES instrument. Water vapour DIAL measurements for an example flight leg of such a jet stream overpass are presented in the following.

The left part of Figure 3.9a (up to a distance of roughly 800 km) illustrates the water vapour distribution north of the jet stream (see magenta isopleths) where the aircraft flew above the low-located tropopause within the LS. HALO then crossed the pronounced jet stream with wind velocities of more than $90\,\mathrm{m\,s^{-1}}$. Near the jet core, the tropopause altitude jumps from 6.5 to 14 km within a few kilometres flight distance. The dynamical tropopause (2 PVU contour line) also displays the ascent of the tropopause and a corresponding tropopause fold that extends along inclined isentropes into the mid-troposphere. In the right part of Figure 3.9a, the air mass located to the south of the jet stream exhibits high tropopause altitudes exceeding the flight level by roughly 2 km, so that measurements are restricted to tropospheric air. Along the entire cross section, the highest specific humidity is observed at the lowest levels in the UT, ranging from $10^{-2}\,\mathrm{g\,kg^{-1}}$ to occasionally more than $10\,\mathrm{g\,kg^{-1}}$. The tropospheric air to the south of the jet stream has an increased humidity content compared to the air north of the jet stream. In the LS, specific humidity values lower than $10^{-2}\,\mathrm{g\,kg^{-1}}$ are frequently observed. At a first glance, the specific humidity curtain of ERA5 (Figure 3.9b) is very similar to the observations. However, the ERA5 humidity field appears to be smoother, particularly in the presence of strong horizontal water vapour gradients, for instance, near the jet stream and meso-scale filaments.

Differences between observations and model, calculated by applying Equation 3.6, are shown for the vertical section in Figure 3.9c. Reddish regions indicate an overestimation of humidity by ERA5, while bluish areas represent an underestimation. High positive and negative values of the bias alternate below the tropopause. In the LS, a coherent region of positive values is detected between 1 and 3 km above the tropopause, indicating an overestimated humidity that extends over the entire part north of the jet. At the highest altitudes, beyond 3 km above the tropopause, the moist bias is smaller. In order to study

the systematic nature of the diagnosed LS moist bias and its vertical structure, a statistic of all observations in tropopause-relative coordinates is performed.

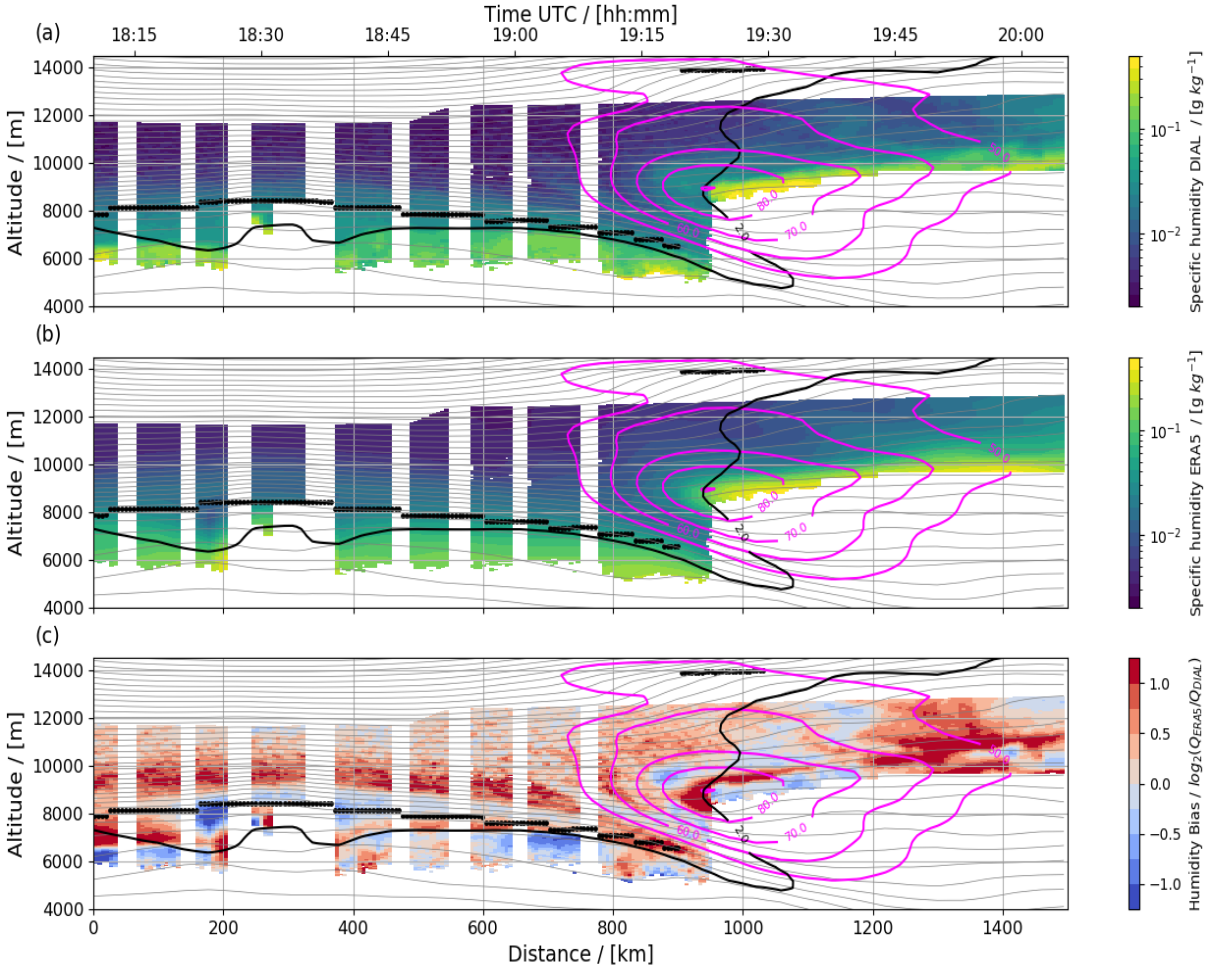

Figure 3.9: Vertical cross sections of (a) the DIAL specific humidity (colour shading, g kg^{-1}), (b) ERA5 specific humidity (colour shading,g kg^{-1}), and (c) the corresponding humidity bias (colour shading) on 1 October 2017. Panels (a)–(c) are superimposed by ERA5 fields of the potential temperature (grey contours, $\Delta\theta = 3$ K), the isopleths of the wind speed (magenta contours, in m s^{-1}), and the thermal (thick black dots) and the dynamical tropopause (2 PVU, black isoline).

3.3 Statistical analysis of the LS moist bias

3.3.1 Vertical structure

For all 32,905 profiles from the 41 flights, the average profiles of specific humidity and the humidity bias are presented in Figure 3.10. The moisture profiles of WALES and ERA5 show an exponential decline of specific humidity in the UT, ranging from about $5 \times 10^{-1}\,\mathrm{g\,kg^{-1}}$ at the lowest levels to approx. $3 \times 10^{-2}\,\mathrm{g\,kg^{-1}}$ at the tropopause. The strongest vertical gradient occurs in a layer of 0.5 to 1 km above the tropopause. Beyond, a less pronounced decline of water vapour extends until 4 km above the tropopause, followed by a vertical constant specific humidity of about $3.5 \times 10^{-3}\,\mathrm{g\,kg^{-1}}$. There is a high level of agreement between the ERA5 and WALES specific humidity profiles, particularly in the UT, although ERA5 appears to be moister at all altitudes. For both data sets, the median and arithmetic mean profiles of specific humidity slightly vary from each other. The median line is slightly shifted towards drier humidity values, most pronounced in the UT. Figure 3.10a also demonstrates a high data availability throughout the entire UTLS. The number of observations is highest between -5 and +1 km around the tropopause, with two local maxima at -2 km and roughly 1 km. Note that these two peaks in data availability are related to the typical flight altitudes, either above or below the main transatlantic air traffic routes (Schäfler et al., 2018), and the maximum data coverage close to the aircraft. Above the tropopause, the number of observations continuously decreases and roughly halves per kilometre altitude. At 4 km above the tropopause, about 3000 observations are available. The large number of observations above the tropopause allows for a robust evaluation of ERA5 humidity in an altitude region, where knowledge gaps concerning the bias have been identified (see RQ1.1).

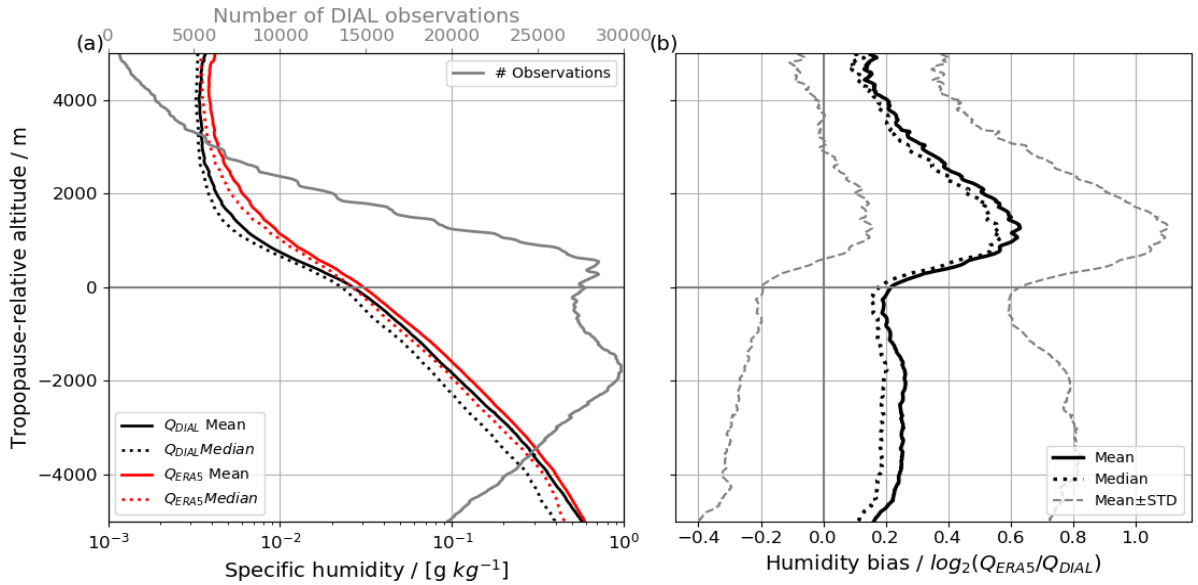

Figure 3.10: Tropopause-relative (a) vertical profiles of the DIAL (black lines) and the ERA5 (red lines) mean (solid) and median (dotted) specific humidity and the number of observations (grey). Note the log-scale notation of the x axis. (b) Mean/median bias (solid/dotted lines) and standard deviation (dotted grey lines).

The higher moisture values in the ERA5 data become apparent in the vertical profile of the humidity bias (Figure 3.10b) that is weakly positive (0.2; 15 %) in the UT and associated with a high standard deviation. This is a result of strong positive and negative bias values, as seen for example in the case study (Figure 3.9c). The weakest bias of 0.2 (< 15 %) is reached at the tropopause level. Above, the vertical moisture gradient is stronger in observations, leading to a significant overestimation of humidity in the LS up to 4 km above the tropopause. The bias increases to a maximum of 0.63 (55 %) at 1.3 km altitude above the tropopause. Beyond, the bias reduces by roughly 0.2 per 500 m up to 4 km altitude above the tropopause, where it is approx. 0.2 (15 %). At the highest altitudes (> 4 km above the tropopause), a weak and vertically nearly constant bias is observed. At the tropopause as well as above, the standard deviation is significantly reduced compared to the UT. Mean and median profiles of the humidity bias slightly differ, but these differences are very small compared to the magnitude of the bias. The maximum mean and median biases are 0.63 (55 %) and 0.58 (49 %), respectively.

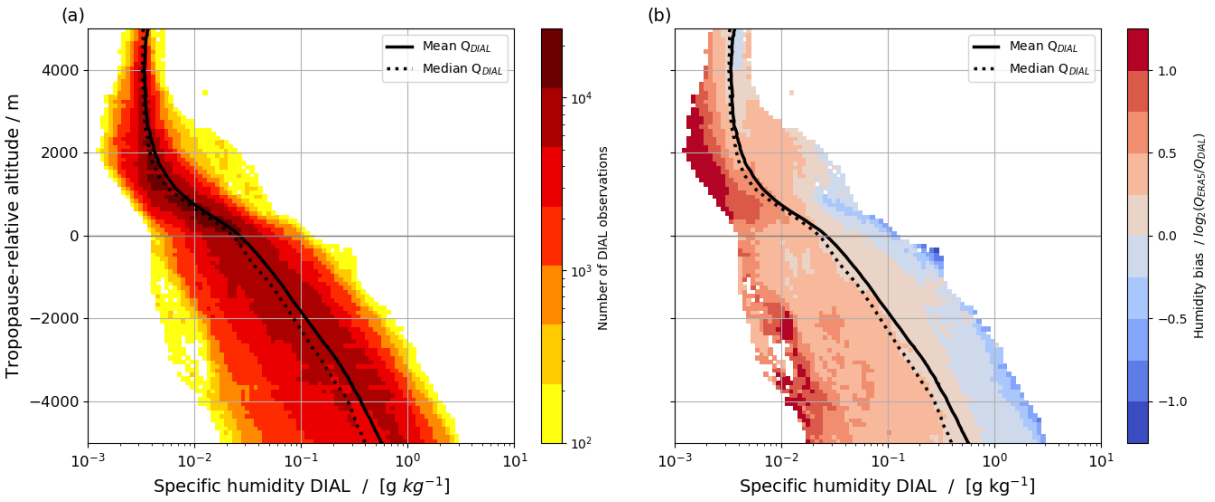

Figure 3.11: Binned distribution of DIAL specific humidity observations relative to the thermal tropopause coloured by (a) the number of observations per bin and (b) the bin-average humidity bias. Solid thick black (dotted) lines in (a) and (b) show mean (median) values per altitude bin. Bin sizes are 100 m and 0.01 g kg^{-1} of $\log_{10}(q_{DIAL})$. Please note the logarithmic abscissa and colour bar in (a).

To better illustrate the variability of the water vapour observations in the vertical, Figure 3.11 shows the number of data and the mean bias in bins of tropopause-relative altitude and specific humidity. Figure 3.11a indicates a broader distribution of water vapour observations in the UT compared to the LS. A small number of unusually low humidity values (< 10^{-2} g kg^{-1}) are detected below the tropopause (-4 to -1 km); on the other hand, some data that show high specific humidity (> 10^{-2} g kg^{-1}) are detected at approx. 1–3 km above the tropopause. These observations are related to incorrectly assigned tropopause altitudes that were not removed by the applied MPV filtering (Sect. 3.1.3). However, these remaining outliers are tolerable as they have a negligible impact on the statistics. Throughout the UT, a weak positive bias is detected in the bins of highest data availability. At the edges of the distribution, the highest humidity values show a negative

bias, while the lowest humidity values stand out due to a positive bias (Figure 3.11b). It is found that this is related to a narrower distribution of ERA5 humidity compared to the observations (not shown). The low number of observations at the edges should be noted here. In the LS the positive bias is higher and most pronounced up to 3.5 km above the tropopause and at very low specific humidity values. The positive bias reduces towards highest altitudes (> 3 km above the tropopause) of the LS, although the reduced data coverage has to be kept in mind (Figure 3.10a).

3.3.2 Synoptical and seasonal variability

In this section, the variability of the LS bias between flights, campaigns, and tropopause altitudes is investigated according to RQ1.2. Figure 3.12a shows the observed humidity distribution within a 3 km layer above the tropopause, i.e. the area of the strongest LS moist bias. The observed humidity values of all flights range from 1×10^{-2} to 4×10^{-3} g kg^{-1}, and their IQR strongly varies between the individual flights, which presumably relates to differences in the flight level, the tropopause altitude, and the synoptic situation. During summer (CIRRUS-HL) and autumn campaigns (NAWDEX and WISE), the range of observed humidity is larger compared to the winter campaign (NARVAL).

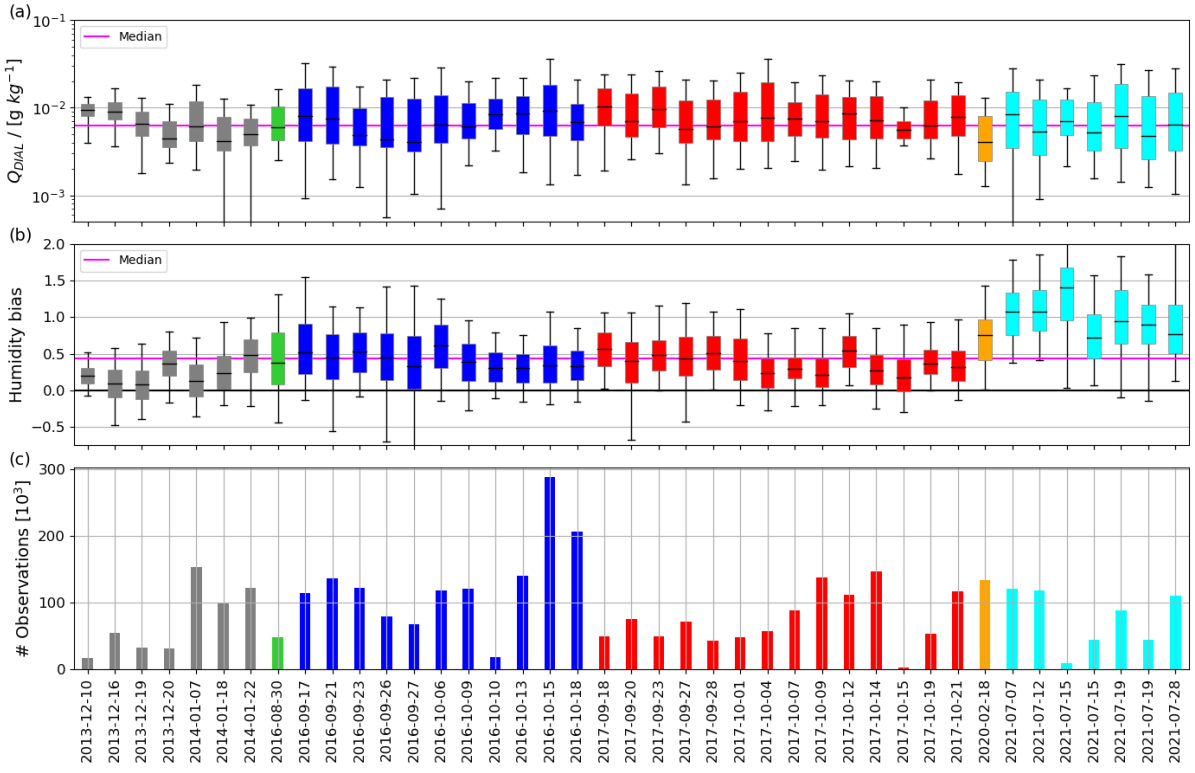

Figure 3.12: Distributions of (a) the observed humidity and (b) the humidity bias and (c) the number of data points for the layer 0 to 3 km above the tropopause. The average observed humidity and bias for all flights is given by the magenta lines in (a) and (b). The boxes in (a) and (b) define the IQR located around the median (black), and the whiskers illustrate the 5th and 95th percentile. The different campaigns are colour-coded as in Figure 3.2.

It is furthermore noticeable that intra-campaign variations (i.e. synoptic variability) of observed humidity exceed the seasonal variability. Per flight, the median LS bias (Figure 3.12b) varies from 0.2 (15 %) to 1.4 (164 %), but a positive bias is detected for each flight. Whereas the magnitude of the bias shows no obvious correlation with the LS moisture distribution, the moist bias appears to be smaller in winter (NARVAL) compared to autumn (NAWDEX and WISE). Interestingly, the moist bias during the CIRRUS-HL summer campaign is remarkably strong. The number of observations that is available for each flight is strongly variable between a few and several hundred thousand (Figure 3.12c).

The average profile of the bias and the number of observations for campaigns with an increased data coverage is shown in Figure 3.13. The data availability is very different across the campaigns (Figure 3.13b). During NAWDEX and WISE, a large number of observations is present between -5 km below and 4 km above the tropopause. CIRRUS-HL provides approximately half as much data at each altitude, except for altitudes beyond 3 km above the tropopause, where few data are available. Due to frequent low flight levels during the seven NARVAL flights, only a small number of observations is available beyond 1 km above the tropopause. The general structure with a pronounced positive bias, a local minimum at the tropopause, and a decrease towards the highest altitudes is apparent for all campaigns (Figure 3.13a), although significant differences can be identified across the campaigns. For the autumn campaigns in two successive years (NAWDEX and WISE), a similar shape of the bias is observed across the entire profile. The maximum moist bias is located at approximately the same altitude, and a similar decrease beyond this maximum is observed. However, the magnitude of the LS moist bias is slightly higher for NAWDEX (0.6, 50 %) compared to WISE (0.5, 40 %). During summer (CIRRUS-HL), a stronger moist bias is detected exceeding 1.2 (130 %) at its maximum. Compared to autumn, the summer bias is increased by a factor of 2–3. During winter (NARVAL), the LS moist bias is small (0.3, 23 %) and not substantially higher than the UT bias, but the limited representation due to the low number of observations should be noted here.

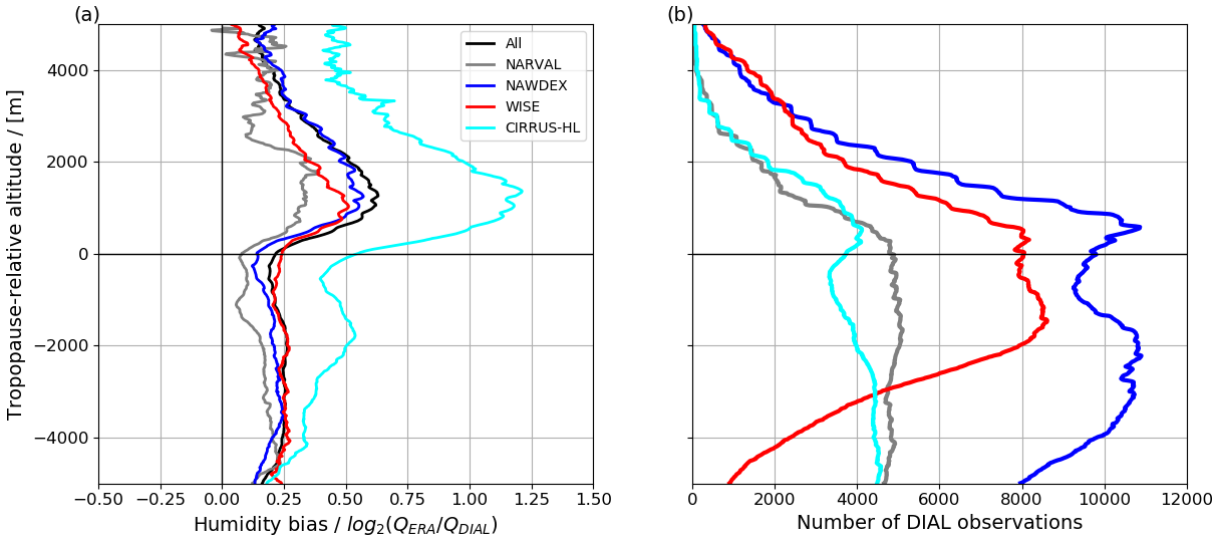

Figure 3.13: Tropopause-relative vertical profiles of (a) humidity bias and (b) number of observations for the different campaigns (colour-coded as in Figure 3.2). The black line in (a) represents the multi-campaign average (Figure 3.10b).

In addition, it is examined whether the observed vertical structure of the moist bias is sensitive to different synoptic situations. For this investigation, the DIAL profiles are classified by their corresponding tropopause altitude. Lower tropopauses are typically associated with trough situations and high tropopauses occur above ridges. For each category the corresponding average bias profile and the number of observations is given in Figure 3.14. The vertical structure of the bias (Figure 3.10b) is reproduced for each tropopause altitude interval. No systematic differences between the bias profiles can be revealed in the UT. Interestingly, each category shows an increased moist bias of comparable magnitude as well as a decrease above, although its vertical position relative to the tropopause is different. The maximum bias is located higher for low tropopause altitudes, while profiles with a large tropopause altitude show a maximum closer to the tropopause. For instance, the maximum bias for low tropopauses (< 8 km) is located at 2 km above the tropopause, while for the category with highest tropopauses (12–14 km), the maximum value is found at 1 km. The number of data points illustrates that each category exhibits a reasonable number of observations (Figure 3.14b).

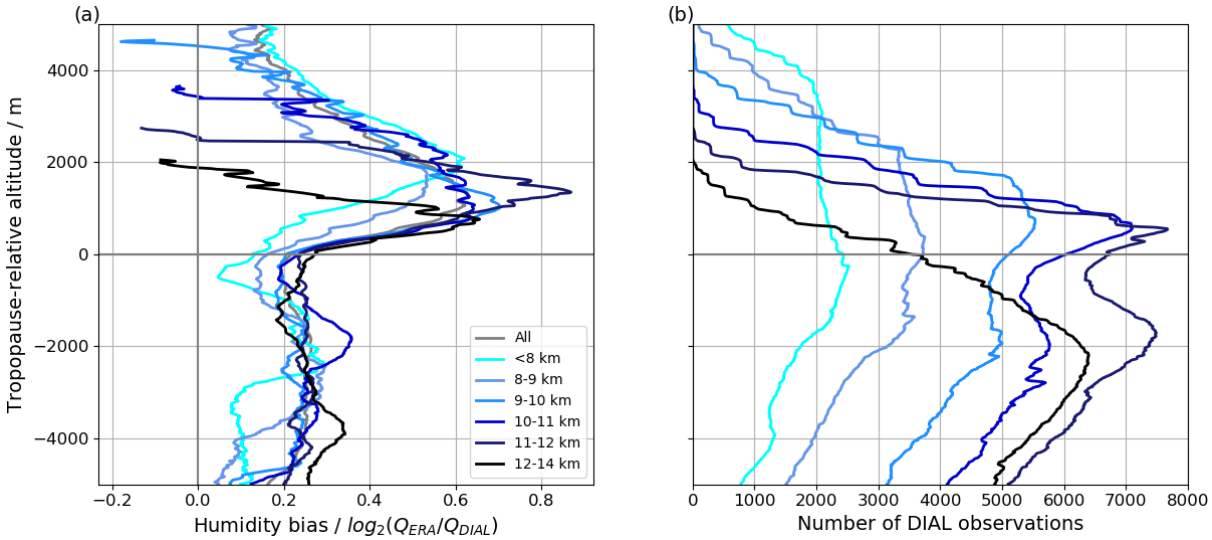

Figure 3.14: Tropopause-relative profiles of the (a) humidity bias and (b) number of observations for different intervals of observed tropopause altitudes (colour-coded).

3.4 Vertical structure of the humidity bias related to mixing processes

This analysis aims to evaluate possible causes of the LS moist bias in the ECMWF modelling system, motivated by the hypothesis, that the LS moist bias may be related to the representation of mixing processes (RQ1.3). For this purpose, collocated ozone and water vapour observations collected during four WISE research flights, which provide suitable data coverage, are evaluated. First, the observed ozone distribution for the same case study as introduced in Sect. 3.2 is shown in Figure 3.15a. Note that the ozone and water vapour observations are given as volume mixing ratios (VMR) in the following. The distribution of ozone is opposite to that of water vapour, with the lowest concentrations ($VMR_{O_3} < 100$ ppb) in the troposphere and an increase with altitude across the tropopause to $VMR_{O_3} > 500$ ppb. Note the filamentary structures of increased ozone values in the LS and the ozone-rich air which is transported downward within the tropopause fold (see more details in Schäfler et al. (2023)).

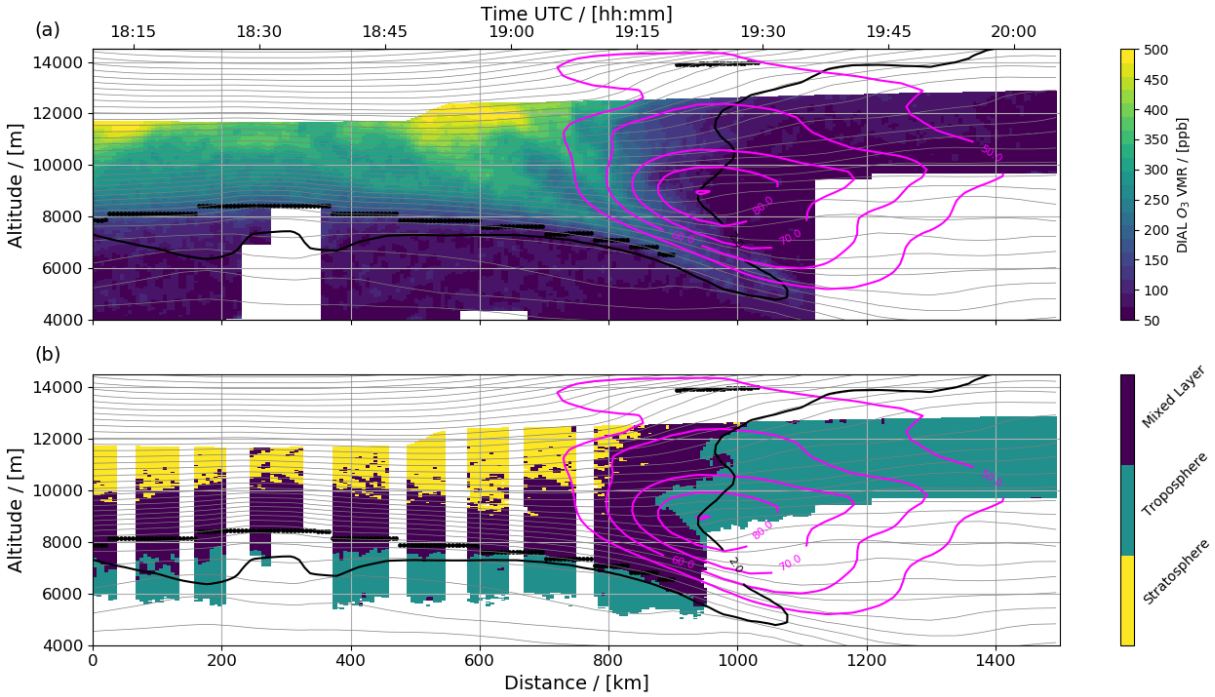

Figure 3.15: Vertical cross sections for the case study presented in Figure 3.9 but for (a) DIAL ozone volume mixing ratio (in ppb) and (b) air mass classes derived from the DIAL water vapour and ozone measurements.

Following the approach by Schäfler et al. (2021), the collocated water vapour and ozone observations for four WISE flights are illustrated in tracer–tracer (T–T) phase space in Figure 3.16, and three classes of observations are identified based on the characteristic distributions (e.g. Pan et al., 2004). First, tropospheric observations are characterised by low VMR_{O_3} (typically < 100 ppb) and a large spread of VMR_{H_2O}. Second, high VMR_{O_3} at low VMR_{H_2O} (< 6.5 ppm or $< 4 \times 10^{-3}$ g kg^{-1}) is assigned to lower-stratospheric

air. Third, a class with intermediate chemical characteristics ($VMR_{H_2O} > 6.5$ ppm and $VMR_{O_3} > 100$ ppb) is attributed to mixed air masses that experienced mixing between the troposphere and stratosphere. Although Schäfler et al. (2021) suggested a careful selection of the threshold for individual flights, here constant values are used for all four WISE flights, which is sufficient to fundamentally identify the three air masses. Sensitivity tests with slightly varied thresholds have shown only a little impact on the distribution of the classes in geometrical space. Such a re-projection of the air mass classification from T–T space to geometrical space with a coherent distribution of the three classes is shown in Figure 3.15b. Observations below the tropopause are predominantly assigned to tropospheric air, while the uppermost data to the north of the jet stream are classified as stratospheric air. South of the jet stream, where the flight altitude is below the tropopause, only tropospheric air is detected. In between the tropospheric and the stratospheric air, the mixed air mass is following the tropopause in a 2–3 km thick layer, which appears to be vertically deeper in the tropopause fold.

