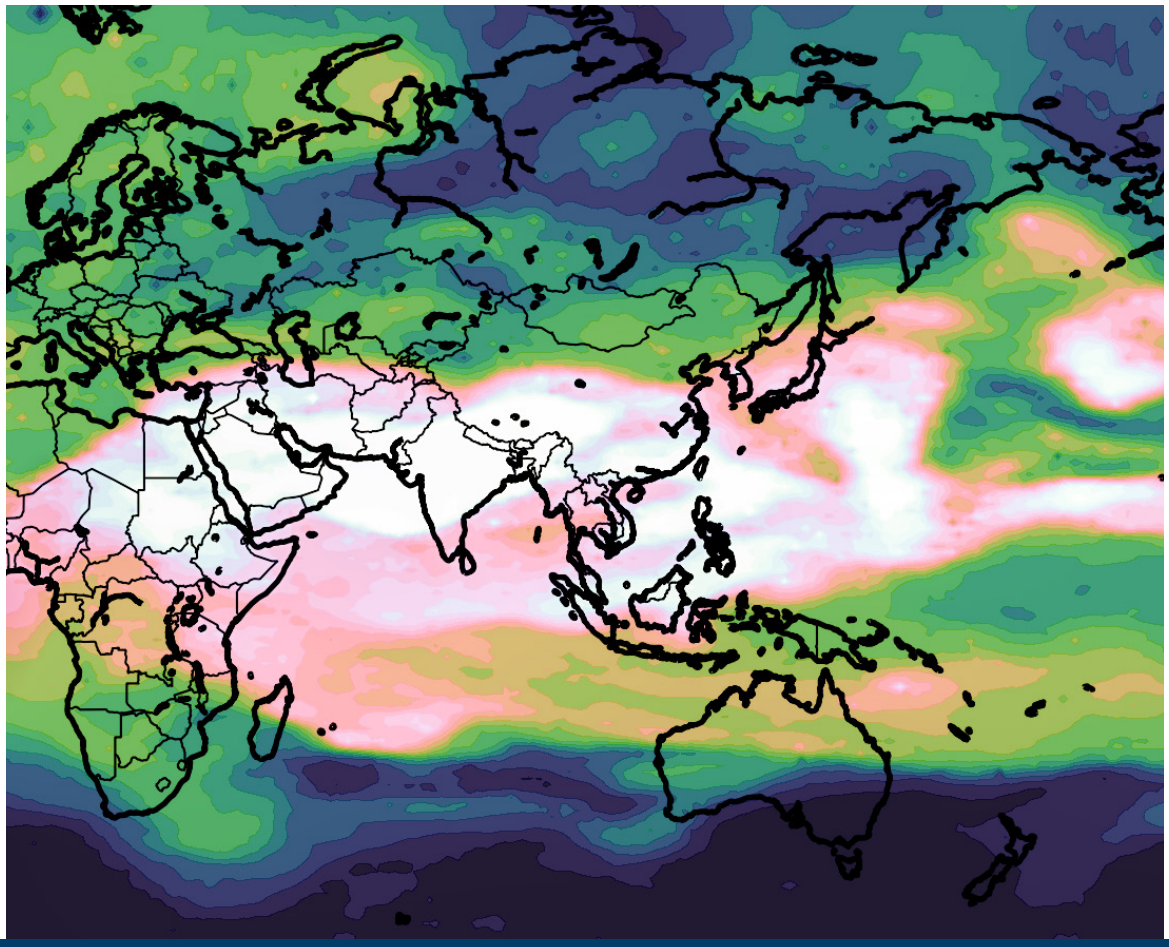

Identifying and quantifying mixing processes in filaments using computer vision and Artificial Intelligence

Jan Kaumanns

Energie & Umwelt / Energy & Environment
Band / Volume 747
ISBN 978-3-69367-029-3

Mitglied der Helmholtz-Gemeinschaft

JÜLICH
Forschungszentrum

Schriften des Forschungszentrums Jülich
Reihe Energie & Umwelt / Energy & Environment Band / Volume 747

Forschungszentrum Jülich GmbH
Institute of Climate and Energy Systems (ICE)
Stratosphäre (ICE-4)

Identifying and quantifying mixing processes in filaments using computer vision and Artificial Intelligence

Jan Kaumanns

Schriften des Forschungszentrums Jülich
Reihe Energie & Umwelt / Energy & Environment Band / Volume 747

ISSN 1866-1793 ISBN 978-3-69367-029-3

Bibliografische Information der Deutschen Nationalbibliothek.
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der
Deutschen Nationalbibliografie; detaillierte Bibliografische Daten
sind im Internet über http://dnb.d-nb.de abrufbar.

Herausgeber und Vertrieb:	Forschungszentrum Jülich GmbH Zentralbibliothek, Verlag 52425 Jülich Tel.: +49 2461 61-5368 Fax: +49 2461 61-6103 zb-publikation@fz-juelich.de www.fz-juelich.de/zb
Umschlaggestaltung:	Grafische Medien, Forschungszentrum Jülich GmbH
Druck:	Grafische Medien, Forschungszentrum Jülich GmbH
Copyright:	Forschungszentrum Jülich 2026

Schriften des Forschungszentrums Jülich
Reihe Energie & Umwelt / Energy & Environment, Band / Volume 747

D 468 (Diss. Wuppertal, Univ., 2026)

ISSN 1866-1793
ISBN 978-3-69367-028-6 (Print)
ISBN 978-3-69367-029-3 (E-Book)

Vollständig frei verfügbar über das Publikationsportal des Forschungszentrums Jülich (JuSER)
unter www.fz-juelich.de/zb/openaccess.



This is an Open Access publication distributed under the terms of the Creative Commons Attribution License 4.0, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

For my beloved Freya

Quales in vita fuimus, tales in morte erimus.

Dolor amorque meus erga te aeterni sunt.

Abstract

The spatial distribution of chemical trace gases in the stratosphere is essentially governed by complex transport and multi-scale mixing processes. These tracer fields influence the stratosphere's radiative budget, drive various climate feedbacks, and modulate the dynamical coupling between different atmospheric compartments. However, because many key mixing processes occur on sub-grid scales, global climate models heavily rely on parametrizations of these processes. To date, these parametrizations remain a primary source of uncertainty in our models, therefore necessitating novel and innovative observational strategies of mixing processes for model validation. Direct observations of mixing processes are notoriously difficult, as these processes unfold over extended timescales, and spatially resolved, consecutive measurements of individual, moving air parcels are extremely difficult to obtain. Promising events to study mixing are filaments, which are generated by horizontal Rossby wave breaking. Characterized by intense localized shear and strain, these dynamically active structures act as physical hotspots where irreversible mixing occurs rapidly, thus providing opportunities to measure mixing processes directly. To optimize the information yield of such measurements, this thesis utilizes various advanced machine learning approaches tailored to the physical needs of the different measurements and provides a data-driven approach to mixing processes.

On the mesoscale, mixing processes as observed in a dedicated double flight experiment during the Probing High Latitude Export of air from the Asian Summer Monsoon (PHILEAS) campaign are analysed. Using high-resolution, 3-D tomographic multi-species observations from the glsgloria, we track air parcels of a filament of Asian summer monsoon outflow in the through their trajectories on two consecutive days. The dynamically disturbed situation requires a new way to identify and characterize the air types involved in the mixing. By exploiting the internal structure of the data, we derive a data-driven classification using Gaussian Mixture Model (GMM)capable of identifying the air types and subsequently the mixing processes involved without prior knowledge of the atmospheric state. This classification inherently reflects ongoing mixing and expresses mixing as a change in classification for the revisited air parcels. The concept is validated on additional data products obtained during the measurement flights to show its universal applicability. The concept is further generalized using Deep Embedding Clustering (neural network) (DEC) neural networks, which minimize the assumptions put into the classification.

On the synoptic scale, we explore the capabilities of measuring mixing processes in filaments of tropical air within the stratospheric surf zone using satellite instruments. Using synthetic observations based on the design specifications of Changing Atmosphere Infra-Red Tomography Explorer (CAIRT) – a tomographic infrared explorer concept originally proposed to the European Space Agency (ESA) – simulated by the Chemical Lagrangian Model of the Stratosphere (CLaMS), we utilize a conceptually similar parametrization of the observed atmospheric state to identify filaments as anomalous structures and obtain time series of matched trajectories from these observations. The parametrization allows us to both describe the spatial structure and the mixing dynamics of the filaments at once. Using convolutional neural networks, we demonstrate how information about the mixing processes contained in these time series can be distilled and used to make accurate predictions of the evolution of these air parcels.

Finally, on the planetary scale, we investigate the residual circulation driven by mixing processes through the concept of Age-of-Air. Since Age-of-Air is a pure model value, it cannot be directly observed and must instead be derived from observable quantities such as long-lived tracers. By using Multi-Layer Perceptrons (MLPs), we provide highly accurate and precise predictions of Age-of-Air derived from long-lived tracers, exceeding conventional methods. We optimize these MLPs with respect to anticipated measurement noise, input selection and out-of-domain predictions, such that reliable predictions many years into the future are possible. Combining our insights collected throughout these studies, we propose a novel, data-driven representation of Age-of-Air based on advanced physics-informed, mixture-of-experts autoencoders, optimized towards the usage of real measurements for climate model validation.

Zusammenfassung

Die räumliche Verteilung von chemischen Spurengasen in der Stratosphäre wird hauptsächlich durch komplexe Transport- und Mischungsprozesse auf verschiedenen Skalen bestimmt. Diese Spurengasverteilungen beeinflussen das Strahlungsbudget der Stratosphäre, treiben verschiedene Klimarückkopplungen und formen das dynamische Zusammenspiel der verschiedenen atmosphärischen Gebiete. Da jedoch viele Mischungsprozesse auf Skalen unterhalb der räumlichen Auflösung der Klimamodelle ablaufen, sind diese stark abhängig von Parametrisierungen dieser Prozesse. Derzeit sind solche Parametrisierungen eine Hauptquelle für Unsicherheiten in unseren Modellen, weshalb neuartige und innovative Messansätze erstrebenswert für die Validierung der Modelle sind. Direkte Beobachtungen der Mischungsprozesse sind bekanntermaßen schwierig zu realisieren, da sich Mischungsprozesse auf verschiedenen, teilweise großen Zeitskalen abspielen und einzelne, sich bewegende Luftpakete nur außerordentlich schwer zu verschiedenen Zeitpunkten messen lassen. Ein vielversprechendes Phänomen sind atmosphärische Filamente, die durch das horizontale Brechen von Rossby-Wellen entstehen. Filamente sind großen, lokalen Spannungen und Schwerungen ausgesetzt, durch die starke, irreversible Mischungsprozesse eingeleitet werden, sodass dort Mischung direkt beobachtet werden kann. Um die meiste Information aus solchen Messungen zu extrahieren, werden in dieser Dissertation verschiedene, maßgeschneiderte Ansätze aus dem Bereich des Maschinenlernens eingesetzt und eine datengestützte Betrachtungsweise von Mischungsprozessen präsentiert.

Auf der Mesoskala werden Mischungsprozesse aus Messungen eines speziellen Doppelflug-Experiments analysiert, welches während der PProbing High Latitude Export of air from the Asian Summer Monsoon"(PHILEAS) Kampagne durchgeführt wurde. Hochaufgelöste, dreidimensionale tomographische Messungen von vielen verschiedene Spurengasen des "Gimballed Limb Observer for Radiance Imaging of the Atmosphere"(GLORIA) Instrumentes werden genutzt, um einzelne Luftpackete aus einem Filament des asiatischen Sommermonsun-Ausfluss über ihre Trajektorien zu verfolgen und an zwei aufeinanderfolgenden Tagen zu messen. Eine Klassifikation der beobachteten Luftmassen, basierend auf einem Bayes'schen Gauß'schen Mischungsmodell, welches die interne Struktur der Messdaten ausnutzt und nicht an Vorwissen über die physikalische Situation gebunden ist, wird präsentiert. Aus dieser Klassifizierung folgen Mischungsprozesse

ganz natürlich als Änderungen in der Klassifizierung der mehrfach gemessenen Luftpackete. Dieses Konzept wird an weiteren Datenprodukten, die während der Messflüge aufgenommen wurden, erneut demonstriert und seine universelle Anwendbarkeit gefestigt. Durch den Einsatz von Deep Embedding Clustering (DEC) neuronalen Netzwerken wird dieser Ansatz weiter verallgemeinert, sodass nur noch ein Mindestmaß an Annahmen in die Klassifizierung einfließen.

Auf einer synoptischen Skala werden die Möglichkeiten untersucht, wie Mischungsprozesse in Filamenten von tropischer Luft in die stratosphärische Surf-Zone anhand von Satellitendaten beobachtet werden können. Mittels synthetischer Beobachtungsdaten, die das der European Space Agency (ESA) vorgeschlagene CAIRT-Instrument aus Daten des Chemical Lagrangian Model of the Stratosphere (CLaMS)-Modells imitieren, wird eine ähnliche Form der Parametrisierung vorgestellt, um Filamente in diesen Daten als anomale Strukturen ausfindig zu machen. Diese Parametrisierung erlaubt es, sowohl die räumliche Struktur als auch das dynamische Verhalten der Filamente gleichzeitig zu beschreiben. Trajektorien werden für Luftpackete berechnet und mit diesen Beobachtungsdaten zugeordnet, wodurch Zeitreihen für einzelne Trajektorien gewonnen werden. Es wird gezeigt, wie mit Hilfe von konvolutionellen neuronalen Netzwerken die Information über die Mischungsprozesse aus diesen Zeitreihen extrahiert werden kann, und wie diese Netzwerke verlässliche Vorhersagen für individuelle Trajektorien treffen können.

Auf einer planetaren Skala wird schließlich die Restzirkulation, welche durch die Mischungsprozesse in der Atmosphäre angetrieben wird, anhand des Konzepts des Luftalters betrachtet. Da das Luftalter eine reine Modellgröße ist, die in der Natur nicht direkt gemessen werden kann, muss diese aus anderen, messbaren Größen wie langlebigen Spurengasen abgeleitet werden. Multi-Layer Perceptrons (MLPs) werden benutzt, um äußerst akkurate und präzise Vorhersagen des Luftalters basierend auf langlebigen Spurengasen zu machen, welche konventionellen Methoden übertreffen. Diese MLPs werden bezüglich der erwarteten Messunsicherheiten, der ausgewählten Eingangsgrößen und der Vorhersagen jenseits der Trainingsdomäne optimiert, sodass verlässliche Vorhersagen viele Jahre in die Zukunft möglich werden. Die Erkenntnisse, die im Laufe dieser Dissertation gewonnen wurden, gipfeln in einer neuartigen, datengestützten Größe, welche mittels vorgeschrittenen, physikalisch informierten, Mischung-von-Experten Autoencoder bestimmt wird und das Luftalter akkurat widerspiegelt, wodurch eine unmittelbare Validierung von Klimamodellen mittels echter Messgrößen möglich wird.

Contents

1 Introduction **1**

1.1 General motivation 1

1.2 Stratospheric transport and mixing 3

1.3 Objectives and outline of the thesis 10

2 Theory and concepts **12**

2.1 Conservation laws and isentropic coordinates 12

2.2 Lagrangian kinematics, filamentation, and mixing 14

2.3 Rossby waves 15

2.4 Machine learning 18

2.5 Data structures 18

2.5.1 Similarity and cluster structure 18

2.5.2 Dimensionality reduction 20

2.6 Evaluation metrics and concepts 22

2.7 Neural networks 23

3 Filaments of Asian monsoon air **26**

3.1 Data products and the PHILEAS campaign 27

3.2 Classification concept 32

3.3 Analysis of mixing in 3-D tomographic retrievals 34

3.3.1 Classification 37

3.3.2 Explicit mixing 43

3.3.3 Localization 46

3.3.4 Tracer changes 47

3.4 Analysis of 2-D curtains 49

3.5 Deep embedding clustering using AI . . . 56
3.6 Chapter conclusion . . . 58

4 Filaments in the lower stratosphere 60
4.1 Parametrization . . . 63
4.2 Dynamics . . . 68
4.3 Filament detection . . . 71
4.4 Geometrical analysis . . . 73
4.5 Simulated satellite observations . . . 76
4.5.1 Adaptations for satellite data . . . 76
4.5.2 Revisit times . . . 80
4.6 Prediction of filament evolution . . . 83
4.6.1 Recursive prediction models . . . 88
4.7 Chapter conclusion . . . 91

5 Beyond filaments – mixing in the general circulation 92
5.1 Dataset preparation . . . 95
5.2 Supervised MLPs . . . 97
5.2.1 Prediction performance and influence of noise . . . 98
5.2.2 Feature selection and training domain . . . 104
5.2.3 Predicting the future . . . 110
5.3 Unsupervised autoencoders – an alternative approach . . . 113
5.4 Chapter conclusion . . . 123

6 Summary and outlook 125

7 Danksagung 129

Appendix A 130
A.1 Numerical calculation of trajectories . . . 130
A.2 Matching of trajectories . . . 131
A.3 Modified DBSCAN algorithm . . . 133
A.4 Convolution . . . 134
A.5 NARX models . . . 136
A.6 Time series of trajectories used for filament detection validation . . . 136

A.7 Regularization scheme 140
A.8 MOE AE setup and configuration 142
A.9 Gaussian mixture model 144
A.10 DBSCAN 146
A.11 Properties of the PCA with respect to clustering 147
A.12 Effect of topology on clustering 148
A.13 Visualization of the manifold formed by Long–Lived Tracers (LLTs) 150

List of figures **165**

List of tables **180**

Bibliography **181**

Chapter 1

Introduction

1.1 General motivation

The face of our planet is shaped by the Earth's atmospheric circulation. It redistributes heat, moisture and greenhouse gases across the globe and ultimately is responsible for the climate system we live in today. The atmospheric circulation is not only governed by pure advection, but also by a myriad of mixing processes happening all over the atmosphere, from large-scale Rossby wave breaking in the surf zones to the chaotic mixing in eddies. Transport and mixing are intimately intertwined and do not occur in isolation from each other. We differentiate between different forms of transport: Advection and convection describe the horizontal and vertical transport through the bulk movement of air masses. Stirring describes deformation, stretching and folding of air masses typically along potential vorticity (PV) contours and usually induces mixing, which refers to the chaotic, irreversible change of an atmospheric state. The irreversibility of mixing imprints a directionality on the evolution of the atmosphere, in contrast to the other forms of transport, which are reversible in principle. A good example of the connection of stirring and mixing is the formation of filaments induced by breaking of the planet-scale Rossby waves maintained by the Coriolis force. Inside these filaments stirring enhances the surface between originally distinct air masses, smooths PV contours and induces mixing through small-scale turbulence.

While the advection components of the global circulation are, in principle, mostly well understood and have solid theoretical foundations (e.g. Holton et al., 1995; Bönisch et al., 2011), the mixing processes involved in the circulation are less so (Haynes and Shuckburgh, 2000a,b; Riese et al., 2012; Garny et al., 2014; Ploeger et al., 2015; Curbelo and Linz, 2025; Konopka et al.,

2025). To date they are a source of great uncertainty for climate models. Most mixing processes occur on scales below the resolution of the model, requiring parametrizations of mixing processes, which can differ substantially from model to model. Eulerian models are prone to numerical diffusion (e.g. von Neumann and Richtmyer, 1950; Plumb and Ko, 1992; D'Isidoro et al., 2010), which acts as a form of artificial mixing, whereas many Lagrangian models rely on inter-particle interaction (e.g. McKenna et al., 2002b,a; Stohl et al., 2005). Some models combine both concepts into a hybrid approach (e.g. Kaas et al., 2013; Marro et al., 2015; Engdahl et al., 2017). The mixing of greenhouse gases in particular is sensitive to deviations, as even small changes in mixing ratios can significantly influence the global radiative budget (Riese et al., 2012; Charlesworth et al., 2023). How well climate models manage to capture the true mixing processes is still subject to debate (Hoppe et al., 2014; Charlesworth et al., 2020; Konopka et al., 2022). Incomplete understanding of the different couplings in the atmosphere can have undesirable outcomes. It was thought in the 1950s that atomic detonations in remote, desert regions would not negatively affect the population. Yet, years later the radioactive fallout of the same detonations was observed all over the northern hemisphere. The reason was that the detonations injected radioactive material into the stratosphere, where it distributed globally and was transported back into the troposphere at high latitudes through the Brewer-Dobson circulation (BDC) (Machta, 1959; Feely and Spar, 1960). In the 1970s, it was believed that chlorofluorocarbons (CFCs) as chemically inert, nonflammable gases would be harmless if released into the atmosphere, which was found to be only partially true. While these species are inert in the troposphere, they lead to the creation of chlorine radicals through photolysis. These radicals lead to the catalytic depletion of ozone in the stratosphere, which in turn led to the formation of the ozone hole (Molina and Rowland, 1974; Farman et al., 1985). The extraordinary chemical stability of these species ultimately becomes their most harmful trait; because their only significant sink is the destruction in the upper stratosphere, the slow transport through the stratosphere causes these substances to decay over many decades, until they are eventually converted into reservoir species and removed from the stratosphere (Stolarski and Cicerone, 1974; Rowland et al., 1976; Keene et al., 1983). Cases such as these underscore the importance of understanding the stratospheric circulation and its transports. Research into it is necessary to avoid similar occurrences in the future.

1.2 Stratospheric transport and mixing

For a general overview, first the vertical structure of the global atmospheric circulation system (Fig. 1.1a) is laid out before focusing on the mixing and transport processes (Fig. 1.1b).

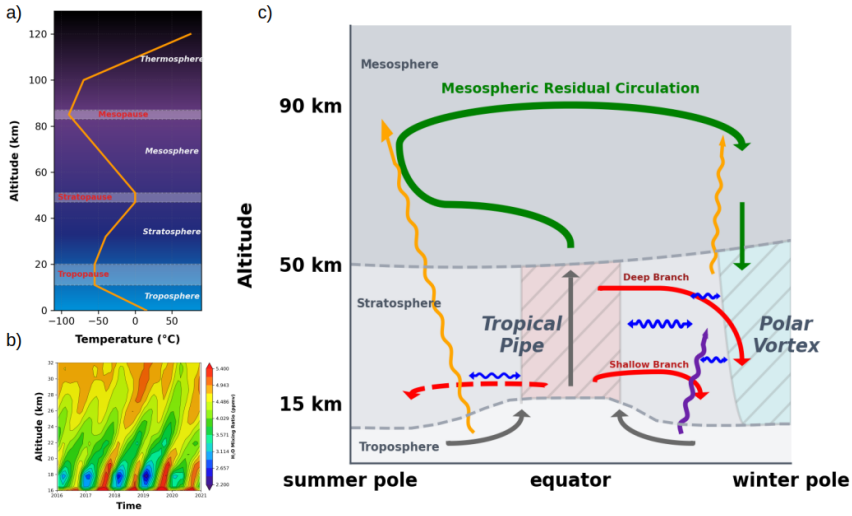

Figure 1.1: The vertical structure of the atmosphere and the atmospheric circulation with respect to mixing. Panel a: Vertical temperature profile. Beginning in the troposphere (0 km), temperature decreases at an average rate of 6.5 K/km until the tropopause ($\approx 15\,km$), beyond which the lapse rate becomes positive due to absorption of radiation by ozone. Temperature increases with height until the stratopause ($\approx 55\,km$) is reached, and the lapse rate becomes negative again in the mesosphere. At the mesopause ($\approx 85\,km$) the lapse rate becomes positive again due to a thin atmosphere and absorbed high-energy radiation. Panel b: Time series of zonal mean water vapour in the tropics (from $-15°$ to $15°$ latitude), also known as the tape-recorder. Due to the asymmetry in wave activity between the hemispheres, the tropopause has lower water vapour content in boreal winter. This freeze-dried air can be seen in the zonal mean signatures as they propagate upwards like an imprint on a tape. The vertical wind speeds can be estimated from the slope of the water vapour signatures and amount to 0.2-0.5 mm/s. Based on data by Davis et al. (2016). Panel c: Summary of the global circulation. The heart of the circulation is the Brewer-Dobson Circulation (BDC), which is driven by Rossby waves (purple arrows). The momentum deposited by breaking of these waves induces mixing (blue arrows), which drives upwelling of tropical air (grey arrow) and horizontal transport in the extratropics (red arrows). The tropical air is confined to the tropical pipe (red shading), from where air is transported dominantly towards the winter pole and the polar vortex (blue shading). Gravity waves (yellow arrows) drive the overturning mesospheric circulation (green arrows), which transports tropical air across the mesosphere towards the polar vortex.