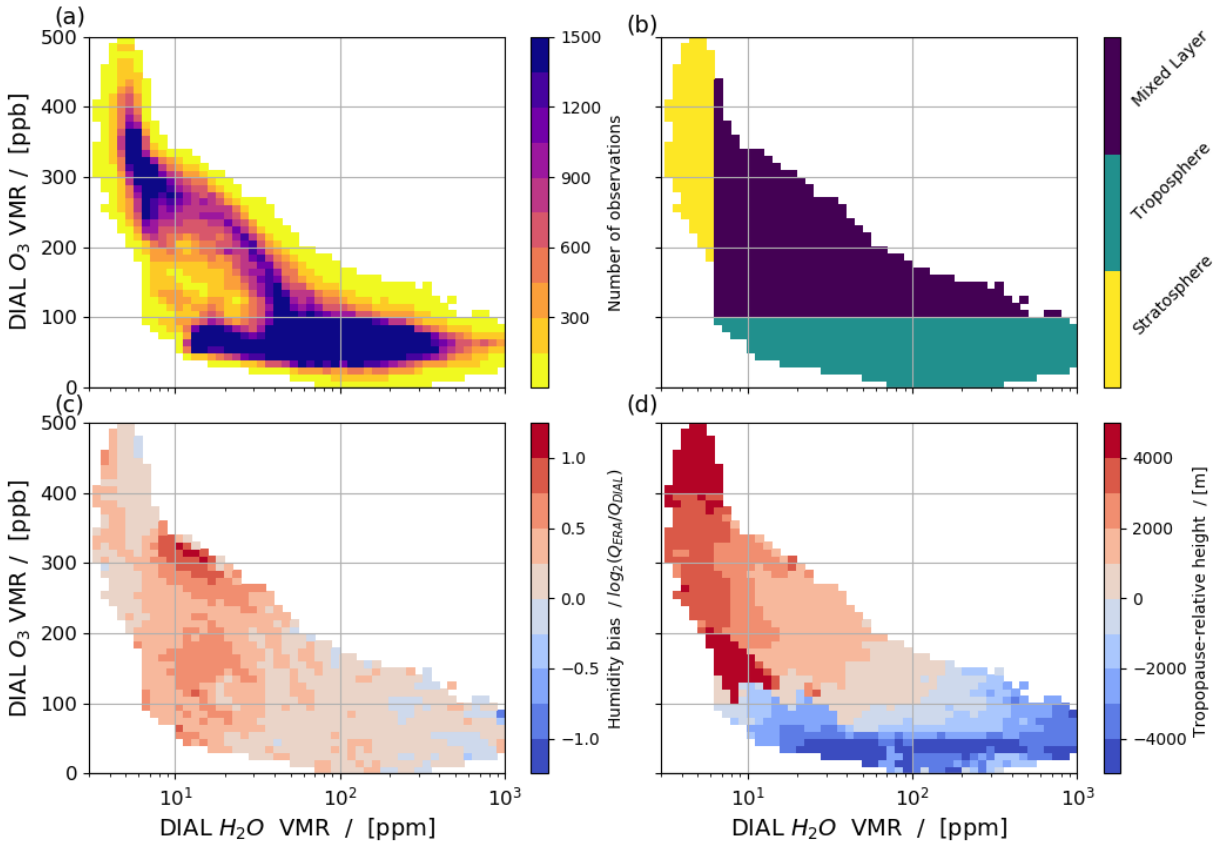

Figure 3.16: Binned distribution of water vapour and ozone observations in T–T space for four WISE flights coloured by bin-average (a) number of collocated DIAL observations, (b) type of classified air mass with troposphere ($VMR_{O_3} < 100$ ppb and $VMR_{H_2O} > 6.5$ ppm), mixed air ($VMR_{O_3} > 100$ ppb and $VMR_{H_2O} > 6.5$ ppm), and stratosphere (> 100 ppb VMR_{O_3} and < 6.5 ppm VMR_{H_2O}), (c) humidity bias and (d) tropopause-relative altitude. Bin sizes are 10 ppb for VMR_{O_3} and 0.05 ppm for $\log_{10}(VMR_{H_2O})$.

For each bin in T–T space, the average humidity bias and the mean tropopause-relative altitude are displayed in Figure 3.16. The humidity bias is weak for both tropospheric and stratospheric air (Figure 3.16c and Figure 3.16c), ranging mostly between -0.25 and 0.25. In the mixed air class, the humidity bias is most pronounced (>0.25), particularly where the VMR_{H_2O} is below 40 ppm. In the tropospheric and stratospheric air, a stronger positive/negative bias is indicated for lower/higher VMR_{H_2O}, which is associated with the sharper humidity distribution in ERA5. Figure 3.16d displays the tropopause-relative height, which is the vertical distance to the calculated ERA5 tropopause, for each bin. Across the mixed air class, an increase of the tropopause-relative altitude is visible, corresponding to a decrease of VMR_{H_2O} and to an increase of VMR_{O_3}. At low VMR_{H_2O} (<10 ppm) and low VMR_{O_3} (100–200 ppb), the transition of tropopause-relative altitudes is more abrupt, which is related to variability of the tropopause altitude across the jet stream, for example, as visible in the uppermost part of Figure 3.15b. When comparing the tropopause-relative height with the distribution of the bias (Figure 3.16), it is noticeable that the average bias is particularly increased between 1 and 3 km, where it ranges from 0.5 (40 %) up to 1.25 (137 %). In contrast, the mean bias is weak beyond 3 km above the tropopause in the LS.

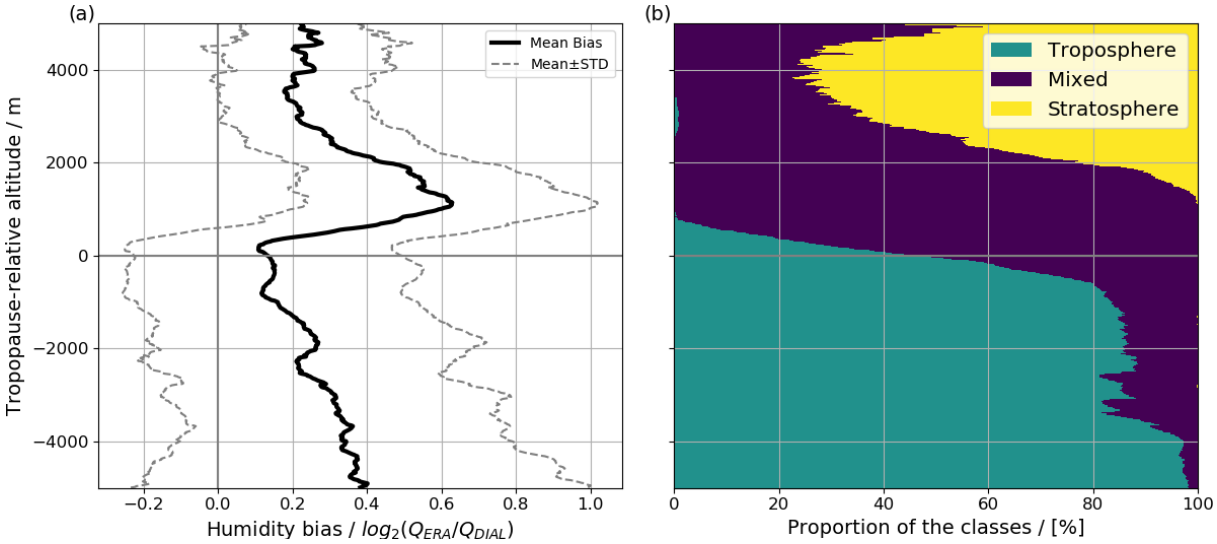

Figure 3.17: (a) Tropopause-relative profile of the mean (thick black line) and standard deviation (dotted thin grey lines) of the humidity bias and (b) relative proportion of the individual air mass classes for four WISE flights.

The average vertical profile of the moist bias for the WISE flights (Figure 3.17a) is similar to the full data set (Figure 3.10b) at the tropopause and in the LS, i.e. a local minimum is found at the tropopause (0.1; 7 %) and a pronounced maximum of 0.62 (54 %) peaking at about 1 km above the tropopause. The tropospheric part of the profile, however, is almost constant in the full data set (0.2–0.25) but decreases with increasing altitude in the WISE data (0.4–0.1). Figure 3.17b shows the relative proportion of the individual air masses at a given tropopause-relative altitude and thus provides information about the connection between the vertical structure of the moist bias and the air mass classes. In the entire UT, the tropospheric air provides the largest contribution of more

than 80 % up to 500 m below the tropopause. Across the tropopause, the proportion of tropospheric air rapidly decreases with altitude in accordance with a rapid growth of the fraction of mixed air. This is accompanied by an increase of the moist bias, and the altitude of the largest bias (1–2 km above the tropopause) coincides with the maximum relative contribution of the mixed air class (> 90 %). Above, the relative fraction of stratospheric air grows, while the moist bias reduces and reaches constant values (0.2) at approx. 4 km above the tropopause with a 65–85 % share of stratospheric air. Note that contributions of mixed air below the tropopause and at altitudes > 4 km above the tropopause may be related to falsely detected tropopause altitudes (see discussion in Sect. 3.6) or situations of complex tropopause structure (e.g. as shown in the second part of Figure 3.15b).

3.5 Moist bias in ERA5 and IFS analyses for a convective case

The correlation between the vertical structure of the LS moist bias and the mixed air layer is a strong indication that mixing processes in the IFS contribute to the formation of the moist bias. Multiple mixing processes influence the composition of this mixed air layer such as downwelling within the Brewer-Dobson circulation, turbulent transport near the jet or mixing related to deep convection (Schäfler et al., 2021, 2023). However, these processes take place on different spatial and temporal scales, and the methods used in Sect. 3.4 do not allow to identify particular of these mixing processes in order to assess their role in the formation of the LS moist bias. Tropopause-penetrating convection (also termed "overshooting convection") is known to inject large volumes of water vapour into the LS (e.g., Dessler and Sherwood, 2004, Homeyer et al., 2014, Smith et al., 2017). As a sub-grid scale process convection is parameterized in a NWP model. Overestimated overshooting convection is suggested to be a potential source of the LS moist bias (Kunz et al., 2014). This hypothesis may be affirmed by the finding of maximum of the bias in summer (Sect. 3.3) where overshooting convection is observed more frequently in the mid-latitudes.

The LS humidity bias has attracted increased attention at ECMWF in recent years, and major revisions to the convection scheme have been made in recent IFS cycles with the aim of reducing the amount of moisture entering the (lower) stratosphere (Bechtold and Forbes, 2024). Specifically, to limit overestimated convective overshoot, a strengthening of static stability near tropopause has been implemented in a recent IFS update (cycle: Cy47r3). Further details about changes in the IFS are given at: `https://confluence.ecmwf.int/display/FCST/Changes+to+the+forecasting+system`. It is to be expected that the convection updates affect the water vapour distributions in the LS and thus potentially also the moist bias.

In this dissertation, the output of two recent IFS versions with a different convection parameterization are made available, in particular the IFS cycle Cy47r1 (in operation in June 2020) and a pre-operational version Cy47r3 (operationally released in October 2021), the latter including the limitation of overshooting convection. A comparison of both model versions could provide information about the role of convection on the water vapour budget of the LS. A research flight (8 July) of the CIRRUS-HL summer campaign offers an excellent opportunity to investigate this aspect. The synoptic situation over Europe on this day created favorable conditions for strong convection in a large area over northern Italy and the Alps (Schäfler and Rautenhaus, 2023). In this area, the formation of cirrus clouds near the tropopause in the outflow of the convective systems was expected that were the intended target of this research flight. A detailed description of the underlying synoptic situation is provided by Schäfler and Rautenhaus (2023).

Figure 3.18a shows the flight section on the 8 July, where water vapour DIAL measurements were carried out. Between about 15-16 UTC, HALO flew from Oberpfaffenhofen (EDMO) over southern Germany, Switzerland and northern Italy, covering a distance of about 1000 kilometres. The 15 UTC SEVIRI IR satellite at this daytime shows several isolated cloud formations over the Alps (near the flight track) that are associated with convective systems and the accompanying cirrus anvils Figure 3.18b

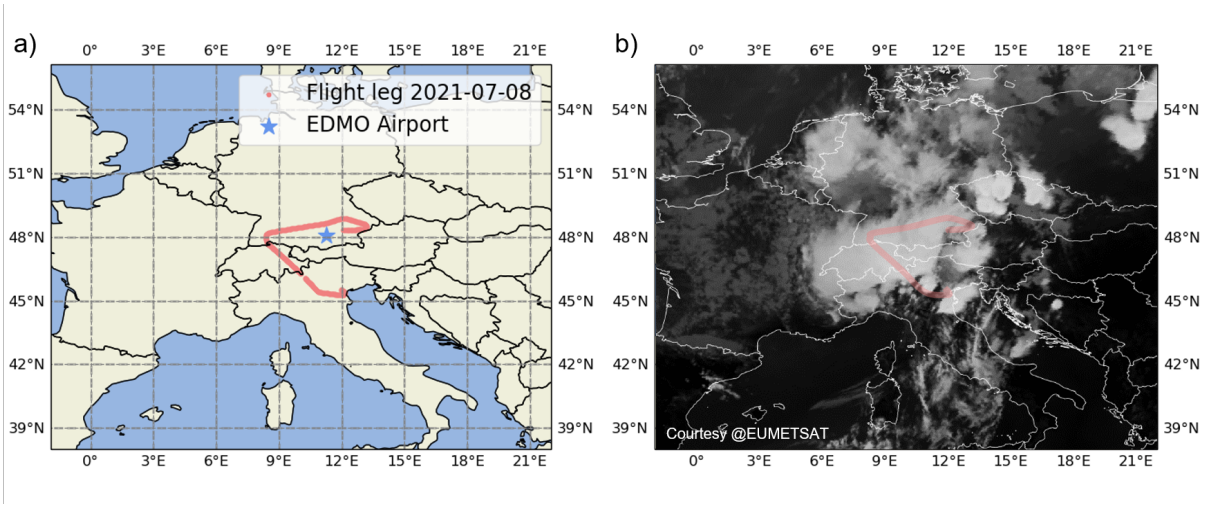

Figure 3.18: a) Flight track (red) with DIAL observations of the CIRRUS-HL research flight on the 08 July 2021 and b) SEVIRI IR 10.8 μm at 15 UTC with flight track (red shading). Bright colors in b) indicate the cold anvils of the high-level cirrus clouds.

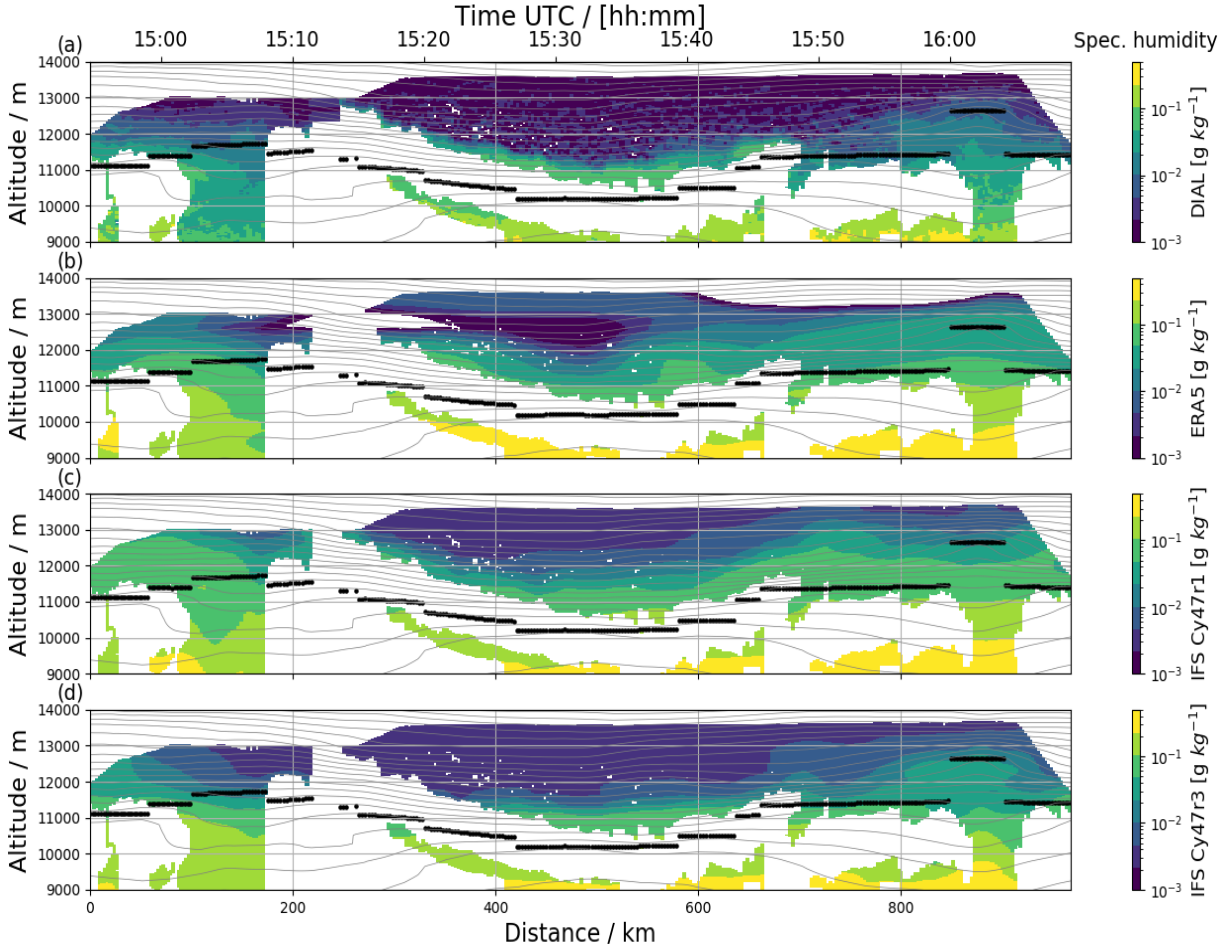

Figure 3.19: Cross sections of specific humidity (colour shading) of a) the DIAL observations, b) ERA5, c) IFS$_{Cy47r1}$, d) preliminary version of IFS$_{Cy47r3}$ on the 8 July 2021. Panels (a)–(d) are superimposed by ERA5 fields of the potential temperature (grey contours, $\Delta\theta = 3\,\mathrm{K}$), and the ERA5-derived thermal tropopause (thick black dots).

First, water vapour cross sections of the DIAL observations and the model output is presented in Figure 3.19. In addition to the IFS output (Cy47r1 and Cy47r3), a cross-section for ERA5 is also presented in order to provide comparability with the findings from Sect. 3.2 and Sect. 3.3. It has to be noted, that the output of Cy47r1 and Cy47r3 (although termed "analyses") is a combination of 6-hourly analysis data, and hourly forecasts.

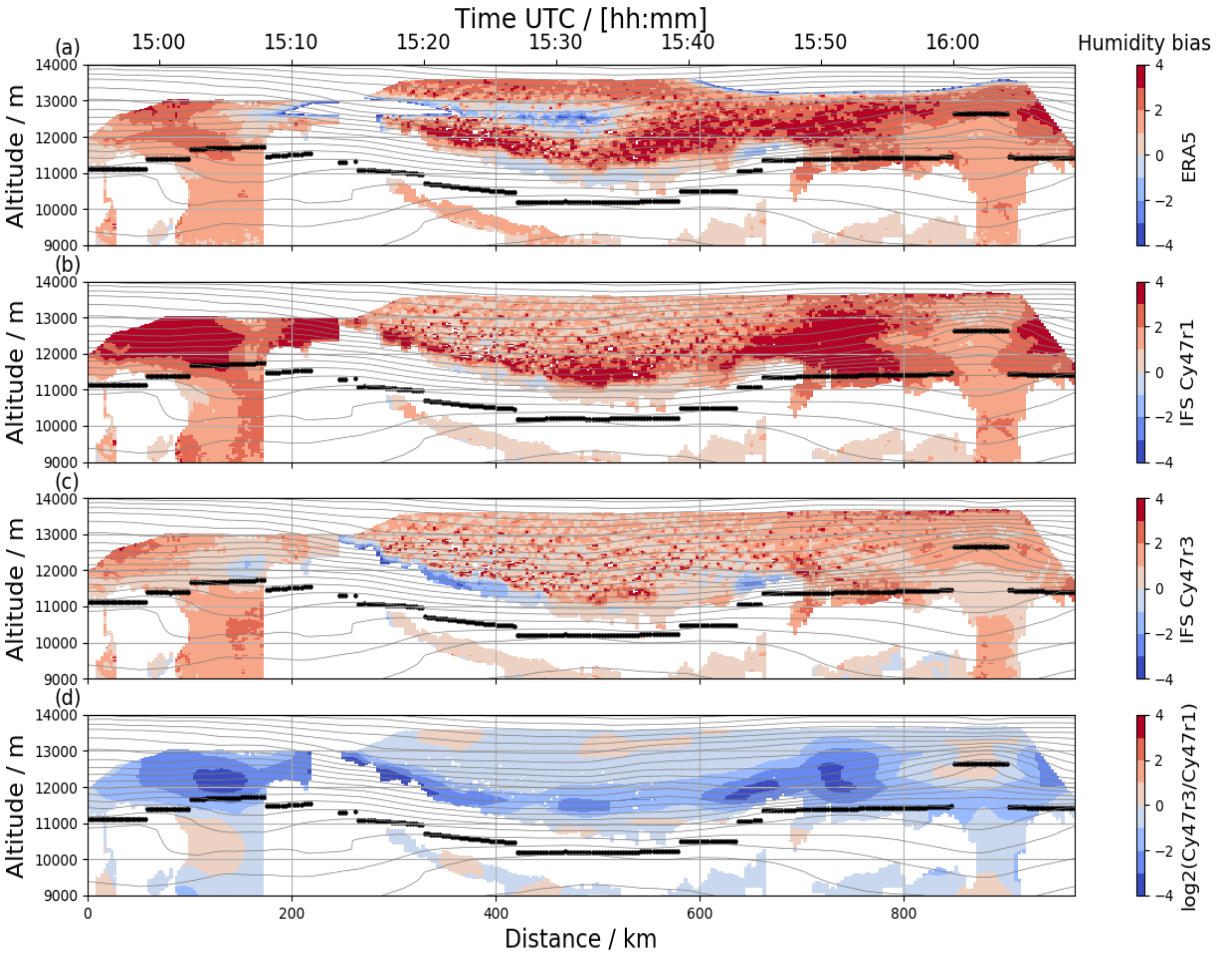

Figure 3.20: Cross sections of the humidity bias (colour shading) of a) ERA5, b) IFS_{Cy47r1}, c) pre-operational version of IFS_{Cy47r3} and d) humidity bias of the different IFS cycles calculated as: $log_2(\mathrm{IFS}_{Cy47r1}/\mathrm{IFS}_{Cy47r1})$ on the 8 July 2021. Panels (a)–(d) are superimposed by ERA5 fields of the potential temperature (grey contours, $\Delta\theta = 3\,\mathrm{K}$), and the ERA5-derived thermal tropopause (thick black dots).

Along the entire leg, HALO flew above 14 km which exceeded the ERA5-derived tropopause altitude by 2-3 km, and consequently, the LS is well-covered by DIAL measurements. The data gaps (white area near tropopause) in the section are related to clouds reflecting the strong convective activity in the region (Figure 3.18b). Each, DIAL and all model cycles show a generally similar distribution of water vapour with largest humidity values ($> 10\,\mathrm{g\,kg^{-1}}$) at lowest altitudes, and a lowest values ($< 10^{-3}\,\mathrm{g\,kg^{-1}}$) at highest altitudes in the LS. It is noted that the moisture distribution varies slightly between the different model runs. Especially above the tropopause (e.g. between 400-800 km

distance) the colouring indicates higher specific humidity in ERA5 and Cy47r1, compared to Cy47r3.

To elaborate on these differences, Figure 3.20a-c shows the logarithmic humidity bias, for ERA5, as well as the analysis Cy47r1 and Cy47r3. In the following, the focus is exclusively on the LS. All model cross sections show positive values of the bias in the LS, which extends along a coherent layer (about 1-2 km above the tropopause) over the entire flight path. ERA5 and Cy47r1 show an enhanced bias values (>2), compared to the lower bias values in Cy47r3 (1-2). In order to compare the different IFS analysis cycles, the logarithmic bias is also derived between Cy47r1 and Cy47r3 as illustrated in Figure 3.20d. Negative values along the flight path in a layer about 1-2 km thick above the tropopause indicate a lower specific humidity in Cy47r3 compared to Cy47r1.

An analysis of this single flight in ERA-based tropopause-relative coordinates is conducted to quantify these humidity differences. Figure 3.21a shows the mean moisture profiles. Both the observations and the modeled moisture profiles show a decrease in specific humidity in the UTLS by three orders of magnitude, with the observations being drier than the model profiles at each altitude. The average profile of the humidity bias is given in Figure 3.21b. All profiles show a mean positive bias (>1) in the entire UTLS, which corresponds to an overestimation of 100 %. In the LS, Cy47r1 and ERA5 have a comparatively stronger bias with values of up to about 3. In contrast, Cy47r3 does not show a pronounced maximum in the LS, and substantially lower values of the bias (about 1), particularly between 1-2 km above the tropopause in the layer. In summary, this case study has shown that ERA5, Cy47r1 and Cy47r3 produce different water vapour concentrations in the LS in an area where enhanced convection was observed. In particular, the IFS version Cy47r3 that includes a limitation of overshooting convection exhibits the lowest specific humidity and the lowest humidity bias in the LS. These findings are discussed in detail at the end of the next chapter.

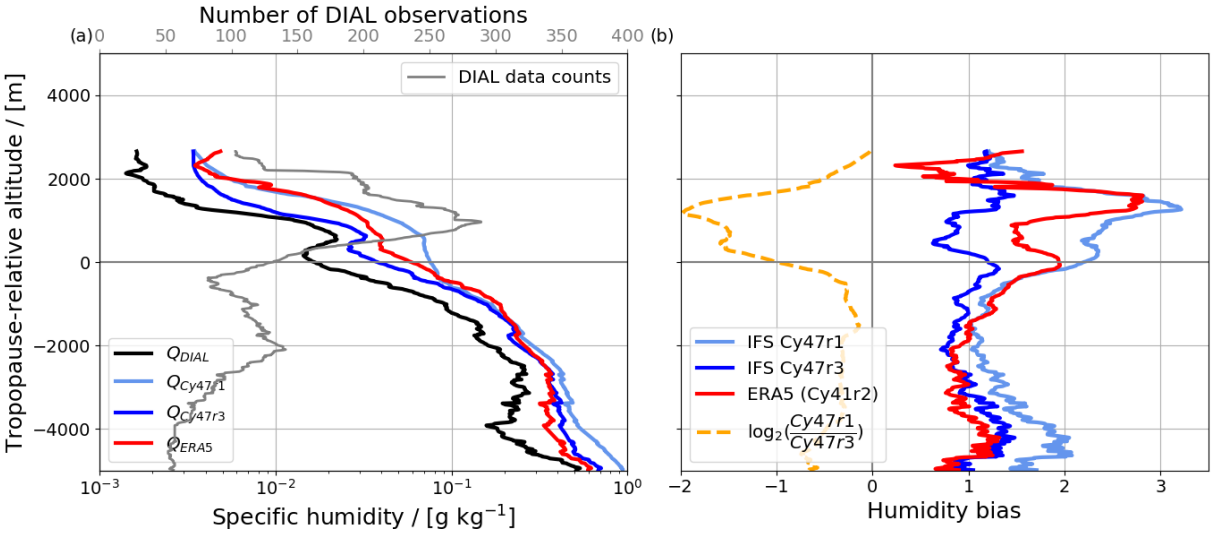

Figure 3.21: Tropopause-relative vertical profiles of (a) the DIAL (black), ERA5 (red), IFS$_{Cy47r1}$ (light blue), IFS$_{Cy47r3}$ (dark blue) and number of DIAL observations (gray, note the top x-axis). (b) shows the log-humidity bias for the individual model cycles (same colouring as a), and the humidity difference log2(IFS$_{Cy47r1}$/IFS$_{Cy47r1}$) (orange dashes).

3.6 Discussion and concluding remarks

A lower-stratospheric moist bias is a major issue that occurs in various global NWP models (Davis et al., 2017) that has attracted increasing interest in recent years (e.g., Kunz et al., 2014, Dyroff et al., 2015, Davis et al., 2017, Kaufmann et al., 2018, Bland et al., 2021). However, there are still open questions regarding the bias, especially with regard to its vertical structure and potential sources of error.

An unprecedented and independent multi-campaign water vapour lidar data set comprising 41 research flights (six campaigns) and roughly 33,000 vertical profiles obtained in the northern hemispheric mid-latitudes during different seasons is used to evaluate the expected moist bias in the ERA5 reanalysis. The observations from the surface up to the LS cover four orders of magnitude and represent typical mid-latitude data for the individual seasons (e.g., Pan et al. (2000); Randel and Wu (2010); Zahn et al. (2014); Kunz et al. (2014); Dyroff et al. (2015); Bland et al. (2021)). The high data availability around the tropopause makes the data set suitable for an evaluation of NWP fields in the UTLS. Although the number of observations reduces considerably towards the highest altitudes (up to 5 km above the tropopause), the data set provides a valuable extension to previous humidity data sets which exhibit increased measurement uncertainties at altitudes larger than 2 km above the tropopause (e.g., Bland et al., 2021).

In the troposphere strong positive and negative biases of small spatial extent are found, which are likely related to insufficiently represented tropospheric transport processes, to model errors of tropospheric processes (e.g. clouds), or to the linear interpolation scheme that may have caused increased differences especially in situations of strong horizontal or vertical moisture gradients. The small positive and vertically almost constant mean bias in the UT, which ranges between 0.2 (15 %) and 0.26 (20 %), confirms earlier findings (Dyroff et al., 2015, Bland et al., 2021). It has to be noted that the UT humidity bias is limited to cloud-free scenes, as DIAL humidity profile observations cannot be retrieved inside or below optical thick clouds. In agreement with Bland et al. (2021), a local minimum of the bias (< 0.2, < 15 %) is found at the tropopause. Above the tropopause, the findings verify a coherent layer of overestimated humidity in ERA5 reanalyses. The magnitude of the maximum bias of 0.63 (55 %) and its altitude of 1.3 km above the tropopause are comparable to previous findings for earlier model cycles of the IFS (Dyroff et al., 2015, Kaufmann et al., 2018, Woiwode et al., 2020, Bland et al., 2021), earlier reanalysis versions (Oikonomou and O'Neill, 2006, Kunz et al., 2014), and other evaluated models (Davis et al., 2017, Bland et al., 2021). Above the maximum bias, in a region where recent studies present diverging results (Dyroff et al., 2015, Woiwode et al., 2020, Bland et al., 2021), the results presented in this analysis reveal a steadily decreasing (with altitude) moist bias that reduces to nearly constant and small positive values comparable to the UT. The high data quality of the DIAL measurements allow a reliable and robust depiction of the bias at the highest altitudes of the UTLS. Furthermore, the magnitude of the LS moist bias exceeds the expected error of the DIAL humidity observations by approx. one order of magnitude, which underlines the significance of our results.

Bland et al. (2021) show that tropopause altitudes are on average about 200 m higher when derived from ECMWF IFS profiles compared to radiosondes, which may impact tropopause-relative moisture distributions and in turn the bias. As no temperature observations are available, this study relies only on ERA5 tropopause altitudes. In a sensitivity analysis, the ERA5 tropopause was artificially lowered by 200 m for each observed and modeled profile. As a result, the mean humidity and bias profiles also shift vertically accordingly, but their values remain the same. Applying a systematic tropopause altitude reduction only on the ERA5 humidity profiles would reduce the tropospheric bias, however, the LS moist bias, although slightly weakened, would persist.

In line with findings of (Bland et al., 2021), who indicated little sensitivity of the moist bias to various atmospheric conditions but revealed a different depth of the moist bias for trough and ridge situations, low tropopause situations (which are typically associated with troughs) exhibit a maximum bias at higher altitudes and a deeper layer of the increased bias compared to high tropopause situations. The magnitude of the moist bias is found to be independent of the tropopause altitude. In addition, a pronounced LS moist bias is detected in the summer (>1.20, $>130\,\%$) which exceeds the diagnosed autumn bias by a factor of 2–3. So far, such a seasonality has only been suggested in (Dyroff et al., 2015). Additional DIAL observations in spring, summer, and winter would be valuable for a more comprehensive study of the seasonality of the vertical bias structure.

For four flights during the WISE campaign, an air mass classification using collocated water vapour and ozone profile data (Schäfler et al., 2021) was applied to separate tropospheric (low ozone and large water vapour mixing ratio), stratospheric (large ozone and low water vapour), and mixed air (intermediate ozone and water vapour). In tropopause-relative coordinates, the vertical structure of the moist bias for the selected cases turned out to be comparable to the multi-campaign LS moist bias, so that these flights are considered to be representative of autumn. The moist bias is found to be increased in the mixed air class representing the ExTL and that the maximum is reached at the altitude where the proportion of mixed air is highest (near 100 %). The decrease of the moist bias above/below is accompanied by a growth of the proportion of stratospheric/tropospheric air. The high correlation in the distribution of the moist bias and the ExTL gives a strong hint at the importance of moisture injection into the LS, either due to numerical diffusion across the tropopause or due to insufficiently modeled transport and mixing processes. As the bias in the ExTL is increased in each of the evaluated WISE flights, we consider systematic uncertainties in the representation of mixing processes to play a key role for the LS moist bias. This is supported by the finding of a deeper bias layer above troughs which are characterised by a thicker ExTL above (Hoor et al., 2002, Pan et al., 2007). In addition, the maximum bias occurs in summer when cross-tropopause mixing is strongest (Hoor et al., 2002), and the bias is reduced in stratospheric background humidity at highest altitudes, which are not directly influenced by mixing processes at the tropopause.

Schäfler et al. (2023) investigate the Lagrangian history of the observed air for the presented WISE case study on 1 October 2017 and find that the ExTL air experienced strong turbulent mixing in the jet stream during 48 h before the observation. They also find that the mixed air (in which the increased bias is identified) shows highly variable origins and transport pathways related to tropospheric weather systems which may be indicative of

the relevance of different mixing processes. Additional collocated observations of ozone and water vapour in different seasons, near active mixing process (e.g. convection), or in the Southern Hemisphere where exchange at the polar jet stream is reduced (e.g., Bowman, 1996) could provide valuable information about the relevance of individual mixing processes and their role in forming the moist bias. The presented results suggest that improving the representation of mixing processes at the tropopause should act to reduce the LS humidity bias.

One of these mixing processes, which are known for the direct tropospheric- stratospheric mass exchange, is overshooting convection (Dessler and Sherwood, 2004, Smith et al., 2017, Khordakova et al., 2021). In a separate analysis, the hypothesis is investigated as to whether the representation of overshooting convection in the IFS influences the LS humidity distribution and thus also the LS moist bias. For this purpose, a research flight during the summer field campaign CIRRUS-HL is selected, providing high data coverage in an area with enhanced convection. In this dissertation, the exceptional opportunity was given to compare the observations with two IFS cycles (Cy47r1, pre-operational Cy47r3) that treat overshooting convection differently. In particular, the Cy47r3 contains an upgraded convection scheme that limits convective "overshoot" (Bechtold and Forbes, 2024).

The humidity cross sections reveal differences in humidity between all IFS versions and observations across the tropopause, but confirm a moist bias in the LS to be present in each cycle. In the tropopause-relative evaluation, all IFS versions show a positive bias in the entire UTLS. ERA5 and Cy47r1 show a particularly high moist bias (2-3) in a 2 km thick-layer above the tropopause. This bias is comparatively higher than the moist bias diagnosed for the entire measurement campaign, which may be related to the low observed humidity values. Cy47r3 does not show a noticeable maximum in the bias profile and a substantial reduction of moisture (and in turn the bias) compared to the Cy47r1 and ERA5.

The finding, that the lowest moist bias is detected for the cycle that limits overshooting convection, is a strong indicator that the upgraded convection scheme may be responsible for the reduced humidity (reduced bias, respectively) in the LS. However, convection is not the only process for water vapour transport from the UT into the LS. In order to verify that convection processes have led to the injection of moisture (and to what extend) into the LS, Lagrange trajectory studies could be helpful to provide information about the origin and transport of the air masses. In addition, this analysis only shows a limited amount of data (only one case study), so a more robust statistics is needed to validate this result. It should be also clarified whether the local input of moisture into the LS through overshooting convection has consequences for the global water vapour budget of the LS. This could be achieved by statistically comparing the lower-stratospheric water vapour content of both model cycles over larger time periods (seasonal means) or spatial scales (zonal means).

The moist bias is a persistent model error relevant to climatological studies for which reanalysis data are widely used as input, and Davis et al. (2017) indicate that various reanalyses overestimate humidity in the extratropical LS. Ploeger et al. (2024) show that LS humidity errors affect atmospheric circulation patterns. Repetition of the study with the upcoming ERA6 reanalysis (that will include an updated convection scheme compared to ERA5) is desirable.

Chapter 4

Analysis of radiosondes influence on the tropopause

The overall goal of the analysis presented in this chapter is to explore the unknown influence of data assimilation on the sharp vertical gradients of the tropopause in the ECMWF IFS. This evaluation is based on radiosonde data assimilation output retrieved from ECMWF's observational data base (ODB, Hersbach et al. (2015)). In general, ODB files contain relevant information on an observation that is used during the data assimilation process such as the observed meteorological quantities, and their departures from first-guess and analysis at the position and time of the observations (i.e., in observation-space; see Figure 2.6). For this study, two radiosonde data sets are retrieved covering the period and research area of the NAWDEX field campaign (Figure 4.1). The first radiosonde data set comprises almost 10000 observations that were operationally assimilated by the IFS and serve for the statistical evaluation of the data assimilation influence on tropopause sharpness and altitude (RQ2.1). For 500 of these radiosondes, which were launched non-operationally on-demand during NAWDEX, a special OSE was carried out at the IFS (details on the OSE are given in Sect. 4.1.1) allowing to attribute the data assimilation influence directly to the radiosondes (RQ2.2). In addition, a model-space case study is carried out to assess the influence of data assimilation on the temperature and wind gradients across the tropopause in a synoptic context (RQ2.3).

4.1 Data and methodology

4.1.1 Radiosonde data sets and Observing System Experiment

Radiosondes are an elementary source of meteorological data for monitoring the weather and climate (Madonna et al., 2020). They measure prognostic quantities (i.e., temperature, pressure, wind components, and tropospheric humidity) with high precision, which can be directly assimilated by NWP models. Although radiosondes only account for a small proportion (about 2 %) of the total assimilated meteorological information, they contribute to a 5 % reduction in 24 h forecast error in the ECMWF IFS in a statistical sense, and thus have a higher impact per observation than e.g., IR radiances (Pauley and Ingleby, 2022). Radiosondes furthermore serve as anchor observations to correct bi-

ases in other assimilated observations (satellite data) in a process called variational bias correction, which highlights their important role in the data assimilation system (e.g., Carminati et al., 2019, Francis et al., 2023). Since radiosondes generally ascend to an altitude of 40 km at a rate of five metres per one-second measurement cycle, they provide high-resolution vertical profiles capturing the strong gradients at the tropopause.

Radiosonde data sets in the NAWDEX period

This study is based on about 9200 radiosonde profiles (Figure 4.1a) that were routinely measured at 581 sites covering a wide area between North America and Europe from the subtropics to high latitudes (30°–85°N, 95°W–30°E) during the one-month NAWDEX period (17 September to 18 October 2016). The majority of these observations (96 %) were carried out at 200 land-based stations while the minor share (4 %) were ship-based observations at 381 variable positions across the North Atlantic. In addition to the routine profile observations, about 500 extra radiosondes were launched in the course of the NAWDEX field campaign (Schäfler et al., 2018) (see Figure 4.1b). Over Europe, the extra radiosondes were released on-demand in a variety of synoptic situations, for instance instance, in diabatically active warm conveyor belt flows associated with cyclones or in upper-level ridges associated with blocking situations. Six stations over Canada, upstream of the NAWDEX operation region, released two additional radiosondes per day.

In addition to the radiosonde observations, more than 700 dropsondes were released from research aircraft during the NAWDEX period (mostly in the subtropical and tropical western North Atlantic Schindler et al. (2020)). Due to the limited flight altitude of the aircraft, dropsondes poorly cover the tropopause and the LS. This dissertation therefore focuses on the evaluation of the radiosondes. All radiosonde and dropsonde profiles were made available for operational assimilation at weather centers (Schäfler et al., 2018).

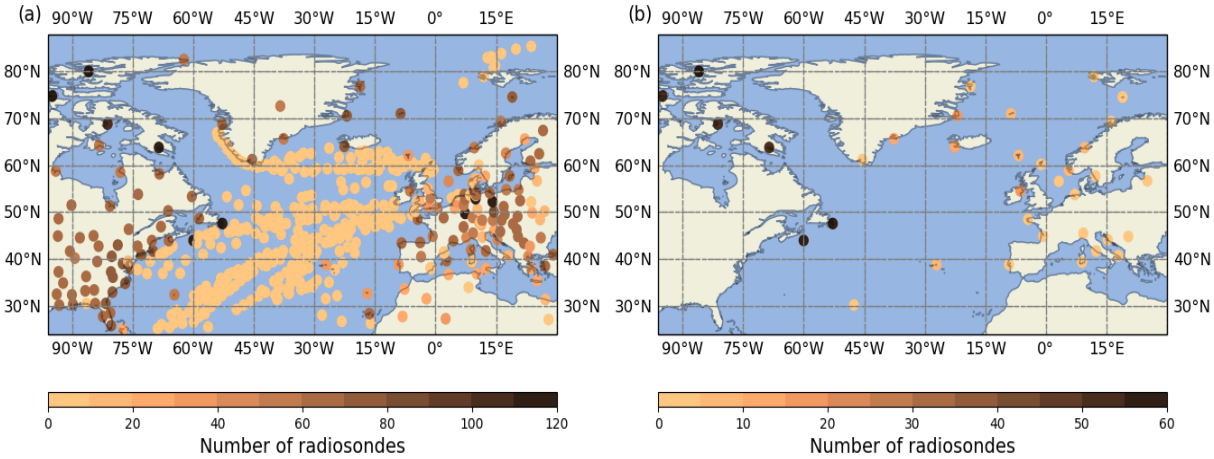

Figure 4.1: Positions of radiosondes assimilated by the ECMWF IFS between 17 September and 18 October 2016 for (a) all radiosondes (9729), and (b) the subset of 497 non-operational radiosondes launched during NAWDEX. The colouring denotes the number of assimilated profiles at a particular site. The scale of the colourbar changes between (a) and (b).

It has to be noted, that radiosonde profiles are not assimilated by the IFS at their fully measured vertical resolution but at a reduced number of levels (roughly 50–350), which

depends on the reporting type (alphanumeric, BUFR, etc.) the individual stations used for the data transmission to the Global Telecommunications System (GTS) (Ingleby and Edwards, 2015, Ingleby et al., 2016). About 65 % of the assimilated radiosonde profiles used in this work have a low vertical resolution (< 100 data points per profile), while 35 % of the radiosondes exhibit up to roughly 400 levels (Figure 4.2a). Accordingly, the distribution of the average vertical distance of neighboring data points in the UTLS (in this study defined as ± 3 km around the tropopause) shows a bi-modal shape and has maxima at approx. 100 m and 400 m (Figure 4.2b).

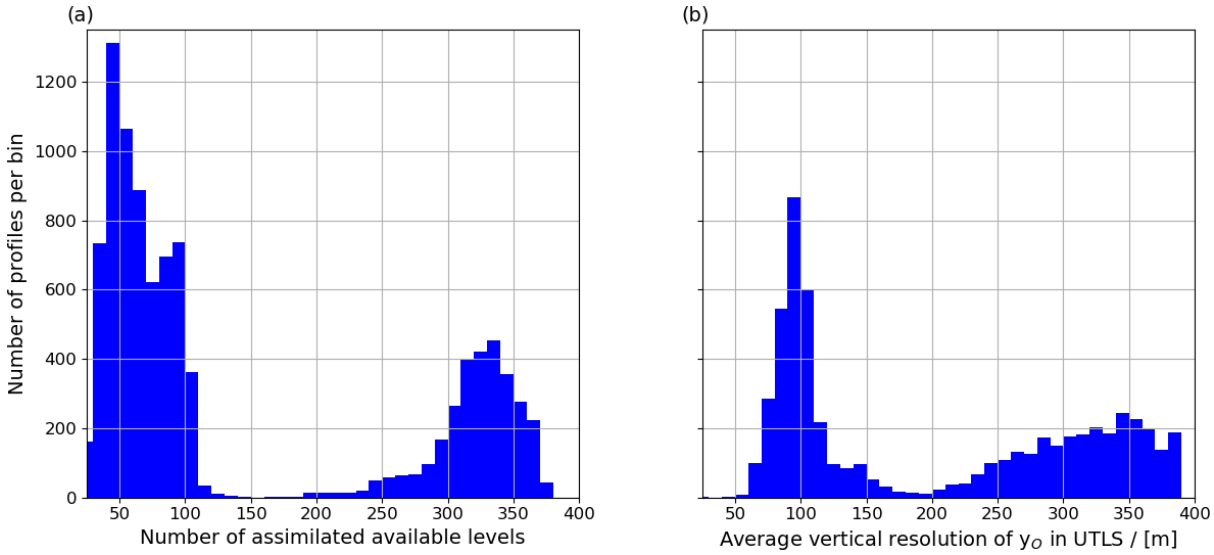

Figure 4.2: Number of radiosondes per for (a) number of assimilated vertical levels and (b) the given average vertical resolution in the UTLS.

Observing System Experiment set up

To investigate the influence of the extra radiosonde observations during NAWDEX a dedicated OSE was performed with the IFS (Schindler et al., 2020). The cycled OSE covers the whole NAWDEX campaign period (17 September to 18 October) and uses IFS model cycle 43r1 (Cy43r1; ECMWF (2016b)), which became operational in November 2016. The triangular-cubic-octahedral grid (TCo1279) used in Cy43r1 provides a horizontal resolution of about 9 km and 137 vertical sigma-hybrid model levels that range from the surface up to approx. 80 km. The vertical resolution of the model levels is highest in the planetary boundary layer (few metres between model levels) decreases with altitude. At typical mid-latitude tropopause altitudes (6-15 km) the vertical grid spacing is about 300 m (e.g., Schäfler et al., 2020). The incremental hybrid 4D-VAR data assimilation scheme used at ECMWF assimilates observations available in a 12 h time window to update a prior short range forecast in order to determine the analysis (see Sect. 2.2). More details about the implementation of 4D-VAR in the IFS are given in Rabier et al. (2000) or in the IFS documentation (ECMWF, 2016a). As in the operational ECMWF system, the B-matrix for the experiments is based on a blended combination of a climatological estimate and an estimate from an EDA. The cycled OSE comprises two separate model runs. The control run (CTR) considered all routine and extra radiosondes as well as the

dropsondes launched during NAWDEX. The denial (DEN) run excluded and passively monitored all extra NAWDEX observations in a region over the North Atlantic (25°–90°N; 82°W–30°E). In addition, a 25 member EDA experiment was conducted at lower horizontal resolution (TCo639 about 18 km) for both experiments. More details on the OSE design are given in Schindler et al. (2020).

4.1.2 Processing of the radiosonde data assimilation output

To prepare the statistical analysis of the data sets, several important processing steps must be applied on the raw ODB data (illustrated in Figure 4.3). These include 1) radiosonde profile selection as some of the radiosondes have not reached the tropopause altitude, and 2) the determination of the geometric height from the pressure coordinates provided by the feedback files. Then, the (thermal) tropopause altitudes are derived separately for background, observation and analysis to conduct a tropopause-based evaluation similar to Sect. 3.

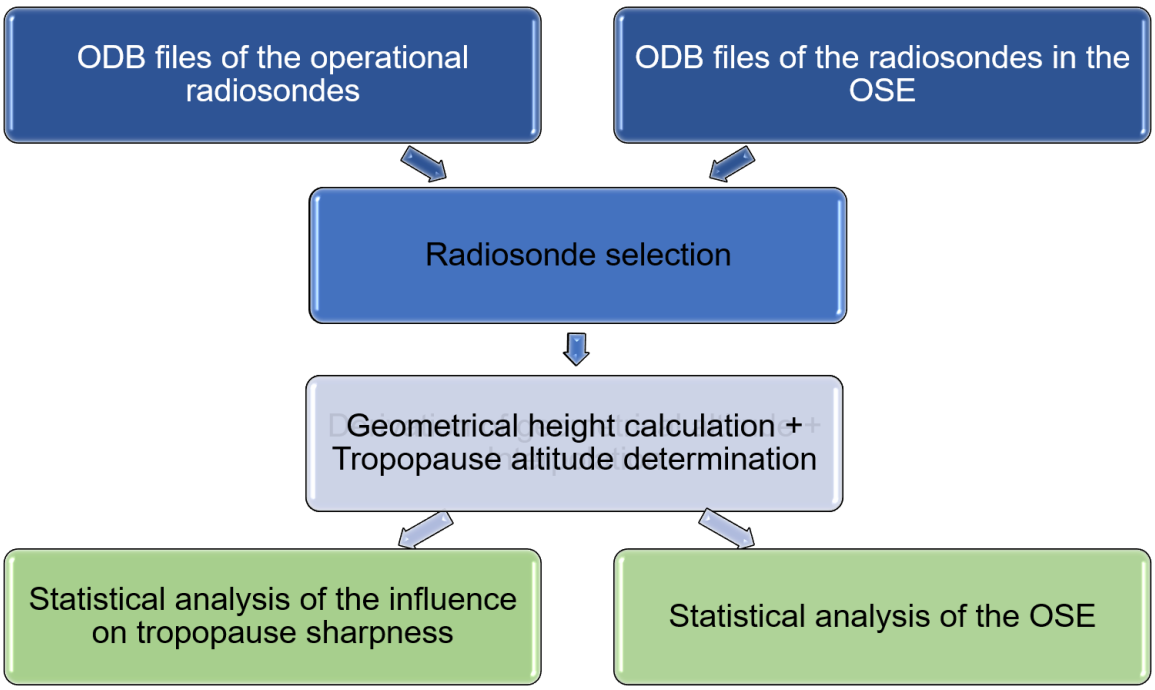

Figure 4.3: Data processing scheme of the ODB files.

Profile selection and conversion from pressure to altitude levels

The observation-space feedback files of the operational radiosondes (9200 sondes) and the passive radio are retrieved from the ECMWF's ODB. These files contain the temperature, wind and humidity data from the assimilated radiosonde observations (y_O). For each variable, the ODB-files also provide departures from the background and analysis state. All data is given as profile data using pressure as vertical coordinate. In the following the observation-space model equivalents i.e., analysis and background are termed y_{FG} and y_{AN}, respectively. From the requested files, only those profiles are considered that are actively assimilated. Some profiles (about 1 %) which do not provide temperature

and wind data above the 540 hPa level (approx. 5 km) are excluded from the statistical analysis. This level serves as a starting point for tropopause detection which is explained later in this section.

The pressure coordinates are converted into geometric altitude via the hydrostatic relation similar to Sect. 3.1.2 by applying Equation 3.2. The geometric height of the individual levels can be determined through the cumulative sum of the layer thicknesses for each profile (Equation 3.3). It should be noted that the lowest pressure level of the profiles in the feedback files is not exactly located at the surface. The height of the lowermost level ($z_{k=0}$-$z_{sfc,IFS}$) of the radiosondes can be estimated through the barometric formula (Equation 4.1) by assuming a US-standard atmosphere,

$$z_{k=0} - z_{sfc,IFS} = \frac{R * 0.5 * (T_{US,sfc} + T_{k=0})}{g} \frac{p_{US,sfc}}{p_{k=0}}, \tag{4.1}$$

where the surface temperature $T_{US,surface} = 288.15\,\mathrm{K}$, and the surface pressure is $p_{US,surface} = 101325\,\mathrm{Pa}$ are taken from the US standard atmosphere. The surface altitude $z_{sfc,IFS}$ is provided by the feedback files.

Temperature and wind measurements in the profile can be available at different vertical pressure levels. The described derivation of the height coordinates is therefore carried out separately for wind and temperature. The temperature and wind profiles of the observations, the background and the analysis, are linearly interpolated in the vertical onto a uniform, equidistant 10 m vertical grid. The potential temperature (Θ) and the squared static stability N^2 are then computed from the temperature profile through

$$\theta = T \cdot \left(\frac{p_0}{p}\right)^{\frac{R_L}{c_p}} \tag{4.2}$$

$$N^2 = \left(-\frac{g}{\theta}\right) \cdot \left(-\frac{d\theta}{dz}\right) \left[\frac{1}{s^2}\right], \tag{4.3}$$

with the specific heat capacity of air at constant pressure c_p and $\frac{d\theta}{dz}$ being the vertical gradient of θ.

From the wind profile, the vertical wind shear of the horizontal wind speed (hereafter referred to as wind shear) is calculated as follows:

$$\text{wind shear} = \frac{d|\vec{u}|}{dz}, \tag{4.4}$$

where $\frac{d|\vec{u}|}{dz}$ is the vertical gradient of the magnitude of the horizontal wind vector $\vec{u}$.

It is noticed that radiosonde humidity measurements are only assimilated by the IFS within the troposphere and at an air temperature greater than -40 °C, due to high sensor uncertainties at low humidity and low temperatures as they are typical in the LS (e.g., Pauley and Ingleby, 2022). Hence, this study only considers temperature and wind data.

Thermal tropopause determination

The goal is to investigate the influence of the radiosondes on the tropopause by using tropopause-relative coordinates. Analogous to Sect. 3.1.3, the lapse-rate based (thermal) tropopause (LRT) is chosen which, by definition, points to the sharp transition of thermal

stratification from the UT to the LS (Birner et al., 2002, Tinney et al., 2022). Thermal tropopauses are determined individually for the observations (referred to as LRT_{y_O}), the background ($\mathrm{LRT}_{y_{FG}}$), and the analysis ($\mathrm{LRT}_{y_{AN}}$) by following the WMO definition outlined in Sect. 3.1.3. The tropopause-relative altitude is derived by applying Equation 3.5.

The distribution of the observed tropopause altitudes for all 9729 profiles (Figure 4.4a) has a bi-modal shape in the altitude range of 6–18 km with peaks at 11.5 and 15.5 km. The left mode represents profiles with a high frequency (75% of the profiles) of LRT altitudes at 10–14 km which is typical for the midlatitudes in autumn (Hoffmann and Spang, 2022). Its broad spectrum is related to the variability in the mid-latitude tropopause altitude in different synoptic situations, e.g., in ridges and troughs (Hoerling et al., 1991). The right mode (LRT > 14 km; 25 % of the profiles) with its smaller maximum indicates profiles in the subtropics. The LRT_{y_O} distribution for the additional NAWDEX radiosondes (Figure 4.4b) does not exhibit a corresponding second peak, due to the low number of releases at stations with a latitude below 40°N.

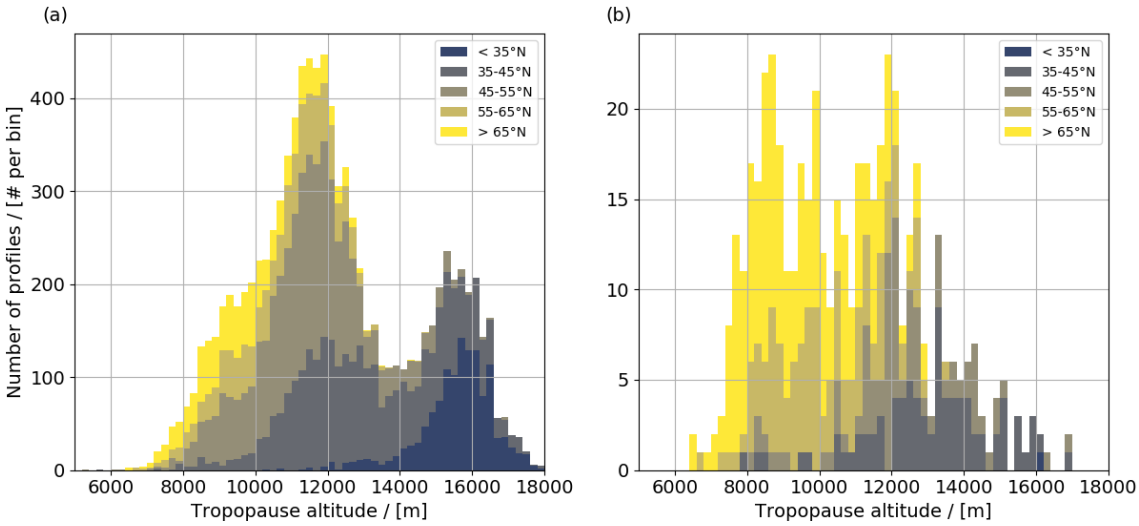

Figure 4.4: Stacked distribution of the observed thermal tropopause with 0.2 km bin size for all 9729 radiosondes. The colouring shows the latitude of the radiosonde stations (10° bins).

Tropopause-relative distributions of observed temperature and wind

Figure 4.5 presents the mean vertical profiles of observed temperature, N^2, wind speed, and wind shear profiles averaged in LRT_{y_O}-relative (with respect to the observed tropopause) coordinates. These profiles outline the main characteristics of the mid-latitude tropopause that are known from climatology (e.g., Birner et al., 2002, Grise et al., 2010, Hoffmann and Spang, 2022): above a linearly decreasing temperature in the troposphere (about $7\,\mathrm{K\,km^{-1}}$), a temperature minimum of about 213 K is reached at the LRT_{y_O}. Above the tropopause, a distinct temperature inversion (0–1.5 km above LRT_{y_O}) is observed, followed by an isothermal temperature profile in the stratosphere (up to about 5 km above the LRT). This change in stratification results in a rapid jump of N^2 (from 2 to $6.5\times 10^{-4}\,\mathrm{s^{-2}}$ at the LRT_{y_O} altitude. Wind speed continuously increases with altitude in the troposphere up to a maximum (about $23.5\,\mathrm{m\,s^{-1}}$) at roughly 1 km below LRT_{y_O}. Corresponding to the distribution of wind speed, the vertical shear of wind speed is positive up to the

wind speed maximum, then abruptly decreases beyond and reaches a distinct minimum (-5× $10^{-3}\,s^{-1}$) at about 300 m above LRT_{y_O}. Note that the presented data set of 9729 radiosondes provides a high data coverage (blue line in Figure 4.5a) in the UTLS.

A separate analysis of extratropical (LRT < 14 km) and subtropical (LRT > 14 km) observations reveals similar shapes of tropopause-relative distributions for the extratropical and the overall data. However, the subtropical mean profiles exhibit lower temperatures in the entire UTLS, a weaker temperature inversion in the LS, and no wind maximum located near the tropopause.

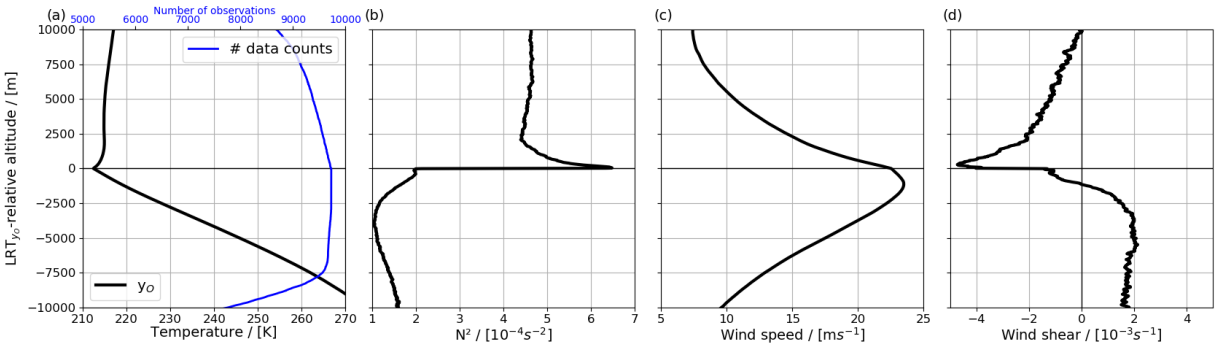

Figure 4.5: LRT_{y_O}–relative mean profiles of (a) temperature (black) and number of data (blue), (b) N^2 and (c) wind speed and (d) wind shear using 9729 radiosondes.

Statistical metrics to evaluate the observation impact

The observation-space background (or first-guess, y_{FG}) and analysis (y_{AN}) states are derived from the observations (y_O) and their departures from the first-guess (dep_{FG}, referred to as innovation) and the analysis (dep_{AN}, hereinafter residuals). These departures are defined as follows:

$$dep_{FG} = y_O - y_{FG}, \tag{4.5}$$

$$dep_{AN} = y_O - y_{AN}. \tag{4.6}$$

The observation-space increment is defined as the analysis minus the background state and shows whether a quantity has been increased or decreased in the data assimilation cycle:

$$Increment = y_{AN} - y_{FG}. \tag{4.7}$$

The increments, innovations and residuals (visualized in Figure 2.6) are the principal diagnostic metrics for the subsequent statistical analysis to examine the influence of radiosondes in the data assimilation system.

4.2 Innovations, residuals and increments for an example case

In order to firstly picture the vertical distribution of the key metrics, Figure 4.6a-c shows the temperature innovations, increments and residuals for the mid-latitude station Iqaluit (Canada) during the NAWDEX period. Iqaluit is selected as it comprises a high number of radiosonde profiles (#114, at 6-hourly intervals) and outlines the high tropopause altitude and wind speed variability related to the changing synoptic situations typical of the extratropics. Several strong jet stream events with wind speeds occasionally exceeding 45 m s^{-1} passed over the station, accompanied by high variability of the LRT 7-13 km.

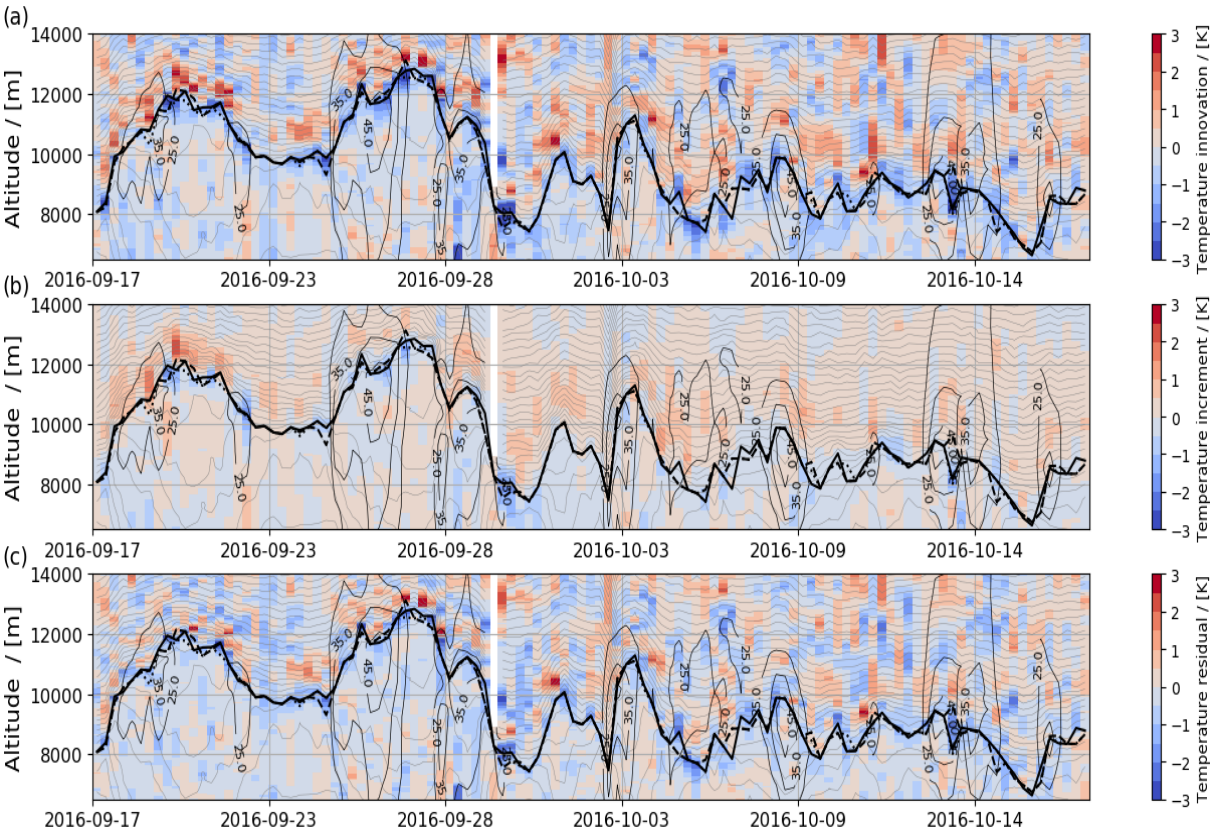

Figure 4.6: Time series (17 September–18 October 2016) of (a) innovations, (b) increments and (c) residuals of temperature (colour shading) at the Canadian station Iqaluit (63.75°N, 68.53°W) in geometrical coordinates. The panels are superimposed by the observed θ (thin grey lines, $\Delta\theta = 4$ K) and wind speed (thin black contouring). Black lines illustrate the altitude of the tropopauses with LRT$_{y_O}$ (solid), LRT$_{y_{FG}}$ (dashed) and LRT$_{y_{AN}}$ (dotted).