The vertical structure of the atmosphere is as follows: The lowermost part of the atmosphere is the troposphere. It is the wet, convective and most turbulent layer of the atmosphere, and is characterized by strong vertical, turbulent, diabatic mixing. The troposphere is well mixed with

respect to long-lived tracers and PV, which is quite low (around < 1.5 PVU [1]) compared to the rest of the atmosphere. The temperature in the troposphere decreases with height, freezing out most of the water vapour content, until an inflection point is reached, after which the temperature begins to increase with height again. The boundary separating the troposphere and stratosphere is the tropopause. It acts as a transport barrier between the two layers due to the conservation of PV in isentropic mixing processes and is visible in many trace gas species as a sharp gradient. Due to the higher radiation influx in the tropics (compared to the poles), among other effects, the tropopause lies much higher in the tropics than in the extratropics. The interaction between filaments and the tropopause as a transport barrier is the focus of Chap. 3.

There exist multiple commonly used tropopause definitions. The cold point tropopause (CPT) indicates the temperature minimum at the inflection point (Highwood and Hoskins, 1998). The World Meteorological Organization (WMO) defines the (thermal) tropopause as the lowest point where the lapse rate is reduced to 2 K / km and does not exceed this value for the next 2 km (WMO, 1957). The (dynamical) tropopause is defined by the PV gradient, often as the boundary of 2 PVU or some value depending on the potential temperature (Kunz et al., 2013). The tropopause is also reflected in many trace fields, each of which can act as a chemical tropopause. Around the tropopause lies the Upper Troposphere – Lower Stratosphere (UTLS) region, which is one of the most important regions in the atmosphere. In the UTLS, a large share of the exchange between troposphere and stratosphere is occurring, and this exchange of tracers and energy has profound effects on the radiative budget of the atmosphere. This exchange is essential for the shape of our climate (Riese et al., 2014).

The tropopause marks the transition into the stratosphere, the next layer of the atmosphere. Incoming ultraviolet radiation is absorbed by the ozone within the stratosphere, which warms the stratosphere with increasing height. Due to this temperature inversion, the stratosphere is much more stratified than the troposphere, with higher PV (> 2 PVU) and much less (vertical) mixing. The stratosphere extends upwards to around 50 km, until the stratopause, indicated by a local temperature maximum, marks the beginning of the mesosphere. In contrast to the stratosphere, the mesosphere is much less stratified and is subject to strong winds. Beyond the mesopause lies the thermosphere, where temperatures increase sharply with height due to the absorption of high-energy solar radiation.

[1] 1 PVU = $10^{-6} \frac{\mathrm{K \cdot m^2}}{s \cdot \mathrm{kg}}$

If we consider the zonal mean water vapour in the tropics (Fig. 1.1b), we see a recurring pattern over time of high and low water vapour. During boreal winter, increased Rossby wave activity in the northern hemisphere shifts the tropical tropopause upwards, freezing the water contained therein out and led to the formation of these dry features. Over time these features appear to propagate upwards. This structure is known as the tape-recorder model, and provides valuable insights: First, there is a slow upwelling inside the tropics, which extends into the stratosphere. Second, the fact that these features persist for up to 1.5 years tells us that the tropical stratosphere, despite considerable zonal wind speeds, is well isolated from the extratropical stratosphere.

The tropical upwelling is part of a wider stratospheric circulation, known as the Brewer-Dobson Circulation (BDC). It was first postulated after it was found that the maximum of stratospheric ozone does not lie in the tropics, where ozone is predominantly produced, but rather in the extratropical stratosphere (Brewer, 1949). The BDC is the main circulation pattern in the middle atmosphere and is the main mechanism by which the distribution of trace gases is shaped. Rather than describing the rapid, instantaneous three-dimensional flow, the BDC is a zonally averaged, residual circulation, with effective transport times on the order of magnitude of several years, depending on the individual transport pathway. The BDC is driven by the momentum deposited by atmospheric wave drag through the downward-control principle rather than convection. Its vertical structure and its components are shown in Fig. 1.1c:

Rossby waves, which are excited primarily in the troposphere due to temperature gradients between landmasses and oceans, can propagate upward into the stratosphere only under suitable background-wind conditions. In particular, vertical propagation requires westerly winds whose strength lies within a wave-dependent range. This condition acts as a dynamical filter: waves with shorter horizontal scales (wavenumbers > 10) are inhibited in the lower stratosphere, whereas planetary-scale waves with longer horizontal scales (wavenumbers ≤ 4) can reach the middle and upper stratosphere (Charney and Drazin, 1961; Strube et al., 2020; Baikhadzhaev et al., 2025). When the waves break or approach critical levels, they deposit momentum into the mean flow and decelerate it. This in turn weakens the Coriolis force, such that the resulting pressure gradient induces a meridional, residual circulation.

As mentioned before, air is uplifted in the tropics as part of the BDC. This leads to the formation of a stable, mostly isolated (see Fig. 1.1b) vertical column of tropical air, known as the tropical pipe. The tropical air inside the tropical pipe has low PV compared to the extratropical stratosphere, such that direct mixing is heavily suppressed due to PV conservation. However, if a wave

breaks and deposits momentum into the background flow, this momentum can smooth out the PV gradient locally such that transport between the extratropical stratosphere and the tropical pipe becomes possible. In the lower stratosphere – where those Rossby waves with larger wavenumbers break – this leads to horizontal transport out of the tropical pipe known as the shallow branch of the BDC. The tropical pipe 'leaks' into the extratropical stratosphere. As Ploeger et al. (2021b) have shown, the transit times inside the shallow branch are strongly dependent on seasonal effects, with stronger upwelling during boreal winter and subsequent 'flushing' (increased horizontal transport) in boreal summer. Typical transit times inside the shallow branch range from 0.5 to 1.5 years. Rossby waves with lower wavenumber can propagate into the middle stratosphere and eventually break due to steadily increasing amplitude. Their deposited momentum slows down the polar vortex, inducing a meridional, residual circulation known as the deep branch (see Fig. 1.1c). The transit times through the deep branch range from 2 to 5 years, mostly due to the vastly increased distance an air parcel needs to travel. In contrast to the shallow branch, whose driving waves exist all year, the deep branch only exists in the winter hemisphere, where the wave activity is strongest due to strong, poleward temperature gradients. The broad section of the middle and upper stratosphere in the middle latitudes (between 25° and 60°) is called the surf zone (Palmer et al., 1986). The breaking of these low wavenumber waves leads to turbulent distortions of the tropical pipe and eventually the formation of atmospheric filaments of tropical air. This section is characterized by isentropic mixing between the tropical pipe and the extratropical stratosphere. The surf zone and the filaments therein is the focus of Chap. 4. The momentum deposited by a breaking wave not only distorts and smooths the dynamical properties such as PV, but also induces spatial distortion through shear. Segments of tropical air are drawn out and isolated from their original air masses and brought into contact with the extratropical stratospheric background. Initially these filaments show sharp spatial gradients and are a visual indicator for turbulence and onset of mixing. These filaments can persist visually distinct for many days, although the strong turbulence induced by the wave breaking leads to gradual dissolution.

The surf zone is particularly pronounced in the winter hemisphere, where both the tropical pipe and the polar vortex act as transport barriers, which are constantly eroded by the wave breaking. Consequently, the surf zone can be recognized by smoothed PV gradients, smoothed tracer fields and sharp gradients at its borders. This wave breaking is the major driver of the BDC, inducing the residual upwelling in the tropical pipe and the poleward motion. In the surf zone, young air from the tropical pipe is mixed with older air from the vortex. The BDC closes when the warmer,

tropical air is cooled at higher latitudes and descends back into the troposphere.

Another type of atmospheric wave, the so-called gravity waves, are oscillations of air parcels with buoyancy as their restoring force. These waves are unaffected by the Charney-Drazin criterion, because they have a different dispersion relation. In contrast to Rossby waves, gravity waves can propagate well into the increasingly thin mesosphere, where their momentum upon breaking drives the overturning mesospheric circulation towards the boreal polar vortex. Winds in the stratosphere are always easterly in the summer hemisphere and westerly in the winter hemisphere (due to solar heating). Since gravity waves can only propagate against the wind, the stratosphere filters out easterly waves in the summer hemisphere and westerly waves in the winter hemisphere. In the summer hemisphere, the momentum of the breaking westerly waves induces a residual circulation towards the equator; in the winter hemisphere, the breaking easterly waves likewise induce a poleward residual circulation. The resulting circulation is an overturning circulation from the summer pole to the winter pole (see Fig. 1.1c).

In addition to the primary tropical upwelling branch of the BDC, regional circulation systems play a crucial role in troposphere-to-stratosphere transport. Specifically, the Asian Summer Monsoon (ASM) acts as a major subtropical conduit during summer, bypassing the slow tropical ascent by rapidly injecting young air masses directly into the lower stratosphere, which fundamentally alters local chemical compositions and Age-of-Air distributions.

To describe the transport and the mixing of air within the BDC, the concept of Age-of-Air (AoA) was introduced (Hall and Plumb, 1994). It describes the lag time of an air parcel since it left the troposphere. The age of an air parcel is technically a spectrum of the ages of all air constituting the parcel and therefore changes during mixing. It is common to refer to the mean age of an air parcel as AoA. With respect to AoA the tropical pipe consists of young air, which ages slowly with altitude due to the upwelling. The shallow branch therefore transports generally younger air compared to the deep branch, with younger air being mixed with older air inside the surf zone. Mixing with older air effectively ages air masses additionally, such that the apparent AoA is older than a pure transport would suggest. The concept of AoA is the focus of Chap. 5.

The surf zone in particular can act as a 'trap' for air parcels, as they can neither reenter the tropical pipe nor the polar vortex horizontally (Charney and Drazin, 1961; Baikhadzhaev et al., 2025). An additional contribution to AoA is air passing through the mesospheric circulation, which can yield AoA of up to six years. The concept of AoA illustrates the coupling between the net circulation transport and the accompanying mixing processes. Similar structures are likewise

visible in the distribution of long-lived tracers.

Developments over the last decade regarding the BDC have led to the *apparent* 'Age-of-Air Paradox': simulations based on reanalysis data (Ploeger et al., 2021b) and comparisons between many different models (Butchart et al., 2010; Garny et al., 2011; Abalos et al., 2021; Garny et al., 2014) consistently show a long-term strengthening of the BDC. Garfinkel et al. (2017) point out that while the BDC accelerates over long periods, the mid-stratosphere in the mid-latitude northern hemisphere does not accelerate in the ensemble mean since the late 1980s and even decelerates in one ensemble member. Potential effects are indeed asymmetrical with respect to hemisphere and season (e.g. Ploeger and Garny, 2022). The shallow branch is projected to accelerate more strongly compared to the deep branch (Lin and Fu, 2013). Polvani et al. (2019) attribute most of the strengthening of the BDC to ozone depleting substances, which would suggest some deceleration with decreases in these substances. On the other hand, the AoA derived from observed trace gas species (Engel et al., 2009b; Stiller et al., 2012; Engel et al., 2017; Garny et al., 2024b, such as) consistently shows a trend towards older AoA, indicating an apparent deceleration of the BDC. The 'paradox' stems from the asymmetry of the BDC, which many models did not adequately consider, such that the differences between models and observations are rather due to local transport variations. Nonetheless, it was found that the BDC is more complex than previously assumed. To understand this complexity, resolving mixing processes in the BDC is key. The general consensus is that increases in AoA are likely to be attributed to increased wave activity and changes in the wave breaking (Austin and Li, 2006; Ploeger and Garny, 2022; Garny et al., 2024b; Orbe et al., 2025; Voet et al., 2025, see). The distributions of chemical species (or AoA in models) have become the main diagnostics by which the BDC is investigated. Rather than considering the mean AoA of an air parcel, there have been efforts to resolve AoA in terms of the AoA spectrum of the air parcel (Hall and Plumb, 1994; Garny et al., 2024a) and with respect to contributions due to transit and due to mixing (Dietmüller et al., 2018; Ploeger et al., 2019b; Garny et al., 2024b).

Further study requires access to real observations. It was found that the parametrization of these processes in climate models are still a major source of uncertainty (McFarlane, 1987; Gettelman, 2023), which couples back into other work reliant on accurate models. Direct observation of explicit mixing processes has proven to be extraordinarily difficult. For example, the MATCH campaign conducted by von der Gathen et al. (1995) between 1991 and 1992 performed ozone measurements with hundreds of weather balloons equipped with ozone probes at different loca-

tions in the northern polar region (Rex et al., 1997). By tracking the trajectories of these probes, new balloons were launched where trajectories of previous measurements would intersect with the measurement station. This way changes along individual trajectories due to photochemistry were isolated from changes due to mixing. Very few similar experiments were conducted, particularly because weather balloons cannot be guided and matching a previously measured air parcel is a matter of considerable chance, thus requiring a large number of probes. Simultaneously observing many different chemical tracers can significantly constrain localized mixing processes, as these mixing processes must necessarily be reflected in each species (in the absence of chemistry). Infrared sounders are ideally suited for this, as their remote sensing measurements allow measurements even in turbulent filaments. Filaments are particularly well suited to study mixing processes, as the irreversible, molecular mixing is proportional to the square of the contour length of the filament (Haynes and Shuckburgh, 2000a,b). Filaments in the UTLS are accessible via aircraft, while filaments in the surf zone require satellite observations. In either case, the spatial resolution of the instrument is key. Air parcels must be sufficiently resolved to differentiate between different air parcels. An aircraft can be guided towards the air parcels of interest and their outflow, making dedicated self-match experiments possible. Whether a satellite instrument can similarly measure air parcels at subsequent times depends on its global coverage and track width. Moreover, handling complex data and correlating many quantities at once requires advanced algorithms and methods. Artificial intelligence, which is emerging in many scientific fields, is a well-suited tool for such tasks. Particularly neural networks can be utilized to discover complex structures in the data itself, allowing one to gain insights into mixing processes without relying on a rigid theoretical framework.

1.3 Objectives and outline of the thesis

Therefore, the first question this thesis aims to answer is: **Given a suitable classification of air masses, can spatially resolved, consecutive measurements of the same air parcels be used to derive the mixing processes inside filaments in the UTLS in a physically meaningful way?** Such consecutive measurements can be obtained by dedicated measurement flights conducted by aircraft. A meaningful classification is required to identify the initial and final chemical states, from which the changes due to mixing can be derived. AI methods could prove valuable in this due to their capabilities to correlate even high-dimensional data in a model-agnostic approach.

The second question this thesis aims to answer is whether a comparable observation-based approach can be transferred from aircraft measurements in the UTLS to satellite observations of tropical-air filaments in the lower-stratospheric surf zone. In this setting, filaments are identified as spatially anomalous structures in multi-tracer observations. By associating observations with the same air parcels at consecutive times, changes in their classified state can be used as a measure of mixing. In other terms: **Can satellite observations identify and repeatedly sample filaments in the lower-stratospheric surf zone such that their mixing processes can be quantified? How sensitive is this capability to the tracer selection, the definition of filamentary anomalies, and the satellite measurement geometry?** The chemical composition of the surf zone is dominated primarily by long-lived tracers, so the representation used to identify filaments differs fundamentally from the air-mass classification applied to the chemically more heterogeneous UTLS observations in Chap. 3. In addition, the sparse spatial sampling, revisit interval, and viewing geometry of a satellite constrain both the probability of observing a filament and the possibility of associating observations with the same evolving air parcels. Where sufficiently long time series can be obtained, this thesis further examines whether neural networks can extract information on mixing from these observations and predict the evolution of individual air parcels.

Finally, accurate estimates of mean AoA from observations of long-lived trace gases are essential for evaluating representations of stratospheric transport in climate models. Conventional tracer-based approaches are limited by the complex and non-linear influence of transport and mixing on the relationships between individual trace gases and AoA. The third question this thesis therefore aims to answer is: **To what extent can neural networks improve estimates of stratospheric mean AoA relative to conventional tracer-based methods by learning the**

multivariate, nonlinear relationship between long-lived trace gases and AoA? Can this relationship also be used to construct an observation-based representation of AoA without requiring model AoA as a training target? The trained models are further analysed to assess the influence of individual trace gases, measurement noise, the choice of input features, and the robustness of the inferred relationships outside the training domain. This also tests whether the information contained in observed trace-gas distributions can support a meaningful, data-driven representation of stratospheric transport and mixing.

This thesis is structured as follows: Chap.2 provides a brief summary of the most important concepts for understanding mixing and a brief introduction into essential AI concepts used for the less traditional methodology in the various chapters. Chap.3 analyses the mixing processes observed by the Gimballed Limb Observer for Radiance Imaging of the Atmosphere (GLORIA) during a self-matching experiment conducted over two days in the 2023 PHILEAS campaign. It shows how mixing processes can be derived from a model-agnostic point of view, and demonstrates the validity of the approach through application of the methodology to additional datasets. Chap.4 takes the general concept described in Chap. 3and applies it to simulated observations of the Earth Explorer 11 candidate CAIRT. It describes a formalism through which filaments can be identified, and tracks air parcels over the span of two weeks. It discusses how mixing processes can be derived from observations compared to ideal model data and demonstrates how neural networks can be used to infer information about the mixing processes from these spatially sparse observations. Chap. 5 considers the derivation of mean AoA from long-lived trace gas species using neural networks. Its focus lies on the predictive performance of the networks, the choice of input parameters and the robustness against measurement noise. It closes with the demonstration of a concept to derive a model-agnostic measure of AoA from observations only as an alternative proxy for AoA. Chap.6 summarizes the contents of this thesis and provides a brief outlook into future research into the field of mixing processes.

Chapter 2

Theory and concepts

The study of the mixing processes in filaments requires a robust understanding of the kinematics of fluids and the physical principles governing it. The following chapter also utilizes dedicated machine learning concepts with which most atmospheric scientists may not be intimately familiar. This chapter provides both a brief overview of the core concepts used to describe mixing in the atmosphere and a selection of concepts from the field of machine learning that are relevant for each chapter. Each of the following chapters does provide additional information about machine learning concepts which are uniquely relevant for that chapter.

2.1 Conservation laws and isentropic coordinates

The dynamics of air parcels in the stratosphere are conveniently described in isentropic coordinates. The potential temperature

$$\theta = T \left(\frac{p_0}{p} \right)^{\frac{R}{c_p}} \tag{2.1}$$

is the temperature that an air parcel of temperature T and pressure p would attain if it were brought adiabatically to a reference pressure p_0. Here, R denotes the specific gas constant for dry air and c_p the specific heat capacity at constant pressure. The material tendency of potential temperature is given by

$$\frac{\mathrm{d}\theta}{\mathrm{d}t} = \frac{\theta}{c_p T} Q, \tag{2.2}$$

where Q denotes the diabatic heating rate. In the absence (or negligibility) of diabatic heating, potential temperature is conserved following the motion of an air parcel. Although radiative heat-

ing and cooling are not strictly zero in the atmosphere, they are sufficiently weak over the short timescales relevant to the predominantly adiabatic mixing processes considered here. Air-parcel motion and mixing can therefore be approximated as occurring along surfaces of constant potential temperature, referred to as isentropic surfaces. Accordingly, θ provides a physically meaningful vertical coordinate for the analysis of stratospheric transport and mixing.

Potential temperature is not the only approximately conserved quantity. Under adiabatic and frictionless conditions, the Ertel potential vorticity is also materially conserved:

$$PV = \frac{1}{\rho}\zeta_a \nabla\theta \tag{2.3}$$

with the absolute vorticity ζ_a. The change of PV is then given by:

$$\frac{dPV}{dt} = \frac{1}{\rho}\zeta_a \nabla\dot{\theta} + \frac{1}{\rho}\nabla \times \mathbf{F} \cdot \nabla\theta \tag{2.4}$$

where $\boldsymbol{F}$ denotes the friction forces. Since θ is conserved for adiabatic processes and friction forces can be neglected, Eq. 2.4 yields $\frac{dPV}{dt} = 0$, i.e. PV must also be conserved. These conserved quantities are essential for the structure of the atmosphere, and for mixing processes.

Lastly, it is beneficial to define another coordinate: To filter out reversible, wave-induced spatial fluctuations of some conservative tracer field (C), the flow-following coordinate equivalent latitude (ϕ_e) is introduced. Instead of relying on geographic coordinates, ϕ_e is defined by mapping the deformed area ($A(C)$) enclosed by a specific tracer contour onto an unperturbed, equivalent spherical cap centered at the pole:

$$\phi_e(C) = \arcsin\left(1 - \frac{A(C)}{2\pi R_E^2}\right) \tag{2.5}$$

where R_E denotes the Earth's radius, and $A(C)$ represents the total hemispheric surface area where the tracer value exceeds (or falls below, depending on the gradient) the threshold C.

2.2 Lagrangian kinematics, filamentation, and mixing

Atmospheric motion can be described from either an Eulerian or a Lagrangian perspective. In an Eulerian framework, atmospheric quantities are evaluated at fixed positions in space. In a Lagrangian framework, an air parcel is followed along its trajectory. The latter is particularly useful for the analysis of mixing, because it directly relates changes in tracer composition to the history and deformation of individual air parcels.

For an arbitrary atmospheric quantity C, the material derivative relates the two descriptions:

$$\frac{\mathrm{D}C}{\mathrm{D}t} = \frac{\partial C}{\partial t} + \mathbf{v} \cdot \nabla C,$$

where $\mathbf{v}$ denotes the velocity field. Given $\mathbf{v}$, the trajectory $\mathbf{x}(t)$ of an air parcel is obtained from

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{v}\left(\mathbf{x}, t\right). \tag{2.6}$$

In the following, the wind field is regarded as prescribed; the forces that generate the flow are therefore not considered explicitly.

To describe the deformation of an air mass, one can consider two initially neighbouring air parcels at positions $\mathbf{x}$ and $\mathbf{x} + \delta\mathbf{x}$. Their relative displacement evolves over time according to:

$$\frac{\mathrm{d}\delta\mathbf{x}}{\mathrm{d}t} = \mathbf{v}(\mathbf{x} + \delta\mathbf{x}, t) - \mathbf{v}(\mathbf{x}, t) \approx \mathbf{J}\,\delta\mathbf{x},$$

where $\mathbf{J} = \nabla\mathbf{v}$ is the horizontal velocity-gradient tensor. For an incompressible, large-scale flow on an isentropic surface, $\mathbf{J}$ can be decomposed into a symmetric deformation tensor $\mathbf{D}$ and an antisymmetric rotation tensor $\mathbf{W}$:

$$\mathbf{J} = \mathbf{D} + \mathbf{W}.$$

The velocity-gradient tensor $\mathbf{J}$ determines how the separation between neighbouring trajectories evolves. Its rotational component $\mathbf{W}$ changes the orientation of an air-parcel ensemble without changing its shape. Its deformational component $\mathbf{D}$, consisting of strain and shear, changes the ensemble shape by stretching it in one direction and compressing it in another. When deformation dominates over rotation, initially compact air masses are stretched and folded into thin filaments. The separation of neighbouring trajectories then grows approximately exponentially along the

stretching direction:

$$|\delta\mathbf{x}(t)| \propto |\delta\mathbf{x}(0)| \exp(\lambda t),$$

where λ denotes a Lyapunov exponent. This process is commonly referred to as stirring. However, stirring alone does not change the tracer composition of an individual air parcel. In the absence of chemical sources and sinks, the evolution of a tracer C is described by the advection–diffusion equation:

$$\rho \frac{\mathrm{D}C}{\mathrm{D}t} = \nabla \cdot (\rho \kappa \nabla C),$$

where ρ is air density and κ is the molecular or effective diffusivity. Stretching a filament increases the tracer gradients between the filament and its surroundings. Diffusive transport therefore becomes increasingly effective and irreversibly smooths these gradients, leading to mixing.

For a tracer contour of length L, the effective diffusivity associated with this deformation can be expressed as:

$$\kappa_{\mathrm{eff}}(\phi_e, t) = \kappa \left(\frac{L(\phi_e, t)}{L_{\min}(\phi_e)} \right)^2,$$

where $L_{\min}(\phi_e) = 2\pi R_E \cos\phi_e$ is the length of the corresponding undisturbed contour at equivalent latitude ϕ_e. Thus, filament elongation strongly enhances effective mixing even when the underlying diffusivity is small (Haynes and Shuckburgh, 2000a).