The temperature innovations (Figure 4.6a), i.e. the difference between observed and background temperature, are roughly between -3 and 3 K, with the largest innovations occurring near the tropopause. Positive innovations (i.e., cold background bias) tend to occur in the lowermost stratosphere, while negative innovations (i.e., warm bias) predominate in the altitude range of the tropopause. The temperature increments (Figure 4.6b) are weaker (between -2 and 2 K) compared to the innovations, but also follow

the tropopause with negative increments at tropopause and positive increments in the LS. Thus, they appear to counteract the background biases. The temperature residuals (Figure 4.6c) show that the differences between the observations and the analysis are locally as large as the innovations. However, particularly in the vicinity of the tropopause, the residuals appear to be reduced compared to the innovations which would indicate an improved temperature in the analysis trough data assimilation. To obtain a robust evidence of the data assimilation influence on the representation of the tropopause, increments of temperature and wind are analysed in tropopause-relative coordinates.

4.3 Statistical assessment

4.3.1 Average influence

Figure 4.7 presents LRT_{yO}-relative average profiles of temperature, N^2, wind speed, and vertical wind shear for the 9729 radiosondes and their model equivalents. The minimum temperature detected in a layer of ±500 m around LRT_{yO} is overestimated (by up to 1 K) in the background profiles (Figure 4.7a) confirming a warm temperature bias at the tropopause. In the LS above, the background temperature increases less strongly which results in a cold model bias between 0.5 and 2 km above the tropopause. The weaker thermal gradients in the background are accompanied by an underrepresentation of the amplitude and sharpness of the N^2 jump across the tropopause (Figure 4.7b). Wind speed is underestimated in the background throughout the UTLS (Figure 4.7c), with a maximum underestimation ($0.5\,m\,s^{-1}$) between –1 km and 0.5 km. The rapid decrease in wind shear above the wind maximum towards the lowermost 1 km layer of the LS is less pronounced in the background. Figure 4.7a–d show that the analysis is drawn towards the observations for all parameters at any altitude in the UTLS. The slightly sharper tropopause structure reveals a positive influence of data assimilation on the representation of the tropopause in the analysis.

Figure 4.7e–h shows the vertical structure of the increments. The temperature increments (Figure 4.7e) imply a cooling (up to –0.25 K), between –1 km and +0.5 km around the LRT_{yO}, i.e., the altitude range of the warm bias. In the LS, a warming of up to 0.25 K between 0.5 and 2 km above LRT_{yO} counteracts the cold bias in the model background. This impact on the temperature distribution results in negative N^2 increments ($-0.15 \times 10^{-4}\,s^{-2}$) in a 1.5 km thick layer below the LRT_{yO} and between 1 and 2 km above LRT_{yO} (Figure 4.7f). In the 1 km layer above LRT_{yO}, N^2 increments are positive with a distinct maximum of about $0.3 \times 10^{-4}\,s^{-2}$ at approx. 0.5 km. Wind increments are predominantly positive in the entire UTLS (Figure 4.7g), which indicates a wind speed increase in the analysis. The wind increments are stronger in the UT than in the LS peaking ($0.25\,m\,s^{-1}$) at the altitude of the wind maximum and the strongest underestimation in the background (1 km layer below LRT_{yO}, Figure 4.7c). Increments of wind shear (Figure 4.7h) are positive in the 2 km below, and negative in the 1 km layer above the LRT_{yO} (minimum at 500 m above LRT_{yO}). Between 1.5 to 3 km above LRT_{yO}, wind shear increments are positive and of comparable magnitude to the UT.

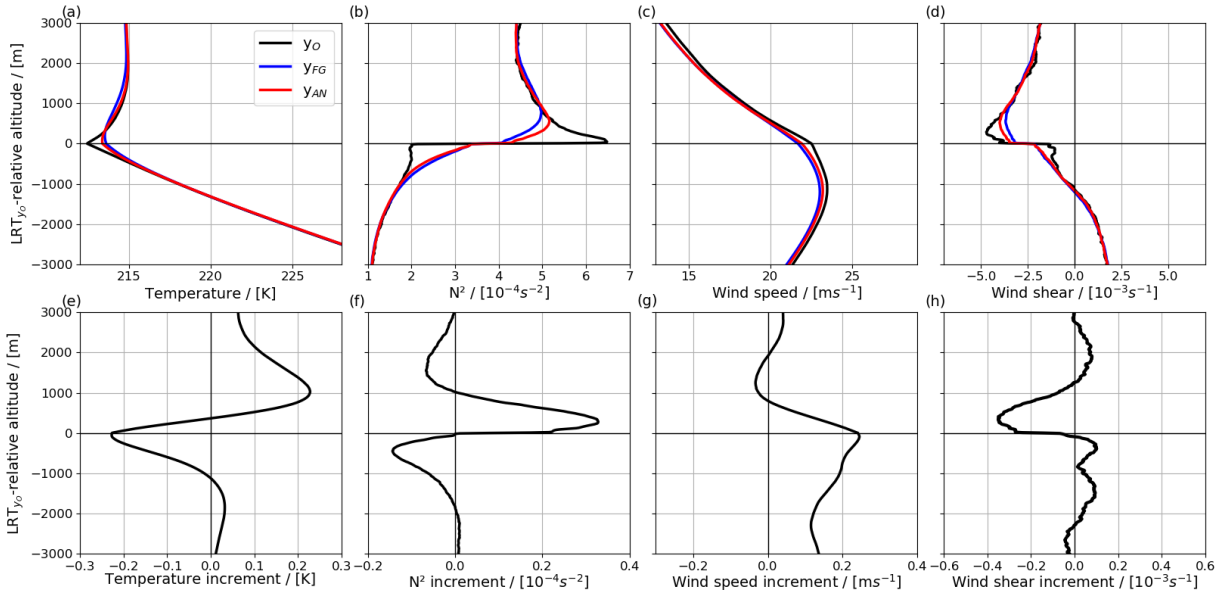

Figure 4.7: LRT$_{yO}$–relative distributions of (a) temperature, (b) N^2 (c) wind speed and (d) wind shear and the respective increments (e–h) averaged for the 9729 profiles of observations (black), background (blue) and analysis (red).

4.3.2 Sensitivity of the influence to the choice of LRT-relative coordinate

Figure 4.7 presented profiles of the parameters and increments relative to the observed tropopause. However, Figure 4.6 demonstrates that the observed tropopause altitude may differ from the background (and analysis). This raises the question of which LRT-relative view is the most suitable reference for evaluating tropopause structure. In the following, the distributions are presented in different LRT-relative reference systems to discuss their significance for evaluating tropopause sharpness.

Figure 4.8 illustrates the profiles of temperature and N^2 in the observations (Figure 4.8a, e), background (Figure 4.8b, f) and analysis (Figure 4.8c, g) with respect to the observed LRT$_{yO}$, the background LRT$_{yFG}$ and the analysis LRT$_{yAN}$, respectively. In each case the lowest tropopause temperature as well as the strongest temperature inversion and jump in N^2 occurs when the "own" LRT is used. This is particularly obvious for the observed profile relative to LRT$_{yO}$ (see black curve in Figure 4.8a, e). In addition, the background and analysis profiles have the lowest tropopause temperature and strongest inversion when viewed relative to LRT$_{yFG}$ (medium red curve in Figure 4.8b) and LRT$_{yAN}$ (light blue curve in Figure 4.8c), respectively.

Figure 4.8d and h show the increments referenced to the different LRT-relative coordinates. Each of the LRT reference systems confirms a cooling near the LRT, a warming in the LS above (Figure 4.8d), and an increase of N^2 just above the LRT (Figure 4.8h). The differing LRT altitudes of the individual profiles in y_O, y_{FG} and LRT$_{yFG}$ result in small differences in the magnitude of the increments for the different LRT–relative coordinates. The increments are smallest when referenced to LRT$_{y_{FG}}$.

As the LRT$_{own}$-relative distributions provide highest sharpness, LRT$_{own}$-relative in-

crements (grey line in Figure 4.8d, h) are considered to further analyze the influence on tropopause sharpness. These increments in LRT_{own}-relative coordinates, which are calculated as $y_{AN}(\mathrm{LRT}_{yAN}) - y_{FG}(\mathrm{LRT}_{yFG})$ and ideally remove effects on the average increments from differing LRT_{yFG} and LRT_{yAN} altitudes, have a comparable structure in the LS. However, they show only a slight cooling (< 0.1 K) at the tropopause and an increasing warming with decreasing altitude in the troposphere, which does not agree with increments in geometrical and LRT_{yO} space (e.g., Figure 4.6 and Figure 4.8a). The warming in the troposphere is a systematic temperature bias that is caused by the tropopauses detected at different altitudes (either in LRT_{yFG} or in LRT_{yAN} or both, Figure 4.8a).

To emphasize the role of differing tropopause altitudes on the distribution of the increments, the four types of increments are shown for cases with similar LRT altitudes (within ± 100 m), which are almost identical (see overlapping dotted lines in Figure 4.8d). An analysis of LRT_{own}-relative increments is not further pursued because such increments are determined after shifting the profiles with respect to the own LRT which does not correspond to real changes to the model background field in geometrical space (when the LRT_{yFG} and LRT_{yAN} differ). This analysis is nonetheless presented to emphasize the sensitivity of cross-tropopause distributions and their increments to the choice of the LRT reference and the impact of systematic LRT altitude differences. LRT_{yO} provides the most realistic representation of the tropopause altitude (Figure 4.6a, Figure 4.8a) and is used in the following to analyze the influence on tropopause sharpness in this study. In addition, the influence on the tropopause altitude is studied relative to the LRT_{yFG} (Sect. 4.4).

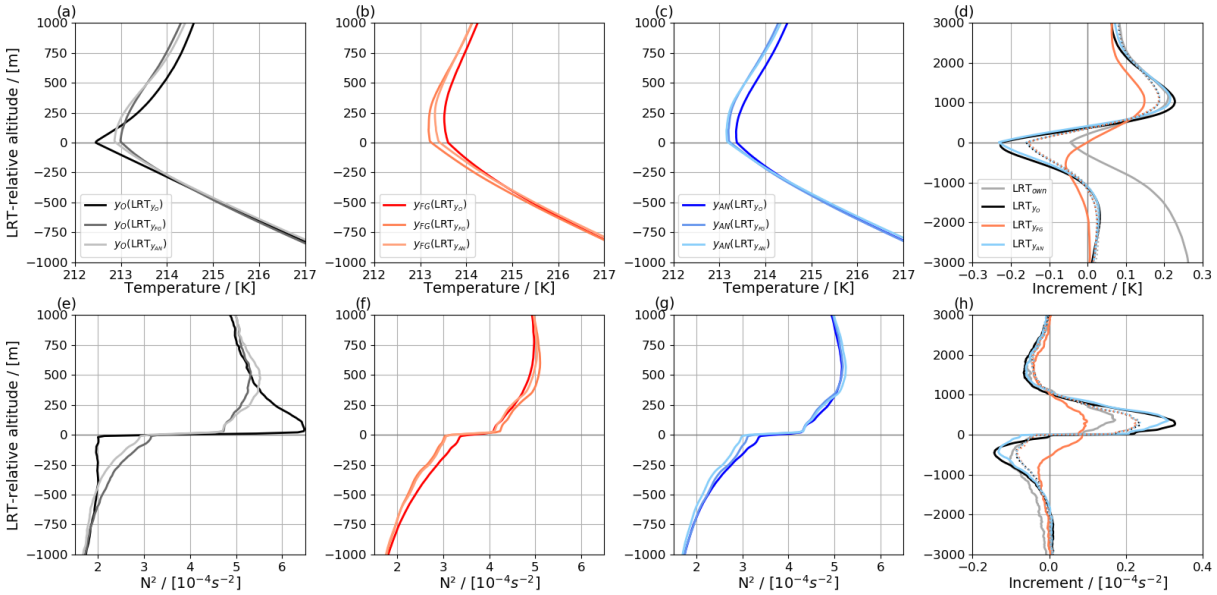

Figure 4.8: Mean profiles of temperature (a–c) and N^2 (e–g) for observations (a, e;), background (b, f) and the analysis (c, g) relative to LRT_{yO}, LRT_{yFG} and LRT_{yAN}, respectively (colour coded). The panels d) and h) show the associated increments. In addition, increments using the own LRTs are shown (grey, calculated as $yAN(\mathrm{LRT}_{yAN}) - (\mathrm{LRT}_{yFG})$, for details see text). The dotted lines in d) and h) represent the 3712 profiles with the LRT altitudes of observations, background and analysis being within ± 100 m (note that dotted lines overlap).

4.3.3 Influence on tropopause sharpness

The previous results indicated an increase in sharpness in the analysis with suggested high temporal variability in the increments (Figure 4.6) that is likely influenced by particular dynamical situations. Figure 4.9a illustrates the distribution of the observed maximum squared static stability (N^2max) in the 3 km above the LRT_{yO}, which is a common indicator for tropopause sharpness (Birner et al., 2006, Pilch Kedzierski et al., 2015). N^2max shows a uni-modal, positively skewed distribution ranging from 3 to 30×10^{-4} s^{-2} with largest frequency (> 200 profiles per bin) of between 6–12 $\times 10^{-4}\,s^{-2}$ and lowest frequency (< 50 profiles per bin) for $5 \times 10^{-4}\,s^{-2} < N^2\text{max}$ and $N^2\text{max} > 15 \times 10^{-4}\,s^{-2}$. The quartiles of this distribution are used to classify the data into the smoothest (N^2max,Q00–Q25), the intermediate (N^2max,Q25–Q75) and the sharpest (N^2max,Q75–Q100) tropopause cases.

The observed profiles (Figure 4.9b) display that the sharp class has the lowest tropopause temperature and the strongest inversion with the largest jump in N^2. In contrast, the smoothest tropopauses exhibit a higher tropopause temperature, a weaker temperature inversion, and a lower amplitude in N^2 at the LRT_{yO}. The intermediate class depicts a comparable tropopause structure to the full data set average described in the previous section. In agreement with these findings the observed mean tropopause altitude for the sharp and smooth classes are 12750 m and 11580 m, respectively, which suggests that the sharp (smooth) tropopauses can be related to ridge (trough) situations characterised by high (low) tropopause altitudes.

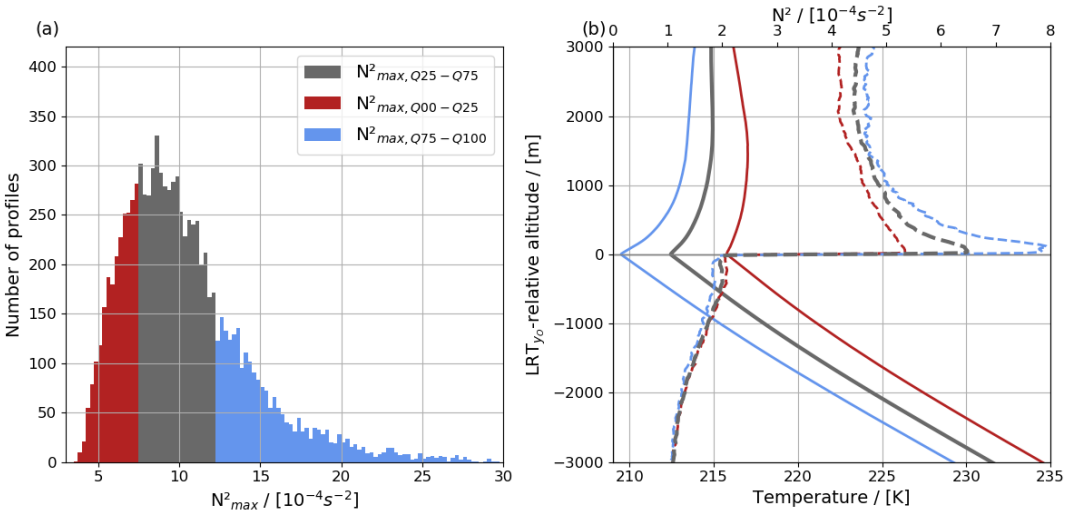

Figure 4.9: (a) Distribution of N^2max as observed in the 3 km layer above LRT_{yO} with bin size $0.5 \times 10^{-4}\,s^{-2}$. Quartiles of the $N^2{}_{max}$ distribution in (a) are used to classify tropopause sharpness: smoothest (red; Q00–Q25), intermediate (grey; Q25–Q75) and strong sharpness (blue; Q75–Q100). (b) the corresponding LRT_{yO}–relative mean profiles of observed temperature (solid) and N^2 (dashed).

For each class of N^2max, the mean vertical profile of innovation (Equation 4.5), increment (Equation 4.7) and residual (Equation 4.6) for temperature and N^2 relative to LRT_{yO} are presented in Figure 4.10. We first focus on the intermediate tropopause sharpness. In the UT, the temperature innovations are weak, negative and vertically nearly

constant (Figure 4.10a, about –0.1 K) before they reach a minimum of about –1.2 K at LRT_{yO} indicating a warm bias at the background tropopause. Above, the innovations strongly increase and become positive at 0.5 km above the LRT_{yO} before a maximum cold bias of 0.3 K is reached at 0.8 km altitude. The temperature increments (Figure 4.10b) correspond to the findings in Figure 4.7 with the negative increments around the LRT_{yO} counteracting the warm bias and the positive increments above decreasing the cold bias. In the ± 0.5 km around LRT_{yO} large N^2 innovations between –2 to $3 \times 10^{-4}\,s^{-2}$ illustrate the strong underestimation of tropopause sharpness in the background (Figure 4.10d). The average positive (above LRT_{yO}) and negative (below LRT_{yO}) N^2 increments (Figure 4.10e) for the intermediate profiles agree in shape and magnitude to the structure of N^2 increments given in Figure 4.7. Apparently, they lead to a sharpening of the tropopause. Increments are much smaller than the innovations (20 % for temperature and 10 % for N^2) which explains that the vertical structure of the innovation is preserved in the residuals (Figure 4.10c, f). For the smooth and sharp classes (blue and red lines in Figure 4.10), innovations, increments, and residuals have a similar vertical distribution but show weaker and stronger amplitudes respectively. For instance, temperature increments are about –0.3 K (–0.1 K) at the tropopause for the sharp (smooth) class and about 0.3 K (0.1 K) for the maximum above, in the LS. The influence is stronger where the background biases are strongest.

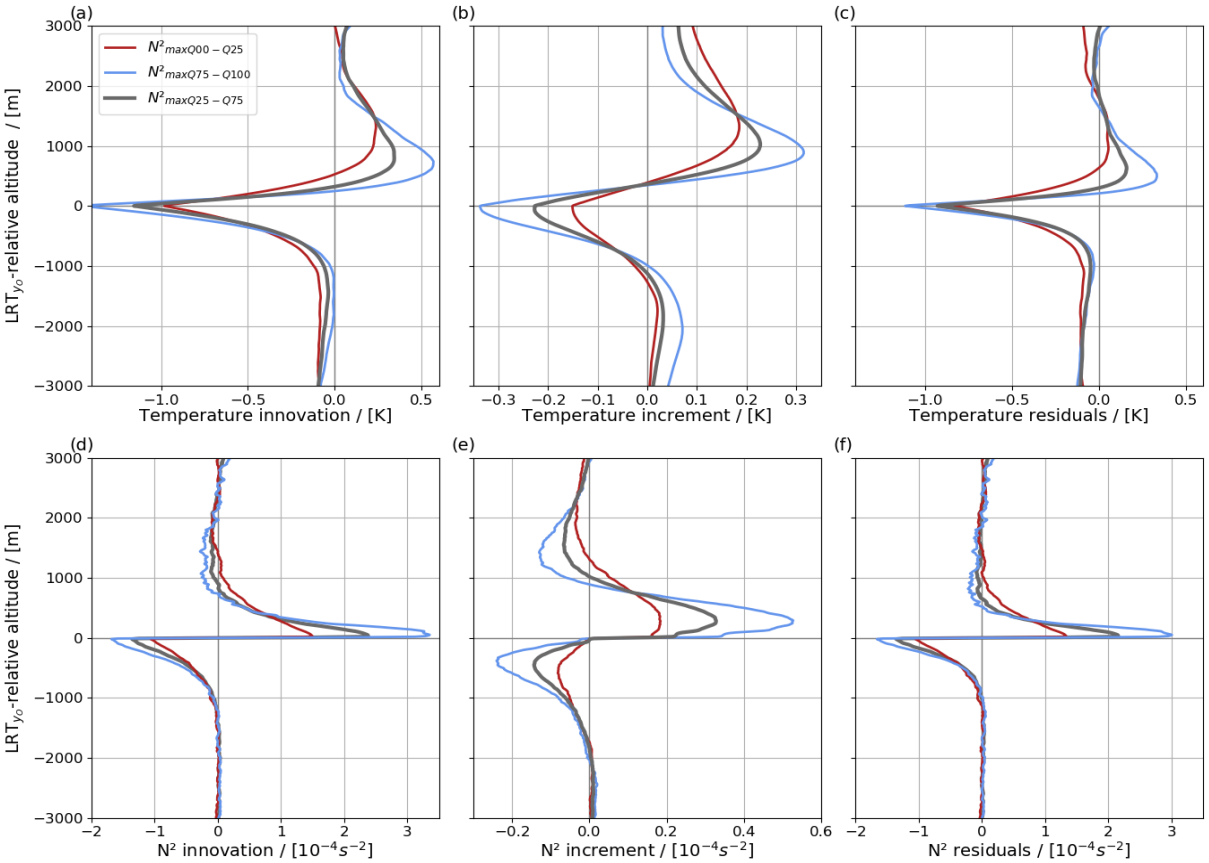

Figure 4.10: LRT_{yO}–relative mean profiles of innovations, increments, and residuals for (a–c) temperature and (d–f) N^2 for the classes of N^2max defined in Figure 4.9.

Figure 4.11 illustrates the variability in wind speed in the radiosonde data set. Average wind speeds in a layer of ±3 km around LRT_{yO} range from 0 to 60 m s^{-1} with the highest frequency between 5–25 m s^{-1} (Figure 4.11a). Quartiles of layer-mean wind speed divide the data set into weak (windQ00–Q25), intermediate (windQ25–Q75) and strong (windQ75–Q100) winds. The weak wind class shows vertically fairly constant low wind speeds (< 10 m s^{-1}). While the intermediate class exhibits a comparable shape of the mean wind profile as the full data set (Figure 4.7), the strongest wind class depicts a pronounced wind maximum (> 40 m s^{-1}) at 1 km altitude below the LRT_{yO} expressing strong jet stream winds. While wind shear in the weak wind class is small, the intermediate and strong wind classes have a decrease in vertical wind shear from positive values below the wind maximum to negative values above it, with a peak in the LS. All classes show an increased reduction in wind speed above the LRT which is associated with a step change in vertical wind shear.

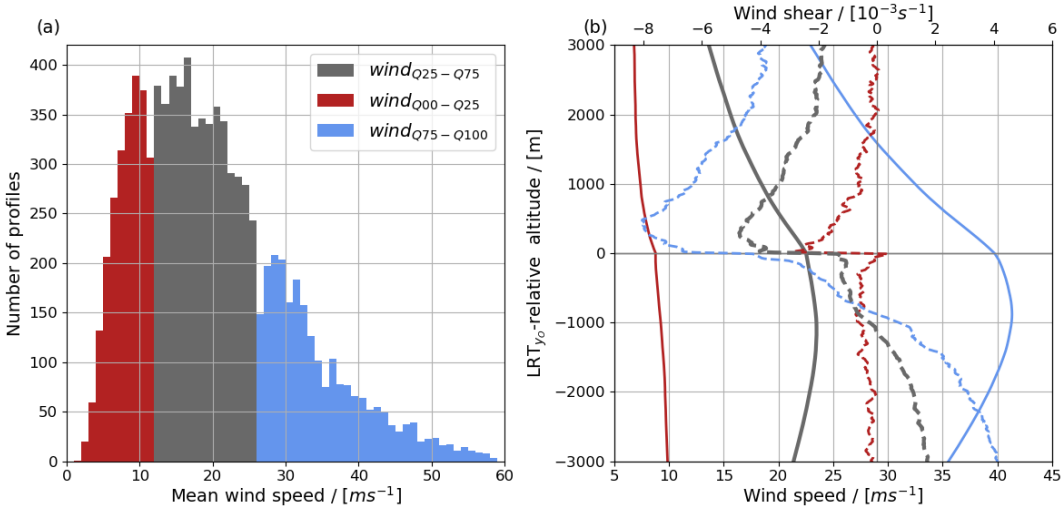

Figure 4.11: (a) Distribution of mean observed wind speed in the ±3 km above and below LRT_{yO} with 1 m s^{-1} bin size. Quartiles of the distribution in (a) are used to distinguish wind classes: weakest wind (red; Q00–Q25), intermediate wind (grey; Q25–Q75) and strongest wind (blue; Q75–Q100). (b) The corresponding LRT_{yO}-relative mean profiles of observed wind speed per class.

The mean profiles of innovation, increment and residual for each class are shown in Figure 4.12. The positive wind innovations of all classes across the UTLS express the underestimated wind speeds in the background (Figure 4.11a, see also Figure 4.7). Innovations in the UT are generally larger than in the LS and peak at the tropopause. Maximum innovations range between 0.5 m s^{-1} for the weak and 1.2 m s^{-1} for the strong wind class. The predominately positive wind speed increments throughout the UTLS (Figure 4.11b) represent a wind increase in the analysis which is largest in the 500 m layer below LRT_{yO} ranging between 0.1 m s^{-1} for the weak and almost 0.5 m s^{-1} for the strongest wind class. The positive wind speed residuals (Figure 4.11c) show that a slow wind bias remains in the analysis; however, the weaker residuals than the innovations observed for each class point to an improvement. For the strongest winds the innovations are reduced by roughly 40v%. In the layer with increased negative shear directly above

the LRT (Figure 4.11d), there are particularly strong negative shear innovations for the class of the strong winds, which are associated with increased negative shear increments. After the assimilation, wind shear residuals show little variability across all classes.

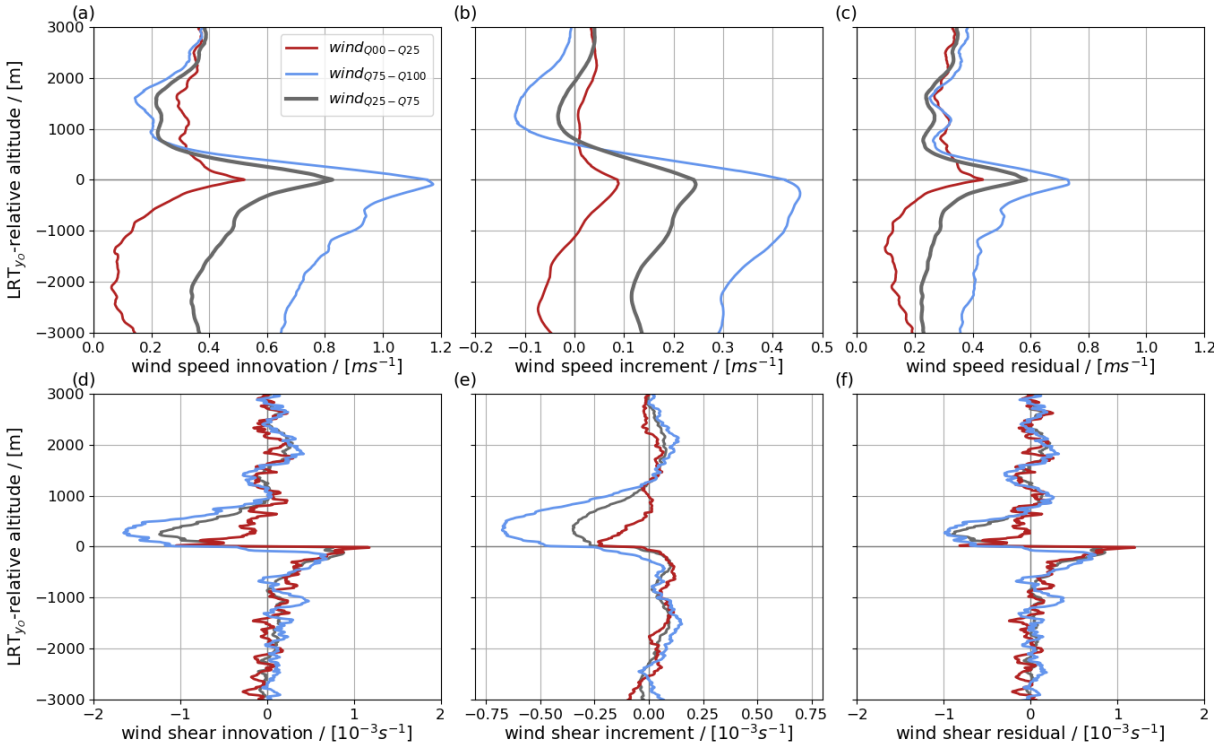

Figure 4.12: LRT_{yO}–relative mean profiles of wind speed of (a) innovations, (b) increments and (c) residuals per classes defined in Figure 4.11.

4.4 Influence on the tropopause altitude

In the following, it is investigated whether the modification of the temperature profile (Figure 4.7e) affects the tropopause altitude. Hereafter, LRT altitude differences between the observations and the background ($LRT_{yO} - LRT_{yFG}$) are referred to as "LRT innovations" according to Equation 4.5. LRT altitude differences between the observations and the analyses are referred to as "LRT residuals" ($LRT_{yO} - LRT_{yAN}$; Equation 4.6), respectively. An overview of LRT innovations for the entire data set is given in Figure 4.13a providing a symmetric normal distribution centered near zero (–26 m). For 43 % of the profiles, the LRT innovations are in range of ± 100 m. For about 10 % of the profiles, LRT differences are larger than 1 km. In order to prevent an impact of mis-detected LRTs (see discussion in Sect. 4.6), which are most likely for unusually large LRT differences, the following evaluation is restricted to 8778 profiles (about 90 %) which provide LRT innovations within ± 1 km.

For these 8778 profiles, the distributions of LRT innovations and residuals are shown in Figure 4.13b and c, which are colour-coded at different intervals of LRT innovations. The grey interval reflects LRT innovations within ± 100 m, while bluish colours represent profiles where the observed LRT is higher than the background and reddish colours vice versa. Comparing the colour distributions in Figure 4.13b and Figure 4.13c allows interval changes to be identified, and only a small fraction of profiles change the intervals. The distribution of LRT innovations shows a clear maximum near zero with a frequency corresponding to about 3000 profiles per bin, and the frequency decreases towards the edges of the distribution to approx. 10 profiles per bin. The distribution of LRT residuals (Figure 4.13c) shows a slightly increased number (+15 %) of profiles within ± 100 m indicating an improved tropopause altitude in the analysis and this improvement is confirmed by a slightly narrower shape of the Gaussian fit of the LRT residuals compared to the LRT innovations.

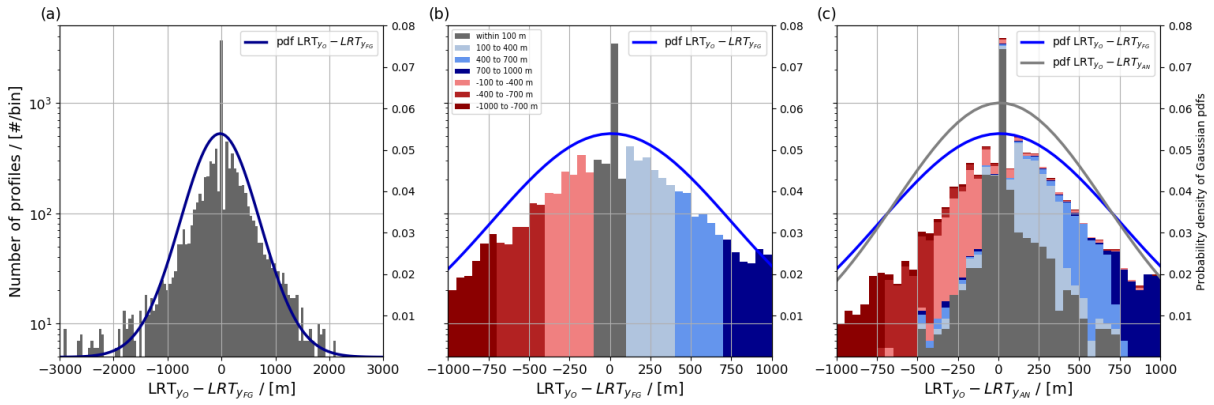

Figure 4.13: Distribution of (a) LRT innovations in the full data set, (b) as in (a) but for the range ± 1000 m and (c) LRT residuals. The colour coding reflects intervals of LRT innovations shown in (b) and is reused in (c) to visualize the LRT altitude change in LRT_{yAN} (for details see text). Gaussian PDFs are given in lines of dark blue, blue and grey, respectively, for 50 m bins. Note the log-scale of the y-axis.

Table 4.1: Profile counts, mean LRT_{yO} altitude, innovation, residual, and improvement for the intervals of LRT innovation (Figure 4.13). Improvement is defined as $|\mathrm{LRT}_{yO} - \mathrm{LRT}_{yFG}| - |\mathrm{LRT}_{yO} - LRT_{yAN}|$, so positive values reflect an improved LRT altitude in the analysis.

Intervals [m]	Number [#]	LRT_{yo} [m]	Innovation [m]	Residual [m]	Impr. [m]
All profiles	9729	12226	-26	12	79
-1000 to -700	209	12806	-812	-420	371
-700 to -400	473	12425	-525	-277	227
-400 to -100	1309	11963	-237	-121	76
-100 to 100	4196	12064	2	18	-33
+100 to 400	1637	11706	222	175	36
+400 to 700	627	12456	517	349	153
+700 to 1000	227	13607	830	555	271

For the intervals of LRT innovations, Table 4.1 provides the number of profiles, the mean LRT_{yO} altitude, and the mean innovation, residual and improvement. Except for the interval with the smallest innovation (± 100 m; grey interval), the average innovation is larger than the residual which implies a vertical shift of LRT_{yFG} towards LRT_{yO}. The overall positive influence is supported by the depicted improvements, defined as the absolute difference of innovation and residual. Interestingly, both, the LRT altitude shift and also the improvement grow with increasing distance between LRT_{yO} and LRT_{yFG}.

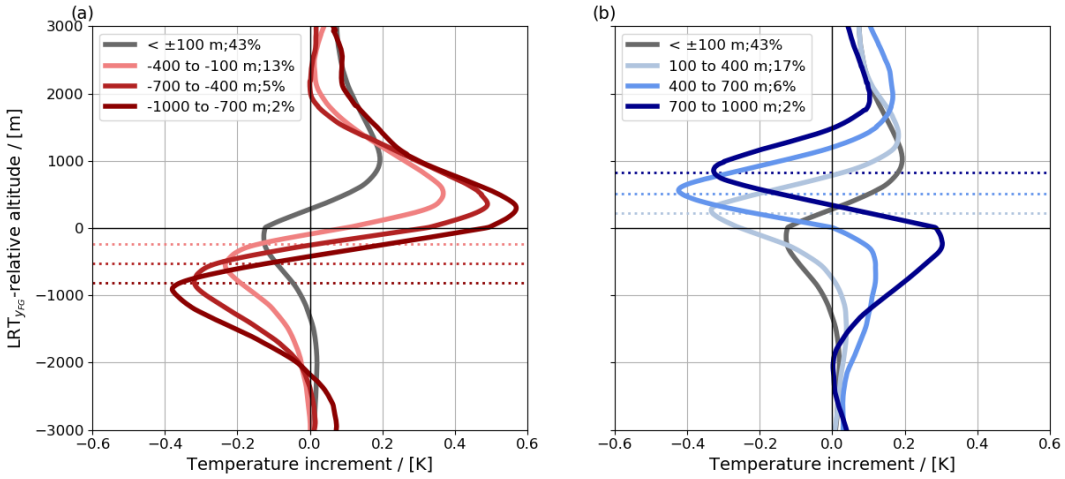

Figure 4.14: Mean temperature increments with respect to LRT_{yFG}–relative altitude for intervals of LRT innovations (colour coding, following Figure 4.13). (a) Negative LRT innovations ($\mathrm{LRT}_{yO} < \mathrm{LRT}_{yFG}$, in red) and (b) positive LRT innovations ($\mathrm{LRT}_{yO} > \mathrm{LRT}_{yFG}$, in blue). The grey lines show the increment for LRT innovations within ± 100 m. In (a) and (b) the averaged LRT_{yFG}–relative altitude of LRT_{yO} for each interval is depicted by the dotted lines.

To understand how the LRT altitude changes are related to the changes of the background temperature profile, the average temperature increments for the individual intervals of LRT innovations are presented with respect to the LRT_{yFG}-relative altitude

(Figure 4.14). For small LRT innovations (within ± 100 m, grey line in Figure 4.14a,b), the temperature increments are negative at the LRT (–0.15 K) and positive in the LS (0.2 K) (analogous to the sharpening influence, Sect. 4.2) which does not lead to major changes in the LRT altitude in this interval (see Table 4.1). With increasing LRT innovations, the positions of the peaks in the LRT_{yFG}-relative increments are vertically shifted. In case of the negative LRT innovations (Figure 4.14a), which means that the observed LRT_{yO} is located lower than the background LRT_{yFG}, we observe positive increments (warming, 0.3–0.6 K) at and above LRT_{yFG} and negative increments (cooling, 0.2–0.4 K) below LRT_{yFG}. As the strongest negative increments agree with the altitude of LRT_{yO} (dotted lines in Figure 4.14a), peaks in the increments are shifted downwards and show slightly higher maxima for increasing negative LRT innovations (red profiles in Figure 4.14a). In contrast, positive LRT innovations (i.e., LRT_{yO} located above LRT_{yFG}; Figure 4.14b) exhibit negative increments (–0.3 to –0.4 K) above LRT_{yFG}, and positive increments below the LRT_{yFG}. Here, the increment peaks are shifted upwards for more positive LRT innovations (blue profiles in Figure 4.14b) in agreement with the altitude of the observed tropopause.

One example radiosonde profile (Orland, 23 September 06 UTC) is taken to illustrate how the modification of the background temperature profile in the data assimilation process influences the tropopause height in the analysis, Figure 4.15. For each temperature profile in Figure 4.15a, the tropopause is detected at its respective temperature minimum. Interestingly, LRT_{yFG} is located about 400 m lower than LRT_{yO}, and correspondingly a higher background temperature is found in the LS (Figure 4.15b) which is also represented by the strong temperature innovations (up to -3 K) in the 1 km layer above the background tropopause (Figure 4.15c). The temperature increments (Figure 4.15c) in this corresponding layer show a cooling of the analysis leading to smaller (only up to -1 K) residuals. This altitude shift of the increments is comparable to the shifted average increments presented in Figure 4.14b. Through the improved analysis temperature profile, LRT_{yAN} is shifted upwards towards LRT_{yO}.

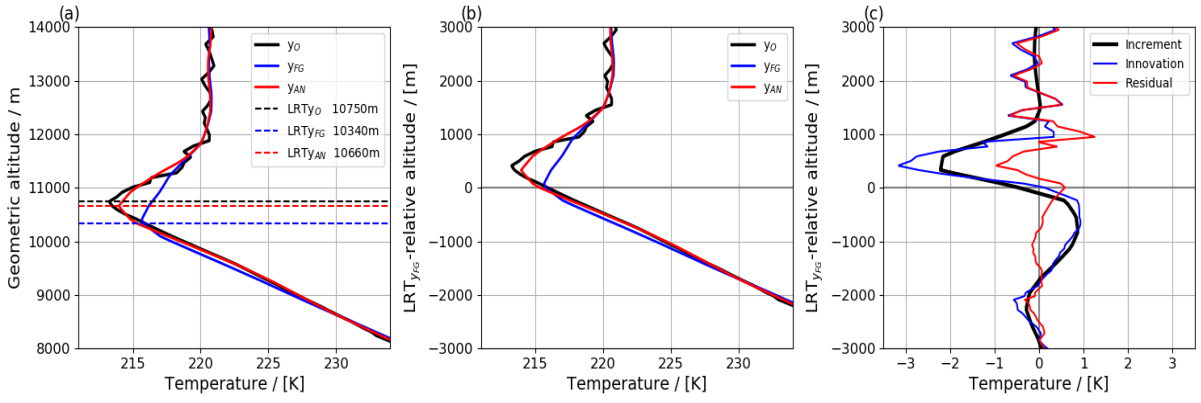

Figure 4.15: Temperature profiles of y_O (black), y_{FG} (blue), y_{AN} (red) in a) geometrical altitude, b) LRT_{yFG}-relative coordinates, and c) innovation (blue), residuals (red) and increments (black) at Orland station (Norway, 63°4'N,9°4'E) for the radiosonde launch on 23 September 06 UTC.

4.5 Influence in the Observing System Experiment

The presented results reveal that data assimilation sharpens the tropopause at the location of the radiosondes which provides a strong indication that this influence is related to the information contained in the radiosondes. However, a potential contribution of other observations cannot be excluded. To find out the extent to which the radiosondes are responsible for the sharpening, an OSE with radiosondes is conducted at the ECMWF. In this OSE, the NAWDEX radiosondes are assimilated in a CTR run and excluded from assimilation (but passively monitored) in a DEN run. The NAWDEX radiosondes are processed according to the same scheme as in the previous section (pressure filtering, interpolation, tropopause derivation). The tropopause altitudes are derived for the observations as well as for the background and analysis of CTR and DEN runs.

The average profiles of observed temperature and N^2 of the 497 NAWDEX profiles in the OSE (Figure 4.16) are comparable to the average profiles of the 9729 radiosondes (Figure 4.7) with a similar magnitude of the jump in N^2 and an alike decrease in wind shear across the tropopause. The average minimum temperature at the tropopause (Figure 4.16a) is slightly higher (by $\sim 2\,\mathrm{K}$), which is related to the observation locations of the NAWDEX radiosondes at higher latitudes where the tropopause is typically lower and warmer. The additional radiosondes show a pronounced wind maximum below LRT_{yO} of about $29\,\mathrm{m\,s^{-1}}$ (Figure 4.16c), which, compared to the lower wind speeds in the complete data set (Figure 4.7c), indicates the occasionally strong jet streams in the focus of the NAWDEX campaign (Schäfler et al., 2018).

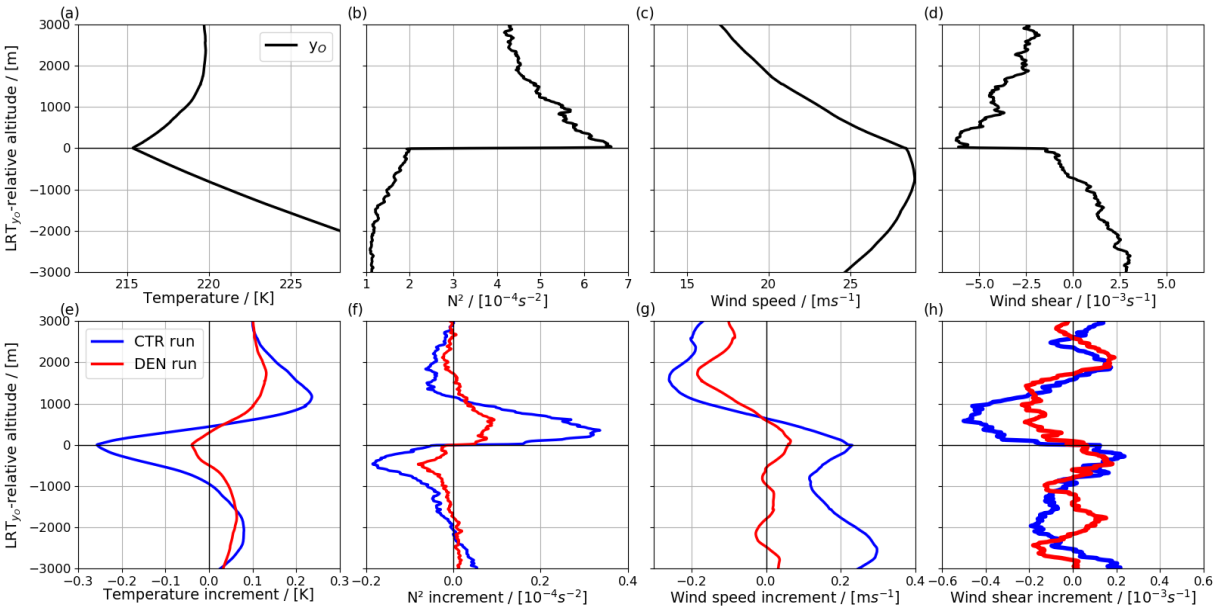

Figure 4.16: LRT_{yO}–relative mean profiles of the (a) temperature, (b) N^2, (c) wind speed and (d) wind shear as observed by the 497 NAWDEX radiosondes and the respective increments of (e) temperature, (f) N^2 and (g) wind speed and (h) wind shear for the CTR (blue) and DEN (red) experiment.

The increments for temperature, N^2, wind speed, and wind shear for the subset of 497 additional NAWDEX radiosondes are presented in Figure 4.16e-f. The profiles of the CTR run show a vertical structure comparable to the complete data set (the 9729 profiles, Figure 4.7): Temperature increments are negative (-0.25 K) around the observed tropopause and positive in the LS (1-2 km above LRT_{yO}). Accordingly, the CTR N^2 increments are positive within a 1 km-thick layer just above LRT_{yO} with a maximum of $0.3 \times 10^{-4}\,s^{-2}$, and negative ($-0.2 \times 10^{-4}\,s^{-2}$) in a 1 km layer beneath LRT_{yO}, and beyond this layer. The vertical structure of the CTR wind speed increments for the NAWDEX radiosondes also agrees with the complete data set (compare increments for the strongest wind class, Figure 4.11), with a positive increment $\pm 0.2\,m\,s^{-1}$ in the UT and a negative increment in the LS. Wind shear increments in the CTR run are also similar, but the minimum wind shear just above the LRT_{yO} shows slightly lower values ($0.5 \times 10^{-3}\,s^{-1}$) compared to Figure 4.7h.

The increments of temperature, N^2, wind speed, and shear in the DEN run (Figure 4.16) are weaker at each altitude, but tend to show a similar vertical distribution. This implies that the main contribution of tropopause sharpening and influence on wind comes from the assimilated radiosondes, but the non-zero DEN increments indicate that other observations also influence tropopause structure in the same direction. This may either be due to the remote impact of operational radiosondes or to dropsonde observations of which a larger number were deployed during NAWDEX (see Schindler et al. (2020)). Further contributions of assimilated aircraft observations and GPS-RO data are also conceivable.

The findings from the OSE analysis provide a valid indication of the positive influence of the radiosondes on the temperature and wind gradients at the tropopause. In addition (as in Sect. 4.4), the influence of radiosonde data assimilation on the tropopause height in the CTR and DEN experiments is analysed in the following. Table 4.2 contain the established metrics to evaluate the influence on the tropopause altitude. First, Table 4.2 shows substantially smaller residuals than innovations corresponding to a positive influence (i.e., positive improvement) whose order of magnitude corresponds to the results from Sect. 4.4. Second, it is noted that also the DEN indicates an improvement of the analysis tropopause altitude, however, this influence is clearly weaker compared to the DEN.

Table 4.2: Profile counts, mean LRT_{yO} altitude, innovation, residual, and improvement for the intervals of LRT innovation as in Table 4.1, for the CTR (DEN) experiment.

Intervals [m]	Number [#]	LRT_{yo} [m]	Innovation [m]	Residual [m]	Impr. [m]
-1000 - 700	5	10542	-786 (-462)	-390 (-308)	308 (162)
-700 - 400	18	12065	-559 (-410)	-226 (-342)	324 (44)
-400 - 100	46	11209	-235 (-150)	-92 (-148)	105 (30)
-100 - 100	223	10419	7 (15)	29 (29)	-39 (-30)
+100 - 400	109	10203	223 (217)	166 (198)	42 (27)
+400 - 700	32	10798	502 (406)	297 (340)	205 (66)
+700 - 1000	9	10452	798 (720)	427 (640)	371 (80)

Figure 4.17 illustrates the temperature increments with respect to the background tropopause separately for the CTR (solid lines) and the DEN run (dashed) for the different classes of tropopause altitude innovations. For each class, the CTR profiles show stronger temperature increments (i.e, higher maxima and lower minima) compared to the DEN run. The different strength of the temperature increments between CTR and DEN run provides an explanation for the different tropopause altitude improvement. The different strength of the temperature increments between CTR and DEN runs provides an explanation for the corresponding different tropopause altitude improvement. However low number of the profiles for the different classes has to be mentioned. Nevertheless, this evaluation shows that the positive influence on the tropopause height can be attributed to radiosonde data assimilation.

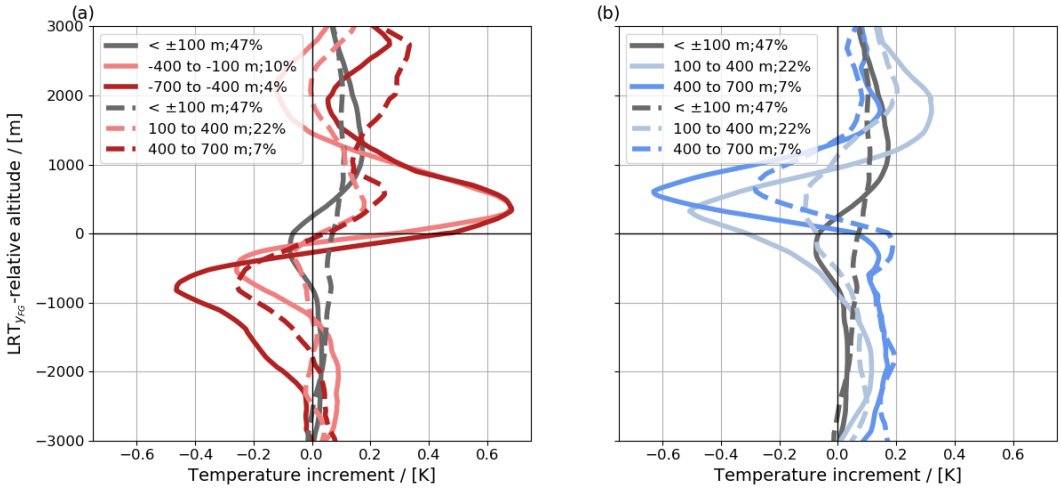

Figure 4.17: Mean LRT$_{yFG}$–relative temperature increments for intervals of LRT innovations as applied in Figure 4.14. (a) Negative LRT innovations (LRT$_{yO}$ < LRT$_{yFG}$, in red) and (b) positive LRT innovations (LRT$_{yO}$ > LRT$_{yFG}$, in blue). The grey lines show the increment for LRT innovations within ±100 m. In (a) and (b) the averaged LRT$_{yFG}$–relative altitude of LRT$_{yO}$ for each interval is depicted by the dotted lines.

4.6 Case study on the influence of data assimilation in model-space

In Sect. 3.2-Sect. 3.5 it is found that the assimilated radiosondes have a positive influence on the tropopause altitude and sharpness in observation-space, i.e., the location where the radiosonde observations where taken. Additionally, the DEN experiment shows a (weaker) tropopause sharpening at the location of the radiosondes, even though these are excluded from the assimilation, which indicates that other assimilated data (e.g. from distant radiosondes, aircraft, GPS-RO) positively influence the analysis tropopause. This raises the question to what extent the sharpening effect on the temperature and wind gradients near the tropopause has an influence in the model beyond the observation location, in the surrounding region. A sharpening effect would be conceivable, since the flow-dependent B-matrix used in the hybrid 4D-VAR scheme of the IFS distributes the information of the assimilated observations horizontally and vertically in space and time. Furthermore, a variable influence (i.e., stronger sharpening at sharp tropopause, stronger wind increments at high wind speeds) is revealed Sect. 3.4 which poses the question of whether the influence is related to the dynamic situation.

In order to investigate the non-local influence of data assimilation on the tropopause, model-space data allow to view the influence of data assimilation on the synoptic scale, as they supply the output of the analysis and the background directly on the model grid. The model-space files provide temperature and wind fields at a $0.125° \times 0.125°$horizontal grid and on 137 sigma-hybrid pressure levels from which the corresponding altitude of each level is derived as in Sect. 3.1. The vertical interpolation and tropopause derivation is then conducted at each grid point as for the observation-space profiles (see Sect. 4.1). The model-space influence is analysed for a case study on 25 September 2016 at 00 UTC, which represents a typical synoptic situation over Europe and where an increased number of radiosonde observations were carried out during NAWDEX (Schäfler et al., 2018).

The distribution of the geopotential height at 500 hPa (Figure 4.18a) shows a pronounced large-scale ridge (i.e. high geopotential) that is located over central and northern Europe. In the entire area of the ridge, the tropopause is located at high altitudes, roughly between 12 and 13 km altitude. At the western flank of the upper-level ridge, the strong jet stream is directed far to the north. Upstream a distinct trough (low geopotential) is located over the northern Atlantic, with an associated area of low surface pressure extending from Iceland towards the British Islands and further to the Bay of Biscay. The surface low is associated with the extratropical cyclone called "VLADIANA" which has formed in this region in recent days (Oertel et al., 2019). Within the trough, tropopause altitudes are located substantially lower (i.e., below 10 km). On the southeast flank of the trough a cold front running roughly along the west coast of England (6°W). Upstream (west of the front), a colder air mass (equivalent potential temperature $\theta_E < 280\,\mathrm{K}$) is transported southestwards. Ahead of the cold front, within the warm sector, comparably moist and warm air ($\theta_E > 285\,\mathrm{K}$) is transport northward. This northward flow is associated with a WCB lifting air from the boundary layer to the tropopause leading to enhanced cloud formation and precipitation (e.g., Harrold, 1973). WCBs are therefore regions of high diabatic activity due to the latent heat release of water vapour caused by phase changes in clouds. In model data, WCBs can be identified by Lagrangian backward trajectories as

described in Madonna et al. (2014) and Heitmann et al. (2024). A WCB is identified if an air parcel experiences an ascent of more than 600 hPa within 48 h. Regions of ascent (air between 800-500 hPa) and outflow (air < 400 hPa) within the WCB can be distinguished through such Lagrangian trajectories and are masked in white and black contours in Figure 4.18b, respectively. In a 3D analysis of VLADIANA, Beckert et al. (2023) shows that air within the WCB rises in the warm air sector along the cold front and follows both the anti-cyclonic and the cyclonic sides of the jet stream in the upper troposphere. Accordingly, the ascent (black contour, between 42°-60°N, 3-6°W) and outflow at jet stream level splits around (66°N,12°E) (white contour) (Figure 4.18). A detailed overview of the formation and development of VLADIANA is given in (Oertel et al., 2019).

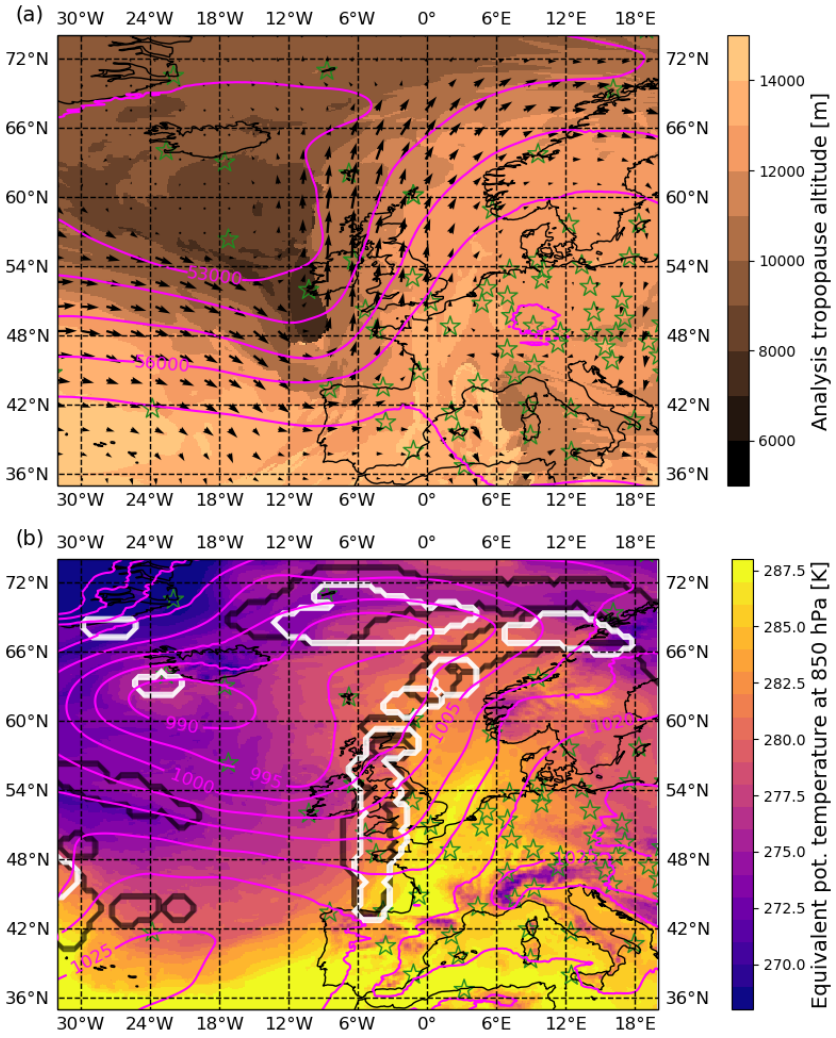

Figure 4.18: IFS analysis for 25 September 2016 at 00 UTC: (a) Tropopause altitude (colouring), geopotential height ($\Delta\phi = 100$ m, magenta contours), and wind (black arrows) at 300 hPa. (b) Equivalent potential temperature at 850 hPa (colouring), sea-level pressure (magenta contours, $\Delta P = 10$ hPa contours), and WCB ascent (white contour) and outflow (black contour) regions as derived from backward trajectories (Madonna et al., 2014). Green stars in panel a) and b) show radiosondes launched between 24 September 15 UTC and 25 September 03 UTC.

Figure 4.18 also shows the locations of the radiosondes (green stars) that were actively assimilated in the two short assimilation windows (15-21 UTC, 21-03 UTC) used in the SCDA to produce the background and the analysis (Sect. 2.2). The distribution shows that each of the dynamic features described are captured by radiosonde observations, with the majority of observations being over Europe covering the front side of the trough and the ridge.

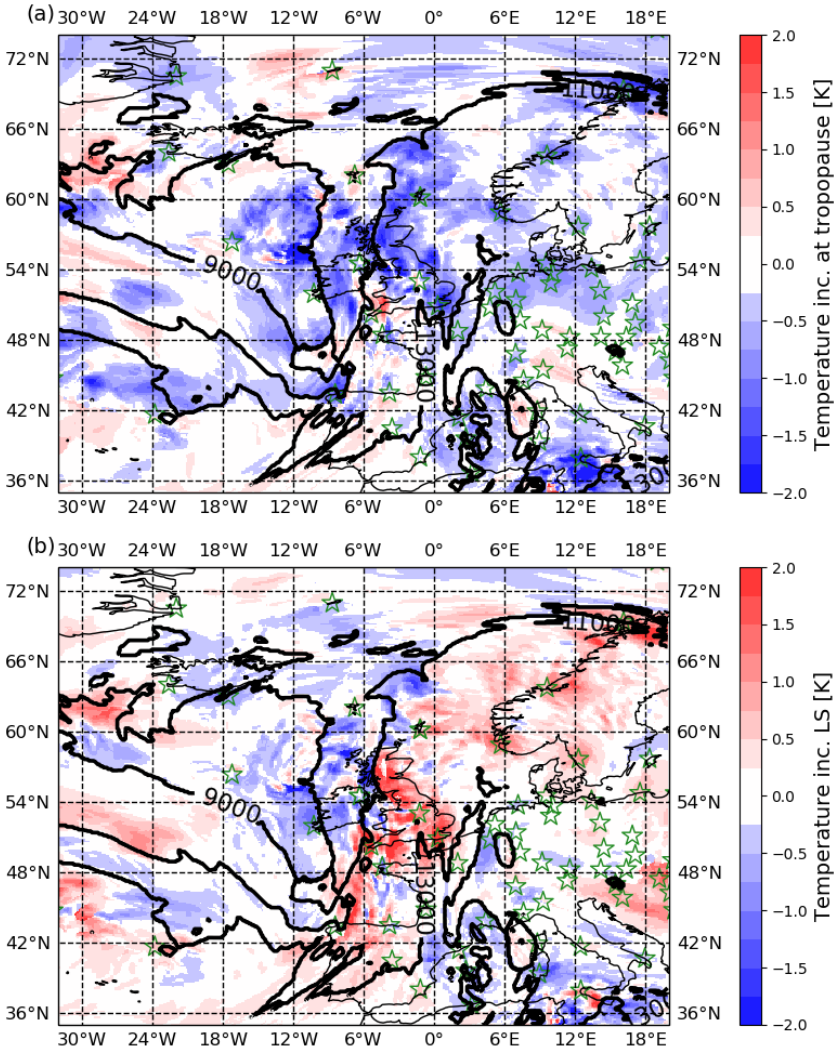

Figure 4.19: Tropopause-relative average temperature increments for (a) the ± 200 m layer across the analysis tropopause, and (b) the layer between 800 and 1200 m above the analysis tropopause on 25 September 2016 at 00 UTC. (a) and (b) are superimposed by isolines of the analysis tropopause height (9, 11 and 13 km).

The data assimilation influence in model-space is discussed based on tropopause-relative temperature and wind increments. In this evaluation, the analysis tropopause is selected as reference for the tropopause as the radiosonde observations do not provide tropopause altitudes for the whole area. In the observation-space analysis of the temperature increments (Sect. 4.3), a pronounced cooling is found in a ± 200 m layer around the

tropopause and a warming in the LS, which is strongest between about 800-1200 m above the tropopause. For these altitude layers, the model-space temperature increments are averaged and shown in Figure 4.19.