2.3 Rossby waves

The main drivers of the global circulation are atmospheric waves. There are gravity waves, which are oscillations in buoyancy, and planetary waves – also known as Rossby waves – which are a direct result of the conservation of PV. If an air parcel is displaced to the north, the planetary vorticity increases. This is described by the β-parameter:

$$\beta = \frac{\partial f}{\partial y} = \frac{2\Omega \cos\phi}{R_E} \tag{2.7}$$

where ϕ denotes latitude. Due to the conservation of the total vorticity, the air parcel's relative vorticity must decrease to compensate. This results in an anticyclonic eddy, which pushes the air parcel southwards. Once sufficiently displaced southwards, the same process happens in the

opposite direction. This leads to the formation of a wave. The Rossby wave (in 3-D) must satisfy:

$$\frac{dq}{dt} = 0 \qquad \text{with} \qquad q = \nabla_h^2 \psi + \frac{\partial}{\partial z}\left(\frac{f_0^2}{N^2}\frac{\partial \psi}{\partial z}\right) + \beta y \tag{2.8}$$

where q denotes the quasi-geostrophic vorticity, ψ denotes the geostrophic stream function derived from the horizontal winds, f_0 denotes the Coriolis parameter and N is the Brunt–Väisälä frequency:

$$N = \sqrt{\frac{g}{\theta}\frac{\partial \theta}{\partial z}} \tag{2.9}$$

which describes the vertical stability of a stratified fluid. Assuming a harmonic wave, a Rossby wave has the dispersion relation ω:

$$\omega = \bar{u}k - \frac{\beta k}{k^2 + l^2 + \frac{f_0^2}{N^2}m^2} \tag{2.10}$$

where ω denotes the angular frequency of the wave, k, l, m denote the zonal, meridional and vertical wavenumbers, respectively, and $\bar{u}$ denotes the mean zonal wind. The magnitude of the wave vector is given by $K^2 = k^2 + l^2 + m^2$. The phase velocity is a vector with components:

$$c_x = \frac{\omega}{k} = \bar{u} - \frac{\beta}{k^2 + l^2 + \frac{f_0^2}{N^2}m^2} \tag{2.11}$$

$$c_y = \frac{\omega}{l} = \bar{u}\frac{k}{l} - \frac{\beta k}{l\left(k^2 + l^2 + \frac{f_0^2}{N^2}m^2\right)} \tag{2.12}$$

$$c_z = \frac{\omega}{m} = \bar{u}\frac{k}{m} - \frac{\beta k}{m\left(k^2 + l^2 + \frac{f_0^2}{N^2}m^2\right)} \tag{2.13}$$

These components provide a few important insights relevant for this thesis: Since $\frac{\beta}{k^2+l^2+\frac{f_0^2}{N^2}m^2}$ is always positive, we get $c_x < \bar{u}$. A Rossby wave propagates westwards with respect to the mean background flow. In the winter stratosphere, the wind $\bar{u}$ is eastwards, such that the Rossby waves can become stationary with respect to the ground. The group velocity, which indicates the

direction in which the wave transports its energy, is likewise a vector with components:

$$c_{gx} = \frac{\partial \omega}{\partial k} = \bar{u} + \frac{\beta\left(k^2 - l^2 - \frac{f_0^2}{N^2}m^2\right)}{\left(k^2 + l^2 + \frac{f_0^2}{N^2}m^2\right)^2} \tag{2.14}$$

$$c_{gy} = \frac{\partial \omega}{\partial l} = \frac{2\beta k l}{\left(k^2 + l^2 + \frac{f_0^2}{N^2}m^2\right)^2} \tag{2.15}$$

$$c_{gz} = \frac{\partial \omega}{\partial m} = \frac{2\beta k m\left(\frac{f_0^2}{N^2}\right)}{\left(k^2 + l^2 + \frac{f_0^2}{N^2}m^2\right)^2} \tag{2.16}$$

In order for a Rossby wave to propagate upwards (and reach the stratosphere, where it can deposit its momentum), both k and m must share the same sign. Therefore, lines of constant phase must lean to the west with increasing height. A Rossby wave can only propagate upwards if m^2 is a real, positive number. For a stationary wave ($\omega = 0$), which is the most important driver for filamentation in the stratosphere, Eq. 2.10 yields:

$$m^2 = \frac{N^2}{f_0^2}\left(\frac{\beta}{\bar{u}} - (k^2 + l^2)\right) > 0 \tag{2.17}$$

This relationship forms the basis of the Charney–Drazin criterion. Setting the threshold for vertical propagation at $m^2 > 0$ defines the upper limit for the mean zonal wind speed:

$$0 < \bar{u} < U_c \quad \text{with} \quad U_c = \frac{\beta}{k^2 + l^2} \tag{2.18}$$

While the original formulation by Charney and Drazin (1961) includes an additional atmospheric stratification term ($\frac{f_0^2}{4N^2H^2}$) in the denominator of U_c due to upper-boundary conditions in a semi-infinite atmosphere, this simplified version captures the essential filtering mechanism. Given a specific background wind $\bar{u}$, the criterion effectively blocks smaller-scale Rossby waves with large horizontal wavenumbers k, l, allowing only planetary-scale waves to propagate vertically through the stratosphere.

2.4 Machine learning

Machine learning is a commonly employed and powerful tool in modern science. It can analyse data in a way a human scientist cannot, be it due to the complexity of the data or due to the sheer volume of it. This thesis approaches many traditional problems in atmospheric science from a data-scientific point of view. Every chapter begins with the introduction of niche concepts specific to the content of that chapter. The philosophy behind these concepts, however, is mostly the same. This section provides a brief overview of shared concepts to establish a solid understanding for the various problems that can occur when utilizing machine learning.

2.5 Data structures

The term *structure* is deeply ingrained in our everyday language, and we have an intuitive understanding of it. Formally, structure describes the constellation and relationships of a complex system. Structure is an inherent property of the system and is generally not exclusive. We differentiate between different kinds of structure, and these structures are not exclusive, such that a system may have many structures at the same time. From an information-theoretic perspective, structure is the key that unlocks predictability, and therefore enables new knowledge. The methods used in this thesis exploit different types of underlying data structure, which are summarized below.

2.5.1 Similarity and cluster structure

Apples and oranges are not the same. This popular idiom conveys a lot about similarity structure: Apples (A) and oranges (O) can be characterized by a myriad of different properties, such as colour, shape or taste. We consider A distinct from O if there exist some properties by which we can differentiate between them. For example, A may have many naturally occurring colours, but they are never orange. We could use their natural colour to differentiate between A and O. On the other hand, no two oranges are completely identical, neither by colour nor by shape or any other property. And yet we can look at a basket of oranges and claim: these are all oranges. Similarity therefore depends on the attributes used to compare and on the strictness of these comparisons. Formally, given a set of N elements x_n, we consider two elements x_i, x_j to be similar if

$$\|x_i, x_j\| < c \tag{2.19}$$

where $\|\cdot\|$ denotes some metric with respect to some qualities and c denotes a threshold we define. In Chap. 4 this similarity structure is used to identify coherent spatial structure in potentially sparse data. Likewise, we consider x_i, x_j to be dissimilar if their similarity does not meet c. The next type of structure directly follows from this argument: If we began putting all fruits which are similar to a randomly picked first fruit into the same basket, and all others into another, we would observe that all fruits in our first basket are similar to each other, and that every fruit in that basket is dissimilar to any fruit in the other basket. We could then randomly pick a fruit out of the second basket and repeat the process with a third basket, and so on. In our apple-and-oranges example we would end up with two baskets of fruit, which are all similar within the basket and as per construction dissimilar to all fruits not part of their basket. Data which can be partitioned into different groups (or clusters) has *cluster* structure. The number of clusters in the data depends on the threshold c and the metric $\|\cdot\|$ used to partition it. Formally, a clustering algorithm which partitions a set X of N elements x_n into K clusters C_k minimizes the within-cluster variability (Sum-of-Squares):

$$SS_{\text{within}} = \sum_{k=1}^{K} \sum_{\mathbf{x}_i \in C_k} \|\mathbf{x}_i - \boldsymbol{\mu}_k\|^2 \tag{2.20}$$

while simultaneously maximizing the inter-cluster variability:

$$SS_{\text{between}} = \sum_{k=1}^{K} |C_k| \cdot \|\boldsymbol{\mu}_k - \boldsymbol{\mu}_{\text{global}}\|^2 \tag{2.21}$$

where $|C_k|$ denotes the cardinality of the cluster k, i.e. the number of points contained within that cluster, and $\boldsymbol{\mu}_k$, $\boldsymbol{\mu}_{\text{global}}$ denote the centroids:

$$\boldsymbol{\mu}_k = \frac{1}{|C_k|} \sum_{\mathbf{x}_i \in C_k} \mathbf{x}_i, \quad \boldsymbol{\mu}_{\text{global}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i \tag{2.22}$$

In Chap. 3 this concept is applied to derive a suitable classification of air masses from their internal data structure alone and to express mixing processes in terms of changes in similarity.

Clustering algorithms, which determine a suitable partition of the data, may provide either a hard partitioning by assigning each data point to one cluster, or a soft partitioning by assigning each data point a probability for each cluster. We differentiate between density-based clustering, such as the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm used in Chap. 4, which uses fixed similarity thresholds but a variable number of clusters, and centroid-based clustering, such as the Gaussian Mixture Model (GMM) used in Chap. 3, which

uses a fixed number of clusters but with variable similarity thresholds. Both algorithms are explained in App. A.10 and App. A.9, respectively. Clustering falls into the category of unsupervised machine learning. There are two important lessons to draw from similarity and cluster structure: First, we can approach a problem from a model-agnostic point of view by simply considering the clusters found within the data. Even if we had no concept of A or O, we would still be capable of deducing that A and O are two different kinds of fruit. Second, by simply sorting data into clusters we can deduce the number of clusters in the data without needing to know this number beforehand. The number of distinct types of fruit in our initial fruit basket corresponds to the number of baskets needed to divide them into groups of similar fruit. Formally, the silhouette score (Rousseeuw, 1987)

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i),\, b(i)\}} \tag{2.23}$$

of some data point x_i can be used to determine the number of clusters within some data, as it describes the dissimilarity of a point to all points outside of its cluster relative to the similarity between itself and other points inside of its cluster given some cluster assignment. Here, $a(i)$, $b(i)$ denote the mean distances to all points inside its own cluster and outside of its own cluster, respectively. For datasets, the mean value of the silhouette scores is taken.

2.5.2 Dimensionality reduction

An important, but highly philosophical distinction is that between what is referred to as the observational space and the phase space. When handling data we must recognize that the data is technically an object from an observational space. The insights we aim to derive from this object, however, are contained in the phase space, not necessarily the observational space. It is best understood with a simple example: Let us assume we have a piece of paper that someone crumpled up (maybe in frustration over the nature of observational spaces). We can see that this ball of paper is a 3-D object. We could assume that we require at least three independent coordinates to fully describe it, since it exists in 3-D observational space. If we instead begin to unfold the paper it will eventually unravel into a 2-D sheet of paper. While the transformation by which we turn the 3-D paper ball back into the 2-D sheet of paper is highly non-linear, it is simply given by the inverse of each motion used to crumple it. No matter how hard we may try, we cannot further reduce its dimensionality without tearing it up, thus changing its nature (topology). The sheet of paper has an *intrinsic dimensionality* of 2, which corresponds to its representation in phase

space and is an intrinsic property of it. Crumbling it up has resulted in a 3-D representation in observational space, but by its very nature it is still but a 2-D object (if we neglect its thickness of course). Similarly, a phenomenon or a process may be captured in some higher–dimensional observational space, but could be represented by a much lower–dimensional phase space. In general, the components of the observational space are not components of the phase space. Instead, it is common that all components of the observational space are required to derive the components of the phase space. The observational space of an object could also be incomplete, such that parts of the phase space are not included. A more complete observational space is therefore generally desirable. Let X be a metric space and $S \subseteq X$ a subset. For any $d \in [0, \infty)$ and $\delta > 0$, the d–dimensional Hausdorff pre-measure (or content) of S at scale δ is defined as:

$$\mathcal{H}^d_\delta(S) = \inf \left\{ \sum_{i=1}^{\infty} (\operatorname{diam} U_i)^d : \bigcup_{i=1}^{\infty} U_i \supseteq S, \operatorname{diam} U_i \leq \delta \right\} \tag{2.24}$$

where $\operatorname{diam} U_i$ denotes the diameter of the covering set U_i, defined as the greatest distance between any two points in U_i. The d–dimensional Hausdorff measure $\mathcal{H}^d(S)$ is obtained by taking the limit as the resolution scale approaches zero:

$$\mathcal{H}^d(S) = \lim_{\delta \to 0} \mathcal{H}^d_\delta(S) \tag{2.25}$$

As d varies, the value of $\mathcal{H}^d(S)$ jumps from ∞ to 0 at a critical threshold. This critical value defines the Hausdorff dimension $\dim_{\mathrm{H}}(X)$ of the set, formally expressed as the infimum of all dimensions for which the measure vanishes:

$$\dim_{\mathrm{H}}(S) := \inf\{d \geq 0 : \mathcal{H}^d(S) = 0\} \tag{2.26}$$

which is equivalently the supremum of all d for which $\mathcal{H}^d(S) = \infty$. Intuitively, it is the highest dimension in which the measure of an object is still finite. Our sheet of paper (even crumpled up in 3-D) has a vanishing measure in 3-D (very small volume due to its minimal thickness), a finite measure in 2-D (a finite surface) but an infinite measure in 1-D (the length of a line completely covering the entire sheet is infinite). Thus, the sheet of paper has a Hausdorff dimension of 2.

Generally, even non-integer values for the Hausdorff dimension are not only possible, but rather the norm. In practice there exist several methods to estimate the effective dimensionality of data. One way is to consider the Fisher separability (Albergante et al., 2019), which determines

the effective dimensionality such that points within the lower–dimensional subspace $\boldsymbol{X}$ are more similar to each other than they are to points within the higher–dimensional space $\boldsymbol{Y}$ following:

$$X^\top X > X^\top Y \quad \text{for all } Y \neq X \tag{2.27}$$

As intrinsic dimensionality is an inherent property, the choice of method is usually not too critical and different approaches will yield roughly (qualitatively) similar results. If the intrinsic dimensionality of some data is found, a transformation into the minimal phase space needs to be found. A traditional form of dimensionality reduction is the Principal Component Analysis (PCA, Pearson, 1901; Hotelling, 1933; Jolliffe and Cadima, 2016a), which performs linear rotations of high dimensional data and projects the data onto a set of its principal components such that the variance of the data is best preserved. The PCA is used in Chap. 3 to simplify observational data. Since not all data can be reduced to its effective dimensionality linearly, autoencoders provide highly non-linear transformations. An autoencoder is a neural network in which two components – the encoder and the decoder – attempt to find a lower–dimensional embedding of the data (encoder) such that the original data can be accurately reconstructed from that embedding (decoder). An overview of neural networks is given in Sect. 2.7. Autoencoders are a functional part of another neural network architecture used in Chap. 3 and are used in Chap. 5 to derive an alternative definition of Age-of-Air.

2.6 Evaluation metrics and concepts

In predictive modelling, uncertainty is fundamentally divided into two distinct categories: *aleatoric* uncertainty, which arises from inherent and irreducible randomness in a system, and *epistemic* uncertainty, which stems from a lack of knowledge, data or an insufficient model. If we attempt to predict whether it will rain tomorrow based on temperature measurements, there is a fundamental (aleatoric) limit to how well tomorrow's rain can be predicted from temperature alone. If our measurements are also imprecise, e.g. if we manually read an analogue thermometer, we will induce additional epistemic uncertainty, making our predictions even less likely to succeed.

Formally, a predictive model $\mathcal{M}$ should be both *accurate* and *precise*: given a predicted value $Y_p = \mathcal{M}(X)$ and a corresponding reference value (or true value) Y_r, we consider $\mathcal{M}$ accurate if $\epsilon_1 = \|Y_p - Y_r\|$ is sufficiently small, where $\|\cdot\|$ denotes a metric such as the Euclidean metric. The accuracy ϵ_1 is typically larger than zero, and reflects the aleatoric uncertainty of the prediction

problem. We consider a model to be precise if the prediction of a small variation δ of an input $Y_\delta = \mathcal{M}(X - \delta)$ is close to the prediction Y_p, i.e. if $\epsilon_2 = \|Y_\delta - Y_p\|$ is small as well. Mathematically the model is required to be continuous, which is usually the case in practice. An over-fitted model is highly accurate, but tends to have poor precision, since it does not reflect any X outside of its training domain. For a model which is both accurate and precise, we get $\|Y_\delta - Y_r\| \to \epsilon$, where ϵ is small. We define this as the *performance* of a model, which is not an established expression, but is convenient. A common metric to measure the average predictive capability of $\mathcal{M}$ is the root-mean-square Error (RMSE):

$$\text{RMSE}(A, B) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(A_i - B_i)^2} \tag{2.28}$$

or the mean absolute error (MAE):

$$\text{MAE}(A, B) = \frac{1}{N}\sum_{i=1}^{N}|A_i - B_i| \tag{2.29}$$

where A, B denote the quantities which are compared against each other. In terms of the RMSE the accuracy of a model can be expressed as RMSE(Y_p, Y_r), the precision as RMSE(Y_δ, Y_p) and the performance as RMSE(Y_δ, Y_r).

2.7 Neural networks

A neural network is a machine-learning model consisting of multiple interconnected nodes (neurons) arranged in layers following the example of biological brains such that the network can 'learn' a function or relation from provided training data. The main advantage of a neural network is that no prior knowledge of the function that it to be learned is needed, neither in terms of parameterization, nor form nor the relevance of the inputs provided. All information about the function is derived from the training data itself. If reference (true) values are provided, which the neural network must attempt to predict during training, it is referred to as a supervised neural network. If no such reference values are given it is referred to as an unsupervised neural network. The majority of neural networks are supervised networks.

Each neuron of a neural network performs a simple mathematical operation. The core idea is that each neuron must correlate the state of each neuron that feeds into it via trainable weights

$\mathbf{w}_i$ and figure out what information is relevant for the correct prediction. In addition, each neuron usually has a trainable offset b, which indicates how important each neuron is for the prediction. Before training, each neuron is initialized with random weights and offsets. Given some input values X and some reference values Y, each neuron passes a value to each neuron in the next layer which it is connected to via:

$$z = \boldsymbol{w}^\top \boldsymbol{x} + b \tag{2.30}$$

$$y = \phi(z) \tag{2.31}$$

where $\mathbf{x}$ are the outputs from all connected neurons in the previous layer, $\mathbf{w}$ is the part of the weight matrix $\mathbf{W}$ of the respective connection, b is the offset of the neuron and ϕ is an activation function which introduces a non-linearity, such as the *Rectified Linear Unit* (ReLU):

$$\phi(z) = \mathsf{max}(0, z) \tag{2.32}$$

The output of the neural network is then compared to Y via some loss function such as the RMSE (Eq. 2.28). The loss function is used to judge the accuracy of the prediction. Random weights will generally not accurately predict the correct Y. To determine how much each weight in each neuron has contributed to this wrong result, the loss is passed backwards through the network via backpropagation: At the output layer, the loss error δ is simply calculated as the gradient of the loss function$L$:

$$\boldsymbol{\delta}^{[L]} = \frac{\partial L}{\partial \mathbf{y}^{[L]}} \odot g'^{[L]}\left(\mathbf{z}^{[L]}\right) \tag{2.33}$$

where $\odot$ denotes the Hadamard product. For each preceding layer, the local gradient of the loss is calculated following:

$$\boldsymbol{\delta}^{[l]} = \left[\left(\mathbf{W}^{[l+1]}\right)^T \boldsymbol{\delta}^{[l+1]}\right] \odot g'^{[l]}\left(\mathbf{z}^{[l]}\right) \tag{2.34}$$

the gradients for the weights **W** and offsets **b** are then given by:

$$\frac{\partial L}{\partial \mathbf{W}^{[l]}} = \boldsymbol{\delta}^{[l]}\left(\phi^{[l-1]}\right)^T \tag{2.35}$$

$$\frac{\partial L}{\partial \mathbf{b}^{[l]}} = \boldsymbol{\delta}^{[l]} \tag{2.36}$$

During each training step, the network parameters are updated using gradient descent according

to the following scheme:

$$\mathbf{W}^{[l]} \longleftarrow \mathbf{W}^{[l]} - \eta \cdot \frac{\partial L}{\partial \mathbf{W}^{[l]}} \tag{2.37}$$

$$\mathbf{b}^{[l]} \longleftarrow \mathbf{b}^{[l]} - \eta \cdot \frac{\partial L}{\partial \mathbf{b}^{[l]}} \tag{2.38}$$

where η denotes the learning rate. Since the weights and offsets are iteratively updated to minimize the training loss, the model's predictions will slowly approach the desired reference values Y. It is common to train a neural network in batches. Weight updates are only performed based on the average loss of the batch, which mitigates over-fitting. The training data can be passed multiple (but not arbitrarily many) times through a neural network. Each pass of the entire training data is referred to as an epoch.

Chapter 3

Filaments of Asian monsoon air

The extra-tropical Upper Troposphere – Lower Stratosphere (UTLS) is the first compartment of the atmosphere where mixing through filamentation is prevalent. In the UTLS, wave breaking can induce direct, adiabatic exchange between tropospheric and stratospheric air, which is essential for the global distribution of trace gases, the radiative budget and for the shape of the global climate (Gettelman et al., 2011; Manney et al., 2017; Seguel et al., 2025; Riese et al., 2025). However, very few explicit observations of these mixing processes exist.

In this chapter, observational data from the 'Probing High Latitude Export of air from the Asian Summer Monsoon' (PHILEAS) campaign conducted in 2023 are used to study the mixing processes of a filament of Asian Summer Monsoon (ASM) air near a tropopause fold inside the UTLS. These filaments are well suited to studying mixing processes, because they have a distinct chemical composition which contrasts with the background, and are located at altitudes where airborne measurements can easily be performed. Further, they contain many different trace gas species in sufficiently high mixing ratios such that even complex mixing processes can be resolved. The following analysis consists of two major parts: in the first part, 3-D observations of an ASM filament are analysed in a double flight experiment. The dynamically disturbed situation near a tropopause fold necessitates a new classification approach, which will be used to identify mixing processes and reveal the roles of the different air types in these mixing processes. The second part transfers the classification concept to other data products and explores its capabilities to reveal the structure of these data as well. For these data, a physically similar situation is observed through a different, larger set of trace gas species Finally, the approach is generalized using more complex machine learning methods for air type classification.

The analysis aims at answering the following questions:

- Which chemically distinct air masses are present in the double-flight observations, and how can their composition and likely origin be characterized?
- How does the chemical composition of trajectory-matched air parcels change between the two measurements, and what do these changes reveal about the mixing processes and air masses involved?
- How robust and transferable is the proposed data-driven classification across different data products, tracer selections, and classification architectures?

3.1 Data products and the PHILEAS campaign

The Probing High Latitude Export of Air from the Asian summer monsoon (PHILEAS, Riese et al. (2025)) aircraft measurement campaign was conducted between June and September 2023. It was a joint project headed by the Forschungszentrum Jülich and the University of Mainz with a broad scientific payload onboard the High Altitude and Long Range research aircraft (HALO). A major scientific interest of the PHILEAS campaign was to characterize the transport pathways (both advection and mixing) and the timescales of polluted and moist air from the Asian summer monsoon (ASM) into the extratropical UTLS. The Asian summer monsoon (ASM) is one of the most important circulation patterns of the northern hemisphere. Important source regions of anthropogenic pollutants and greenhouse gases are located within the ASM area, which directly affect a large portion of the global population. From early June until late September the large-scale, upper-tropospheric ASM anticyclone is formed near the Tibetan Plateau (Krishnamurti and Bhalme, 1976) due to strong heating and convection during summer (Li et al., 2005). The ASM anticyclone extends to the lower stratosphere and is confined to the south by a tropical easterly jet and to the north by a subtropical westerly jet (Randel and Park, 2006). The air inside the anticyclone is characterised by lower Potential Vorticity (PV) values (see Randel and Park, 2006) compared to its surrounding air masses, enhanced mixing ratios of tropospheric trace gases such as water vapour (e.g. Rosenlof et al., 1997), methane (e.g. Park et al., 2004) and anthropogenic pollutants (e.g. Park et al., 2007; Baker et al., 2011), as well as low mixing ratios of trace gases of stratospheric origin (Randel and Park, 2006; Konopka et al., 2010). These characteristic tracer mixing ratios are sharply confined inside the ASM anticyclone.

Towards the end of the boreal summer the anticyclone weakens, which leads to increased out-

flow of ASM air into the wider atmosphere. Trajectory calculations have shown that air masses originating from regions bordering the ASM anticyclone, e.g. the tropical–western Pacific region (Bian et al., 2012), also contribute significantly to the composition of this outflow (Chen et al., 2012; Bergman et al., 2013; Vogel et al., 2016). The overall transport pathways and their significance are subject to a long-standing debate (Bourassa et al., 2012; Ploeger et al., 2017; Yan et al., 2019; Bian et al., 2020; Zhang et al., 2021). Rossby wave breaking leads to so-called eddy shedding (Hsu and Plumb, 2001; Popovic and Plumb, 2001; Garny and Randel, 2013), through which filaments of ASM air are formed with complex spatial structures of trace gases and often sharp gradients (e.g. Polvani and Plumb, 1992; Vogel et al., 2011; Bacmeister et al., 1999). These filaments are transported eastwards and westwards, where they enter the northern mid-latitudes. Here, they become an effective pathway for mixing tropospheric air into the lower stratosphere on timescales of a few days (Ungermann et al., 2016).

This injected ASM air likely has profound effects on the global climate. For example, Ploeger et al. (2013) found that the injection of water vapour originating from the ASM is a key factor in moistening in the lowermost stratosphere. Tropospheric pollutants such as methane further amplify this effect by producing stratospheric water vapour via oxidation (Riese et al., 2006; Rohs et al., 2006). Even small changes in this pathway affect the upper tropospheric and lower stratospheric water vapour budget (Solomon et al., 2010; Riese et al., 2012) and subsequently the radiative balance of the UTLS. The eastward-travelling ASM filaments are of particular interest because they traverse the relatively clean Pacific area, which provides an excellent background for detecting their chemical signature (Vogel et al., 2016; Pan et al., 2022; Bak et al., 2025). Consequently, phase II of the PHILEAS campaign was based in Anchorage, Alaska, from where all scientific flights in this phase were performed. In this chapter, the data products of the Gimballed Limb Observer for Radiance Imaging of the Atmosphere (GLORIA) (Riese et al., 2014; Friedl-Vallon et al., 2014) and supporting simulations by the Chemical Lagrangian Model of the Stratosphere (CLaMS; e.g. McKenna et al., 2002b,a; Konopka et al., 2010; Pommrich et al., 2014) are used.

These simulations include surface-origin tracers, which are artificial tracers that indicate the contribution of different surface source regions to an air parcel (Vogel et al., 2026). Surface-origin tracers are initialized every day in the model boundary layer (≈ 2–3 km above Earth's surface, considering orography) and are subsequently transported into the free atmosphere over the course of the simulation. The percentage of a surface-origin tracer in an arbitrary air parcel indicates the extent to which the considered air parcel originates from the respective region. These surface–origin regions are shown in Fig. 3.1: The following surface-origin tracers in the ASM region (Fig. 3.1)

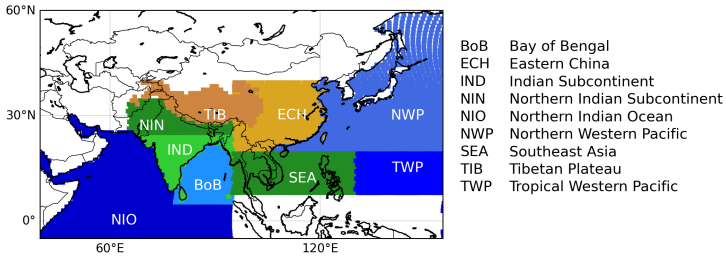

Figure 3.1: Map of surface-origin tracers used in the global 3-D CLaMS simulation for 2023.

are considered in this chapter: Northern INdian subcontinent (NIN), Indian Subcontinent (IND), Tibetan Plateau (TIB), Eastern China (ECH), South East Asia (SEA), North-West Pacific (NWP), and Tropical Western Pacific (TWP). Furthermore, the sum of the following surface-origin tracers, referred to as the "South Asia" tracer, is used as a proxy for air contributing to the ASM: "ECH" + "TIB" + "NIN" + "IND" + "Bay of Bengal (BoB)" + "Northern Indian Ocean (NIO)".

The GLORIA instrument (Riese et al., 2014; Friedl-Vallon et al., 2014) represents the first airborne realisation of the infrared limb imaging technique. It consists of a Michelson interferometer, which projects the observed radiation onto a 256×256 pixel infrared detector array of which 128 vertical by 48 horizontal pixels are used. The interferometer is calibrated by in-flight measurements (Ungermann et al., 2022). The instrument measures thermal radiation emitted by atmospheric constituents in limb viewing geometry. The main contribution of measured radiation stems from the lowest altitude points (tangent points) along the line of sight. The range of vertically resolved altitudes is limited upwards by the flight altitude and downwards by the water vapour continuum, saturated spectral signatures of further species, clouds and aerosols. The field of view covers 4° vertically and 1.5° horizontally. For the measurements discussed here, a spectral sampling of 0.2 cm^{-1} is used. The spectral range from 770 to 1440 cm^{-1} is covered. GLORIA generally looks to the right with respect to the flight direction. The interferometer is

mounted in a gimbal frame to stabilize the line of sight against aircraft movements. Further, the instrument can be panned to allow viewing directions from 45° to 132° horizontally with respect to the airplane heading. The emission spectra derived from these measurements are characteristic of the different chemical species composing the observed air masses, since individual species have characteristic absorption lines in these spectra proportional to the mixing ratio of the species. By solving an inverse radiative-transfer model, the chemical composition of the observed air masses can be derived (Ungermann et al., 2011, 2015). These retrievals yield 2-D observations, which are referred to in the following as *curtains*. If the aircraft is flying along a closed path, ideally a hexagon with an approximate diameter of 400 km (Ungermann et al., 2010a) around an air mass while the instrument is panned, measurements are obtained which allow us to perform a 3-D tomographic retrieval of the enclosed air mass (Ungermann et al., 2010b). Such 3-D tomographic retrievals are technically more demanding, such that fewer trace gas species can be retrieved. These 3-D observations are key for the first part of this chapter.

A large filament of ASM air was predicted to be sufficiently close to the coast of Alaska to be reachable by plane on 10 and 11 September 2023 (UTC). On these days a double-flight experiment (flights PH13 and PH14) was conducted measuring air near the edge of the filament (Fig. 3.2a) and the outflow of that previously measured air (Fig. 3.2b). Forecasts by the high–resolution ICON model predicted increased mixing ratios of several anthropogenic pollutants (not shown) such as peracetyl nitrate PAN inside this filament, as well as a tropopause fold near the outer edge of the filament on 10 September 2023.
In order to perform a 3-D tomographic retrieval of the filamented air, a hexagonal flight path segment (referred to in the following as hexagon 1, indicated by the red path in Fig. 3.2a) was included near the end of PH13 on the edge of the ASM filament. Forward trajectories for those air parcels inside hexagon 1 were calculated using a fourth–order Runge–Kutta scheme (see App. A.1).

A second hexagonal flight path segment (referred to in the following as hexagon 2, indicated by the green path in Fig. 3.2b) was included in the follow-up flight PH14 such that parts of the outflow of the air parcels in hexagon 1 were revisited (see Fig. 3.2c). Inside hexagon 2 a cloud cover at around 10 km altitude partially limited the measurements of GLORIA towards lower altitudes, resulting in a section of the second hexagon volume not being observed. In the following, air parcels which can be connected through trajectories between the two observed volumes are referred to as 'matched'. Due to the high spatial resolution of GLORIA, it is possible to study the changes and mixing that each matched air parcel undergoes between the two measurement flights.

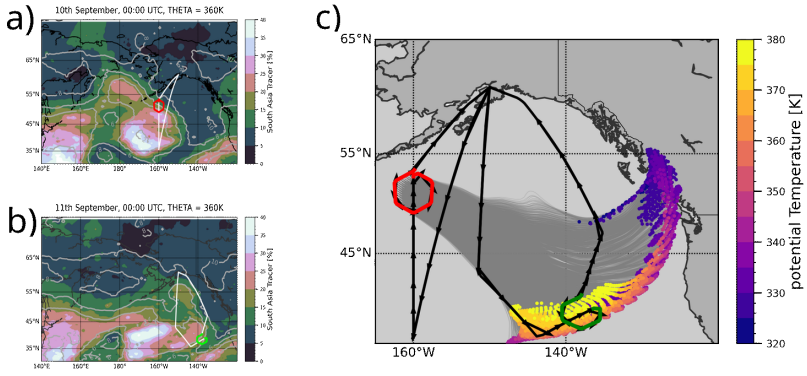

Figure 3.2: Summary of the double flight experiment. Panels a, b: CLaMS South Asia tracer (= 'ECH' + 'TIB' + 'NIN' + 'IND' + 'BoB' + 'NIO', see Fig.3.1 at 360 K potential temperature. PV values are shown as gray isolines. Panel a: Initial flight on 10 September 2023 UTC. A large filament with increased mixing ratios of South Asian air is projected to be in range of HALO research flights from the base in Anchorage. The first hexagon is located towards the end of the flight near the edge of the filament and is marked in red. Panel b: Follow-up flight on 11 September 2023 UTC. The filament 'rolls over' southeastwards. The second hexagon (marked in green) is located inside the outflow from the first hexagon at that time. Panel c: Forward trajectories of hexagon 1. The trajectories are shown as gray lines, beginning inside hexagon 1 at 01:35 UTC and ending at 01:50 UTC on the next day. The trajectories were started at every point on the retrieval grid and were driven by ERA5 (1 °x 1 °) reanalysis. Potential temperature of end points is shown in colour-code.

The first part of this chapter focuses on these 3-D tomographic retrievals and the matches between them. For both flights 2-D retrievals exist at all times, which are the focus of the last two parts of the chapter.

3.2 Classification concept

Traditionally, air types in observations are distinguished via trace gas thresholds, i.e. a water vapour threshold to define tropospheric air Weigel et al. (e.g. 2012), tracer-tracer correlations (Plumb, 2007; O'Shea et al., 2013) or by conceptual arguments such as the tropopause (Kunz et al., 2011). In any case, prior knowledge of the atmosphere is required. For dynamical and potentially perturbed situations such as tropopause folds these concepts begin to break down. Strong mixing may shift thresholds and affect otherwise clear correlations due to changes in tracer gradients. Vertical distortions can lead to (locally) unexpected tropopauses. In particular, the chemical composition of the ASM outflow in this exact state of dissolution has not yet been measured – a key motivation for the PHILEAS campaign. We must consider that the composition of this outflow may be complex due to many different source regions potentially contributing to it. We therefore approach the classification problem from a *model-agnostic* point of view based on the considerations in Sect. 2.5.1.

We can differentiate between different types of air if they have properties by which they differ from each other. For example, each distinct air type involved is likely to have a unique chemical composition. Otherwise, we would not consider them to be distinct. Observations of the same air type therefore are similar in phase space, and lead to the formation of clusters. By identifying the clusters contained within our data and analysing their chemical composition we can thus obtain information about the observed air types. The number of air types then simply follows by counting these clusters, for example using the silhouette score (Eq. 2.23). It is not required to know the nature of these air types or the number of air types beforehand – it is only important that they appear chemically distinct. The only criterion is that sufficiently many observations are available for well–distinguishable clusters to be formed. Clusters are generally easier to separate in higher dimensions, such that as many non-collinear trace gases as possible should be considered. The model-agnostic classification used in the following is illustrated in Fig. 3.3: During pre-processing, each dataset (measurements or model data) is scaled to zero mean and unit variance individually following:

$$\hat{x_i} := (x_i - \mu_i)/\sigma_i \tag{3.1}$$

where μ_i, σ_i are the mean and the standard deviation of that tracer, respectively . This ensures that each constituent has comparable numerical magnitude while preserving correlations. Therefore, different datasets can be compared regardless of their scales and potential offsets up to

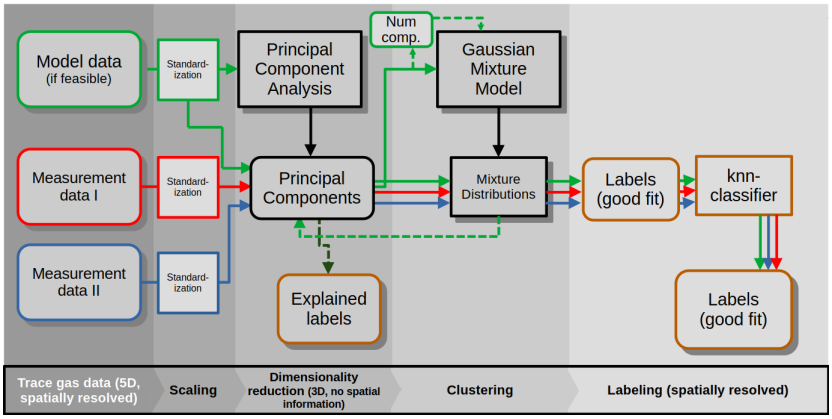

Figure 3.3: Schematic outline of the data pipeline. The workflow goes from left to right. Model data and their data products are indicated in green. Measurement data and their data products are indicated in blue and red depending on which measurement. Transformations and classifiers that are not explicitly dependent on a dataset are marked in black. Transformations and classifiers that explicitly depend on a specific dataset are outlined in that dataset's colour. Transformations and classifiers which are applied to multiple datasets individually are outlined in brown.

linear effects due to e.g. either climatological changes or instrument biases.

The next step is to perform a dimensionality reduction. Manifold structure may allow us to express the data in a lower–dimensional space, greatly improving numerical stability. For this classification we use Principal Component Analysis (PCA), see Sect. 2.5.2. The cut-off dimension is determined as described in that section, yielding 3-D manifolds for the 3-D tomographic retrievals and 7-D manifolds for the curtains. A discussion of the approximate equivalence of clustering in such a PCA manifold can be found in App. A.11.

The clusters in the manifolds (still representing different types of air) generally correspond to clusters in the original data and can be handled as being sampled from Gaussian distributions, which can be inferred by fitting a GMM, see App. A.9. Its basic assumption about the data is that the data can be accurately represented as random samples from different Gaussian distributions. The GMM determines means, standard deviations and weights (indicating the frequency by which a particular distribution is sampled from) for each cluster. The number of clusters in the manifold, which is the major input parameter of a GMM, can be estimated using the Silhouette score (Eq. 2.23) and is then chosen ad hoc. Any other parameter is determined during fitting.

It provides *probabilistic* cluster assignments, i.e. determines the probabilities that a point belongs to any of the assumed distributions, see Eq. A.28. Since air types naturally have variability in their composition – both due to atmospheric variability and measurement noise – they are well

represented by such a distribution. Moreover, mixing of air types naturally leads to the creation of intermediate, mixed states whose composition lies somewhere between the mixing partners. For such mixed states probabilistic labels are the most useful. A hard label can be applied by assigning the most likely distribution, see Eq. A.29. Such labelling is particularly useful if the clusters lie well separated such that the probabilities for all but a single cluster vanish. Labelling artefacts as discussed in App. A.6 must also be addressed accordingly to ensure coherent labels. These labels (and their probabilities) are assigned point-wise for every measurement point.

By applying the inverse PCA transformation to the mean and the covariance matrix of each derived distribution, the average composition (*explained labels* in Fig. 3.3) and the correlation between the trace gas species in the associated air type can be derived. These allow us to interpret the clusters in the abstract manifold in physically meaningful terms.

3.3 Analysis of mixing in 3-D tomographic retrievals

Before we dive into the analysis of the 3-D tomographic retrievals we should try to visualize these data products. Fig. 3.4 shows the spatial structure of the 3-D tomographic retrieval (as described in: Ungermann et al., 2010a, 2011; Krasauskas et al., 2019) of the first hexagon: The foundation

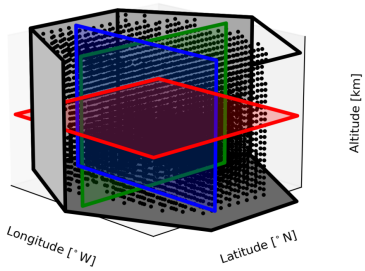

Figure 3.4: Geometry of a 3-D tomographic retrieval. All data points are contained within a hexagonal prism, whose foundation is given by the flight path. These data points consist of those points within the retrieval volume with high measurement influence (black dots). For further analysis, the cross-sections indicated in red, blue and green are presented to represent the total volume.

of the hexagonal prism is given by the flight path. The prism is around 400 km in diameter and spans around 6 km vertically and is visually magnified in Fig. 3.4. The GLORIA instrument measures into the hexagon from each flight leg under alternating angles as described above. Based

on these measurements the tomographic retrieval solves the inverse problem of finding a (smooth) atmospheric state that best fits the observed spectra. The points with strong measurement sensitivity are shown as black dots. At each such black dot a trace gas mixing ratio for each retrieved trace gas species is determined. These points make up the retrieved data as 3-D resolved tracer fields.

The trace gas species retrieved for these 3-D tomographic retrievals are water vapour (H_2O), the pollution tracer peroxyacetyl nitrate (PAN), the long-lived greenhouse gas dichlorodifluoromethane ($CFC-12$), ozone (O_3) and nitric acid (HNO_3). The a priori values for the retrieval are taken from different sources, such as the climate models CLaMS, Copernicus Atmosphere Monitoring Service (CAMS) or products offered by the European Centre for Medium-Range Weather Forecasts (ECMWF). Since 2-D visualizations convey information more concisely and reduce the visual clutter of full 3-D visualizations, we will represent the tracer fields by several interpolated cross-sections, which are shown in Fig. 3.4 for constant altitude in red, for constant longitude in green and for constant latitude in blue rectangles.

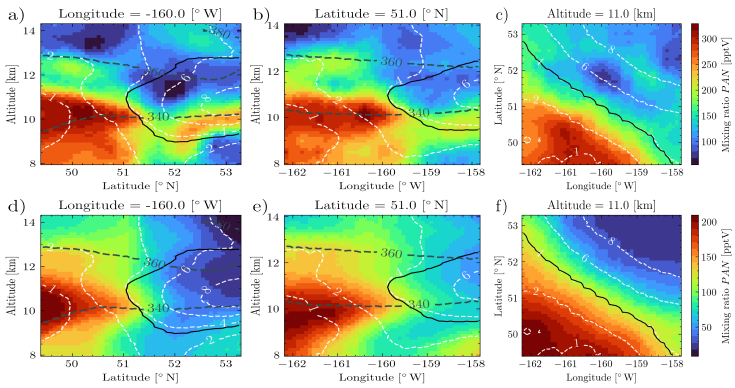

Figure 3.5: PAN *mixing ratios inside the first hexagon on 10 September 2023 (upper row: GLORIA retrievals; lower row: CAMS model values). Shown are the cross-sections indicated in Fig. 3.4. PV isolines are plotted as white dashed lines, potential temperature isolines are given as gray dashed lines. The dynamical tropopause adapted from Kunz et al. (2015) is indicated as a black line in all panels.*

One such tracer field is shown in Fig. 3.5. The first row shows the retrieved mixing ratios of PAN interpolated from the aforementioned cross-sections; the second row shows interpolations for the same cross-sections but for the CAMS model instead, which has a 1° by 1° horizontal resolution and 250 m vertical resolution in the relevant region.

For all panels the dynamical tropopause is calculated for September according to the climatological values in Kunz et al. (2015) on the relevant isentropes – based on PV values from ECMWF

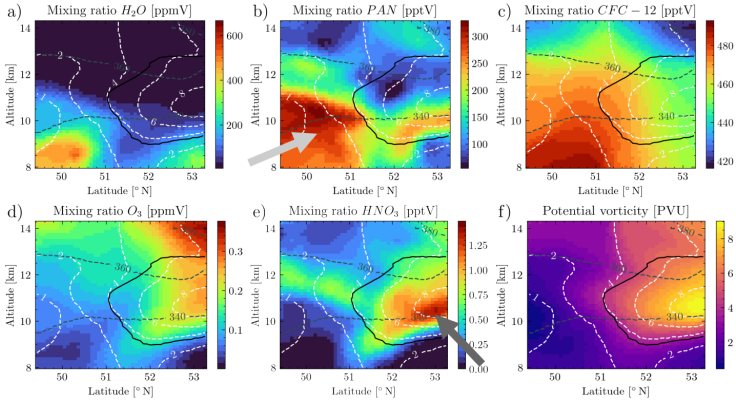

Figure 3.6: ***(a-e)****: Retrieved trace gas species inside the first hexagon on 10 September 2023 UTC. Extraordinary regions, such as the maximum signature in PAN and HNO_3, which contain apparent inconsistencies, are marked by the bright and dark gray arrows, respectively.* ***f)****: Potential vorticity derived from ERA5. Shown is the green cross-section in Fig. 3.4. The dynamical tropopause and PV-isolines are indicated as black and white lines, respectively. The potential temperature is indicated by the gray dashed lines in each panel.*

ERA5 reanalysis data – and is shown as a black line. Additionally, PV isolines are shown as white lines; isentropes are shown as gray lines.

Increased mixing ratios of PAN are indicators for the outflow of the Asian summer monsoon. The filament is located on the low–latitude / western side of the hexagon at altitudes between 13 km down to at least 8 km (Fig. 3.6 b, bright arrow) in accordance with the flight planning predictions. It contrasts with the surrounding air through steep gradients. The general shape and size of the filament are in good agreement with the model predictions, although the predicted mixing ratios are up to 50% lower than the observations. Smaller–scale features are not adequately resolved by the model, and therefore small–scale mixing is unlikely to be correctly resolved as well. In particular, the measurements show much stronger gradients than e.g. the CAMS simulation. This indicates that model values may not be suitable to use for the classification, which is why the PCA is determined based on the retrieval data instead.

The low values of PAN on the high latitude / north-eastern side of the hexagon suggest stratospheric air, which extends downward to around 11 km. This stratospheric air does not extend as low as the CAMS model had predicted. The PV values mostly reflect this structure; a 3-PVU isoline would encapsulate most of the distinct features well. There is a distinct feature of increased PAN on the north-eastern side of the hexagon at around 10 km, which lies inside the stratospheric air mass/intrusion according to the dynamical PV-tropopause (panel e, dark arrow). The other observed trace gases show similar patterns, but with several apparently conflicting features,

as shown in Fig. 3.6 as an example in the blue cross-section (Fig. 3.5 a/d). The filament is visible in the retrieved $\mathrm{H_2O}$, but only extends upward to around 11 km, which is around 1 km lower than PAN would suggest. This can be explained by the freezing out of water near the tropopause. $\mathrm{O_3}$ and $\mathrm{HNO_3}$ reflect the filament as well, with low values localized around the same region, but the low O_3 values would suggest an extent to much higher altitudes. $\mathrm{O_3}$ in particular shows the stratospheric intrusion at the tropopause fold, where stratospheric air extends downwards to 10 km altitude. $\mathrm{HNO_3}$ shows a layer of increased mixing ratios in the upper part of the filament, but reveals low values again above that layer, indicating a more tropospheric character of these air masses. The feature of increased PAN to the north-east is completely missing in $\mathrm{H_2O}$, while both stratospheric tracers show increased values in this region as well. This already indicates mixing of different air masses, e.g. of aged tropospheric air and fresh stratospheric air, as will be elaborated on in the next section. $\mathrm{CFC-12}$, on the other hand, shows very little structure at all (cf. the narrow range of the colour scale), except for lower values around this feature.

The observations therefore indicate that there is active mixing between the UT and LS occurring inside the intrusion. However, the dynamical situation is not adequately characterised by either the dynamical tropopause (shown in Fig. 3.6), or the thermal tropopause (not shown here) nor any single or pair of tropospheric and stratospheric tracers. Since all tracers reflect the physical reality, there must exist a classification which resolves the apparent contradictions. It is therefore useful to include as many trace gas species as possible in the classification to maximize the information yield.