The temperature increments near the tropopause (Figure 4.19a) display large coherent structures of negative increments (cooling), clearly dominating the small areas of weak positive increments. A particular strong cooling (up to -2 K) is found over almost the entire British Isles in regions associated with the outflow of the WCB at high tropopauses, but also at the cyclonic side of the jet stream (low tropopauses). Figure 4.19b shows strong positive and negative temperature increments in the LS. The positive increments in the LS predominate particularly at high tropopause altitudes on the anti-cyclonic side of the jet stream from Spain to Scandinavia, while negative LS increments predominate in the trough for the lowest tropopauses (i.e., near Ireland).

The vertical structure of the temperature increments (i.e. cooling at the TP, warming in the LS) diagnosed in the observation-space, corresponds to a tropopause sharpening, which is reflected in positive N^2 increments in a 1 km thick layer above the tropopause. The model-space increments of N^2 in this layer illustrated in Figure 4.20a show, that positive increments dominate over negative increments and confirm the tropopause sharpening. This sharpening is most pronounced on the anti-cyclonic side of the jet stream, especially in areas of the warm sector and the WCB outflow. These areas of strongest sharpening, where the tropopause is particularly high and sharp (see $N^2 > 8 \times 10^{-4}\,s^{-2}$ contour in Figure 4.20b), also show the strongest LS warming (compare Figure 4.19b). A further finding in observation-space is the positive wind increments (i.e., an acceleration of wind speed) in the 1 km layer coinciding with the altitude of the strongest jet stream winds. In this layer, in model-space (Figure 4.20b) large-scale structures of positive and negative wind increments are found that are aligned with the flow direction (see wind arrows Figure 4.18a). Interestingly, there are strong negative increments ($<$-3 m s^{-1}) along the wind maximum of the north-oriented section of the jet stream (from the Bay of Biscay to the Faroe Islands), while positive increments (2-3 m s^{-1}) predominate east of the jet stream. Interestingly, the wind increments located in the ridge (i.e., over central and northern Europe) form larger, coherent structures than smaller-scale increments in the trough (over the Atlantic).

To further investigate the influence on temperature and wind in the region of increased increments over the British Isles in context of the synoptic situation, a longitude-height cross section of the increments averaged between 54°N and 58°N are shown in Figure 4.21. In this cross section, the jet stream is orientated northward and thus lies perpendicular to the cross-section, with a wind maximum at roughly 7°W and 9 km altitude. Low tropopauses (about 9 km) associated to the trough are located to the west of the jet stream (10°-20°W), and high tropopause to its east (5°-20°E) related to the ridge (approx. 12-13 km). The sloping isentropes located below the jet stream indicate the cold front ahead of the trough. The dense isentropes above the tropopause represent the LS. Along the cross section, the negative temperature increments (Figure 4.21a) show a cooling at the tropopause which is particularly marked on the anti-cyclonic side of the jet stream (5°-0°W) and at high tropopauses in the ridge. Above the tropopause, a remarkable strong LS warming is found above the jet stream (12°-1°W, 1-3 km above the tropopause). Above the low tropopauses in the trough the lowermost 4 km of the LS are cooled. Below the

tropopause in the UT, it is noticed the western and lower part of the jet (3°-5°W, 7-8 km altitude) is evenly cooled by almost 2 K. Besides stronger (positive and negative) temperature increments are found near the cold front, and in the lowermost 2 km layer of the troposphere (associated with the boundary layer).

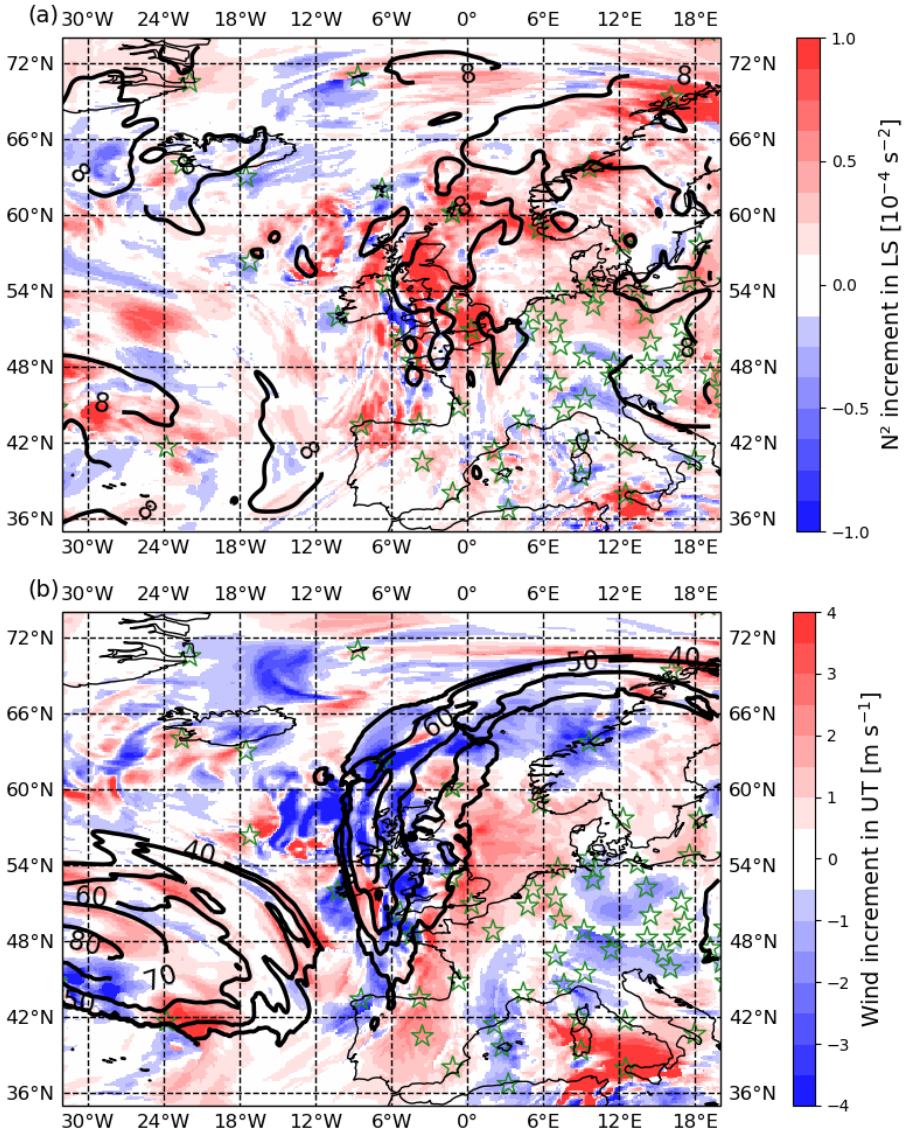

Figure 4.20: Tropopause-relative average increments for (a) N^2 increment in the 1 km layer above the analysis tropopause, and b) wind speed increment in the 1 km layer below the analysis tropopause on 25 September 2016 at 00 UTC. Black contours shows in (a) analysis profiles with sharp tropopauses ($N^2 > 8 \times 10^{-4}\,s^{-2}$) and in (b) iso-contours of wind speed ($10\,m\,s^{-1}$ spacing).

The strong positive and negative wind increments show a spatially more variable influence (Figure 4.21b), with the strongest wind increments (-4 to $4\,m\,s^{-1}$) located in the jet stream and in its immediate vicinity (10°-0°W), and near the cold front. The wind increments in the troposphere are generally more localized in the trough (i.e., near the cold front) than in the ridge. In the western part of the jet stream (increment up to

roughly -4 m s^{-1} between 7-10 km altitude) wind speed is strongly reduced, while a strong increase is shown on its eastern flank (2-3 m s^{-1}; 7-12 km altitude). This indicates a slight shift of the jet stream to the east, which is reflected in the easterly shifted isopleths.

In Figure 4.21a-b the locations of actively assimilated radiosondes (between 51°-61°N) are displayed. There is no clear indication that the temperature and wind increments at the radiosonde locations are generally stronger over the entire profile. However, similar patterns of temperature or wind increments can be identified locally at the positions of the radiosondes. For example, between 5°-17°E a continuous cooling is found at the tropopause, as well as a warming at about 2 km altitude in the troposphere, whereby 10 radiosondes were assimilated in this area.

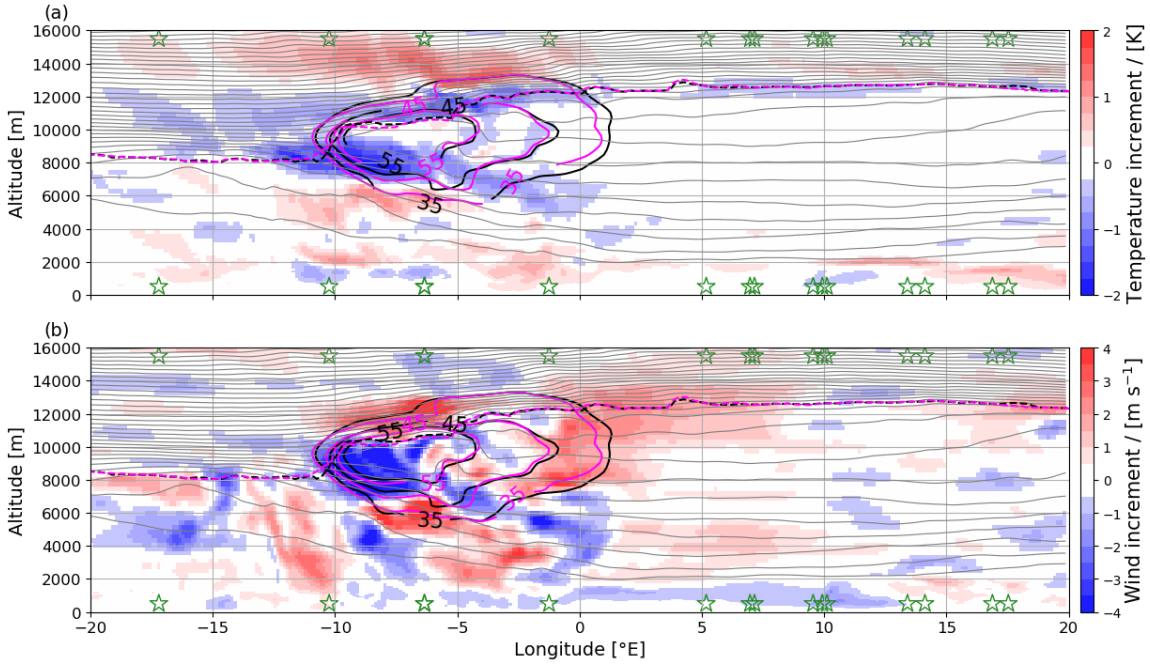

Figure 4.21: Longitudinal-altitude section of a) temperature increment and b) wind speed increment averaged between 54°N-58°N. Panels are superimposed by average isentropes ($\Delta\theta = 4$ K), and isopleths of wind speed (35, 45, 50 m s^{-1}) for background (magenta) and analysis (black). The dashed magenta (black) lines show the background (analysis) thermal tropopause. Radiosondes actively assimilated between 51°-61°N are illustrated by green stars.

To be able to compare the model- and observation-space influence on the tropopause, a statistical analysis of the increments in tropopause-relative coordinates is conducted for the map section. In the following, temperature, N^2 and wind speed profiles (Figure 4.22) and their increments (Figure 4.23) are presented in the tropopause-relative metric. Black curves show averages for all grid point locations of the map section (all, black). A variable influence is found in observation-space for differing tropopause sharpness and wind situations, averages are calculated for classes of tropopause sharpness (blue = sharp; red = weak) and wind speed (strong wind = blue; weak wind = red) analogously to Sect. 4.3. In addition, a further class is introduced that includes profiles related to the outflow of the WCB (grey), as the previous results suggested a stronger influence in this area.

Figure 4.22a-c show the mean profiles of temperature, N^2 and wind in tropopause-relative coordinates for the N^2 classes, representing differently sharp tropopauses, as described in Sect. 4.3. All temperature and N^2 profiles show the typical characteristics of the tropopause and agree with the observation-space analysis for the whole NAWDEX period in terms of a distinct temperature minimum, a N^2 jump at the tropopause and a temperature inversion above. The profile of the sharp (smooth) class shows a colder (warmer) tropopause, and a stronger (weaker) amplitude of N^2. Concerning wind, the strong wind class, shows a wind profile with a distinct maximum of 52 m s^{-1} just below the tropopause representing the jet stream. The profiles assigned to the WCB outflow show a slightly higher temperature at the tropopause, but a stronger temperature inversion compared to the average over all grid points (black), which is also reflected in a slightly stronger N^2 maximum at the tropopause. The WCB wind profile shows a wind maximum in the UT (of approx. 37 m s^{-1}) which lies in between the maxima of the strong wind class and the average of the full case study.

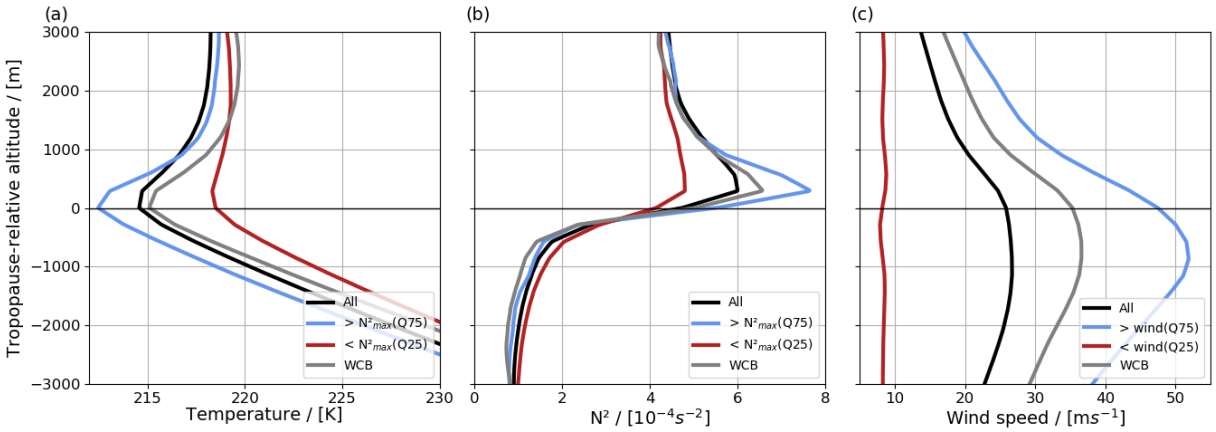

Figure 4.22: Tropopause relative profiles of a) temperature, b) N^2, and c) wind speed. The different colors illustrate averages for the entire map (All; black), and for the WCB outflow area (grey). Temperature and N^2 profiles are shown for sharpest (blue; $N^2_{max,Q75}$) and smoothest (red; $N^2_{max,Q25}$) classified tropopauses. In (c) wind profiles are provided for the strong (blue; $\text{Wind}_{mean,Q75}$) and weak wind (red; $\text{Wind}_{mean,Q25}$) class.

The increments for the different classes are shown in Figure 4.23. All profile classes show a similar vertical distribution of temperature and N^2 increments (Figure 4.23a-b), with a negative increment (cooling) at the tropopause, a positive increment (warming) in the LS, and an increase in N^2 (sharpening) in a 1 km thick layer just above the tropopause. The influence on temperature is stronger for the sharper classified tropopauses. Interestingly, the largest temperature increments are found in the WCB category with a constant cooling (-0.2 K) in the UT, but also an increased warming (0.3 K) approximately between 1-2 km above the tropopause.

Concerning wind speed (Figure 4.23c) the vertical distribution of the wind increments is less similar between the classes. The profiles representing the average over all grid points show entirely positive wind increments (indicating an acceleration) with a maximum of almost 0.2 m s^{-1} in a 300 m layer just above the tropopause. The increments related to the WCB outflow (profiles located east of the jet) are negative in the entire

UT with a maximum (-0.3 m s^{-1}) roughly at 1 km below the tropopause at the level of the maximum winds. Above the tropopause, the WCB increments are weakly positive (approx. 0.1 m s^{-1}) and nearly constant in the LS. For the strong wind class, increments are mainly negative in the UT (up to -0.4 m s^{-1} at jet stream level), and increase upward to a maximum of 0.6 m s^{-1} just above the tropopause. The increments of the class with weak winds show consistently positive values, which are, however, very variable in the profile (between 0 and 0.4 m s^{-1}) with a maximum in the LS (between 1-2 km above the tropopause).

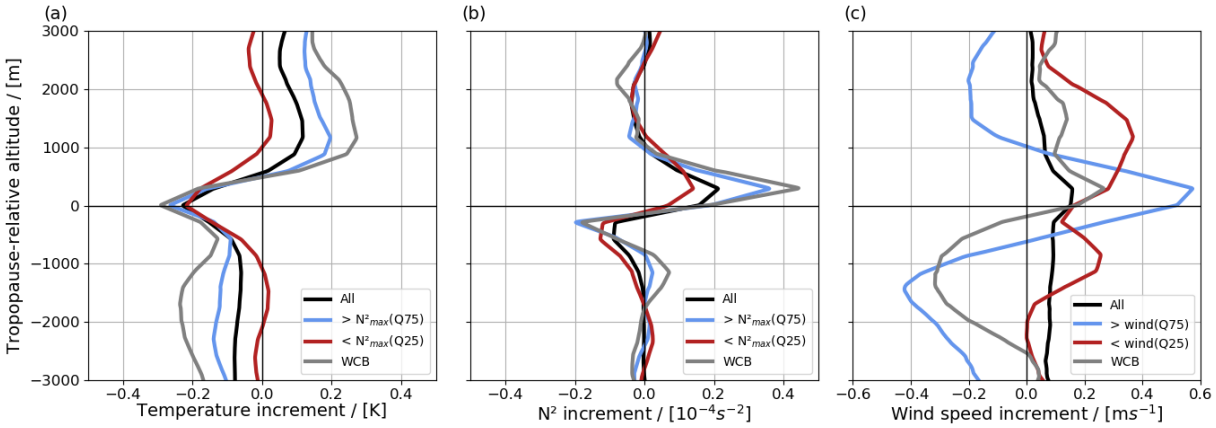

Figure 4.23: Tropopause-relative increments of a) temperature, b) N^2, and c) wind speed for the classes as in Figure 4.22.

4.7 Discussion and concluding remarks

Sect. 4 investigates the influence of data assimilation on the extratropical tropopause in the ECMWF IFS. The evaluation is based on 9729 radiosondes, measured during a one-month autumn 2016 of which 500 are non-operationally launched in the course of the NAWDEX field campaign. The radiosonde data set provides profiles of temperature, static stability, wind, and wind shear that are observed at almost 600 locations covering Europa, the eastern part of North America, and the northern Atlantic. The (thermal) tropopause altitudes derived from the radiosonde temperature profiles lie predominantly in a broad range between 7 and 16 km, which is considered representative of the synoptic situations (i.e., troughs and ridges) characterizing the mid-latitudes and the autumn season (Hoerling et al., 1991, Birner et al., 2002, Bland et al., 2021).

To study the representation of the tropopause structure in the IFS, the observations, and the model equivalents (i.e., analysis and background profiles in observation-space) are transferred to tropopause-relative coordinates. The average profiles of the radiosonde parameters show the characteristic features that are typical of the mid-latitude (e.g., Birner et al., 2002, see also description in Sect. 1.1), and summarized in the following: The temperature profile reaches a minimum at the tropopause, followed by a roughly 1-2 km thick temperature inversion, and a isothermal LS above it. N^2 is low in the UT, and high in the LS with a marked jump of N^2 in between located at the tropopause. Wind speed increases with height in the UT, up to a maximum about 1 km below the tropopause. Above this, and in the LS, the wind decreases constantly with altitude. Correspondingly, vertical wind shear is positive up to the altitude of the wind maximum, and then rapidly drops across the tropopause. Each of these outlined features is also contained in the tropopause-relative profiles of the background, and the analysis tropopause, however, the vertical distributions of the parameters are notably smoother in the background and analysis states (details follow on the next page).

Furthermore, the tropopause height is derived in the background and analysis states. Since the tropopause can be detected at a different height in observation, analysis and background, this raises the question of which tropopause is the most suitable reference for the statistical evaluation. It is found that the tropopause-relative distributions in observations, background, and analysis vary in the different tropopause-relative coordinates due to slightly varying individual tropopause altitudes. Such tropopause altitude differences may result from either misdetections caused by slight temperature fluctuations in the upper troposphere or from differences in the three-dimensional temperature distribution, e.g., in the vicinity of the jet streams. Since the origin of these tropopause altitude differences cannot clearly be identified and they affect the evaluation of the influence on tropopause altitudes (section 4.4), in this study, only those profiles are considered with a tropopause differences in the range of ± 1 km. The sharpest tropopause in the observations, background, and analysis occur when viewed with respect to the "own" tropopause. However, it turned out that the influence on sharpness cannot be assessed in own-tropopause-relative coordinates as different tropopause altitudes in the background and analysis would shift the profiles against each other and thus cause a spurious tropospheric temperature bias. For the evaluation on the data assimilation influence on tropopause sharpness, the observed tropopause is used as reference, as it provides

the most accurate representation of the tropopause. To investigate the influence on the tropopause altitude, increments with respect to the background tropopause are used to study the influence to changes in the background profile. This shows that the tropopause reference system needs to be purposefully and carefully selected, and uncertainties related to tropopause detection need to be considered.

Data assimilation diagnostics i.e., innovation, residuals, and increments are used to compare the tropopause representation in background, analysis and observations with the goal to study the influence of data assimilation on the tropopause sharpness. The statistical comparison of the parameters in the observations, in the background and in the analysis shows systematic deficiencies in the model states resulting in an underestimation of the tropopause sharpness: The temperature innovations indicate a warm bias at the tropopause and a cold bias in the LS. In this layer of the sharp reversal in thermal stratification, N^2 is overestimated (underestimated) by the background below (above) the tropopause confirming findings by Birner et al. (2002). The temperature increments on average draw the background temperature towards the observations by decreasing the tropopause temperature by approx. 0.25 K and increasing it with a similar magnitude in the LS above, which is in line with Radnoti et al. (2010). The accompanied increase in N^2 ($0.3 \times 10^{-4}\,s^{-2}$) in a 1 km layer above the tropopause and a decrease in N^2 ($0.2 \times 10\,10^{-4}\,s^{-2}$) in the uppermost troposphere is equivalent to tropopause sharpening, which is consistent in shape and magnitude with that of Pilch Kedzierski et al. (2016). Although the increments clearly reduce the background temperature biases, the influence is rather small (10 %) compared to the innovations' magnitude. The remaining LS cold bias in the analysis (0.2 K) corresponds to previous assessments (Radnoti et al., 2010) and is driven by radiative cooling due to water vapour (e.g., Shepherd et al., 2018), which is systematically overestimated at those levels (Sect. 3). The warm bias (1 K) at the tropopause in the IFS was related to the finite vertical resolution of the IFS incapable of fully resolving the tropopause (Ingleby et al., 2016), the assimilation of warm-biased aircraft data at tropopause flight levels (Ingleby, 2017) and the moist bias in the LS of the IFS (Bland et al., 2021). The magnitude of the warm bias (about 1.2 K) at the tropopause is about two to three times stronger than the corresponding warm bias reported in (Bland et al., 2021). This difference may be related to the vertical smoothing of the radiosonde profiles in Bland et al. (2021), which could lead to a higher tropopause temperature (König et al., 2019). Large N^2 biases (–2 to $3 \times 10^{-4}\,s^{-2}$) in the analysis are found in the ± 0.5 km layer around the tropopause. In addition, it is found that the magnitude of tropopause sharpening depends on the dynamic situation. For sharper tropopauses, which are typically higher and thus colder predominantly occurring in situations of upper-level ridges (Hoerling et al., 1991, Pilch Kedzierski et al., 2015), temperature (and also N^2) increments, innovations and residuals are larger.

Positive wind innovations (approx. $1\,m\,s^{-1}$ near the tropopause) show a slow wind bias in the background, particularly for the wind maximum, which is in line with Schäfler et al. (2020) and Lavers et al. (2023). The observed vertical wind shear profile is characterised by positive values below and negative values above the wind maximum, as well as by a sharp increase in negative shear across the tropopause. The enhanced (negative) shear in the 1 km layer above the tropopause in the observations is also present in the ECMWF which is consistent with previous findings (Schäfler et al., 2020, Kaluza et al., 2021). Its

magnitude, however, is considerably weaker in the background and analysis as compared with the observations. Positive wind speed increments in the UT with a peak at the tropopause ($0.2\,\mathrm{m\,s^{-1}}$) are found, leading to a corresponding acceleration of wind speed and nearly unchanged winds in the LS. The wind shear increments are positive just below the tropopause and negative in a 1 km layer above the tropopause. The generally positive influence of data assimilation on the wind profile at all altitudes is depicted by smaller residuals than innovations for all wind speed classes. However, it is found that high wind speed situations are characterised by an increased wind bias in the background around the tropopause (underestimation of $1.2\,\mathrm{m\,s^{-1}}$) which is reduced by about 40 % in the analysis. This affirms Schäfler et al. (2020) who speculated that large wind errors near the jet stream in IFS short-range forecasts are reduced in the analysis through data assimilation. The stronger positive impact on wind for high wind speed situations was recently demonstrated by Lavers et al. (2023).

This evaluation is certainly affected by representativeness errors when comparing point measurements to grid-average NWP output values as discussed in (e.g., Weissmann et al., 2005, Hodyss and Nichols, 2015, Janjić et al., 2018). Such an effect could be partially addressed through vertical averaging of the profiles; however, the vertical resolution of the assimilated data (100–400 m; see Figure 4.2) is already close to the model grid spacing in the UTLS (300 m). Another contribution to the representativeness error is the horizontal displacement of the radiosonde during its ascent, which can exceed 100 km from the station in strong wind situations. It should be mentioned that the used assimilation system does not consider the horizontal drift. Voggenberger et al. (2023) show that the consideration of radiosonde drift can lead to a reduced representativeness error and thus to a stronger influence of the observations on the analysis in the current cycle.

In addition to the influence of sharpening, a positive influence of data assimilation is revealed on the representation of the tropopause height in the analysis, which is closer to the observations than the background. In cases of tropopause altitude differences (innovation > 100 m), the analysis shows a systematically improved tropopause altitude whereby the improvement grows with increasing innovation. The vertical shift in the temperature increments relative to the background tropopause agrees with the resulting tropopause altitude changes in the analysis: if the observed tropopause lies below the background tropopause, the region below is cooled, which leads to a lower LRT in the analysis. In contrast, if the observed LRT is located above the background, the region above is cooled, which, on average, shifts the analysis LRT upwards. Bland et al. (2021) and Schmidt et al. (2010) show that local changes in a temperature profile in the UTLS affect tropopause altitude. For instance, cooling of the LS and warming of the UT leads to higher tropopause altitude in models. The opposite effect is true in cases of cooling of the UT and warming of the LS.

The analysis of a subset of 497 NAWDEX profiles considered in a data denial OSE allowed the sharpening influence and the positive influence on the tropopause altitude to be attributed directly to radiosondes data assimilation. The CTR run increments which assimilated NAWDEX radiosondes showed a shape and magnitude similar to those of the full data set. The increments in the DEN run, where the non-operational radiosondes were only passively monitored, are much weaker, but the positive and negative increments were pointing in the same direction as those of the CTR run. This provides evidence that the

assimilation of the radiosondes provides the major contribution to the increments (and thus to the positive influence on tropopause sharpness and altitude), which most likely holds true for the entire data set of the presented results. The non-zero increments in the DEN run might be related to the assimilation of other observations, for instance, GPS-RO data (Pilch Kedzierski et al., 2016), or to the contribution of the routine radiosondes and aircraft data that are assimilated in the same assimilation time window at a nearby location. A more sophisticated OSE with more observations and different observation types to be denied would be required for a deeper investigation of this effect.

The approach of assessing an OSE in observation-space allows only to evaluate the influence of the observations on temperature and wind distributions at the observation location. In hybrid 4D-VAR schemes the B-matrix spreads information of assimilated observations also horizontally and vertically in space. This poses the question as to which extent the sharpening influence on the temperature and wind gradients in the UTLS affects tropopause sharpness not only locally but also in the surrounding region in the model. To address this question, a model-space case study is used to study the influence of data assimilation beyond the location of the radiosondes. This case represents typical NAWDEX situation with a ridge over Europe, a trough over the Atlantic with diabatic active regions (fronts, WCB), and a strong jet stream allowing to view the influence of data assimilation also in the relation to different dynamic features.

The temperature, N^2 and wind increments are first averaged for each profile in certain vertical layers in which the observation-space has shown a systematic influence of data assimilation. At the tropopause, dominating negative temperature increments indicate a cooling (of up to -2 K) almost throughout the entire map section. In the LS, positive increments show a strong warming especially on the anti-cyclonic side of the jet stream, while the cyclonic side is rather cooled. The influence on the temperature is reflected in a widespread sharpening of the tropopause which is mirrored in the predominant positive N^2 increments in the LS. This sharpening particularly pronounced in the diabatically active regions associated with the WCB outflow, in which a strong cooling the tropopause coincides with a marked warming of the LS. The geographical distribution of UT wind increments, show a particularly strong influence in the vicinity of the jet stream. The altitude-longitude cross-section reveals dipole-shaped increments with an strong decrease of wind in the jet stream's core and on its western flank, while wind speed at its eastern flank is increased. This indicates a slight eastward shift (about 0.25°; 20 km) of the jet stream in the analysis through data assimilation. The strong influence on the jet stream could indicate the presence of strong wind errors in this regions, as strong innovations were found in observation-space. The overall spread tropopause sharpening demonstrates that the influence diagnosed in the observation-space is not only a local phenomenon. It has to be noted here, that model-space increments reflect the joint influence caused by the assimilation of all observations, which means that the non-local sharpening character cannot be attributed to the radiosonde solely which was indicated by the non-zero increments in the OSE.

The model-space increments are additionally presented in tropopause-relative coordinates. The corresponding temperature and N^2 increments averaged over the full map section confirm a cooling at the tropopause, a warming of the LS, and thus, a sharpening of the tropopause which was diagnosed in observation-space. Furthermore, positive

wind increments across the entire UTLS show an acceleration of wind with a maximum in a 300 m layer just above the tropopause. This presented model-space influence on temperature and wind roughly halved in magnitude compared to the observation-space increments.

The tropopause-relative increments are also analysed for different tropopause classes (sharp/smooth tropopauses; strong/weak winds), and separately for profiles associated with the WCB outflow. There is a stronger influence on temperature and N^2 for particularly sharp tropopauses. A strong influence on wind is detected for the strongest wind class representing regions of the strong jet stream. Interestingly, the negative wind increments at the altitude of the wind maximum show a weakening of the jet, while positive increments at the tropopause above indicate an acceleration. The tropopause-relative increments averaged over the profiles assigned to the WCB outflow show the strongest sharpening of the tropopause. It is striking that mean wind profile in the WCB is slowed in the entire UT.

The observation-space analysis has shown that the influence (i.e., the increments) is especially strong when the errors in the background are also increased. However, the model-space data does not provide information on background error, which makes it more difficult to understand the influence of data assimilation. Nevertheless, the particularly strong influence on temperature and wind in the WCB outflow could be an indication of potentially increased background errors in this region. Air masses associated with the WCB outflow have experienced strong diabatic modification in the recent history e.g. through cloud microphysics and radiative processes. These processes are parameterized, and thus potential sources of errors in the background forecast.

Although the model-space case study represents a typical synoptic situation in the mid-latitudes, an extension of this study with more cases is required to find a robust evidence of a systematic influence in certain dynamic features (e.g. jet stream or regions influenced by diabatic processes, such as WCBs). Schäfler et al. (2024) demonstrate for dry intrusions how the data assimilation influence in certain atmospheric features can be particularly addressed in observation-space. A key advantage here is that the influence can also be viewed in the context of innovations and residuals. However, in order to robustly evaluate the observation-space influence on the jet stream or WCBs, a large number of observations and careful matching with the targeted structures is necessary.