3.3.1 Classification

The classification analysis described in Sect. 3.2 is applied to both 3-D tomographic retrievals. We choose to determine the manifold based on the retrieval data due to the aforementioned discrepancies between model values and observations. It should be noted that spatial information such as longitude, latitude or potential temperature were not considered in the classification. We count five distinct clusters using the Silhouette score (Eq. 2.23). The GMM is thus trained assuming five mixture components. Labels are calculated as the most likely associated cluster following Eq. A.29 while applying label corrections as described in App. A.6. Analysis of the mixing processes, which depend on well-defined pre-mixing states, requires that both 3-D retrievals are transformed and classified with the same parameters. From hexagon 1 it is derived that a 3-D manifold consisting of the first three principal components of that retrieval is adequate. The

manifolds for both hexagons and their classifications are shown in Fig. 3.7. The full manifolds

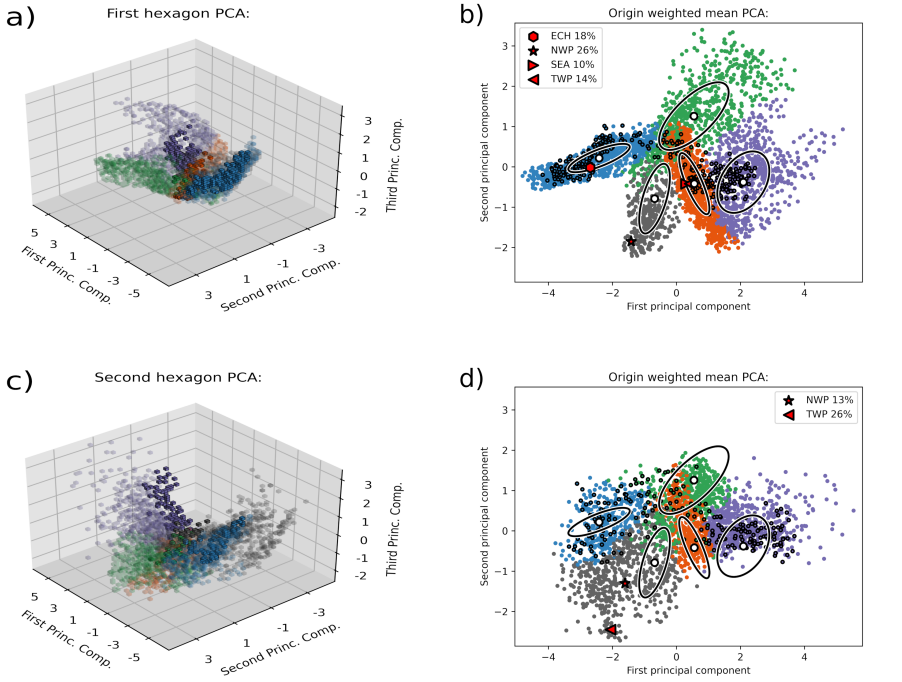

Figure 3.7: GMM classification based on the first three principal components. ***(a, c)****: Full manifold of hexagon 1 and hexagon 2 consisting of all three principal components.* ***(b, d)****: First two principal components of the manifold of hexagon 1 and 2, respectively, for better visualization. All panels show the labels in colour-code. For each Gaussian distribution a 1-σ environment is indicated by ellipses. The PCA of the mean of each measured trace gas weighted by the relevant surface-origin tracers of the CLaMS simulations is shown in red markers. The red markers for 'SEA' and 'TWP' in panel b) lie close to each other and basically overlap (both in 3-D and 2-D). The percentages shown are the maximum percentage of each respective surface-origin tracer. In all panels, the classes found by the GMM are shown in colour. Air parcels that were observed in both hexagons are indicated by a black outline. For each class the cluster centre is indicated by white dots.*

are shown in Fig. 3.7a, c for hexagon 1 and 2, respectively. Fig. 3.7b, d shows only the first two principal components for visualization purposes. The labels are shown in colour-code. Those air parcels, whose trajectories end inside the second hexagon at the time of measurement (see Fig. 3.2c), are marked by their black outline.

The choice of five classes is well reflected in hexagon 1 with five distinct clusters visible in Fig. 3.7b. The classes lie well separated from each other and the labelling is unambiguous for the majority of points, with approximately 1% of points requiring label adjustment. The shape of the manifold in Fig. 3.7a already reveals much information about the physical situation. Clus-

ters in the manifold generally remain clusters in the full space, since the PCA retains distances. Each principal component is spanned by a linear combination of the original trace gases. Thus, clusters positioned along different principal components generally have distinct chemical compositions. Similarly, clusters whose covariances are similar are likely closer in chemical composition, especially if both clusters lie inside the same plane, i.e. share a (sub)manifold, where they may differ in magnitude, but less in composition.

Given that hexagon 2 was measured in a dynamically similar, but temporally later situation, its manifold deviates from hexagon 1 in several ways. The blue and gray clusters are much less distinct from each other. One could argue for merging them into one cluster; for consistency, however, we do so only visually. The green and orange clusters also converge in hexagon 2, implying that mixing resulting in similar air masses is occurring. The effect is most visible for those points having positive values for the principal components. Unfortunately, no air parcels lying in this region were revisited and consequently these changes were not observed directly. It should be noted that while the orange cluster in hexagon 1 shares a large border with the violet cluster and might be considered an extension of the latter, the local correlations in both clusters are quite different, meaning key chemical components may be similar, but show distinct differences in their proportions.

For each class, the cluster centres (shown as white dots in Fig. 3.7b, d) represent the mean values of points in each class and can be considered representative for each class, as will be explored further in the following. The GMM not only provides the position of each cluster centre, but also provides a measure of the shape of each distribution via the covariance matrix of the data in each cluster. Similar to these representative points for each class, representative points for the contribution of each Asian monsoon source region can be constructed by taking the weighted mean of the observations weighted by the surface-origin tracers. These values are then transformed and shown within the PCA in Fig. 3.7b and 3.7d as red markers. The positions of these points within the manifolds indicate the source region at the Earth's surface which has the largest impact on the respective classes.

The explained labels, i.e. the inverse transformed cluster centres, provide a physical interpretation for the currently purely abstract clusters within the manifold. The intrinsic dimensionality of each cluster, here determined by the Fisher separability (see Sect. 2.5.2), expresses the chemical complexity of each air type. The resulting characteristic chemical composition of each class is shown in Fig. 3.8a in percentiles of the total distribution.

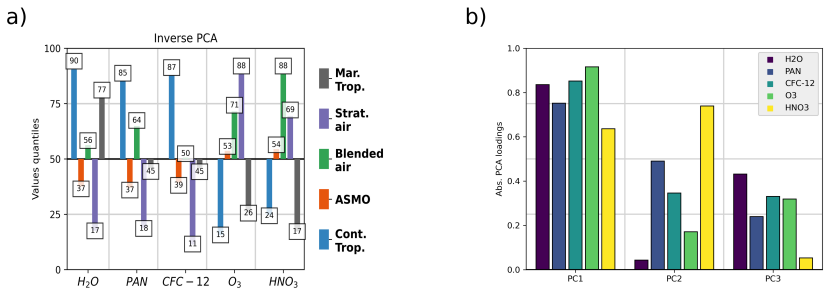

Figure 3.8: Panel a): Chemical composition of each cluster centre as obtained by the inverse PCA transformation. The chemical composition of each cluster is shown as percentiles of the total distribution, i.e. a value of 20 indicates that 20% of all observed values are lower than this value. Panel b): Loadings of each principal component. The length of each bar indicates the importance of that tracer to that principal component.

Combining the chemical fingerprint of each cluster and the information derived from the surface-origin tracers, the following characterisation of each air type can be derived:

- Continental, tropospheric class: the blue class shows by far the highest mixing ratios for tropospheric tracers, especially the pollution tracer PAN, and the corresponding lowest values of both stratospheric tracers. Since this air is tropospheric in nature, high water vapour and high pollutants suggest an origin within the continental regions of the ASM. The surface-origin tracers link this type of air to the East China region (see Fig. 3.7b). Its intrinsic dimensionality is around 1.69, indicating high similarity structure and little variability. In the following, this air type is referred to as continental, tropospheric air (CT).

- Maritime, tropospheric class: the gray class, as expected, shows a similar chemical profile to the CT air, with the difference that the pollution tracers are much lower. This air is tropospheric in nature and likely stems from less populated or less industrialized regions within the ASM. Indeed, the surface-origin tracers link this air type to the Western Pacific region. Its intrinsic dimensionality of 2.75 is higher than that of its continental counterpart, which indicates significant influence due to mixing. In the following, this air type is referred to as maritime tropospheric air (MT).

- Stratospheric class: the violet class has strong stratospheric character indicated by the highest ozone values and the lowest mixing ratios in every tropospheric tracer. The fact that no surface-origin tracer is linked to this air type strongly suggests that this air type is stratospheric air and will be referred to as such (S). Its intrinsic dimensionality of 2.78

is rather high, indicating no purely stratospheric character, but rather some influence of tropospheric air.

- Asian summer monsoon outflow: the orange class shows a unique constellation of moderately low values across both tropospheric and stratospheric tracers, with none of them being exceptionally low. According to the surface-origin tracers there are two regions linked to it, namely South East Asia (SEA) and the Tropical Western Pacific (TWP). Recalling the information we summarized about the ASM, air from the TWP is largely collected at the outer edge of the ASM anticyclone, where it could have undergone mixing with SEA air. Work by Vogel et al. (2026) shows this behaviour inside the ASM, but analysis from the chemical composition alone is vague at this point. Further arguments should and will be made in the following. Its intrinsic dimensionality of 1.90 indicates average complexity. For now this air is referred to as Asian summer monsoon outflow (ASMO or simply A).
- Blended class: the green class shows another unique composition of moderately increased tropospheric tracers (except for $\mathrm{CFC}{-}12$) and increased stratospheric tracers. Its composition can be understood as a mixture of tropospheric and stratospheric air. This air type is a strong candidate for identifying mixing processes. Such mixed air may not necessarily be unique, as different mixtures of different sources may yield a similar composition. We will refer to this air type as blended air (B) instead of mixed air to distinguish between it and the other mixing processes. Its low intrinsic dimensionality of 1.80 would support this characterisation, as mixing generally occurs along a manifold (i.e. a mixing line).

A PCA also allows one to study how much each chemical tracer contributes to each of the various principal components. The so-called *loadings* L, shown in Fig. 3.8b, are the correlation coefficients between each principal component and the respective tracer. A variable contributes much to the structure of the manifold if the absolute value of its loading is large ($|L| > 0.8$) and contributes very little if it is close to zero, i.e. if removing that variable would not change the PCA much. For this classification we find major contributions for every trace gas species to the first (and most important) principal component. This indicates that information of every trace gas species meaningfully contributes to that component. The second principal component depends most strongly on $\mathrm{HNO_3}$ and PAN, which are the two tracers that showed inconsistencies in Fig. 3.6 via isolated regions of increased mixing ratios. These inconsistencies are not reflected in any other tracer. They contribute most to the mesoscale structure of the data. Finally, the third component only has smaller contributions from water vapour, $\mathrm{CFC}{-}12$ and ozone, which are the tracers that

separate the CT/MT classes from B/S classes.

The classification was derived from the chemical composition only and did not have direct information on the three-dimensional spatial distribution of the different trace gases / classes within the hexagon. If the classification and its interpretation are meaningful, the spatial distribution of the classes is expected to be consistent with this interpretation.

The spatial structure of the classification is shown in Fig. 3.9:

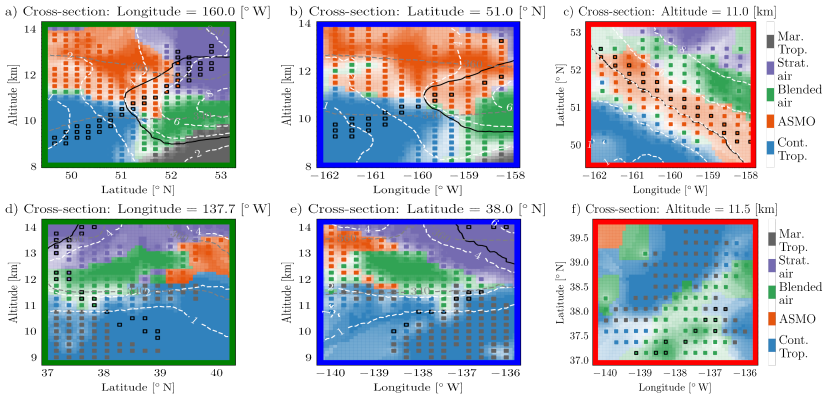

*Figure 3.9: Class distribution inside **(a, b, c)** hexagon 1 and inside **(d, e, f)** hexagon 2. Cross-sections are identical to Fig. 3.5a-c. The dynamical tropopause is indicated as black line. PV isolines are shown in white. Measurement points within each cross-section are shown as solid markers. Matched points are indicated by black borders. Class assignment is indicated in colour-code, with high class alignment shown in more opaque colours. In hexagon 2, the background of both tropospheric classes is displayed in blue as if they were a single class. For clarity gray labelled points are still shown in gray. The background shows the classification applied to the trace gases interpolated in each cross-section. Spaces containing few measurement points are more uncertain.*

Panels a) to c) show the class distributions of hexagon 1 along the same cross-sections used in Fig. 3.5, panels d) to f) show the corresponding cross-sections in hexagon 2. The background in between measurements in these cross-sections is interpolated from the retrieved values and positions. These interpolated values are then transformed and classified as described before. A high probability for the class assignment (Eq. A.28) is shown in bright colours; lower probabilities (corresponding to labelling ambiguity) are shown in pastels or white. Regions which contain few measurements are most influenced by interpolation, but generally reflect the class structure well due to the compactness of the manifold. Air parcels revisited between the two flights are indicated by black outlines. Air types that were identified by their compact clusters in phase space, similarly form compact, coherent sets in real space, despite no spatial information being used.

The tropospheric classes CT and MT are at the lowest altitudes and are mostly located below the

2-PVU isoline, which is commonly used as an upper limit for the troposphere. This behaviour is most consistent in hexagon 2, where both tropospheric classes lie intertwined, suggesting similar to Fig. 3.7d that mixing of maritime and continental tropospheric air occurs. In the following, both tropospheric classes are considered functionally the same (and referred to simply as the tropospheric class (T)) and are indicated in blue colour. Conversely, the stratospheric class is located in the highest altitudes. Due to the tropopause fold in hexagon 1, the vertical structure is distorted, which is correctly captured in the classification, in which the CT class and the S class lie diametrically opposed. When we compare the location of the ASMO class to Fig. 3.2, we see that the ASMO class coincides with the location of the filament in both hexagons, further justifying the identification of the class with ASMO air. In hexagon 1, an intrusion of ASMO air into the stratosphere is seen (panel b). This suggests in-mixing of ASMO air directly into the lower stratosphere. The blended class is located bordering the ASM air filament between the troposphere and the stratosphere, mostly at the tip of the stratospheric intrusion. Visually it appears as if the blended class acts as a partially dissolved extension of the intrusion. This characterisation would be in line with its chemical composition. As stated beforehand, this classification is not necessarily unique, and the blended class could theoretically also be the result of mixing of ASMO air with stratospheric air if the chemical balance holds. Neither point can be substantiated further from this analysis alone and will be revisited later.

Active mixing between classes is directly indicated by reduced probabilities of class assignment, as the transition from one class into another crosses the regions of low class assignment probability. Considering these probabilities as shown in pastel and dark colours in Fig. 3.9, mixing is occurring near the edges of most classes in either hexagon. Edges with still high probabilities imply only recent contact between those air masses, such that mixing has not yet sufficiently occurred. The strongest ongoing mixing is occurring in hexagon 1 between the CT class and the A/B class. In hexagon 2, however, the strongest ongoing mixing is between the blended class and its bordering classes. For those air parcels which were observed in both hexagons the mixing processes can be derived explicitly by comparing their chemical composition in both hexagons.

3.3.2 Explicit mixing

The matching experiment was designed to observe the same air parcels on two successive days (self-match flight). In total, 396 such matched air parcels were observed based on their trajectories, corresponding to around 10% of all observed air parcels.

Changes in the chemical composition of these matched air parcels give insight into the mixing processes, as little to no change is expected over the course of 24 h for any of the observed trace gases. Any such observed changes in composition are therefore most likely due to mixing. The GMM classification directly implies different types of mixing processes via the changes in class assignments between hexagon 1 and 2. Such transitions between classes are summarized in the confusion matrix shown in Fig. 3.10: A confusion matrix summarizes each possible combination

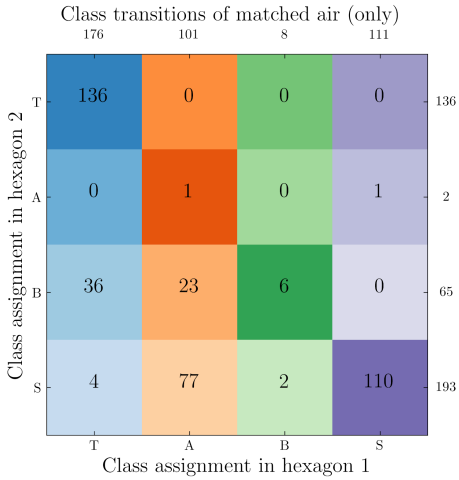

Figure 3.10: Confusion matrix of class assignment of matched air parcels. The number of each transition is shown in each matrix element. The columns indicate the class assignment of all air parcels in hexagon 1; the rows indicate the class assignment of all air parcels in hexagon 2. The air types are indicated by the abbreviations (T, A, B, S). Both tropospheric classes (CT and MT) are combined into one (T). Each possible transition has a unique colour variation of the colour used to encode its class assignment in hexagon 1. The total number of parcels in each class in hexagon 1 and 2 is indicated on each row/columns edges, respectively.

of two instances of labelling, here the labels in hexagon 1 and 2. The column i indicates the label in hexagon 1 (X), the row j indicates the label in hexagon 2 (Y). The number of transitions of that kind is then given by the matrix element (i, j), which in the following is referred to as '$X - Y$'. Each class is abbreviated by its first letter as defined above. A transition $X - Y$ would indicate air of type 'X' is becoming air of type 'Y'. Air type 'Y' is not necessarily the air type 'X' is mixing with, but simply the resulting air type. How many air types are involved in the mixing, i.e. if it is two or even three states mixing into each other, does not follow trivially from Fig. 3.10 and requires considerations of mass conservation. If the mixing is relatively slow, air type 'Y' is likely the air type 'X' is mixing with. In general, not every possible combination is observed. Therefore, most elements are zero. Elements with very few transitions can indicate issues with the trajectories

or with the classification and are not considered significant. However, it is not to say that these transitions are not real. From the confusion matrix qualitative insights into the mixing processes can be derived:
The diagonal entries describe transport rather than mixing. Such transport is most prevalent inside the troposphere ($T - T$) and the stratosphere ($S - S$), with a few such transport transitions in the intrusion ($B - B$).

A large share of blended air is created via the transitions $T - B$ and $A - B$. ASMO air also transitions into stratospheric air via $A - S$ in large quantities. There are no significant transitions $X - A$, where X is any other class, which lead to the creation of ASMO air. This again supports the characterisation as ASMO air, as Asian summer monsoon outflow should be the product of the monsoon and not a product of mixing. The transition $T - S$ sticks out as a direct transition of tropospheric air into stratospheric air without an observed intermediate state, which could be linked to one of many error sources such as the trajectory calculations, issues with the retrieval or the classification itself, as no tropospheric air lies directly bordering the stratospheric air. For this reason the transitions $T - B$ must be localized directly at the intrusion.
The transitions $T - B$ and $A - B$ stand out because when considering the average chemical composition of the classes T, A and B (see Fig. 3.8), neither T air nor the A air contain sufficient stratospheric tracers such that B air could have been created due to mixing from any combination of T, A and B air. The deficit of stratospheric tracers requires at least some S air to make up the difference. A possible explanation would be the transition $A - S$, which has consistent increases in both $\mathrm{O_3}$ and $\mathrm{HNO_3}$. The chemical budget would add up if A air partially mixed with S air, and afterwards mixed with T air creating B air. Such mixing can only occur where the three air types come into direct contact or via intermediate states. Otherwise, an additional source of $\mathrm{O_3}$ and $\mathrm{HNO_3}$ would be required, which is investigated in Sect. 3.3.4.

3.3.3 Localization

Since the 3-D tomographic retrievals retain the full 3-D structure of all their measurements, the spatial structure of the mixing is known as well, which is shown in Fig. 3.11:

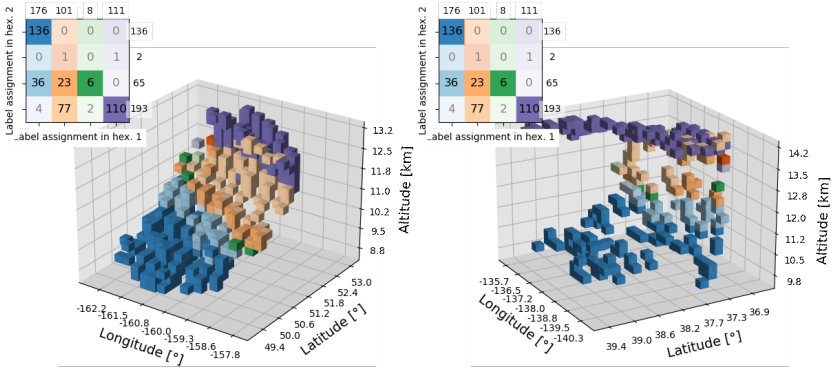

Figure 3.11: Visualization of transition states. Left: trajectory starting points inside hexagon 1. Transition state is indicated in colour-code, see legend. Right: trajectory end points inside hexagon 2. Transition state is indicated in colour-code consistent with Fig. 3.10. Rare and missing transitions are grayed out for visibility.

The mixing processes are structured as follows: the transport $T-T$ is located in the lowermost altitudes purely within the troposphere in both hexagons. It extends up to 9.5 km in hexagon 1 and 11.2 km in hexagon 2, respectively. Above the troposphere lies a layer of blended air from around 9.5 km/10.5 km to 11 km/11.2 km. The blended layer itself consists of two individual layers, which meet at around 9.7 km in hexagon 1. The lower layer is created from the transition $T-B$, the upper layer is created from the transition $A-B$. The transition $A-B$ is notably located at the intrusion – a spatial prerequisite for the mixing required to maintain the chemical balance – and thus makes the associated three-way mixing plausible.

In hexagon 2, the lower blended layer is stacked similarly, but the upper blended layer is much more interwoven with the stratospheric air originating from $A-S$.

Above the blended layer lies a layer of ASMO air A, which transitions into stratospheric air and extends to the stratosphere at around 12 km/13.5 km. Upwards lies the stratosphere with transport within itself. Each transition again forms coherent regions, with the transition $B-B$ being a small exception. Despite the spatial distortion observed in hexagon 1, each transition follows a stratified vertical structure. The exchange between the troposphere and the stratosphere still appears to be mostly vertical, despite the perturbation caused by the tropopause fold, which most likely induces

horizontal mixing. The stratification of the mixing can be explained by the dynamical properties of the atmosphere, which counteract the distortion. The mixing of tropospheric air and stratospheric air via ASMO air occurs over a larger vertical range, which could imply that this distortion is indeed necessary to initiate this tracer exchange. All class transitions are confined to those sections of the hexagons where the involved air masses come into physical contact.

3.3.4 Tracer changes

So far the mixing processes have been explained qualitatively by considering changes in complex correlations, which conform with our understanding of the physical processes. Since for each matched point there are measurements in both hexagons, the direct changes for each observed tracer can be considered as well.

Quantitative insights into the mixing processes can be derived from the autocorrelations in Fig. 3.12, Panels a - g.

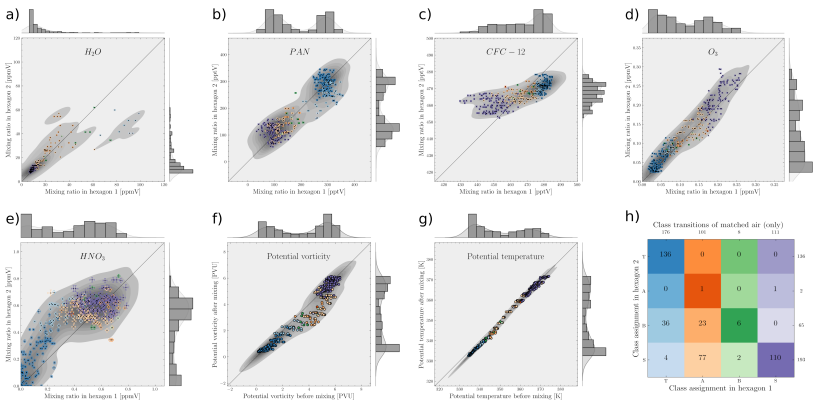

Figure 3.12: ***Panels a) - e)****: Overview of tracer changes between the two hexagons. The identity (no changes) is shown in the background as a line. The different transitions are indicated in colour-code in accordance with the confusion matrix shown again in panel h). Uncertainties due to retrieval parameters are shown as well (small). Histograms of measurements are shown on the edges. A Gaussian kernel density estimation is shown as shading.* ***Panel a)****: The value range for $\mathrm{H_2O}$ is artificially limited to focus on low values.* ***Panel f)****: Changes in potential vorticity (ERA5 data) during mixing.* ***Panel g)****: Changes in potential temperature (ERA5 data) during mixing.*

The autocorrelation of a quantity is the correlation with itself at two different times. The potential temperature (panel g) shows the isentropic nature of the mixing. The potential vorticity (PV, Fig. 3.12f) indicates that the transitions $A - B$ and $A - S$ are induced by turbulent mixing, as their PV changes the most.