Chapter 5

Conclusion

To further improve the weather forecast, both, model errors and systematic errors in the initial conditions in the NWP model must be reduced. The overarching scope of this dissertation is to identify and quantify systematic initial condition errors (i.e. biases) in the ECMWF IFS. In this work, the focus is on the extratropical tropopause and the surrounding layer (the UTLS) where distinct vertical gradients of temperature and wind exist that create sharp meridional PV gradients. These PV gradients act as a waveguide for the Rossby-waves determining the weather downstream in the midlatitudes (Schwierz et al., 2004, Martius et al., 2010, Harvey et al., 2016). To characterise the signature of model biases at the tropopause, high-resolution and precise observations are demanded that are able to capture the sharp cross-tropopause gradients, but the current global observing system lacks such observations. This work makes use of a unique campaign dataset comprising radiosonde temperature and wind observations, and water vapour lidar data that meet these requirements. This data set includes both model-independent and dependent (assimilated) data, offering the opportunity to address open questions on initial condition biases near the tropopause in a combined approach from a novel view.

5.1 Main Conclusions

Evaluation of the lower-stratospheric moist bias in the IFS
The first study focuses on the LS moist bias, a systematic humidity error that has been an unsolved problem in ECMWF analyses and reanalyses for decades. There were still open questions about bias, in particular about its vertical structure and the underlying causes. Currently, there was -up to this thesis- no observation technique available that combines sufficient resolution and measurement precision to address these open aspects. The water vapour DIAL data set used in this analysis provides a high number of precise and highly-resolved humidity profiles with high coverage of the UTLS measured during six research campaigns between 2013-2021 in the midlatitudes. The campaigns cover three seasons (summer, autumn, winter) and a wide range of synoptic situations in the extratropics, which enables to address the seasonal and synoptic variability of the humidity bias. The DIAL data set also holds the advantage of not being assimilated by the ECMWF model, thus allows humidity to be evaluated independently and to tackle three open aspects of the LS moist bias (RQ1.1-RQ1.3) which are recalled below.

- **RQ1.1 What is the vertical structure of the moisture bias in the LS?**
- **RQ1.2 Is the bias variable for different seasons and synoptic situations?**
- **RQ1.3 Is the moist bias related to mixing processes?**

The **first research question** (**RQ1.1**) addressed a potential lower-stratospheric moisture bias in the ERA5, whose vertical structure and magnitude had been debated in recent studies. Based on the DIAL observations, it was revealed that a systematic LS moist bias is also contained in the ECMWF's most recent global reanalysis ERA5. It was demonstrated that the vertical structure of the bias, is characterised by a weak positive bias in the upper troposphere (15–20 %) and a strong overestimation of humidity that reaches a maximum (55 %) at 1.3 km above the thermal tropopause. Above this maximum, a steady vertical decline of the moist bias towards a constant small value (15 %) beyond 4 km above the tropopause is found. The high vertical resolution and precision of the DIAL profiles combined with a large data coverage across the UTLS provide robust evidence of the decline of the moist bias (towards 4 km altitude) where other studies reported opposing results.

The **second research question** (**RQ1.2**) was concerned with the seasonal and synoptic variability of the LS moist bias. Along the individual lidar cross sections, the LS moist bias occurs in coherent and extended structures following the tropopause, whereby it is noticeable that the layer of the strongest bias varies in thickness. A vertically deeper moist bias is found above low located tropopauses (typically related to trough situations) compared to high located tropopauses (associated with ridge situations). Interestingly, a similar magnitude of the moist bias was observed regardless of the tropopause height. The comparison of the moist bias for the different campaigns showed a seasonality of its magnitude with a maximum during summer where it is 2-3 times stronger than in winter. The uncovered seasonal and synoptically variable vertical structure raised the question of whether mixing processes overestimated in the model play a role in the formation of the bias.

To investigate a potential connection between the LS moist bias and mixing processes in the **third research question** (**RQ1.3**), four flights with collocated ozone and water vapour DIAL observations, available during the WISE field campaign, were examined. The collocated measurements allowed the UTLS air masses to be separated in tropospheric, stratospheric and mixed air. For the first-time it was shown, that the strongest LS bias is located at altitudes dominated by the mixed air class, while the tropospheric or stratospheric air exhibits a smaller moist bias. From this correlation it can be clearly deduced that overestimated mixing processes or numerical diffusion shapes the vertical structure of the LS moist bias which corroborated a previous hypothesis on its potential origin.

An obvious process that is known for the extensive vertical, cross-tropopause transport of water vapour from the UT into the LS is overshooting convection. Convection is a sub-grid scale process represented in the model by a parameterization. Overestimated overshooting convection has been suggested as a potential cause of the LS moist bias, but a possible correlation has never been specifically investigated. In a separate evaluation, the DIAL data was used to test whether the LS moist bias is sensitive to the

representation of overshooting convection in the model. For this purpose, a flight from the CIRRUS-HL measurement campaign in summer was selected, during which strong convection was measured over the Alpine region. During this period two versions (i.e. cycles) of the IFS were available, which treat overshooting convection differently. For this individual case, a substantially reduced LS moist bias is found for the IFS cycle in which convective overshoot is limited. The analysis showed that the DIAL data are highly useful to specifically analyse consequences of adjusted parameterizations in NWP models by validating model biases independently.

This work demonstrated the value of the unique water vapour DIAL data to effectively locate and quantify systematic moisture biases at the tropopause. The LS moist bias is a long-known issue in the ECMWF and one of the key systematic errors to be corrected as stated in a report by a stratospheric task force (Shepherd et al., 2018). The results of this work concerning the vertical structure, the seasonal and synoptical variability of the bias and its relation to mixing processes provide a basis towards a reduction the systematic humidity error in the LS, and thus also to correct for the collocated temperature bias in the future.

Influence of assimilated radiosonde observations on the sharpness and altitude of the mid-latitude tropopause

The second study applies data assimilation output with the goal to evaluate the influence of assimilated radiosonde observations on the representation of the tropopause in the ECMWF IFS. A data assimilation system should ideally correct for erroneous vertical distributions of temperature and wind in the background forecast and thus improve the analysis. However, it was unclear whether data assimilation smoothens or sharpens the strong vertical gradients at the tropopause. In this study, the different aspects on the potential influence of radiosonde data assimilation on the sharpness and altitude of the mid-latitude tropopause were addressed (see RQ2.1-2.3 below). For this purpose, a large data set of radiosonde observations during the one-month NAWDEX period in fall 2016 was compared with observation-space background and analysis profiles. Radiosondes record high-resolution vertical profiles of e.g., temperature and wind from the surface up to the stratosphere with high accuracy. This allows to study the sharp temperature and wind gradients at the tropopause, and their representation in NWP background and analysis fields.

- **RQ2.1 How are the vertical temperature and wind gradients at the tropopause represented in the IFS background and what is the influence of data assimilation on the analysis?**
- **RQ2.2 Can the diagnosed influence be attributed to the assimilation of radiosondes?**
- **RQ2.3 What is the influence of data assimilation on the tropopause beyond the observation location?**

In the **first research question** (**RQ2.1**) it was investigated how the vertical temperature and wind gradients at the tropopause as observed by the radiosondes are represented in the IFS background and analysis. The statistical evaluation in tropopause-relative coordinates showed the presence of systematic temperature and wind biases in the background corresponding to a less sharp tropopause compared to the observations. The vertical distribution of the temperature, N^2, wind and wind shear increments revealed that the increments counteract the background biases and, on average sharpen the analysis tropopause. This sharpening is expressed by an cooling (-0.25 K) at the tropopause, a warming (-0.25 K) of the LS and by an increase of N^2 in a 1 km layer just above the tropopause. Furthermore, an acceleration of wind speed ($0.2\,\mathrm{m\,s^{-1}}$) is detected which is most pronounced at the altitude of the highest observed wind speeds. The sharp contrast in wind shear from positive values below and negative values above the wind maximum, and especially in the lowermost 1 km of the LS, were increased by data assimilation. For each parameter, the increments sharpened the tropopause; however, the average influence was small (10-20 %) compared with the magnitude of the background biases.

A further original finding was a variable influence for different dynamical situations. Larger increments, but also larger biases in background and analysis, were related to sharper tropopauses associated with ridge situations (high tropopause), while a weaker influence was observed for smoother tropopauses associated with troughs (low tropopause).

Concerning wind, a large influence was found for strong jet stream wind situations, with a reduction in both, the slow wind bias in the UT by about 40 % and, of a similar magnitude, the wind shear error in the 1 km layer above the tropopause.

In addition, the influence of data assimilation particularly on the tropopause altitude for was addressed for the first time. For about 41 % of all profiles, the observed tropopause altitude agreed with the background and analysis tropopause altitude within ± 100 m. In case of larger tropopause altitude differences, the analysis provided evidence that radiosonde data assimilation, on average, positively influences the tropopause altitudes. This improvement of the tropopause altitude was explained by a systematic shift of the temperature increments relative to the background tropopause, which act to draw the background tropopause towards the observed tropopause.

The **second research question** (**RQ2.2**) was concerned with whether the diagnosed influence can be attributed to the assimilation of the radiosondes. For this investigation an OSE with 497 NAWDEX non-operational radiosonde which were assimilated in a CTR and denied from assimilation in a DEN run, was analysed. The evaluation of the increments in the OSE revealed, that the diagnosed influence on temperature and wind, as well as the positive influence on the tropopause altitude can be mainly attributed to the assimilated radiosondes. Interestingly the substantially weaker, but non-zero increments in the DEN run indicated that other assimilated observations also contribute to the sharpening influence and to the increase of UT winds at the radiosonde locations. This finding provides motivation for an investigation of the non-local influence of data assimilation on the tropopause.

The **third research question** (**RQ2.3**) investigated whether the in observation-space diagnosed local influence of data assimilation on the tropopause is also present beyond the radiosonde locations. For this purpose, tropopause-relative increments were considered in model-space for an example case study with synoptic features representative of mid-latitude weather. The average increments in model-space demonstrated a sharpening of the tropopause to be present even beyond the location of the radiosonde location. The increments in model-space were about half as large as in the observation-space. The evaluation also showed that the influence of data assimilation in model-space is rather highly variable with the dynamic situation. A particularly pronounced tropopause sharpening was shown at the anti-cyclonic side of the jet at high tropopause altitudes. In addition, a strong influence on the jet stream was found, which was slightly shifted by the data assimilation influence in this evaluated case study. Furthermore, a notably strong influence (i.e., strongest sharpening, strong decrease of wind in the UT) was revealed in regions of the WCB outflow associated with high diabatic activity, which is a stark indication of increased model errors in these regions.

5.2 Implications for future research

Overall, the combined approach of independent campaign data and data assimilation diagnostics applied in this work contributes to important advances in the understanding of systematic biases at the tropopause and their impact on the tropopause representation in a state-of-the-art NWP model. The data and methods developed within this work are applicable to further deepen the knowledge on model biases in future.

The high value of the independent DIAL measurements must be emphasised, which are ideally suited to characterise water vapour model biases in regions where other observing techniques are limited. Additional campaign data, for example in spring or in the Southern Hemisphere, could provide further insight on processes that play a role in the formation on the moist bias. The analysis has also shown that the DIAL data can also be used directly to evaluate particular adjustments in parameterization in NWP models as demonstrated for the revision of the convection scheme. DIAL data from future campaign missions such as the North Atlantic Waveguide, Dry Intrusion, and Downstream Impact Campaign (NAWDIC) could and should be used in the future to identify model biases and to track progress in NWP model developments. It has to be noticed that NWP models suffer from the lack of reliable (accurate and vertically resolved) water vapour observations in the entire atmosphere, that could be directly assimilated to further understand systematic moisture errors. The findings of this thesis motivates the great expected benefit of a potential water vapour satellite mission equipped with an active DIAL for the further development of NWP models.

In the second part of the thesis, the evaluation of data assimilation output contributed to a deeper understanding on temperature and wind biases on the tropopause, and on the influence of assimilated radiosonde observations on an improved tropopause representation in the analysis.

The developed novel approach of a feature-based (tropopause-based) analysis was combined with an OSE is also applicable for prospective studies to address model biases. In this work, an increased data assimilation influence is found in certain dynamic situations such as fronts and WCBs which are associated with high diabatic activity and thus, expected to be regions of increased initial condition errors. A careful analysis of model-space or observation-space increments in these regions in a future study could shed light on the representation of diabatic processes and provide the basis for further improving the initial conditions there. In addition, a strong influence of (radiosonde) data assimilation on wind is shown in this dissertation. In fact, there are no global wind profile measurements available to models for the assimilation (at least since the re-entry of the successful Aeolus mission). Hence, it would be worth investigating whether the assimilation of other parameters has a (positive) influence on the wind field. Another interesting target is the Aeolus wind satellite mission, which has been shown to have a large influence especially in the tropics, but not yet in detail for the midlatitudes. A jet stream-relative analysis, combined with an OSE, would allow to better quantify the influence of assimilated Aeolus wind data in the midlatitudes. The systematic biases in the UTLS in ECMWF diagnosed in this work lead to an incorrect representation of the sharp vertical temperature and wind gradients at the tropopause, but data assimilation is able to sharpen these gradients. From this it could be expected that data assimilation also improves the PV

distribution across the tropopause that is highly relevant to the mid-latitude dynamics. Future studies should also focus on the influence of data assimilation on the synoptic-scale PV distribution, and their impact on forecasts to particularly estimate the added value of the assimilated observations for the predictability of weather in midlatitudes.

Appendix A

Contribution of journal publications to this dissertation

This dissertation contains results from the two journal articles: Krüger et al. (2022) and Krüger et al. (2024). In the following table an overview of the contribution of each article to the individual sections of the thesis is given. The listed sections are in part excerpts from the corresponding publications.

Table A.1: Contribution of journal publications to this dissertation.

Krüger et al. (2022)	**Krüger et al. (2024)**
Sect. 1.2-1.3	Sect. 1.2-1.3
Sect. 2.1	Sect. 2.2
Sect. 3.1-3.4, Sect. 3.6	Sect. 4.1-4.5, Sect. 4.7
Sect. 5.1	Sect. 5.1

Abbreviations and Symbols

List of Abbreviations

Abbreviation	Long form
BUFR	**B**inary **U**niversal **F**ormat for the **R**epresentation of meteorol. data
CIRRUS-HL	**CIRRUS** in **H**igh-**L**atitudes
CMAM	**C**anadian **M**iddle **A**tmosphere **M**odel
CTR	**C**on**TR**ol run
DEN	**DEN**ial run
DIAL	**DI**fferential **A**bsorption **L**idar
DLR	**D**eutsches Zentrum für **L**uft- und **R**aumfahrt
ECMWF	**E**uropean **C**entre for **M**edium-Range **W**eather **F**orecasts
EDA	**E**nsemble of **D**ata **A**ssimilation
EDMO	ICAO-Code: Airport Oberpfaffenhofen
ENS	**EN**nsemble forecast **S**uite
EUREC^{4}A	**E**l**U**cidating the **R**ol**E** of **C**loud-Circulation Coupling in Clim**A**te
GFS	**G**lobal **F**orecast **S**ystem
GPS-RO	GPS **R**adio **O**cculation
HALO	**H**igh **A**ltitude and **LO**ng Range Research Aircraft
HIRTRAN	**HI**gh-**R**esolution **TRAN**smission molecular absorption
HRES	**H**igh-**RES**olution
HSRL	**H**igh **S**pectral **R**eso**L**ution
ICON	**ICO**sahedral **N**onhydrostatic Model
IFS	**I**ntegrated **F**orecasting **S**ystem
IR	**I**nfra**R**ed
IQR	**I**nter**Q**uartile **R**ange
Laser	**L**ight **a**mplification by **s**timulated **e**mission of **r**adiation
Lidar	**Li**ght **d**etection **a**nd **r**anging)
LRT	**L**apse-**R**ate based **T**ropopause
LWDA	**L**ong **W**indow **D**ata **A**ssimilation
NARVAL	**N**ext-generation **A**ircraft **R**emote sensing for **VAL**idation studies
NARVAL2	**N**ext-generation **A**ircraft **R**emote sensing for **VAL**idation studies-part **2**
NAWDEX	**N**orth **A**tlantic **W**aveguide and **D**ownstream **EX**periment
NAWDIC	**N**orth **A**tlantic **W**aveguide, **D**ry intrusion, and Downstream **I**mpact **C**ampaign
Nd:YAG	**N**eodymium-**d**oped **Y**ttrium **A**luminium **G**arnet, $Y_3Al_5O_{12}$

Abbreviation	Long form
NWP	**N**umerical **W**eather **P**rediction
METUM	**MET** Office's **U**nified **M**odel
OPO	**O**ptical **P**arametric **O**scillator
OSE	**O**bserving **S**ystem **E**xperiment
PDF	**P**robability **D**ensity **F**unction
PV	**P**otential **V**orticity
PVU	**P**otential **V**orticity **U**nits
RQ	**R**esearch **Q**uestion
SCDA	**S**hort **C**utoff **D**ata **A**ssimilation
SEVIRI	**S**pinning **E**nhanced **V**isible and **I**nfra-**R**ed **I**mager
SHG	**S**econd **H**armonic **G**enerator
TIL	**T**ropopause **I**nversion **L**ayer
TSL	**T**ropopause **S**hear **L**ayer
UTC	**U**niversal **T**ime **C**oordinated
UTLS	**U**pper **T**roposphere and **L**ower **S**tratosphere
UV	**U**ltra**V**iolet
WALES	**WA**ter vapor **L**idar **E**xperiment in **S**pace
WCB	**W**arm **C**onveyor **B**elt
WISE	**W**ave-driven **IS**entropic **E**xchange
WMO	**W**orld **M**eteorological **O**rganization
3D-VAR	**3**-**D**imensional **VAR**iational data assimilation
4D-VAR	**4**-**D**imensional **VAR**iational data assimilation

List of Figures

List of Tables

Bibliography

Abbe, C., 1901: The physical basis of long-range weather forecasts. *Monthly Weather Review*, **29 (12)**, 551–561, https://doi.org/10.1175/1520-0493(1901)29[551c:TPBOLW]2.0.CO;2.

Assmann, R., 1902: *Über die Existenz eines wärmeren Luftstromes in der Höhe von 10 bis 15 km*, Vol. 24. Sitzungsberichte der Königlichen Preussischen Akadademie der Wissenschaften zu Berlin, 495–504 pp.

Bannister, R. N., 2008: A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, **134 (637)**, 1951–1970, https://doi.org/10.1002/qj.339.

Bannister, R. N., 2017: A review of operational methods of variational and ensemble-variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **143 (703)**, 607–633, https://doi.org/10.1002/qj.2982.

Bannister, R. N., H. G. Chipilski, and O. Martinez-Alvarado, 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, https://doi.org/10.1002/qj.3652.

Bauer, P., A. Thorpe, and G. Brunet, 2015: The quiet revolution of numerical weather prediction. *Nature*, **525 (7567)**, 47–55, https://doi.org/10.1038/nature14956.

Bechtold, P. and R. Forbes, 2024: "the convective modification, i.e., the limit of the convective overshoots at the tropospause was provided to the ECMWF by Günther Zängl from DWD", private communication at 2024-04-04.

Beckert, A. A., L. Eisenstein, A. Oertel, T. Hewson, G. C. Craig, and M. Rautenhaus, 2023: The three-dimensional structure of fronts in mid-latitude weather systems in numerical weather prediction models. *Geoscientific Model Development*, **16 (15)**, 4427–4450, https://doi.org/10.5194/gmd-16-4427-2023.

Bhawar, R., P. Di Girolamo, D. Summa, C. Flamant, D. Althausen, A. Behrendt, C. Kiemle, P. Bosser, M. Cacciani, C. Champollion, et al., 2011: The water vapour intercomparison effort in the framework of the Convective and Orographically-induced Precipitation Study: airborne-to-ground-based and airborne-to-airborne lidar systems.

Quarterly Journal of the Royal Meteorological Society, **137 (S1)**, 325–348, https://doi.org/10.1002/qj.697.

Birner, T., 2006: Fine-scale structure of the extratropical tropopause region. *Journal of Geophysical Research: Atmospheres*, **111 (D4)**, https://doi.org/10.1029/2005JD006301.

Birner, T., A. Dörnbrack, and U. Schumann, 2002: How sharp is the tropopause at midlatitudes? *Geophysical research letters*, **29 (14)**, 45–1, https://doi.org/10.1029/2002GL015142.

Birner, T., D. Sankey, and T. Shepherd, 2006: The tropopause inversion layer in models and analyses. *Geophysical research letters*, **33 (14)**, https://doi.org/10.1029/2006GL026549.

Bjerknes, V., 1904: Das Problem der Wettervorhersage, betrachtet vom Standpunkte der Mechanik und der Physik. *Meteor. Z.*, **21**, 1–7, https://doi.org/10.1127/0941-2948/2009/416.

Bland, J., S. Gray, J. Methven, and R. Forbes, 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, https://doi.org/10.1002/qj.4150.

Bolin, B., 1955: Numerical forecasting with the barotropic model 1. *Tellus*, **7 (1)**, 27–49, https://doi.org/10.1111/j.2153-3490.1955.tb01139.x.

Boljka, L. and T. Birner, 2022: Potential impact of tropopause sharpness on the structure and strength of the general circulation. *npj Climate and Atmospheric Science*, **5 (1)**, 98, https://doi.org/10.1038/s41612-022-00319-6.

Bonavita, M., 2014: On some aspects of the impact of GPSRO observations in global numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, **140 (685)**, 2546–2562, https://doi.org/10.1002/qj.2320.

Bonavita, M., E. Hólm, L. Isaksen, and M. Fisher, 2016: The evolution of the ECMWF hybrid data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **142 (694)**, 287–303, https://doi.org/10.1002/qj.2652.

Borne, M., P. Knippertz, M. Weissmann, A. Martin, M. Rennie, and A. Cress, 2023: Impact of Aeolus wind lidar observations on the representation of the West African monsoon circulation in the ECMWF and DWD forecasting systems. *Quarterly Journal of the Royal Meteorological Society*, **149 (752)**, 933–958, https://doi.org/10.1002/qj.4442.

Bösenberg, J., 1998: Ground-based differential absorption lidar for water-vapor and temperature profiling: methodology. *Applied Optics*, **37 (18)**, 3845–3860, https://doi.org/10.1364/AO.37.003845.

Bouttier, F. and P. Courtier, 2002: Data assimilation concepts and methods March 1999. *Meteorological training course lecture series. ECMWF*, **718**, 59, https://twister.caps.ou.edu/OBAN2019/Assim_concepts.pdf, last access: 2024-04-20.

Bouttier, F. and G. Kelly, 2001: Observing-system experiments in the ECMWF 4D-Var data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **127 (574)**, 1469–1488, https://doi.org/10.1002/qj.49712757419.

Bowman, K. P., 1996: Rossby wave phase speeds and mixing barriers in the stratosphere. Part I: Observations. *Journal of Atmospheric Sciences*, **53 (6)**, 905–916, https://doi.org/10.1175/1520-0469(1996)053<0905:RWPSAM>2.0.CO;2.

Bukenberger, M., S. Rüdisühli, and S. Schemm, 2023: Jet stream dynamics from a potential vorticity gradient perspective: The method and its application to a kilometre-scale simulation. *Quarterly Journal of the Royal Meteorological Society*, **149 (755)**, 2409–2432, https://doi.org/10.1002/qj.4513.

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

Chagnon, J., S. Gray, and J. Methven, 2013: Diabatic processes modifying potential vorticity in a North Atlantic cyclone. *Quarterly Journal of the Royal Meteorological Society*, **139 (674)**, 1270–1282, https://doi.org/10.1002/qj.2037.

Craig, G. C., A. H. Fink, C. Hoose, T. Janjić, P. Knippertz, A. Laurian, S. Lerch, B. Mayer, A. Miltenberger, R. Redl, et al., 2021: Waves to Weather: Exploring the limits of predictability of weather. *Bulletin of the American Meteorological Society*, **102 (11)**, E2151–E2164, https://doi.org/10.1175/BAMS-D-20-0035.1.

Davis, S. M., M. I. Hegglin, M. Fujiwara, R. Dragani, Y. Harada, C. Kobayashi, C. Long, G. L. Manney, E. R. Nash, G. L. Potter, et al., 2017: Assessment of upper tropospheric and stratospheric water vapor and ozone in reanalyses as part of S-RIP. *Atmospheric Chemistry and Physics*, **17 (20)**, 12 743–12 778, https://doi.org/10.5194/acp-17-12743-2017.

Dee, D. P., S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. Balmaseda, G. Balsamo, d. P. Bauer, et al., 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the royal meteorological society*, **137 (656)**, 553–597, https://doi.org/10.1002/qj.828.

Dekoutsidis, G., S. Groß, M. Wirth, M. Krämer, and C. Rolf, 2023: Characteristics of supersaturation in midlatitude cirrus clouds and their adjacent cloud-free air. *Atmospheric Chemistry and Physics*, **23 (5)**, 3103–3117, https://doi.org/10.5194/acp-23-3103-2023.

Dessler, A. and S. Sherwood, 2004: Effect of convection on the summertime extratropical lower stratosphere. *Journal of Geophysical Research: Atmospheres*, **109 (D23)**, https://doi.org/10.1029/2004JD005209.

Diamantakis, M. and J. Flemming, 2014: Global mass fixer algorithms for conservative tracer transport in the ECMWF model. *Geoscientific Model Development*, **7 (3)**, 965–979, https://doi.org/10.5194/gmd-7-965-2014.

Dyroff, C., A. Zahn, E. Christner, R. Forbes, A. M. Tompkins, and P. F. van Velthoven, 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, https://doi.org/10.1002/qj.2400.

ECMWF, 2016a: *IFS Documentation CY43R1 - Part II: Data Assimilation*. European Centre for Medium-Range Weather Forecasts, https://doi.org/10.21957/am5dtg9pb, last access: 2024-05-24.

ECMWF, 2016b: *IFS Documentation CY43R1 - Part III: Dynamics and Numerical Procedures*. European Centre for Medium-Range Weather Forecasts, https://doi.org/10.21957/m1u2yxwrl, last access: 2024-05-24.

ECMWF, 2016c: *IFS DOCUMENTATION – Cy41r2 - PART III: DYNAMICS AND NUMERICAL PROCEDURES*. European Centre for Medium-Range Weather Forecasts, https://doi.org/10.21957/83wouv80, last access: 2024-04-26.

Einstein, A., 1916: Zur Quantentheorie der Strahlung. *Physikalische Gesellschaft Zürich*, **18**, 47–62, https://ui.adsabs.harvard.ed/abs/1916PhyGZ..18...47E.

Ertel, H., 1942: Ein neuer hydrodynamischer Wirbelsatz. *Meteorologische Zeitschrift*, **59 (9)**, 277–281.

Esselborn, M., M. Wirth, A. Fix, M. Tesche, and G. Ehret, 2008: Airborne high spectral resolution lidar for measuring aerosol extinction and backscatter coefficients. *Applied optics*, **47 (3)**, 346–358, https://doi.org/10.1364/AO.47.000346.

Evensen, G., F. C. Vossepoel, and P. J. van Leeuwen, 2022: *Data assimilation fundamentals: A unified formulation of the state and parameter estimation problem*. Springer Nature, ISBN:978-3-030-96709-3.

Ferreira, A., J. Castanheira, and L. Gimeno, 2016: Water vapour stratification and dynamical warming behind the sharpness of the Earth's midlatitude tropopause. *Quarterly Journal of the Royal Meteorological Society*, **142 (695)**, 957–970, https://doi.org/10.1002/qj.2697.

Fix, A., F. Steinebach, M. Wirth, A. Schäfler, and G. Ehret, 2019: Development and application of an airborne differential absorption lidar for the simultaneous measurement of ozone and water vapor profiles in the tropopause region. *Applied optics*, **58 (22)**, 5892–5900, https://doi.org/10.1364/AO.58.005892.

Francis, D. J., A. M. Fowler, A. S. Lawless, J. Eyre, and S. Migliorini, 2023: The effective use of anchor observations in variational bias correction in the presence of model bias. *Quarterly Journal of the Royal Meteorological Society*, **149 (754)**, 1789–1809, https://doi.org/10.1002/qj.4482.

Gettelman, A., P. Hoor, L. Pan, W. Randel, M. I. Hegglin, and T. Birner, 2011: The extratropical upper troposphere and lower stratosphere. *Reviews of Geophysics*, **49 (3)**, https://doi.org/10.1029/2011RG000355.

Gimmestad, G. G. and D. W. Roberts, 2023: *Lidar Engineering: Introduction to Basic Principles*. Cambridge University Press, ISBN:978-0-521-19851-6.

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

Grise, K. M., D. W. Thompson, and T. Birner, 2010: A global survey of static stability in the stratosphere and upper troposphere. *Journal of Climate*, **23 (9)**, 2275–2292, https://doi.org/10.1175/2009JCLI3369.1.

Groß, S., M. Esselborn, B. Weinzierl, M. Wirth, A. Fix, and A. Petzold, 2013: Aerosol classification by airborne high spectral resolution lidar observations. *Atmospheric chemistry and physics*, **13 (5)**, 2487–2505, https://doi.org/10.5194/acp-13-2487-2013.

Groß, S., V. Freudenthaler, M. Wirth, and B. Weinzierl, 2015: Towards an aerosol classification scheme for future EarthCARE lidar observations and implications for research needs. *Atmospheric Science Letters*, **16 (1)**, 77–82, https://doi.org/10.1002/asl2.524.

Groß, S., M. Wirth, A. Schäfler, A. Fix, S. Kaufmann, and C. Voigt, 2014: Potential of airborne lidar measurements for cirrus cloud studies. *Atmospheric Measurement Techniques Discussions*, **7 (4)**, 4033–4066, https://doi.org/10.5194/amt-7-2745-2014.

Haiden, T., M. Janousek, F. Vitart, Z. Ben-Bouallegue, and F. Prates, 2023: *Evaluation of ECMWF forecasts, including the 2023 upgrade*. European Centre for Medium-Range Weather Forecasts, https://doi.org/10.21957/d47ba5263c.

Hakim, G. J., 2005: Vertical structure of midlatitude analysis and forecast errors. *Monthly weather review*, **133 (3)**, 567–578, https://doi.org/10.1175/MWR-2882.1.

Harr, P. A., R. L. Elsberry, T. F. Hogan, and W. M. Clune, 1992: Forecasts of north Pacific maritime cyclones with the Navy operational global atmospheric prediction system. *Weather and forecasting*, **7 (3)**, 456–467, https://doi.org/10.1175/1520-0434(1992)007<0456:FONPMC>2.0.CO;2.

Harrold, T., 1973: Mechanisms influencing the distribution of precipitation within baroclinic disturbances. *Quarterly Journal of the Royal Meteorological Society*, **99 (420)**, 232–251, https://doi.org/10.1002/qj.49709942003.

Harvey, B., J. Methven, and M. H. Ambaum, 2018: An adiabatic mechanism for the reduction of jet meander amplitude by potential vorticity filamentation. *Journal of the Atmospheric Sciences*, **75 (12)**, 4091–4106, https://doi.org/10.1175/JAS-D-18-0136.1.

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

Haualand, K. F. and T. Spengler, 2021: Relative importance of tropopause structure and diabatic heating for baroclinic instability. *Weather and Climate Dynamics*, **2 (3)**, 695–712, https://doi.org/10.5194/wcd-2-695-2021.

Heitmann, K., M. Sprenger, H. Binder, H. Wernli, and H. Joos, 2024: Warm conveyor belt characteristics and impacts along the life cycle of extratropical cyclones: case studies and climatological analysis based on ERA5. *Weather and Climate Dynamics*, **5 (2)**, 537–557, https://doi.org/10.5194/wcd-5-537-2024.

Hersbach, H., B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, et al., 2020: The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **146 (730)**, 1999–2049, https://doi.org/10.1002/qj.3803.

Hersbach, H., P. Poli, and D. Dee, 2015: 18 the observation feedback archive for the icoads and ispd data sets. https://www.ecmwf.int/sites/default/files/elibrary/2015/9871-observation-feedback-archive-icoads-and-ispd-data-sets.pdf, last access: 2024-04-21.

Hodyss, D. and N. Nichols, 2015: The error of representation: Basic understanding. *Tellus A: Dynamic Meteorology and Oceanography*, **67 (1)**, 24 822, https://doi.org/10.3402/tellusa.v67.24822.

Hoerling, M. P., T. K. Schaack, and A. J. Lenzen, 1991: Global objective tropopause analysis. *Monthly Weather Review*, **119 (8)**, 1816–1831, https://doi.org/10.1175/1520-0493(1991)119<1816:GOTA>2.0.CO;2.

Hoffmann, L. and R. Spang, 2022: An assessment of tropopause characteristics of the ERA5 and ERA-Interim meteorological reanalyses. *Atmospheric Chemistry and Physics*, **22 (6)**, 4019–4046, https://doi.org/10.5194/acp-22-4019-2022.

Hoinka, K. P., 1999: Temperature, humidity, and wind at the global tropopause. *Monthly Weather Review*, **127 (10)**, 2248–2265, https://doi.org/10.1175/1520-0493(1999)127<2248:THAWAT>2.0.CO;2.

Homeyer, C. R., L. L. Pan, and M. C. Barth, 2014: Transport from convective overshooting of the extratropical tropopause and the role of large-scale lower stratosphere stability. *Journal of Geophysical Research: Atmospheres*, **119 (5)**, 2220–2240, https://doi.org/10.1002/2013JD020931.

Hoor, P., H. Fischer, L. Lange, J. Lelieveld, and D. Brunner, 2002: Seasonal variations of a mixing layer in the lowermost stratosphere as identified by the CO-O3 correlation from in situ measurements. *Journal of Geophysical Research: Atmospheres*, **107 (D5)**, ACL–1, https://doi.org/10.1029/2000JD000289.

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

Houchi, K., A. Stoffelen, G. Marseille, and J. De Kloe, 2010: Comparison of wind and wind shear climatologies derived from high-resolution radiosondes and the ECMWF model. *Journal of Geophysical Research: Atmospheres*, **115 (D22)**, https://doi.org/10.1029/2009JD013196.

Ingleby, B., 2017: *An assessment of different radiosonde types 2015/2016*. European Centre for Medium Range Weather Forecasts Reading, UK, https://doi.org/10.21957/0nje0wpsa.

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

Ingleby, B., P. Pauley, A. Kats, J. Ator, D. Keyser, A. Doerenbecher, E. Fucile, J. Hasegawa, E. Toyoda, T. Kleinert, et al., 2016: Progress toward high-resolution, real-time radiosonde reports. *Bulletin of the American Meteorological Society*, **97 (11)**, 2149–2161, https://doi.org/10.1175/BAMS-D-15-00169.1.

Ingleby, N. B. and A. C. Lorenc, 1993: Bayesian quality control using multivariate normal distributions. *Quarterly Journal of the Royal Meteorological Society*, **119 (513)**, 1195–1225, https://doi.org/10.1002/qj.49711951316.

Isaksen, L., M. Bonavita, R. Buizza, M. Fisher, J. Haseler, M. Leutbecher, and L. Raynaud, 2010: Ensemble of data assimilations at ECMWF.

Janjić, T., N. Bormann, M. Bocquet, J. Carton, S. E. Cohn, S. L. Dance, S. N. Losa, N. K. Nichols, R. Potthast, J. A. Waller, et al., 2018: On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **144 (713)**, 1257–1278, https://doi.org/10.1002/qj.3130.

Jung, T., A. Tompkins, and M. Rodwell, 2003: *Systematic errors in the ECMWF forecasting system*, Vol. 422. ECMWF Reading, UK, https://www.ecmwf.int/sites/default/files/elibrary/2003/10314-systematic-errors-ecmwf-forecasting-system.pdf, last access:27-08-2024.

Jurkat-Witschas, T., C. Voigt, S. Groß, S. Kaufmann, D. Sauer, E. De la Torre Castro, M. Krämer, A. Schäfler, A. Afchine, R. Attinger, I. Bartolome Garcia, Beer, C. G., L. Bugliaro, H.-C. Clemen, G. Dekoutsidis, A. Dörnbrack, A. Ehrlich, S. Grawe, V. Hahn, J. Hendricks, E. Järvinen, T. Klimach, K. Krüger, O. Krüger, J. Lucke, A. E.

Luebke, A. Marsing, B. Mayer, J. Mayer, S. Mertes, M. Moser, H. Müller, M. Pöhlker, U. Pöschl, V. Pörtge, M. Rautenhaus, M. Righi, J. Röttenbacher, C. Rolf, J. Schaefer, M. Schnaiter, J. Schneider, U. Schumann, N. Spelten, F. Stratmann, L. Tomsche, H. Wagner, S. Wagner, Z. Wang, A. Weber, M. Wendisch, H. Wernli, B. Wetzel, M. Wirth, A. Zahn, H. Ziereis, and M. Zöger, 2024: CIRRUS-HL: Picturing the high- and mid-latitude summer cirrus and contrail cirrus above Europe with airborne measurements aboard the research aircraft HALO. *Bulletin of the American Meteorological Society*, submitted to Bulletin of the American Meteorological society in May 2024, currently under review.

Kaluza, T., D. Kunkel, and P. Hoor, 2021: On the occurrence of strong vertical wind shear in the tropopause region: a 10-year ERA5 northern hemispheric study. *Weather and Climate Dynamics*, **2 (3)**, 631–651, https://doi.org/10.5194/wcd-2-631-2021.

Kaufmann, S., C. Voigt, R. Heller, T. Jurkat-Witschas, M. Krämer, C. Rolf, M. Zöger, A. Giez, B. Buchholz, V. Ebert, et al., 2018: Intercomparison of midlatitude tropospheric and lower-stratospheric water vapor measurements and comparison to ECMWF humidity data. *Atmospheric chemistry and physics*, **18 (22)**, 16 729–16 745, https://doi.org/10.5194/acp-18-16729-2018.

Khordakova, D., C. Rolf, J.-U. Grooß, R. Müller, P. Konopka, A. Wieser, M. Krämer, and M. Riese, 2021: A case study on the impact of severe convective storms on the water vapor mixing ratio in the lower mid-latitude stratosphere observed in 2019 over Europe. *Atmospheric Chemistry and Physics Discussions*, **2021**, 1–32, https://doi.org/10.5194/acp-22-1059-2022.

Kiemle, C., M. Wirth, A. Fix, G. Ehret, U. Schumann, T. Gardiner, C. Schiller, N. Sitnikov, and G. Stiller, 2008: First airborne water vapor lidar measurements in the tropical upper troposphere and mid-latitudes lower stratosphere: accuracy evaluation and intercomparisons with other instruments. *Atmospheric chemistry and physics*, **8 (17)**, 5245–5261, https://doi.org/10.1007/978-3-642-30183-4_11.

Klepp, C., F. Ament, S. Bakan, L. Hirsch, and B. Stevens, 2014: The Next-generation Aircraft Remote sensing for VALidation studies (NARVAL) Campaign flight reports using the research aircraft HALO (The NARVAL Campaign Report). https://pure.mpg.de/rest/items/item_2129055/component/file_2129139/content, last access: 2022-06-30.

König, N., P. Braesicke, and T. von Clarmann, 2019: Tropopause altitude determination from temperature profile measurements of reduced vertical resolution. *Atmospheric Measurement Techniques*, **12 (7)**, 4113–4129, https://doi.org/10.5194/amt-12-4113-2019.

Krautstrunk, M. and A. Giez, 2012: The transition from FALCON to HALO era airborne atmospheric research. *Atmospheric Physics: Background–Methods–Trends*, Springer, 609–624, ISBN:978-3-642-30182-7.

Krüger, K., A. Schäfler, M. Weissmann, and G. C. Craig, 2024: Influence of radiosonde observations on the sharpness and altitude of the midlatitude tropopause in the ECMWF IFS. *Weather and Climate Dynamics*, **5 (2)**, 491–509, https://doi.org/10.5194/wcd-5-491-2024.

Krüger, K., A. Schäfler, M. Wirth, M. Weissmann, and G. C. Craig, 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)**, 15 559–15 577, https://acp.copernicus.org/articles/22/15559/2022/.

Kunkel, D., P. Hoor, T. Kaluza, J. Ungermann, B. Kluschat, A. Giez, H.-C. Lachnitt, M. Kaufmann, and M. Riese, 2019: Evidence of small-scale quasi-isentropic mixing in ridges of extratropical baroclinic waves. *Atmospheric Chemistry and Physics*, **19 (19)**, 12 607–12 630, https://doi.org/10.5194/acp-19-12607-2019.

Kunz, A., P. Konopka, R. Müller, L.-L. Pan, C. Schiller, and F. Rohrer, 2009: High static stability in the mixing layer above the extratropical tropopause. *Journal of Geophysical Research: Atmospheres*, **114 (D16)**, https://doi.org/10.1029/2009JD011840.

Kunz, A., N. Spelten, P. Konopka, R. Müller, R. M. Forbes, and H. Wernli, 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 (19)**, 10 803–10 822, https://doi.org/10.5194/acp-14-10803-2014.

Laloyaux, P., M. Bonavita, M. Chrust, and S. Gürol, 2020: Exploring the potential and limitations of weak-constraint 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, **146 (733)**, 4067–4082, https://doi.org/10.1002/qj.3891.

Laloyaux, P., T. Kurth, P. D. Dueben, and D. Hall, 2022: Deep learning to estimate model biases in an operational NWP assimilation system. *Journal of Advances in Modeling Earth Systems*, **14 (6)**, e2022MS003 016, https://doi.org/10.1029/2022MS003016.

Lavers, D. A., R. D. Torn, C. Davis, D. S. Richardson, F. M. Ralph, and F. Pappenberger, 2023: Forecast evaluation of the North Pacific jet stream using AR Recon dropwindsondes. *Quarterly Journal of the Royal Meteorological Society*, **149 (756)**, 3044–3063, https://doi.org/10.1002/qj.4545.

Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, **112 (474)**, 1177–1194, https://doi.org/10.1002/qj.49711247414.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *Journal of atmospheric sciences*, **20 (2)**, 130–141, https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.

Lorenz, E. N., 1969: The predictability of a flow which possesses many scales of motion. *Tellus*, **21 (3)**, 289–307, https://doi.org/10.3402/tellusa.v21i3.10086.

Madonna, E., H. Wernli, H. Joos, and O. Martius, 2014: Warm conveyor belts in the ERA-Interim dataset (1979–2010). Part I: Climatology and potential vorticity evolution. *Journal of climate*, **27 (1)**, 3–26, https://doi.org/10.1175/JCLI-D-12-00720.1.

Madonna, F., R. Kivi, J.-C. Dupont, B. Ingleby, M. Fujiwara, G. Romanens, M. Hernandez, X. Calbet, M. Rosoldi, A. Giunta, et al., 2020: Use of automatic radiosonde launchers to measure temperature and humidity profiles from the GRUAN perspective. *Atmospheric Measurement Techniques*, **13 (7)**, 3621–3649, https://doi.org/10.5194/amt-13-3621-2020.

Magnusson, L., A. Thorpe, M. Bonavita, S. Lang, T. McNally, and N. Wedi, 2013: *Evaluation of forecasts for hurricane Sandy*. European Centre for Medium-Range Weather Forecasts.

Maiman, T. H. et al., 1960: Stimulated optical radiation in ruby. *Nature*, **187 (4736)**, 493–494, https://doi.org/10.1038/187493a0.

Malardel, S., N. Wedi, W. Deconinck, M. Diamantakis, C. Kühnlein, G. Mozdzynski, M. Hamrud, and P. Smolarkiewicz, 2016: A new grid for the IFS. *ECMWF newsletter*, **146 (23-28)**, 321, https://www.ecmwf.int/sites/default/files/elibrary/2016/17262-new-grid-ifs.pdf, last access: 2024-05-04.

Malberg, H., 2007: *Meteorologie und Klimatologie Eine Einfuhrung 5. Auflage*. Springer Berlin Heidelberg New York, ISBN:978-3-540-37219-6.

Martin, A., M. Weissmann, and A. Cress, 2023: Impact of assimilating Aeolus observations in the global model ICON: A global statistical overview. *Quarterly Journal of the Royal Meteorological Society*, **149 (756)**, 2962–2979, https://doi.org/10.5194/wcd-4-249-2023.

Martius, O., C. Schwierz, and H. Davies, 2010: Tropopause-level waveguides. *Journal of the atmospheric sciences*, **67 (3)**, 866–879, https://doi.org/10.1175/2009JAS2995.1.

Oertel, A., M. Boettcher, H. Joos, M. Sprenger, H. Konow, M. Hagen, and H. Wernli, 2019: Convective activity in an extratropical cyclone and its warm conveyor belt–a case-study combining observations and a convection-permitting model simulation. *Quarterly Journal of the Royal Meteorological Society*, **145 (721)**, 1406–1426, https://doi.org/10.1002/qj.3500.

Oikonomou, E. and A. O'Neill, 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: Atmospheres*, **111 (D14)**, https://doi.org/10.1029/2004JD005341.

Pan, L., K. Bowman, M. Shapiro, W. Randel, R. Gao, T. Campos, C. Davis, S. Schauffler, B. Ridley, J. Wei, et al., 2007: Chemical behavior of the tropopause observed

during the Stratosphere-Troposphere Analyses of Regional Transport experiment. *Journal of Geophysical Research: Atmospheres*, **112 (D18)**, https://doi.org/10.1029/2007JD008645.

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

Pan, L. L., E. J. Hintsa, E. M. Stone, E. M. Weinstock, and W. J. Randel, 2000: The seasonal cycle of water vapor and saturation vapor mixing ratio in the extratropical lowermost stratosphere. *Journal of Geophysical Research: Atmospheres*, **105 (D21)**, 26 519–26 530, https://doi.org/10.1029/2000JD900401.

Pauley, P. M. and B. Ingleby, 2022: *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. IV)-Assimilation of in-situ observations*. Springer, 293–371 pp., ISBN:978-3-030-77722-7.

Pilch Kedzierski, R., K. Matthes, and K. Bumke, 2015: Synoptic-scale behavior of the extratropical tropopause inversion layer. *Geophysical Research Letters*, **42 (22)**, 10–018, https://doi.org/10.1002/2015GL066409.

Pilch Kedzierski, R., L. Neef, and K. Matthes, 2016: Tropopause sharpening by data assimilation. *Geophysical Research Letters*, **43 (15)**, 8298–8305, https://doi.org/10.1002/2016GL069936.

Ploeger, F., T. Birner, E. Charlesworth, P. Konopka, and R. Müller, 2024: Moist bias in the Pacific upper troposphere and lower stratosphere (UTLS) in climate models affects regional circulation patterns. *Atmospheric Chemistry and Physics*, **24 (3)**, 2033–2043, https://acp.copernicus.org/articles/24/2033/2024/.

Rabier, F., H. Järvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Quarterly Journal of the Royal Meteorological Society*, **126 (564)**, 1143–1170, https://doi.org/10.1002/qj.49712656415.

Radnoti, G., P. Bauer, A. McNally, C. Cardinali, S. Healy, and P. De Rosnay, 2010: *ECMWF Study on the Impact of Future Developments of the Space-based Observing System on Numerical Weather Precition*. European Centre for Medium-Range Weather Forecasts, https://doi.org/10.21957/skfhvask.

Randel, W. J. and F. Wu, 2010: The polar summer tropopause inversion layer. *Journal of the Atmospheric Sciences*, **67 (8)**, 2572–2581, https://doi.org/10.1175/2010JAS3430.1.

Randel, W. J., F. Wu, and P. Forster, 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, https://doi.org/10.1175/2007JAS2412.1.

Richardson, L. F., 1922: *Weather prediction by numerical process*. University Press.

Rothman, L. S., I. E. Gordon, A. Barbe, D. C. Benner, P. F. Bernath, M. Birk, V. Boudon, L. R. Brown, A. Campargue, J.-P. Champion, et al., 2009: The HITRAN 2008 molecular spectroscopic database. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **110 (9-10)**, 533–572, https://doi.org/10.1016/j.jqsrt.2009.02.013.

Rothman, L. S., D. Jacquemart, A. Barbe, D. C. Benner, M. Birk, L. Brown, M. Carleer, C. Chackerian Jr, K. Chance, L. e. a. Coudert, et al., 2005: The HITRAN 2004 molecular spectroscopic database. *Journal of quantitative spectroscopy and radiative transfer*, **96 (2)**, 139–204, https://doi.org/10.1016/j.jqsrt.2004.10.008.

Saffin, L., S. Gray, J. Methven, and K. Williams, 2017: Processes maintaining tropopause sharpness in numerical models. *Journal of Geophysical Research: Atmospheres*, **122 (18)**, 9611–9627, https://doi.org/10.1002/2017JD026879.

Schäfler, A., G. Craig, H. Wernli, P. Arbogast, J. D. Doyle, R. McTaggart-Cowan, J. Methven, G. Rivière, F. Ament, M. Boettcher, et al., 2018: The North Atlantic waveguide and downstream impact experiment. *Bulletin of the American Meteorological Society*, **99 (8)**, 1607–1637, https://doi.org/10.1175/BAMS-D-17-0003.1.

Schäfler, A., A. Dörnbrack, C. Kiemle, S. Rahm, and M. Wirth, 2010: Tropospheric water vapor transport as determined from airborne lidar measurements. *Journal of Atmospheric and Oceanic Technology*, **27 (12)**, 2017–2030, https://doi.org/10.1175/2010JTECHA1418.1.

Schäfler, A., A. Fix, and M. Wirth, 2021: Mixing at the extratropical tropopause as characterized by collocated airborne H 2 O and O 3 lidar observations. *Atmospheric Chemistry and Physics*, **21 (6)**, 5217–5234, https://doi.org/10.5194/acp-21-5217-2021.

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

Schäfler, A., K. Krüger, A. Oertel, J. Quinting, and S. Raveh-Rubin, 2024: Indication for biases in dry intrusions and the marine boundary layer over the Azores in ECMWF short-term forecasts and analyses. *Geophysical Research Letters*, **51 (15)**, e2024GL109601,https://doi.org/10.1029/2024GL109601.

Schäfler, A. and M. Rautenhaus, 2023: Interactive 3D visual analysis of weather prediction data reveals midlatitude overshooting convection during the CIRRUS-HL field experiment. *Bulletin of the American Meteorological Society*, **104 (8)**, E1426–E1434, https://doi.org/10.1175/BAMS-D-22-0103.1.

Schäfler, A., M. Sprenger, H. Wernli, A. Fix, and M. Wirth, 2023: Case study on the influence of synoptic-scale processes on the paired H 2 O–O 3 distribution in the UTLS across a North Atlantic jet stream. *Atmospheric Chemistry and Physics*, **23 (2)**, 999–1018, https://doi.org/10.519/acp-23-999-2023.

Schindler, M., M. Weissmann, A. Schäfler, and G. Radnoti, 2020: The impact of dropsonde and extra radiosonde observations during NAWDEX in autumn 2016. *Monthly Weather Review*, **148 (2)**, 809–824, https://doi.org/10.1175/MWR-D-19-0126.1.

Schmidt, T., J. Wickert, and A. Haser, 2010: Variability of the upper troposphere and lower stratosphere observed with GPS radio occultation bending angles and temperatures. *Advances in Space Research*, **46 (2)**, 150–161, https://doi.org/10.1016/j.asr.2010.01.021.

Schwierz, C., S. Dirren, and H. C. Davies, 2004: Forced waves on a zonally aligned jet stream. *Journal of the atmospheric sciences*, **61 (1)**, 73–87, https://doi.org/10.1175/1520-0469(2004)061<0073:FWOAZA>2.0.CO;2.

Selz, T., 2019: Estimating the intrinsic limit of predictability using a stochastic convection scheme. *Journal of the Atmospheric Sciences*, **76 (3)**, 757–765, https://doi.org/10.1175/JAS-D-17-0373.1.

Selz, T. and G. C. Craig, 2015: Upscale error growth in a high-resolution simulation of a summertime weather event over Europe. *Monthly Weather Review*, **143 (3)**, 813–827, https://doi.org/10.1175/MWR-D-14-00140.1.

Selz, T., M. Riemer, and G. C. Craig, 2022: The transition from practical to intrinsic predictability of midlatitude weather. *Journal of the Atmospheric Sciences*, **79 (8)**, 2013–2030, https://doi.org/10.1175/JAS-D-21-0271.1.

Shapiro, M., 1980: Turbulent mixing within tropopause folds as a mechanism for the exchange of chemical constituents between the stratosphere and troposphere. *Journal of Atmospheric Sciences*, **37 (5)**, 994–1004, https://doi.org/10.1175/1520-0469(1980)037<0994:TMWTFA>2.0.CO;2.

Shapiro, M., H. Wernli, J.-W. Bao, J. Methven, X. Zou, J. Doyle, T. Holt, E. Donall-Grell, and P. Neiman, 1999: A planetary-scale to mesoscale perspective of the life cycles of extratropical cyclones: The bridge between theory and observations. *The life cycles of extratropical cyclones*, 139–185, https://doi.org/10.1007/978-1-935704-09-6_14.

Shepherd, T. G., I. Polichtchouk, R. J. Hogan, and A. J. Simmons, 2018: *Report on stratosphere task force*. European Centre for Medium-Range Weather Forecasts, https://doi.org/10.21957/0vkp0t1xx.

Smith, J. B., D. M. Wilmouth, K. M. Bedka, K. P. Bowman, C. R. Homeyer, J. A. Dykema, M. R. Sargent, C. E. Clapp, S. S. Leroy, D. S. Sayres, et al., 2017: A case study of convectively sourced water vapor observed in the overworld stratosphere over the United States. *Journal of Geophysical Research: Atmospheres*, **122 (17)**, 9529–9554, https://doi.org/10.1002/2017JD026831.

Stenke, A., V. Grewe, and M. Ponater, 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, https://doi.org/10.1007/s00382-007-0347-5.

Stern, H. and N. E. Davidson, 2015: Trends in the skill of weather prediction at lead times of 1–14 days. *Quarterly Journal of the Royal Meteorological Society*, **141 (692)**, 2726–2736, https://doi.org/10.1002/qj.2559.

Stevens, B., F. Ament, S. Bony, S. Crewell, F. Ewald, S. Gross, A. Hansen, L. Hirsch, M. Jacob, T. Kölling, et al., 2019: A high-altitude long-range aircraft configured as a cloud observatory: The NARVAL expeditions. *Bulletin of the American Meteorological Society*, **100 (6)**, 1061–1077, https://doi.org/10.1175/BAMS-D-18-0198.1.

Stevens, B., S. Bony, D. Farrell, F. Ament, A. Blyth, C. Fairall, J. Karstensen, P. K. Quinn, S. Speich, C. Acquistapace, et al., 2021: EUREC4A. *Earth System Science Data Discussions*, **2021**, 1–78, https://doi.org/10.5194/essd-13-4067-2021.

Stohl, A., P. Bonasoni, P. Cristofanelli, W. Collins, J. Feichter, A. Frank, C. Forster, E. Gerasopoulos, H. Gäggeler, P. James, et al., 2003: Stratosphere-troposphere exchange: A review, and what we have learned from STACCATO. *Journal of Geophysical Research: Atmospheres*, **108 (D12)**, https://doi.org/10.1029/2002JD002490.

Tinney, E. N., C. R. Homeyer, L. Elizalde, D. F. Hurst, A. M. Thompson, R. M. Stauffer, H. Vömel, and H. B. Selkirk, 2022: A Modern Approach to a Stability-Based Definition of the Tropopause. *Monthly Weather Review*, **150 (12)**, 3151–3174, https://doi.org/10.1175/MWRD-22-0174.1.

Trickl, T., H. Vogelmann, A. Fix, A. Schäfler, M. Wirth, B. Calpini, G. Levrat, G. Romanens, A. Apituley, K. M. Wilson, et al., 2016: How stratospheric are deep stratospheric intrusions? LUAMI 2008. *Atmospheric Chemistry and Physics*, **16 (14)**, 8791–8815, https://doi.org/10.5194/acp-16-8791-2016.

Voggenberger, U., L. Haimberger, F. Ambrogi, and P. Poli, 2023: Balloon drift estimation and improved position estimates for radiosondes. *Geoscientific Model Development Discussions*, **2023**, 1–29, https://doi.org/10.5194/gmd-17-3783-2024.

Weissmann, M., R. Busen, A. Dörnbrack, S. Rahm, and O. Reitebuch, 2005: Targeted observations with an airborne wind lidar. *Journal of Atmospheric and Oceanic Technology*, **22 (11)**, 1706–1719, https://doi.org/10.1175/JTECH1801.1.

Weissmann, M. and C. Cardinali, 2007: Impact of airborne Doppler lidar observations on ECMWF forecasts. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, **133 (622)**, 107–116, https://doi.org/10.1002/qj.16.

Weissmann, M., R. H. Langland, C. Cardinali, P. M. Pauley, and S. Rahm, 2012: Influence of airborne Doppler wind lidar profiles near Typhoon Sinlaku on ECMWF and NOGAPS forecasts. *Quarterly Journal of the Royal Meteorological Society*, **138 (662)**, 118–130, https://doi.org/10.1002/qj.896.

Wernli, H. and S. L. Gray, 2023: The importance of diabatic processes for the dynamics of synoptic-scale extratropical weather systems—a review. *EGUsphere*, **2023**, 1–151, https://doi.org/10.5194/egusphere-2023-2678.

Wirth, M., A. Fix, P. Mahnke, H. Schwarzer, F. Schrandt, and G. Ehret, 2009: The airborne multi-wavelength water vapor differential absorption lidar WALES: system design and performance. *Applied Physics B*, **96**, 201–213, https://doi.org/10.1007/s00340-009-3365-7.

Witschas, B., S. Rahm, A. Dörnbrack, J. Wagner, and M. Rapp, 2017: Airborne wind lidar measurements of vertical and horizontal winds for the investigation of orographically induced gravity waves. *Journal of Atmospheric and Oceanic Technology*, **34 (6)**, 1371–1386, https://doi.org/10.1175/JTECH-D-17-0021.1.

WMO, 1957: Meteorology—A three-dimensional science. *WMO Bulletin*, **6**, 134–138.

WMO, 1986: Atmospheric Ozone. Tech. Rep. 16, World Meteorological Organization, https://library.wmo.int/viewer/56592/download?file=wmo_656_en.pdf&type=pdf&navigator=1, last access: 2024-03-28.

Woiwode, W., A. Dörnbrack, I. Polichtchouk, S. Johansson, B. Harvey, M. Höpfner, J. Ungermann, and F. Friedl-Vallon, 2020: Lowermost-stratosphere moist bias in ECMWF IFS model diagnosed from airborne GLORIA observations during winter–spring 2016. *Atmospheric Chemistry and Physics*, **20 (23)**, 15 379–15 387, https://doi.org/10.5194/acp-20-15379-2020.

Zahn, A., C. Brenninkmeijer, and P. Van Velthoven, 2004: Passenger aircraft project CARIBIC 1997–2002, Part I: the extratropical chemical tropopause. *Atmospheric Chemistry and Physics Discussions*, **4 (1)**, 1091–1117, https://doi.org/10.1002/2014JD021687.

Zahn, A., E. Christner, P. van Velthoven, A. Rauthe-Schöch, and C. Brenninkmeijer, 2014: Processes controlling water vapor in the upper troposphere/lowermost stratosphere: An analysis of 8 years of monthly measurements by the IAGOS-CARIBIC observatory. *Journal of Geophysical Research: Atmospheres*, **119 (19)**, 11–505, https://doi.org/10.1002/2014JD021687.

Danksagung

Zunächst möchte ich mich herzlich bei meinen Doktorvätern Prof. Dr. George Craig und Prof. Dr. Martin Weissmann für die engagierte Betreuung meiner Promotion und für die Begutachtung dieser Doktorarbeit bedanken. Ich danke für Euer stetes Interesse an meinem Thema, und für die vielen wertvollen Ratschläge in unseren Treffen. Vielen Dank, dass Ihr mich während der Promotionszeit auf meinen ersten Schritten in der Wissenschaft stets beraten und unterstützt habt. Insbesondere Danke ich Dir Martin, für die Möglichkeit des Forschungsaufenhaltes in Deiner Arbeitsgruppe im schönen Wien und für drei wirklich lehrreiche, konstruktive Zeit.

Ein großer Dank geht an meinen Betreuer Dr. Andreas Schäfler. Lieber Andreas, Du warst mir mit Deiner ermutigenden Art über die ganzen Jahre ein unglaublicher Rückhalt. Du hast mich immer in dem Maße gefordert und gefördert wie es notwendig war. Rückblickend bin ich sehr stolz über das, was ich in den letzten Jahren erreichen konnte, samt Publikationen, Teilnahmen/Vorträge auf Konferenzen und Workshops, und Messkampagnen – dies wäre ohne Deine Unterstützung nicht möglich gewesen. Dein Forschergeist inspiriert und ich konnte durch Dich unheimlich viel Lernen. Es war und ist mir fachlich wie menschlich eine große Freude mit Dir zusammenzuarbeiten! Als Dein erster Doktorand möchte ich Dir sagen: Jeder zukünftige Doktorand kann sich glücklich schätzen, Dich als Betreuer seiner Promotion zu haben.

Während meiner Zeit am DLR wurde ich fachlich als auch menschlich von zahlreichen Kollegen unterstützt. In anspruchsvollen Phasen standet Ihr mir aufmunternd zur Seite, in erfolgreichen Phasen haben wir Fortschritte zusammen gefeiert. Die kritischen Diskussionen um meine "Nasen-Plots", sowie die ausgelassene Stimmung bei unseren Skatabenden, möchte ich nicht missen - es sind tolle Freundschaften entstanden. Für diese Zeit bin ich insbesondere Robert, Sebastian, Christopher, Manu, Sabrina, sowie meine „running mates“ – die „Cloudonauts“, Georgios, Clémantyne, Christian dankbar.

Furthermore, I would also like to take this opportunity to thank all the doctoral students who gave me a wonderful time during my research visit in beatiful Vienna, both in the office (#international#fruitday, #BurritoWednesday), at the Heurigen and all the other events. Dear Lukas, Ioannaki, Kaushi, and Max: I really hope our paths will cross again soon!

I am also grateful to the "Waves-to-Weather" community for the exciting years in this great scientific environment and for the financial support of my research stay in Vienna.

Ein herzliches Dank auch den Postdocs Tobi Selz, Tobi Necker und Philipp Griewank für die Zeit und Geduld, die Ihr Euch genommen habt, um einzelne Aspekte meiner Doktorarbeit zu diskutieren und meine Fragen zu beantworten.

Ein weiterer Dank gilt der gesamten PA-Lidar Abteilung für die positive, kollegiale Arbeitsatmosphäre am Institut und für Euer Interesse am Thema und Fortschritt meiner Promotion. Insbesondere möchte ich mich bei Prof. Dr. Silke Groß, Dr. Martin Wirth, und Dr. Andreas Fix bedanken, die mir ermöglicht haben, an der Cirrus-HL Messkampagne teilzunehmen. Es war eine extrem aufregende und lehrreiche Zeit - und eine riesige Freude in der Flugvorhersage, oder als WALES-Operator zu unterstützen. Ich freue mich auf die weitere Zusammenarbeit mit Euch!

Für intensives Korrekturlesen meiner Doktorarbeit samt vielen wertvollen Ratschlägen danke ich Andreas Schäfler, Martin Wirth, Christopher Geach, Tobias Necker, Robert Reichert, Kimberly, Lilli und Dani.

Auf meinem Weg in die Wissenschaft hatte ich das Glück von einem wunderbaren Umfeld umgeben zu sein. Für jegliche Unterstützung bedanke ich mich von ganzem Herzen bei meinen Schul- und Studienfreunden aus Frankfurt und Karlsruhe. Der allergrößte Dank gilt natürlich meinen Eltern Ina, Thomas, und Holger, sowie meinen lieben Geschwistern Kilian, Kathleen und Kimberly, die mir immer mit Rat und Tat zur Seite standen, und mich zu dem Menschen gemacht haben, der ich bin. Liebe Lilli, Du warst mir auf meiner ereignisreichen und aufregenden Reise in den letzten Jahren ein unglaublicher (emotionaler) Rückhalt, und ich bin froh, Dich in meinem Leben zu haben.