Since the measurements were derived from a retrieval, we estimate their uncertainties by applying a rigorous Monte Carlo approach to determine the influence of different factors on the retrieval result. A detailed overview of the exact methodology can be found in Ungermann (2013) and Ungermann et al. (2022). The uncertainties are generally small compared to the variance in the measurements. Investigating the retrieval diagnostics reveal that the results are dominated by observations, with little influence of the a priori (not shown). Since $\mathrm{H_2O}$ can condense out, it is generally not a well–conserved quantity and therefore not very descriptive for the mixing processes. Such effects appear to be visible for the tropospheric water vapour, which sees significant reductions in between flights. The changes in PAN divide the matched points into two groups, the tropospheric air and all other air types. It stands out that the stratospheric air (in hexagon 2), which was created from the ASMO air (in hexagon 1), contains significantly more PAN than the other stratospheric air, further substantiating the transport of tropospheric tracers into the lower stratosphere via filaments of ASMO air. $\mathrm{O_3}$ shows no such structures and reflects the vertical structure. $\mathrm{HNO_3}$ shows significant increases within the troposphere, with the largest increases in the transition $T - B$, which is to be expected when considering the change in class, but highly unexpected considering the other stratospheric tracer $\mathrm{O_3}$, as this increase cannot be explained by directly mixing with stratospheric air. While $\mathrm{HNO_3}$ shows the largest uncertainties of all retrieved species, such uncertainties cannot explain this increase.
Potential sources of $\mathrm{HNO_3}$ over the eastern Pacific Ocean are mostly limited to lightning storms, which may have occurred around the time of measurement. Tie et al. (2001) state that lightning events can lead to increases of up to 300 pptV in $\mathrm{HNO_3}$ compared to the background, which is of the same order of magnitude to the observed increases. This $\mathrm{HNO_3}$ could then have been mixed into the air masses along the trajectories of the revisited air parcels and thus could explain the anomalous increases. The Geostationary Operational Environmental Satellite (GOES West) satellite provides observations of lightning events using its geostationary lightning mapper. The positions of these events are shown in Fig. 3.13 between 10 September 2023, 00:00 UTC (shortly after the first hexagon) and 11 September 2023, 00:00 UTC (shortly after the second hexagon).

Both hexagons and the trajectories of the air parcels inside hexagon 1 are shown similarly to Fig. 3.2c. Those trajectories that show significant increases in $\mathrm{HNO_3}$ are marked in orange. There are several lightning events close to the trajectories in question so that this is consistent with the hypothesis that these events have influenced the transported air. The increases observed are of the same magnitude of a few hundred pptV as described by Tie et al. (2001) for lightning events.

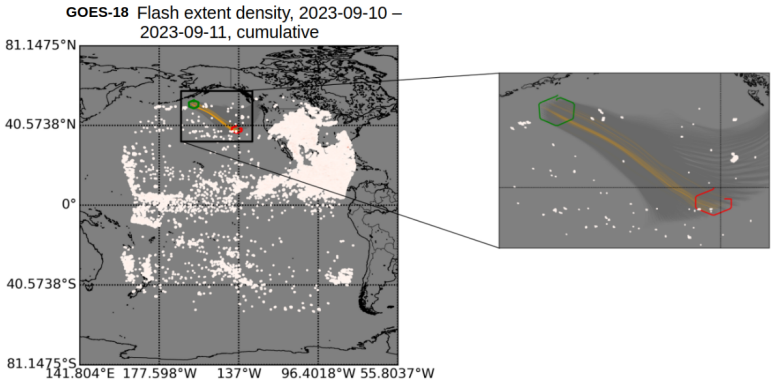

Figure 3.13: Map of all lightning events detected by GOES-18 between the time hexagon 1 and hexagon 2 were measured. The inset focuses on the relevant section including both measurements. Both hexagons are shown in colour, as well as the trajectories with anomalously increased $\mathrm{HNO_3}$.

Further analysis would require vertical resolution of these events, which GOES-18 cannot provide. Assuming that any change in $\mathrm{HNO_3}$ can be explained by lightning events, the classification can be repeated with constant $\mathrm{HNO_3}$. The resulting transitions are however mostly unaffected by this. Most affected are $T - B$, for which 30 of 36 air parcels would still be considered tropospheric, and $A - B$, for which 18 of 23 air parcels would still be considered ASM outflow. Again, this assumes that all changes are due to lightning events and ignores the real changes in $\mathrm{HNO_3}$. Interestingly, the transition $A - S$ is completely unaffected by this, substantiating that ASM outflow is indeed mixing into the lower stratosphere via this mixing pathway.

3.4 Analysis of 2-D curtains

The analysis of the 3-D data products has illustrated how a probabilistic air type classification based on cluster detection can be utilized to identify both the mixing processes and the air types involved in them in a model-agnostic way, even for dynamically distorted situations. So far it was demonstrated only for a single case study. If the concept is more generally applicable, one may ask: Does the concept depend on the explicit choice of trace gas species? Can the spatial coverage be expanded from a single, compact hexagonal volume to, for example, an entire research flight? To answer these questions the 2-D curtains are analysed for both flights PH13 and PH14 in the following. The curtains include many more trace gas species, such that the phase spaces can be expanded considerably. Namely, in addition to the original trace gas species, the species

$\mathrm{CFC}-11$, $\mathrm{HCFC}-22$, $\mathrm{C_2H_6}$, $\mathrm{CCl_4}$, $\mathrm{CH_3OH}$, $\mathrm{ClONO_2}$, $\mathrm{C_2H_4}$ and HCOOH are included:

- $\mathrm{CFC}-11$, $\mathrm{CFC}-12$, $\mathrm{HCFC}-22$, $\mathrm{CCl_4}$ and $\mathrm{ClONO_2}$ are anthropogenic chlorine-containing species, which have their origins in industrial uses and have long lifetimes. These tracers release chlorine in the stratosphere via photolysis, which is a major factor for ozone depletion (e.g. Zuiderweg et al., 2012).
- $\mathrm{C_2H_2}$, $\mathrm{C_2H_4}$ and $\mathrm{C_2H_6}$ are volatile organic compounds (VOCs) created during burning of biomass or fossil fuel and have moderately long lifetimes.
- $\mathrm{CH_3OH}$ and HCOOH are oxygenated VOCs and have similar sources in biomass burning.

Since we already have a good understanding of the physical situation of both flights, we can validate the resulting classifications against this understanding.

When handling many different trace gas species, collinearity must be considered. Functionally similar tracers (e.g. several long-lived CFCs) convey basically the same information, thereby implicitly overweighting that information in the classification. We therefore assess collinearity when selecting tracers for the classification via the variance-inflation factor (VIF) of a trace gas species x_j, which is defined as:

$$\mathrm{VIF}_j = \frac{1}{1 - R_j^2} \tag{3.2}$$

where R_j denotes the regression of x_j on all components x_i with $i \neq j$. It describes how much x_j can be expressed linearly by any x_i. A VIF of one indicates no collinearity, while a VIF with a value greater than $5 - 10$ indicates serious collinearity. Components with high VIFs can be removed iteratively, recalculating the VIFs for the reduced set, until no severe collinearity remains. Since Eq. 3.2 is symmetrical for two collinear components x_i, x_j, either one can be chosen to be removed. It should be noted that removing trace gas species here is used simply as a model-building device. In the presence of measurement noise, this may actually skew the classification towards trace gas species of lower quality, since measurement noise artificially reduces the VIF. Here, this approach is an acceptable trade-off.

For both curtains we find that both $\mathrm{ClONO_2}$ and $\mathrm{C_2H_4}$ are too collinear with the other components and are subsequently excluded. $\mathrm{O_3}$ and $\mathrm{HNO_3}$ are also rather strongly collinear, but are physically sufficiently independent such that both are retained. A total of 11 trace gases are used in the following for either curtain.

The classification is applied as before (see Fig. 3.3) to both curtains. Since there is no reliable way to determine revisited air parcels, we are free to choose different parameters for either

flight. For both flights a 7-D manifold was determined, which can no longer be visualized. For PH13 we find seven clusters within the data, whereas for PH14 only five were found. This is already consistent with the previous analysis, where two pairs of clusters in PH13 merged compared to PH14. For all subsequent classifications, colours are assigned according to the physical interpretation of the respective clusters rather than their arbitrary numerical labels. Clusters with comparable chemical composition to the air masses identified in the previous analysis of the 3-D data products are therefore shown using the same colour. In particular, blue denotes tropospheric air, orange Asian summer monsoon outflow, green blended air, purple stratospheric air, and different shades of these colours indicate chemically similar, but sufficiently distinct air types. Labels and classification confidence are again calculated following Eq. A.29 and Eq. A.28.

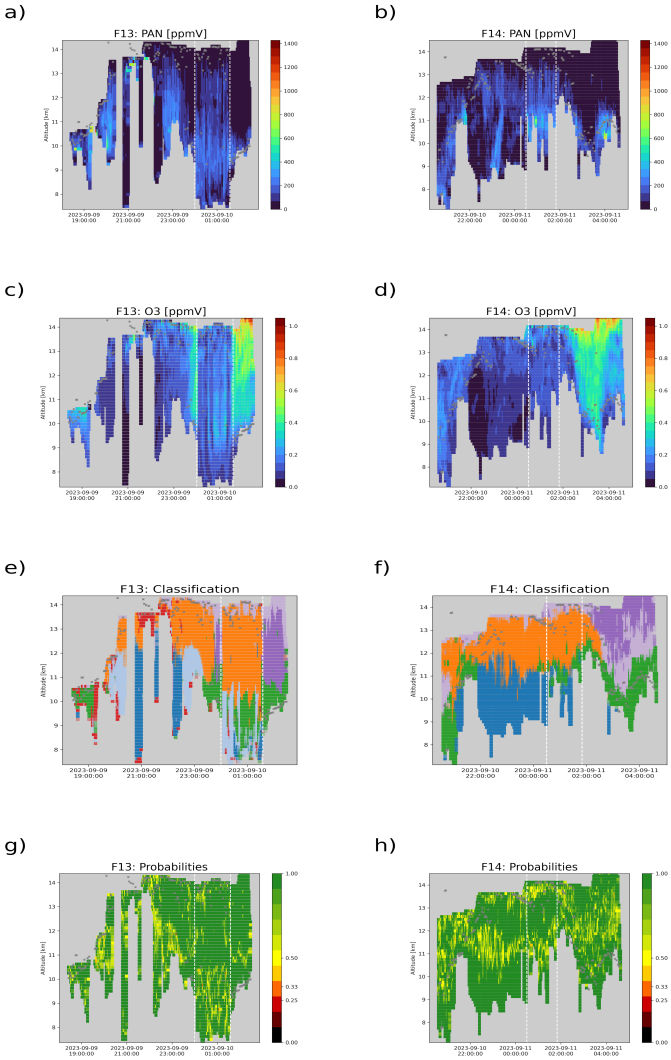

Figure 3.14: Classification of the curtains using GMMs. Panels a, c, e, g (left-hand side): Distribution of PAN, $\mathrm{O_3}$ *, labels and probabilities for PH13, respectively. Panels b, d, f, h (right-hand side): Distribution of* PAN, $\mathrm{O_3}$, *labels and probabilities for PH14, respectively. The labels are shown in colour-code consistent with previous classifications such that similar air types share similar colours. The probabilities are colour-coded such that they indicate label ambiguity (green: unambiguous, yellow: two competing classes, orange: three competing classes ...). The thermal tropopause is shown in gray markers; each hexagon is indicated by the white, vertical lines.*

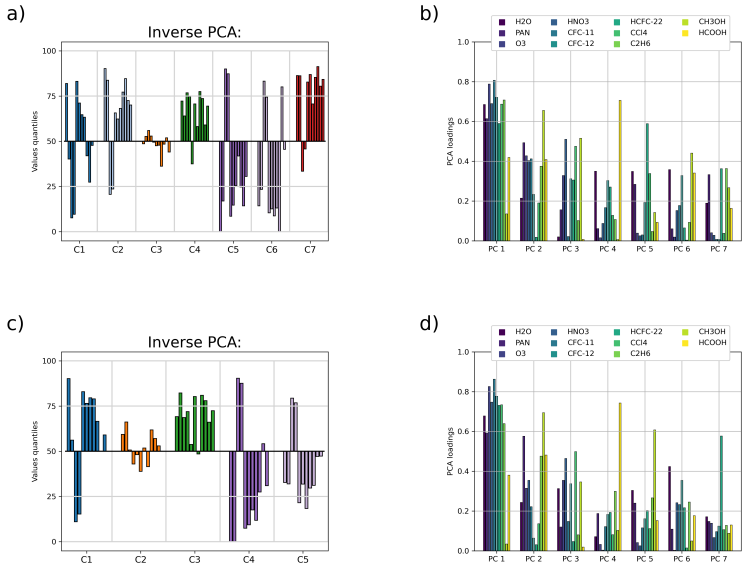

Figure 3.15: Technical information for the GMM classifications of the two 2-D curtains. Panels a,c show the chemical composition of each cluster centre after inverse PCA transformation for PH13 and PH14, respectively; values are shown as percentiles of the respective total distribution. Panels b, d show the corresponding PCA loadings, with the associated trace gas indicated by colour. For cluster-related quantities in panels a, c, colours encode the physically interpreted air-mass category consistently across the 3-D retrievals and both curtains: blue denotes tropospheric air, orange ASM outflow, green blended air, purple stratospheric air, and red strongly polluted air. Different shades denote distinct classes within a category.

The first column in Fig. 3.14 shows values for PH13, the second column for PH14. Since the number of trace gas species is too large to reasonably visualize, the two tracers PAN and $\mathrm{O_3}$ are shown in the first two rows as representative examples. The third row shows the label distribution, where the colour-code is already adjusted such that similar air types (across both curtains and the 3-D analysis, see Fig. 3.15a, c) share the same colours. The last row shows the label probability in a colour-code indicating labelling ambiguity. The thermal tropopause is indicated as gray dots, the respective hexagon is indicated by the white, vertical lines. These segments only show the front–facing surfaces of the 3-D retrieval volume and thus show similar, but not identical situations.

For both flights PAN and $\mathrm{O_3}$ are mostly anti-correlated. For PH13 there are a few minor signatures of increased PAN between 19:00 UTC and 21:00 UTC and a strong increase around 21:30 UTC close to the thermal tropopause. A similar feature of increased $\mathrm{O_3}$ is visible at 19:00 UTC at lower altitudes. Around 21:30 UTC cloud cover limits the vertical resolution of GLORIA such that much information is missing. Past this point there is a feature of somewhat increased

PAN between 10 km and 13.5 km altitude, which is disrupted by clean, apparently stratospheric air (according to $\mathrm{O_3}$) near the edges of the hexagon.

The classification for PH13 (Fig. 3.14e) in conjunction with the chemical composition (Fig. 3.15a) shows the physical situation of the atmosphere: There are two tropospheric air types (shown in deep and light blue). The deep blue type shows less pollution and can be linked to the NIN region. In contrast, the light blue type shows a strong pollution signature and can be linked to the NWP and SEA regions instead, indicating different origins. The difference between both types is visible in PAN. Both types are located at the lowermost altitudes. The labelling probabilities show increased mixing (ambiguity).

The ASM outflow (shown in orange) shows a similar chemical structure of no pronounced features as in the previous analysis and is located exactly where the surface-origin tracers in Fig. 3.2a placed the filament. Contrary to the previous analysis, there is no link between this class and the SEA or TWP region. A reason for this may be the choice of the number of clusters, as reducing the number of clusters to six and identifying clusters between the original and this new classification instance yields a corresponding ASMO cluster with the linkage to the SEA and TWP regions instead.

Between the tropospheric and the stratospheric air lies another layer of blended air (shown in green), with its distinctly mixed chemical signature. The probabilities suggest stronger mixing of air masses in this region, further substantiating the identification.

There are two stratospheric air types (shown in deep and light purple) located at the highest altitudes. In contrast to the deep purple one, the light purple type has a signature of increased $\mathrm{CH_3OH}$. As both air types are stratospheric in character, there are no source regions linked to them. The stratosphere extends downwards to 10.5 km beyond 02:00 UTC in accordance with both PAN and $\mathrm{O_3}$.

Unique to the curtains is a seventh air type in the form of highly polluted air (shown in red), which, while having strong signatures in $\mathrm{H_2O}$ and PAN, also has significant signatures for almost all pollution tracers. This air type coincides with the feature of strongly increased PAN and does not form a coherent structure as far as GLORIA can measure. It is located within the upper troposphere.

Considering the loadings in Fig. 3.15b again, every trace gas species contributes significantly to the first few principal components. The finer structure of the manifold (prevalent in the lower principal components) is given by individual pollution tracers, as most principal components show

at least one peak in these tracers.

In PH14, there are similar structures in PAN, as shown in Fig. 3.14b. There are signatures of high PAN scattered at low altitudes, with a band-like structure of increased PAN visible at altitudes between 11 and 13 km. In the first half of the flight, there are patch-like structures of increased PAN, which are interrupted by patches of very low PAN. Outside of this region there is a seam-like structure at the lowermost observed altitudes. Inside the second hexagon the vertical structure appears rather stratified similar to the previous results. Towards the end of the flight the tropopause is very low, reaching altitudes of 9.5 km, which is also visible in ozone. The $\mathrm{O_3}$ values in Fig. 3.14d mirror this structure. There is a section around 23:00 UTC where both PAN and $\mathrm{O_3}$ are low. The low thermal tropopause around 03:00 UTC is reflected in a large feature of high ozone.

Considering the classification of PH14 in Fig. 3.14f in conjunction with the chemical signatures in Fig. 3.15c, the following physical situation is shown: A single tropospheric air type (shown in blue) with somewhat increased pollution signature can be found at the lowermost altitudes below the thermal tropopause. A clean, tropospheric air type is either no longer present or has insufficient contrast to appear separate. This coincides with the 3-D retrievals, for which the separation between tropospheric classes also vanished for PH14. A highly polluted class, on the other hand, is not present in the data.

The air type with the distinct ASMO signature (shown in orange) is visible as a coherent band at middle altitudes. When we consider Fig. 3.2b, we see that it again coincides with the predicted filament, particularly beyond 03:00 UTC where both the filament and the class vanish. This air type can again be linked to both SEA and TWP regions.

A blended air type (shown in green) with both tropospheric and stratospheric signatures is located between the troposphere and stratosphere. Surface-origin tracers again suggest a linkage to the NWP region.

The final two air types (shown in light and deep purple) show very stratospheric signatures and consequently are located at the uppermost altitudes. They reflect the deep intrusion of $\mathrm{O_3}$ at 03:00 UTC. While both appear very similar, the light purple air type shows slightly increased signatures in the tropospheric tracers. Its lower position may indicate partial, slow mixing with air masses underneath it. The probabilities in Fig. 3.14h indicate some mixing underneath the ASMO filament and in the green / light purple classes past 02:00 UTC.

Considering the loadings in Fig. 3.15d, we see that similarly to the classification of PH13 each

tracer contributes significantly to the first principal components. Lower principal components again show peaks in individual tracers, albeit in different ones.

Across different data products covering different trace gas species the presented classification method is capable of accurately identifying all air types present in the data. The chemical composition of these air types is very consistent across multiple classification instances, and linkage between different origin regions and these air types is consistent across these instances as well. As the approach is model-agnostic, there is no need for any prior information on the physical situation, nor detailed understanding of the physical significance of every chemical species involved. Label probabilities are consistently reduced near the borders of coherent air masses and less well-defined where mixing would be anticipated, thus corroborating their identification with the mixing strength.

The previous analyses show that this form of classification is a powerful way of air mass characterisation. A technical limitation of the presented pipeline is that the PCA is ultimately a linear transformation. Using a GMM as a clustering algorithm therefore assumes that clusters appear Gaussian-shaped in this transformation. Considering the blended class in Fig. 3.7b, we see that this assumption only holds partially. The final question of this chapter is: Are there ways to generalize the approach with respect to dimensionality reduction and clustering, and would they still yield the same results?

3.5 Deep embedding clustering using AI

The fundamental challenge for the classification problem is twofold: finding a suitable manifold, which best represents the original data, and detecting clusters within that manifold. Technically, the shape of a cluster depends on the underlying manifold, the same way the shortest path between two points is a line in flat space, but an orthodrome on a sphere. The concept is illustrated in App. A.12. Even for the simplest clustering algorithms, the labels are usually assigned with respect to some distance metric, which is affected by the shape of the manifold. A suitable manifold should therefore be determined with respect to the clustering algorithm itself.

A DEC neural network solves these two problems in tandem by jointly optimizing a manifold and its clusters to maximize cluster separation. A DEC therefore usually determines highly non-linear manifolds. In keeping with the previous analysis, the distributions used within the DEC's clustering part are assumed to be Gaussian. Both the dimension of the latent space optimized

by the DEC and the number of clusters are adopted from the previous analysis. This choice is justified because the intrinsic dimensionality of the data is an inherent property, and the new clusters are to be compared with the conventional method. Fig. 3.16 shows the classifications found by dedicated DECs:

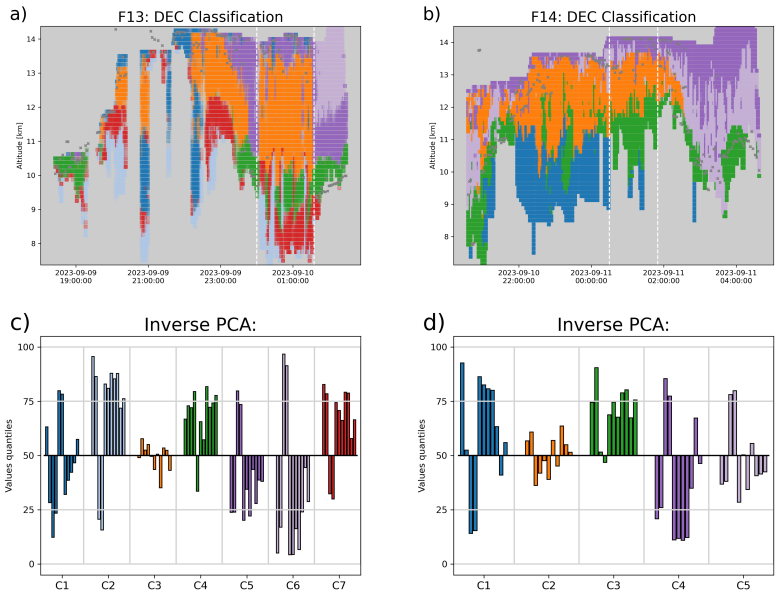

Figure 3.16: Summary of the DEC classifications of the 2-D curtains. The left and right columns show PH13 and PH14, respectively. The upper row shows the spatial distribution of the DEC-derived classes; the lower row shows their chemical composition. Colours are assigned by matching each DEC cluster to the chemically most similar air-mass category identified by the GMM classifications: blue denotes tropospheric air, orange ASM outflow, green blended air, purple stratospheric air, and red strongly polluted air. Different shades distinguish separate but chemically similar classes.

The left-hand side shows the classification for PH13, the right-hand one for PH14. Because cluster labels are arbitrary and are not directly comparable between independent classifications, DEC clusters are assigned the plotting colours of the GMM-derived air-mass categories with the most similar chemical fingerprints and spatial distributions. Comparing the DEC classification to the GMM classification, we find substantial similarities between them. The air types are found in the same regions generally following very similar patterns and accurately reflect prominent physical features such as stratospheric intrusions and low tropopauses. The troposphere, stratosphere and the ASMO filaments all fall into the correct regions, with all air types forming coherent structures. The chemical composition of each cluster (Fig. 3.16c,d) is nearly identical to the com-

positions derived from the GMM model (Fig. 3.15a,c). A few key differences do stand out: The deeply polluted class (red) in PH13 is more prevalent, appearing less as isolated regions but rather as larger, more coherent patches. A major misclassification is visible in PH13 before 22:30 UTC at altitudes ≥ 13 km, which assigns tropospheric air masses in this region. The presence of tropospheric air at these altitudes is rather unphysical given the overall picture. A minor difference in PH14 is the noise inside the ASMO filament, which contains isolated points of blended air. Overall, both concepts provide very similar classifications, despite the fact that the DEC did not rely on any assumptions regarding its manifold.

3.6 Chapter conclusion

This chapter investigated mixing in a filament of ASM outflow observed near a tropopause fold in the UTLS during a dedicated self-matching experiment in the 2023 PHILEAS campaign. 3-D tomographic multi-tracer observations obtained on two consecutive days were combined with trajectory calculations to identify and revisit individual air parcels. The observed filament exhibited a complex chemical and spatial structure that was not adequately represented by the dynamical tropopause or by individual tracer fields alone. This motivated the development of a data-driven, multi-species classification that identifies chemically distinct air types from the shape of their corresponding phase space without requiring prior knowledge of the local atmospheric state. In this double–flight experiment, the classification resolved continental and maritime tropospheric air, stratospheric air, Asian summer monsoon outflow, and a blended air mass with both tropospheric and stratospheric characteristics. The spatial distributions of each class are coherent and physically consistent, despite no spatial information being used in the classification. Using surface-origin tracers, the continental tropospheric class can be linked predominantly to East China, whereas the maritime tropospheric class links to the North–Western Pacific. The Asian summer monsoon outflow class is associated with the contributions from South-East Asia and the Tropical–Western Pacific, consistent with the expected composition of air transported from the edge of the ASM anticyclone.

By comparing the classifications of trajectory-matched air parcels between both observations, explicit mixing pathways could be identified. In particular, ASM outflow air was observed to transition directly into the lower stratosphere, where these air masses came into physical contact. The majority of these matched air parcels initially classified as ASM outflow underwent a transition

into stratospheric air. This pathway transports tropospheric constituents, including the pollution tracer PAN, into the lower stratosphere, and thus is identified as an important pathway with implications for the global climate. A blended air mass with strong tropospheric and stratospheric influences was observed as a coherent, comparatively persistent intermediate state. Its chemical composition indicates that its formation cannot be explained by mixing among the tropospheric and ASM-outflow air masses alone. Together with the observed transitions, this suggests a more complex mixing pathway involving tropospheric, stratospheric, and ASM-outflow air, which is supported by the spatial structure of these air masses.

Changes in the retrieved tracers generally support the transition-based interpretation. The anomalous increase of HNO_3 in a subset of tropospheric air parcels cannot be explained by mixing alone and is consistent with the occurrence of lightning close to the relevant trajectories. A sensitivity study showed that this effect can influence the assignment of some parcels to the blended class, but does not affect the principal transition from ASM outflow to stratospheric air.

The classification is inherently probabilistic and therefore additionally provides an indicator of ongoing mixing through reduced classification confidence at the boundaries between coherent air masses. In order to demonstrate the rather universal applicability of this classification, the same classification was applied successfully to the complementary 2-D data products, which cover a much broader range of chemical species. The chemically and spatially consistent air-mass categories recovered from these independent data products demonstrate that the approach is not tied to a particular retrieval geometry or tracer combination. Finally, the approach is further generalized by using DEC, which does not make any assumptions about the manifold used for the classification. The DEC classification method yields similar results to the PCA approach. Multi-species, data-driven classification therefore provides a robust framework for characterizing complex air-mass structures and for quantifying mixing from repeated observations of the same air parcels.

Chapter 4

Filaments in the lower stratosphere

Moving upwards from the UTLS, the next compartment in which filaments occur frequently is the stratospheric surf zone. Across the subtropical edge of the tropical pipe – the region of upwelling of tropical air in the Brewer–Dobson Circulation (BDC) – large-scale Rossby waves and transient eddies intermittently weaken the subtropical transport barrier by distorting the otherwise strong meridional PV gradients (Plumb and Eluszkiewicz, 1999; Abalos et al., 2021; Gupta et al., 2023; Ivaniha et al., 2026). The resulting horizontal shear and strain stretch long, thin filaments of young tropical air across the climatological subtropical barrier along isentropic surfaces and into the extratropical stratosphere (Polvani et al., 1995; James and Legras, 2009). These filaments are then stirred in the surrounding flow and are eventually irreversibly mixed into it by small-scale turbulence, the breaking of smaller-scale waves, and diabatic processes (Haynes and Shuckburgh, 2000a,b; Gupta et al., 2023). Such filaments are vertically thin (2 – 5 km (e.g. Hegglin et al., 2004; Krüger et al., 2005)), horizontally narrow (50 - 500 km (McKenna et al., 2002b, e.g.)) and can extend up to 10,000 km in length (e.g. Polvani et al., 1995; Krüger et al., 2005). These 'leaks' of tropical air into the extratropical stratosphere are reflected in the horizontal mixing rates diagnosed from age-of-air and long-lived tracer budgets.

The stratospheric surf zone usually begins above the 450 K isentrope, which corresponds to around 20 – 21 km altitude. The contrast between the tropical signature of the filaments and the extratropical stratospheric background is highest at these altitudes (see tape recorder model Mote et al., 1996), making this region particularly well suited for studying the mixing processes in these filaments. Studying structures of this size at these high altitudes requires suitable satellite instruments. The main challenge for studying mixing processes explicitly is to measure the same air

parcels at successive times. Since the orbit of a satellite is fixed, these repeated measurements strongly depend on the measurement geometry of the instrument itself, particularly the swath width.

Mixing processes in the stratosphere were one scientific objective of the 'Changing Atmosphere Infra-Red Tomography Explorer' (CAIRT) proposal submitted to the European Space Agency (ESA) in 2020 (ESA, 2023, 2025). This study is conducted as a proof of concept for the value of such data and is performed based on synthetic CAIRT observations. CAIRT is an imaging infrared limb sounder capable of retrieving 3-D fields of temperature and many trace gas species.

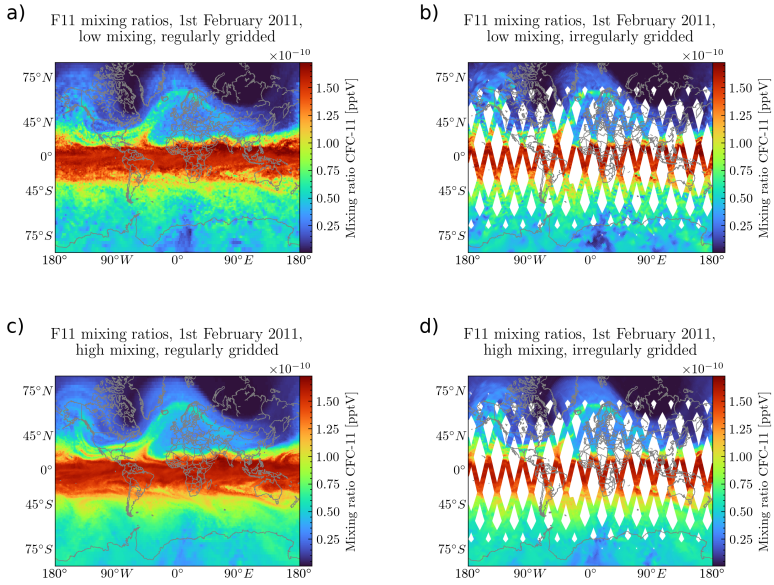

Figure 4.1: Visualization of the different CLaMS simulations and data products. All panels show the spatial distribution of $\mathrm{CFC}-11$ *at 21 km altitude for the 1 February 2011. Regularly gridded data for 12:00 UTC are shown in the first column, the synthetic observations are shown in the second column. The synthetic observations are the cumulative observations over 24 h at individual times. Panels a, b show the control mixing simulation, panels c, d show the high mixing simulation.*

These synthetic observations were created from simulations by the Chemical Lagrangian Model of the Stratosphere (CLaMS; e.g. McKenna et al., 2002b,a; Konopka et al., 2010; Pommrich et al., 2014) using an instrument-emulating sampling scheme: Trace gas values were retrieved only at the current tangent points of CAIRT, i.e. along the orbit and at its measuring frequency of seven measurements per second corresponding to 50 km along-track. The CAIRT instrument

performs 2-D imaging with a horizontal swath of roughly 400 km, a typical sampling of 25 km across-track and about 1 km vertically over a range from 5 to 115 km. It is therefore capable of resolving fine-scale horizontal gradients and large segments of filaments. The final retrieval grid of these measurements is hence along-track and continuous in time. The data structure, compared to conventional simulation grids, which return the whole atmospheric state in typically fixed time intervals, is illustrated in Fig. 4.1.

Mixing is handled in CLaMS by the creation and merging of Lagrangian air parcels where the local deformation of the flow quantified by a Lyapunov coefficient triggers mixing (McKenna et al., 2002b; Konopka et al., 2005; Konopka, P. and Tao, M. and Ploeger, F. and Diallo, M. and Riese, M., 2019). To investigate whether the mixing strength used in the simulations can be resolved from such synthetic observations, two different configurations were used. One simulation was performed using the standard parameter of CLaMS, referred to in the following as the 'control mixing scenario' or the 'low mixing scenario' (Fig. 4.1a, b). A second simulation had the Lyapunov parameter increased to allow for stronger mixing compared to the first simulation. This simulation is referred to in the following as the 'high mixing scenario' (Fig. 4.1c, d). In addition to the synthetic observations, traditional regularly gridded data (at 12:00 UTC for every day) of either mixing scenario are considered as well.

This chapter begins with a classification of the atmospheric state, which is derived from available trace gas species to describe the mixing processes, similarly to the previous chapter. Using this classification, methods for filament detection are presented and applied to the regularly gridded data. The mixing processes are derived by calculating forward trajectories for a set of regularly spaced air parcels using a fourth order Runge-Kutta scheme (see App. A.1) and matching these trajectories with available data. The second part of this chapter adapts the same methodology to the synthetic observations and explores how well these results agree with the earlier findings. Finally, we train convolutional neural networks (Convolutional Neural Network (CNN)s) on the time series derived from matching trajectories with the synthetic data and investigate whether the underlying mixing processes can be captured by the network.

This chapter aims to answer the following questions:

- How can the mixing processes in the stratospheric surf zone be parametrized? Is such a parametrization useful for identifying filaments as spatial anomalies?
- What are typical mixing timescales and how do they vary by region?
- Can the same mixing processes be derived if satellite data is used, instead of regularly gridded data?
- Since such timeseries contain information about the mixing implicitly, can neural networks be used to learn such mixing processes?

4.1 Parametrization

The first step is to identify trace gas species which are suitable for detecting the filaments. Since the filaments typically persist for around one week, chemically inactive species are ideal for separating the mixing processes from chemical processes. For this analysis the trace gas species $\mathrm{CH_4}$, $\mathrm{N_2O}$, $\mathrm{SF_6}$, $\mathrm{CFC{-}11}$, $\mathrm{CFC{-}12}$ and $\mathrm{HCFC{-}22}$ are considered, which all have stratospheric lifetimes ranging from 10 years to 1000 years (Brown et al., 2013; Minschwaner et al., 2013). For convenience, we refer to this set of trace gas species as Long-Lived Tracer (LLT).

Visual inspection of the data (see Fig. 4.1) indicates that filaments can be found at 21 km altitude. We will focus on these filaments to keep the problem relatively simple. Since we investigate isentropic transport, we can limit the classification to those isentropes around 21 km altitude, namely isentropes between 475 K and 580 K. These isentropes are completely contained in the vertical range between 17 km and 26 km. Visual inspection also shows that the spatial structure is very similar for any of the LLTs. Any choice of two LLTs would therefore be suitable to derive a classification, rather than using the full 6-D phase space.

For this parametrization we arbitrarily pick the LLTs $\mathrm{CFC{-}11}$ and $\mathrm{N_2O}$ to demonstrate the process. Their correlation is shown in Fig. 4.2 for the isentropes of 500, 550 and 600 K on the 1 February 2011 for the high mixing scenario: The latitude is shown in colour-code. The point density of this correlation for the 550 K isentrope is shown on the right-hand side by the height of the bins. Clusters in phase space are indicated by numbers. The correlation can be divided into the two hemispheres. On each hemisphere, the respective correlation function f is *compactly supported*,

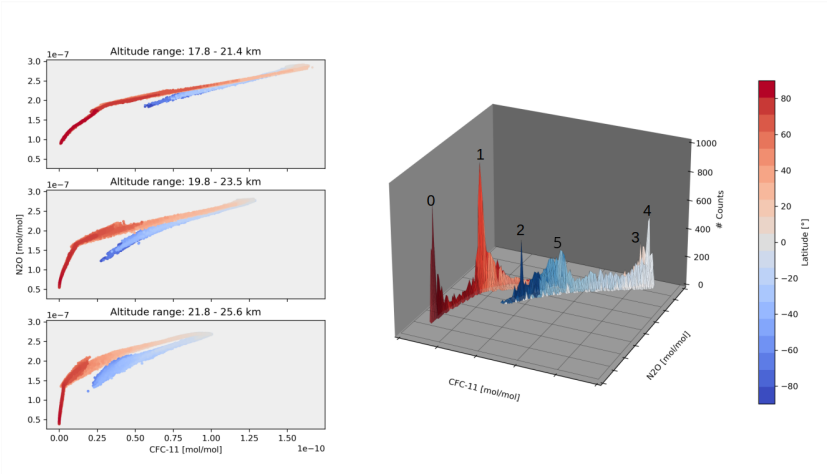

Figure 4.2: Correlation between $\mathrm{CFC}{-}11$ *and* $\mathrm{N_2O}$*. Shown is data for the 1 February 2011 for the control mixing scenario. On the left-hand side correlations for the isentropes of 500, 550 and 600 K are shown, respectively. The latitude is shown in colour-code. On the right-hand side the point density for 550 K is indicated in the vertical axis, with outstanding peaks (clusters) marked by the numbers 0 – 5.*

meaning its support:

$$\mathrm{supp}(f) = \overline{\{x \in X \mid f(x) \neq 0\}} \tag{4.1}$$

is a compact set. Each correlation function contains linear segments, which indicate faster mixing, and curved segments, which represent slower mixing processes as induced by the BDC (Plumb, 2007). As expected, both hemispheres meet both in phase space and in the tropics (points 3, 4). Since the data are for boreal winter, the northern hemisphere contains an additional segment with very low values indicating the polar vortex (point 0). This segment shows strong curvature where it connects to the rest of the correlation at point 1.

We can classify the phase space following similar arguments regarding air mass classification via data structure as described in Chap. 3, where each cluster corresponds to a different air type and the space between clusters is populated by transitional states. Clusters 3, 4 indicate the tropical air of the tropical pipe. Cluster 0 indicates the polar vortex containing very old (chemically depleted) air. Clusters 1, 5 are the northern and southern extratropical stratosphere, respectively. The transitional states between the tropical and extratropical clusters correspond to the traditional mixing lines. Cluster 2 would also be a suitable candidate for the southern hemispheric extratropics, but its narrow shape and its position beyond cluster 5 indicate that this cluster more likely represents the weaker, albeit existing southern polar vortex. The compact support of these lines implies that all net mixing from the tropical pipe into the extratropical stratosphere must proceed

along these lines in phase space. There are, however, physical constraints which make a mapping between real space and phase space a bit more complicated: mixing between two air parcels requires proximity in real space. Their evolution over time then describes a continuous curve in phase space. The presence of transport barriers limits this argument, as there are spatially close air parcels that do not undergo mixing and therefore do not form curves in phase space. This, however, is irrelevant for a classification of mixing processes.

Mathematically, the correlation function of the chemical state an air parcel takes over time is a *compactly supported correlation function*, meaning that its covariance ρ vanishes for distances r beyond some threshold L:

$$\rho(r) = 0, \quad \forall |r| > L \tag{4.2}$$

Air parcels, which are spatially distant, cannot directly influence each other. To clarify the terminology: a correlation with compact support is bound to a limited domain (Eq. 4.1); a compactly supported correlation function describes that elements of this correlation are only influenced by a limited set of these elements (Eq. 4.2). With these constraints, a suitable parametrization can be constructed as follows: Assume that we have an air parcel that begins in a tropical chemical state $\mathbf{x}_0^1$ (e.g. some point in cluster 3) and transitions to an extratropical chemical state $\mathbf{x}_m^1$ (e.g. some point in cluster 1). After some small time interval Δt the air parcel has changed its chemical state due to mixing. According to Eq. 4.1, only a very limited (and varying) number n_1 of chemical states $\mathbf{x}_2^{n_1}$ are possible. Which of these n_i states is actually realized depends on the physical situation in real space, i.e. on the exact motion of the air parcel and the mixing it undergoes. After another small time interval Δt the air parcel could have reached any chemical state $\mathbf{x}_3^{n_3}$, which depends on the choice of $\mathbf{x}_2^{n_2}$ in the last time step. This argument can be repeated until eventually the desired end state $\mathbf{x}_m^0$ is reached after m time steps. This number m of course varies depending on what chain of transitions $\mathbf{x}_0^1, \mathbf{x}_2^{n_1}, .., \mathbf{x}_m^1$ was followed to reach it. By construction, each such chains satisfies compactness (Eq. 4.1) and compactly supported correlation (Eq. 4.2), i.e. is physical in principle. Of course, many of these chains are not actually occurring, and are only hypothetical sequences. Ultimately, trajectory calculations are still needed to identify individual mixing processes. However, a single chain can be representative of a group of air parcels, if each air parcel roughly undergoes the same sequence of transitional states, and is equivalent in phase space. If we are interested only in the net mixing, all loops in the chains can be removed. So far we have allowed that chemical states are reached multiple times, for example if a chain contains the sequence $\mathbf{x}_4^1, \mathbf{x}_5^1, \mathbf{x}_6^1, \mathbf{x}_5^1, \mathbf{x}_6^1$ (without loss of generality we assume the chains consist of the first

choice in each time step, i.e. with superscript 1). The transitions from $\mathbf{x}_6^1$ to $\mathbf{x}_5^1$ and back to $\mathbf{x}_6^1$ are a simple loop and do not help us reach the end state $\mathbf{x}_m^1$. Therefore, we can cut out every loop in our chains. In practice, it suffices to perform binning on the data to obtain sets of points, which by construction are close within at most the correlation length L. This way we can dramatically reduce the number of states to consider when building chains. The final curve $\mathbf{c}(t)$ in phase space, which best connects $\mathbf{x}_0^1$ to $\mathbf{x}_m^1$, can then be determined by finding the shortest chain using Dijkstra's algorithm. The number of entries in each bin can be used as weights to determine the *most likely* chain, favouring transition states which occur more frequently. Since the curve is discrete due to the binning, a Savitzky-Golay filter can be used to smooth c. For mixing of filaments leaking from the tropical pipe (see Fig. 4.1), the starting state $\boldsymbol{x}_0^1$ is cluster 3, 4 and the end state $\boldsymbol{x}_m^0$ is cluster 1, 2, respectively. Intuitively, the curve corresponds to a hike in the mountains from peak 3 to peak 1, while minimizing the elevation gain. One may think that a simple polynomial fit would have sufficed to find c. In practice, however, a polynomial fit has its own issues, which lead to artifacts in the final parametrization. While complex, the described methodology produces a parametrization which meets the high standards required for the following analysis.

Every point in phase space can be characterized by its position along this curve by orthogonal projection onto the curve:

$$t^* = \arg\min_t \|x - c(t)\|^2 \quad \text{s.t.} \quad (x - c(t)) \cdot c'(t) = 0 \tag{4.3}$$

where $c^{'}$ denotes the derivative of c. $c(t)$ is then normalized such that $t^*(\boldsymbol{x}_0^1) = 0$ and $t^*(\boldsymbol{x}_m^1) = 1$. Since the parametrization begins and ends in the respective cluster, it does not convey information about mixing beyond these clusters. Therefore, the parametrization can be clipped to:

$$M(\mathbf{x}) = \begin{cases} 0 & \text{if } t^*(\mathbf{x}) < 0 \\ 1 & \text{if } t^*(\mathbf{x}) > 1 \\ t^*(x) & \text{else} \end{cases} \tag{4.4}$$

We refer to this parametrization M as the *mixing state*. For isentropic mixing, such parametrizations should be performed for each available potential temperature level. Values between these levels are interpolated. Since M does not convey information about the explicit dynamics, both the high mixing and the control mixing scenarios yield the same classification for both regularly gridded data and simulated observations.

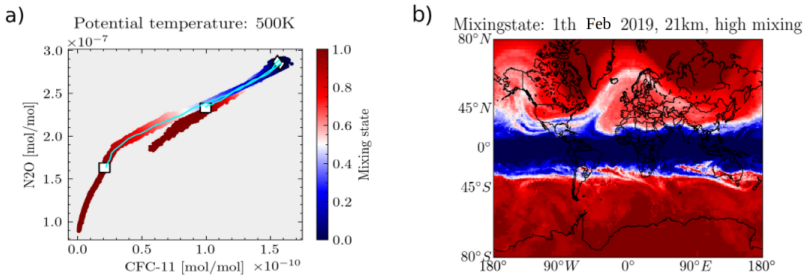

Figure 4.3: Visualization of the mixing state. Panel a): Visualization of the definition according to Eq. 4.3. The mixing state is shown in colour-code. The chain (no filter) is shown in cyan, cluster centres are indicated as white dots. Panel b): Spatial distribution of the mixing state for 1 February 2011 on a level of 21 km altitude.

The mixing state M is shown in Fig. 4.3. It clearly shows the tropical pipe in the tropics and the extratropical lowermost stratosphere as plain coloured regions, and reflects the filaments as anomalous features of the control mixing state, with partially mixed air in white and bright colours. The presented mixing state does not account for any significant variation of the phase space over time. If the position of cluster 1 moved towards cluster 3 (refer to Fig. 4.2), it would affect c and consequently M. We would see a change in M as if the air parcel had undergone mixing, while only the reference system had changed. Changes in cluster position may occur due to imperfect clustering algorithms, changes in sampling as well as real changes of the phase space over time. Noise in the observations could also induce such effects. For these reasons, two variants of mixing state were considered: a *static* mixing state, which determines its parameters at a fixed point in time and keeps them, and a *dynamic* mixing state, which recalculates its parameters for every date. For a period of up to 14 days, the difference between these mixing states lies between 3.6% and 7.4% of the determined value of M on average. For larger periods, these differences become much more pronounced. The different mixing state definitions are only discussed where relevant.

4.2 Dynamics

Rapid exchange is anticipated for medium (≈ 0.5) mixing states (which represent ongoing mixing) and in those regions where breaking Rossby waves induce filamentation. To study the explicit dynamics of this atmospheric state, forward trajectory calculations are performed for all air parcels on a 1° by 1° grid at an altitude of 21 km from 12:00 1 February 2011 until 12:00 14 February 2011. The initial mixing ratios of the different tracers of each tracked air parcel are taken from the initial CLaMS simulation. For each consecutive day at 12:00, the trajectories are matched with the regularly gridded CLaMS simulations of that day, and the mixing ratios of the tracers of each tracked air parcel are transferred from the respective simulation. The ongoing mixing is handled by CLaMS, while the trajectory matching connects this mixing to the tracked air parcels. From these inferred trace gas mixing ratios the mixing state of each air parcel is calculated for every day. This process is non-trivial for irregularly gridded data, as will be explored in Sect. 4.5.

Instead of considering the development of each air parcel for every day individually, we may represent the mixing rate of an air parcel by the slope of a linear fit through the timeseries of its mixing state. This slope represents the *local*, average mixing strength, which we refer to in the following as the *linear trend* of the air parcel. The linear trend is the linearized mixing rate of an air parcel and preserves the mixing strength and the direction of the mixing. These trends are only valid for the considered time period and do not imply anything beyond it. This is an adequate simplification, since we do not anticipate rapid, non-linear changes in the mixing state over the considered time period of two weeks. The dynamics are shown in Fig. 4.4:

The high/control mixing scenario is shown in the first and second row, respectively. The first and second column show the linear trend of the static/dynamic mixing state, respectively. The third column shows the linear trend of the Age-of-Air (AoA) tracer as a pure, unparametrized model quantity as a reference.

The linear trends show a generally consistent picture. There is little or no change in the mixing state in the tropics and at the poles, mostly due to the clipping of the mixing state (see Eq. 4.4). Linear trends, i.e. the linearized, average changing rate, can be calculated similarly to other quantities, such as the AoA, see Fig. 4.4c, f. Compared to the mixing state linear trends, the AoA linear trends indeed show only minor changes at the tropical pipe, with very little change in the extratropical stratosphere. There are distinct regions of high linear trends, i.e. increases in stratospheric character, located at the edges of the tropical pipe. Comparing to Fig. 4.3b these regions coincide with the filaments that can be visibly discerned, which indicates that anomalous spatial structure

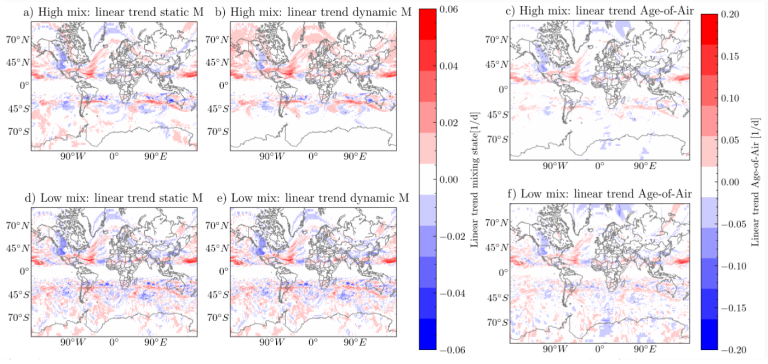

Figure 4.4: Visualization of the mixing dynamics for 1 February 2011. The first row shows values derived from the high mixing scenario, the second row shows values derived from the control mixing scenario. First column: linear trend of the static mixing states. Second column: linear trend of the dynamic mixing states. Third column: linear trend of Age–of–Air.

coincides with high dynamics. There are similarly distinct regions of high negative linear trends located in the same area. These regions are expected, as the strong increases necessitate strong decreases in mixing state due to mass conservation. There are differences between the dynamic and the static mixing state for the high mixing scenario, but essentially none for the control mixing scenario. The dynamic mixing state shows asymmetry between the hemispheres, suggesting that the northern extratropical stratosphere is slowly increasing in stratospheric character. This behaviour is not reflected in the AoA trend, and while hemispheric asymmetry is known to exist, the inconsistencies between the different scenarios suggest one of the aforementioned potential issues the mixing state may be subject to. A likely explanation is a drift in the stratospheric cluster (1) towards the tropical cluster (3), which yields consistent increases in mixing state. Besides this, the dynamic mixing state still accurately shows all other features consistently.

The differences in mixing strength between both mixing scenarios are reflected in the gradients of the linear trends. The control mixing scenario shows patch-like structures in the southern hemisphere where no filaments are visible. These patches are equally visible in Age–of–Air and can be considered real. The edges of the regions with rapid exchange are similarly less sharp. The meridional structure of the linear trends is rather consistent between configurations and can be summarized by their zonal means, which are shown in Fig. 4.5a: As expected, the tropical pipe shows very little change. There are two pronounced peaks of increasing M with minor negative sections in the extratropics at 30°N and 35°S, respectively. These peaks correspond to filaments

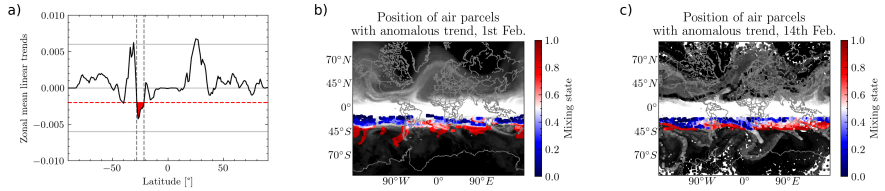

Figure 4.5: Panel a: Zonal means of the linear trends in the static mixing state for the high mixing scenario. A section of anomalous negative trend is indicated by the red dashed line. Panel b - c: Initial positions and final positions of the air parcels associated with the anomalous negative trend, respectively. The current mixing state is shown in colour-code in either panel.

in these latitudes, and consequently Rossby wave breaking. Towards higher latitudes the structure becomes noisy and fluctuating with little amplitude. There is one region at 25°S which stands out with a significant negative linear trend, i.e. an increase in tropospheric character. Inspecting the initial positions of the corresponding air parcels (marked by their mixing state in colour in Fig. 4.5b) shows rather unremarkable structure, with patches of high M parcels in the extratropical stratosphere and patches inside the tropical pipe. However, these air parcels eventually converge and remain close to the tropical pipe (Fig. 4.5c) where they undergo mixing at the boundary. The linear trend indicates turbulent, small-scale mixing outside of any visible filament. This indicates that turbulence may also occur without creating clear, directly visible indicators.

The linear trends can also be considered with respect to the initial mixing state, as shown in Fig. 4.6:

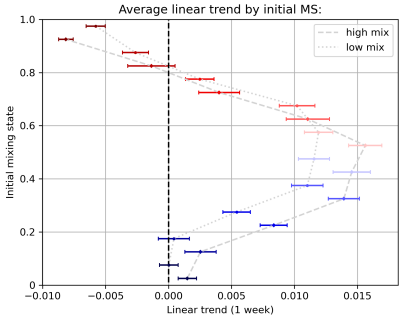

Figure 4.6: Average linear trend divided by initial M. Shown are the high mixing scenario (dashed line) and the control mixing scenario (dotted line) for February 2011. The error bars indicate the standard deviation of the average linear trend of each ensemble. The initial mixing state is also indicated in colour-code.

The linear trends in each initial M bracket (0 – 1) are averaged, with the standard deviation indicated as error bars in Fig. 4.6. There is a clear relationship between the initial mixing state and

Initial M	$0 \rightarrow 0.1$	$0.1 \rightarrow 0.2$	$0.2 \rightarrow 0.3$	$0.3 \rightarrow 0.4$	$0.4 \rightarrow 0.5$
	$0.5 \rightarrow 0.6$	$0.6 \rightarrow 0.7$	$0.7 \rightarrow 0.8$	$0.8 \rightarrow 0.9$	$0.9 \rightarrow 1.0$
Δt next bracket (highmix) [d]	67.0	39.4	12.0	7.2	6.9
	6.4	9.0	24.9	-73.2	-12.3
Δt next bracket (lowmix) [d]	4490.9	230.8	18.5	9.1	8.7
	8.4	9.8	40.3	-38.1	-17.3

Table 4.1: Average transition times between mixing states. Shown are the times in days an air parcel takes on global average to undergo a change in mixing state of 0.1. Positive numbers indicate increases, negative numbers indicate decreases. These estimates are uncertain particularly for a mixing state of zero, as very little change is anticipated.

the linear trend of an air parcel, and consequently between the current mixing state and the current mixing strength. Air parcels with a pure tropospheric/stratospheric state (M < 0.2 and M > 0.8) show very slow changes. Their estimated transition rates are therefore less reliable. The transition states (M > 0.2 and M < 0.8) in contrast show increasingly fast dynamics, with a maximum for $M = 0.5$, since partially mixed air parcels are currently undergoing rapid mixing. This behaviour is reflected in the different mixing scenarios, with the high mixing scenario showing increased mixing strength compared to the control mixing scenario. The time needed to transition into the next bracket of mixing state is summarized in Tab. 4.1. Fig. 4.6 also conveys information about the timescales of potential filament dissolution. Assuming a filament with initial M of 0.3, it would take 14.1 and 17.8 days to reach an M of 0.5 depending on the mixing scenario, and an additional 15.4 and 18.2 days to become visually indistinguishable from the stratospheric background with an M of 0.7. An argument can be made that spatial coherence may already be lost when an M of 0.5 is reached. An observation period of around two weeks therefore is sufficient to observe most of the 'lifetime' of a filament. A partially dissolved filament remains chemically distinguishable up to around one month. The statistic also allows one to directly connect the abstract mixing parameter in CLaMS to the underlying mixing processes. If such a statistic could be derived from actual observations – as will be investigated in the following – the real mixing strengths could be emulated by fine-tuning the CLaMS parameters until a matching statistic could be derived.

4.3 Filament detection

We have seen in the previous section that filaments are visible in their anomalously increased dynamics. They appear as dedicated, isolated regions where air parcels undergo more rapid mixing, indicated by increased linear trends. The diagnosis of the mixing requires trajectory calculations

and matching of these trajectories, which takes computational effort and may not always be trivially possible, as seen in a later section. It would be advantageous to determine filaments by their anomalous spatial structure. The human eye can easily identify the filaments in Fig. 4.3b, but such impressions are subjective. To get a more objective definition of a filament one may utilize the mixing state M. The process is illustrated in the following for the high mixing scenario: The filaments of interest are tropical in nature (i.e. $M < 0.25$), but lie outside of the tropical pipe. There exist many well–established definitions of the tropical pipe, such as through tropical upwelling (Plumb, 1996; Birner and Boenisch, 2011) or tracer distributions (e.g. Stiller et al., 2017). While these definitions would be suitable to use here, the aim of this analysis is to utilize a definition which is maximally consistent with the mixing state used for the rest of the analysis. Further, we strive to use predominantly quantities which can be directly observed by a satellite instrument and do not require climatologies. The boundary between tropical pipe and extratropical stratosphere can be determined from the zonal mean of the mixing state, which is shown in Fig. 4.7a. The boundary of the tropical pipe is the latitude where the zonal mean M reaches a threshold of 0.35. The resulting latitudes are marked by the dashed red lines in both Fig. 4.7a, b. While such boundaries are good for a general overview, they do not take the natural variability of the tropical pipe into account. A more precise boundary can be determined by taking the mean over smaller longitude segments indicated in Fig. 4.7a, b as black markers. A high-degree polynomial is fitted through these positions with Ridge regularization and cyclic boundary conditions, shown as green curves in Fig. 4.7b. Based on the polynomial-fit boundary we determine filaments as *a coherent structure of air parcels with mixing state lower than 0.425 outside of the tropical pipe*. Naturally every filament therefore is bounded by a thin band with a mixing state of 0.5 . The threshold mixing state of 0.425 is chosen such that individual filaments remain separable.

Once areas are defined, the Density-Based Spatial Clustering of Applications with Noise (algorithm) (DBSCAN) clustering algorithm (see App. A.10) is ideally suited to detect these coherent structures. Particularly, the algorithm allows one to determine filaments on irregularly gridded data, which will be relevant in Sect. 4.5 for the synthetic observations. The spacing parameter ϵ is chosen as 1.5, which allows for full diagonal connectivity, and the minimum cluster size is set to 1 to only get ϵ-connectivity. Fig. 4.7c shows the detected filaments as coloured dots. Comparing the position of the spatially detected filaments to the dynamics in Fig. 4.4a, we see that the positions coincide with the regions of strongly positive linear trend. Only the light green filament in the northern hemisphere does not show increased dynamics (stronger changes over time), nor

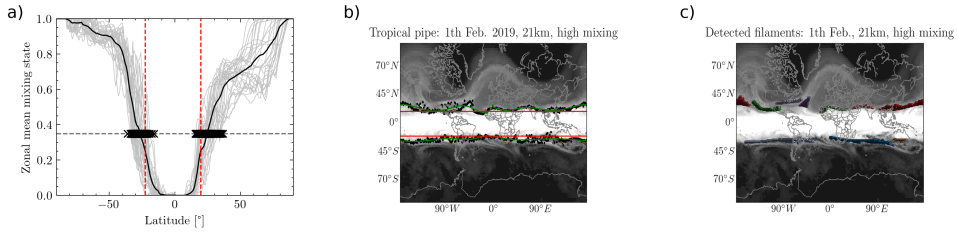

Figure 4.7: Definition of the tropical pipe and the filaments. Panel a): Zonal mean mixing states. The black lines show the zonal means over all longitudes (thick line) and longitude slices (thin lines). The latitudes where the threshold 0.35 is met are indicated by the red vertical lines and the black markers, respectively. Panel b): Spatial distribution of the boundary latitudes. The zonally resolved latitudes are fitted by a high–degree polynomial (green). M is indicated in the background. Panel c): Resulting filaments defined as tropical air outside the tropical pipe. M is indicated in the background.

do parts of the elongated filaments (light/deep blue and deep red). A similar analysis performed on the control mixing scenario yields basically identical filaments. Visually, many regions may not appear very 'filament–like' to the human eye in this snapshot. To underscore the point that these regions indeed behave like filaments, the trajectories of these points are shown in App. A.6 in full. Using this algorithm we can accurately determine filaments as spatial anomalies which coincide with anomalously increased dynamics.

4.4 Geometrical analysis

For the observed filament in Chap. 3, the strongest mixing is observed near the edges of the filament, where the different air types come into direct contact, particularly at the 'tip'. Is this a consistent feature? To characterize the shape of a filament, the following two geometrical properties are used, which are illustrated in Fig. 4.8: We consider a set of points which form a banana, see Fig. 4.8a. The outline of this set of points can be described by its convex hull (Fig. 4.8b, orange line). It acts like a rubber band which contracts around these points, but fails to describe the outline of the points where their shape is not convex. A variation of the convex hull is the *alpha shape* (Fig. 4.8b, red line), which moves the contour closer to the edges of the points. The alpha shape is sometimes referred to as the concave hull, since it allows one to capture concave shapes as well. The alpha shape is not unique and depends on its alpha parameter, which governs how 'tight' the rubber band is drawn around the shape. Any shape can be broken down into a 1–D representation, its *morphological skeleton* (see Fig. 4.8b, green line). The algorithm by Zhang and Suen (1984) is a prominent way to create such a skeleton. From such a skeleton the longest

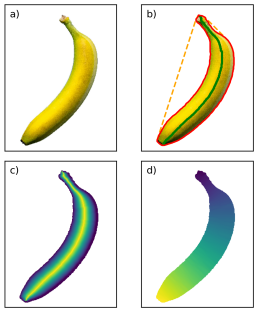

Figure 4.8: Visualization of the geometrical properties of sets of points on the example of a banana. Panel a: All points making up the total banana shape. Panel b: Different geometrical properties. Shown are: convex hull (orange, dashed line), alpha shape (red line) and the skeleton (green line). Panel c: Relative distance (does a point lie close to the shape's edge?). Panel d: Relative length (does a point lie close to the start of the shape?).

path (from 'start' to 'end') can be determined using Dijkstra's algorithm (Dijkstra, 1959). Using the alpha shape and the skeleton for every filament (at any time step) we define the *relative distance* d_r as:

$$d_r := \frac{||x - x_\alpha||}{||x_\alpha - x_s||} \tag{4.5}$$

where x_α is the orthogonal projection of a point x to the alpha shape and x_s is the orthogonal projection of x to the skeleton. The relative distance is shown in Fig. 4.8c. The relative distance describes whether x lies close to the edge of the filament ($d_r \to 1$) or whether x lies close to the skeleton in the centre ($d_r \to 0$). Similarly, given a skeleton $c_s(t)$, we define the *relative length* l_r for a point x as:

$$l_r := \underset{t\in[0,1]}{\arg\min} \big\| x - c_s(t) \big\|. \tag{4.6}$$

i.e. the corresponding parameter t of its orthogonal projection. The relative length is shown in Fig. 4.8d. The relative length describes whether a point x lies close to the start ($l_r \to 0$) or the end ($l_r \to 1$) of a filament.

Since filaments change their shape over time, these quantities need to be recalculated at every time step. If the mixing strength is stronger at the edges of a filament, there should be a positive correlation between M and d_r, and if the mixing strength is strongest at the tip, there should be a positive correlation between M and l_r. A physical argument can be made for both cases: We would anticipate increased mixing at the edges, since here the induced turbulence is strongest. Since the tip of a filament is formed first, it would have the largest exposure to turbulence, and thus would already be more strongly dissolved compared to the rest of the filament. The correlations

between M and d_r, l_r of all points in each filament is summarized in Fig. 4.9:

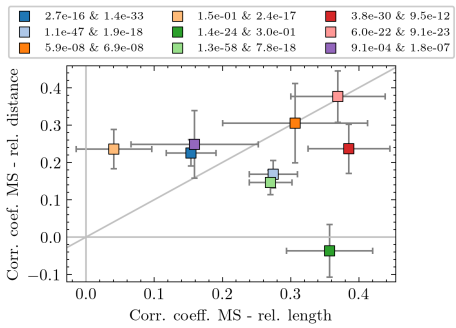

Figure 4.9: Pearson correlation coefficient between mixing state M and geometrical properties of the filaments. The x–axis shows correlation to the position along the filament, the vertical direction shows correlation to the position with respect to the filaments edges. The error bars indicate the limits forming a 95% confidence interval. The p–values of the different regression of d_r, l_r corresponding to a 95% confidence interval are indicated in the legend.

Shown are the Pearson correlation coefficients between the mixing state and the respective geometrical property. The error bars indicate the 95% confidence intervals for the correlation coefficients and are generally asymmetric. The p–value of each correlation is indicated in the legend above. The filament is indicated in the same colour as in Fig. 4.7c. There is a significant correlation between M and *both* d_r and l_r. As expected, the mixing strength – directly correlated to the initial value of M, see Fig. 4.6 – is indeed strongest near the edges of the filament. The tip of each filament also undergoes stronger mixing, such that both hypotheses are supported. Notably, these correlations hold not only for high d_r, l_r, but across all values (not shown). An outlier is the deep green filament (Fig. 4.7c, around 135°W), whose mixing strength correlates strongly with its relative length, but shows almost no correlation to its relative distance. This outlier currently lacks a clear physical explanation.

4.5 Simulated satellite observations

The data considered so far had (1) continuous coverage and (2) regular frequency. The synthetic CAIRT data from Fig. 4.1 reach global coverage during the course of a day and have gaps between the orbit tracks. For these CAIRT observations, the presented methodology therefore needs to be adjusted. The results of this adapted methodology will then be compared to the ideal model results.

4.5.1 Adaptations for satellite data

The calculation of the mixing state does not require any adjustments, since satellite observations may be spatially sparse (see Fig. 4.1b), but are dense in phase space due to the homogeneous coverage. The spatial gaps in Fig. 4.1b are not gaps in phase space. The following analysis steps, however, do require such adjustments:

Trajectory matching with observations

The same trajectories as before are used to match with the synthetic satellite observations. Matching a trajectory with an observation requires both *spatial proximity* and *temporal coincidence*. For the regularly gridded data this is mostly trivial, as for each time step there exist data points everywhere. This is much more complicated for the observations. Formally, given a trajectory $\boldsymbol{x}_{traj}(t)$ and a set of observations $\boldsymbol{x}_{obs}(t)$, the task is to find all times $t^{'}$ for which:

$$||\boldsymbol{x}_{obs}(t_1^{'}) - \boldsymbol{x}_{traj}(t_2^{'})|| < \epsilon \tag{4.7}$$

where ϵ is some matching distance, $\|\cdot\|$ denotes some metric and with $t_1^{'}, t_2^{'} \in [t^{'} - \frac{\delta}{2}, t^{'} + \frac{\delta}{2}]$, with δ denoting a small time period. The issue remaining is to find *all* observations for *all* times which satisfy Eq. 4.7 given some parameters ϵ, δ, i.e. those observations which are at most ϵ away from a trajectory within some time window δ. We have derived an algorithm which can efficiently determine all matches within specified parameters ϵ, δ, which is described in App. A.2. For the rest of this chapter we use the matching parameters $\epsilon = 0.5^{\circ}$ and $\delta = 600\,s$. The spatial matching is performed in longitude and latitude coordinates, which results in a maximum real distance between two matched points that varies between 112 km near the poles and 158 km in the tropics. Since the goal is to find out if points lie close to the trajectories rather than to

determine their distance exactly, this varying distance is an acceptable compromise and does not significantly affect the analysis.

Given a set of (irregularly spaced) matches, the mixing state M can be derived as described above. The linear trends likewise can be calculated from irregularly spaced matches.

Filament detection

To determine filaments as spatially anomalous structures of low M, first the tropical pipe needs to be determined. Assuming relatively slowly changing contours of the tropical pipe, the observations can be accumulated (i.e. over a period of 24 h) to reduce the number of data gaps, see Fig. 4.1b. While data gaps persist, the amount of data is sufficient to determine a tropical pipe defined as before similar to Fig. 4.7a, b. Due to the measuring geometry of the satellite instrument, the data gaps are largest in the tropics, which negatively affects the contour detection and introduces some uncertainty. More challenging is the detection of filamentary structures in incoherent data. The data gaps tend to bisect the filamentary structures and obscure parts of them such that they no longer appear coherent. Furthermore, the filamentary structures can move over time. Since the observations are taken at different times, it is likely that different observations show the same trajectories only at different times. The eventual dissolving of the filaments via mixing also leads to their disappearance, such that it needs to be determined when a filament is no longer visible. We have derived an algorithm which is capable of detecting filaments in the spatially and temporally irregular observations. The core steps of the algorithm are illustrated in Fig. 4.10 and are as follows:

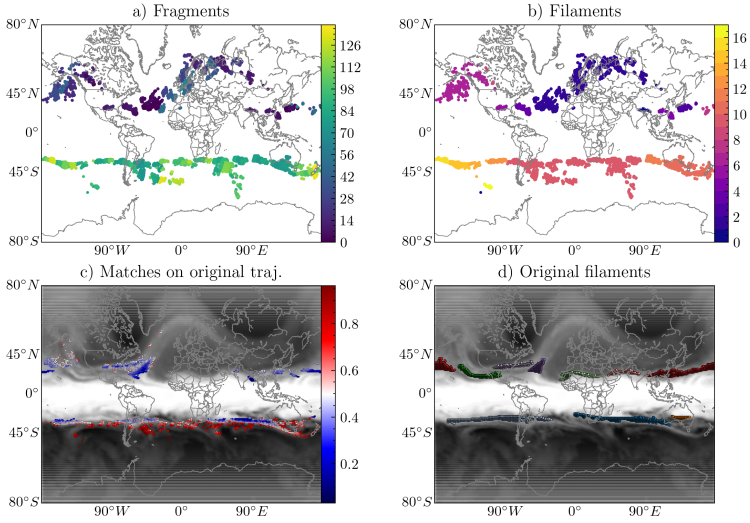

Figure 4.10: The different steps of the filament detection. For a description of the algorithm see the text below. Panel a: Spatially coherent structures (fragments) found among all satellite data. The colour–code indicates the ID of each of the $\sim$ 130 fragments. Panel b: Spatially and temporally coherent structures, formed from the fragments, bridge data gaps. The colour–code indicates the ID of these fragments. Panel c: Initial positions and initial mixing states (shown in colour-code) of all original trajectories matched with the fragment points. Panel d: Positions of the originally detected filaments as a reference.

Filament ID	1	2	3	4	5	6	7	8	9
Fraction of seen points [%]	62.2	69.7	6.7	32.7	23.9	89.0	18.0	19.8	22.4

Table 4.2: Percentages of trajectories belonging to the originally detected filaments that were matched with points identified as filaments in the observational data.

- First, all small–scale, spatially coherent structures of low M (< 0.425) outside the tropical pipe boundary are determined using the DBSCAN algorithm with a small spacing parameter of 1. We refer to these structures in the following as *fragments*, which are shown in Fig. 4.10a. The minimum cluster size needs to be low (e.g. 25 points) to capture sufficiently many fragments, but should be large enough to filter out isolated points. We filter all fragments which do not contain at least one point with sufficiently low M (i.e. 0.3) to focus on fresh filaments rather than partially dissolved ones.

- Second, the fragments are *hierarchically* merged into larger filaments using a modified version of the DBSCAN algorithm. The data gaps necessitate increasing the spacing parameter to bridge these gaps. A large spacing parameter, however, not only allows for connectivity between fragments which are separated by data gaps, but also between fragments which are separated by high M. One may recall that filaments are actually distinguished by these regions of high M. Our modified DBSCAN algorithm checks if two fragments are separated by high–M regions and then suppresses connectivity. The modified DBSCAN algorithm is described in App. A.3. The fragments are merged in temporal order to respect causality. A fragment can only be merged with a single existing filament, and only the few most recent additions of that filament are used to merge with new fragments. We enforce a leniency time after which a filament is considered completed if no new fragments have been merged with it. The resulting filaments are shown in Fig. 4.10b.
- Since these filaments consist of observations at different times (and potentially duplicate observations of the same trajectories), each filament can be reduced to the set of the original trajectories which match with them. Naturally this may remove those parts of the filament which were never matched with any of the trajectories. The fraction of each filament which was seen in the observations varies strongly between almost 90% and a mere 6%. The fractions are shown in Tab. 4.2. The enumeration corresponds to the colour–code in Fig. 4.7c. The starting points of these trajectories and their initial M are shown in Fig. 4.10c.

Compared to the filament detection on regularly gridded data, the air parcels with negative linear trend (see Fig. 4.5b, c) are also recognized as filaments, since these air parcels eventually appear tropical. Comparing the two detections (Fig. 4.10c and d), despite the incoherent coverage of the observations most of the filaments were detected successfully.

If we consider every trajectory which was *revisited*, i.e. was matched with an observation for at least two different times, the dissipation rates of the atmosphere can be derived similarly to Fig. 4.6, since the linear trend does not rely on equally spaced time intervals. To derive reliable estimates for the linear trends, only those trajectories are considered which were matched at least five times. The linear trends for the high mixing scenario are shown in Fig. 4.11a. The dissipation rates are essentially identical to Fig. 4.6, with more structure of negative linear trends in the southern extratropical stratosphere. The linear trends resolved by initial M similarly to Fig. 4.6 are shown in Fig. 4.11b.

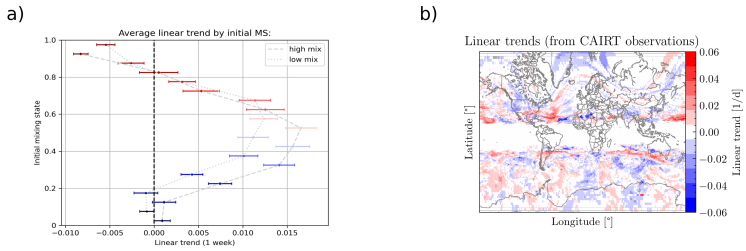

Figure 4.11: Panel a: Linear trends derived from matches on satellite observations. Panel b: Statistical dissipation rates of trajectories sorted by initial mixing state similarly to Fig. 4.6. The dissipation rates are represented by the average linear trend of each group. Positive numbers indicate an increase in mixing state, negative ones indicate decreases. Zero (no change) is indicated as a black, dashed line. Trends are shown for the high mixing scenario (dashed lines) and for the control mixing scenario (dotted lines).

Initial M	0→0.1	0.1→0.2	0.2→0.3	0.3→0.4	0.4→0.5
	0.5→0.6	0.6→0.7	0.7→0.8	0.8→0.9	0.9→1.0
Δt next bracket (highmix) [d]	110.6	89.1	13.4	7.1	6.4
	6.0	8.0	18.9	192.8	-12.1
Δt next bracket (lowmix) [d]	-120.5	-107.7	23.5	9.9	9.0
	8.1	8.8	31.7	-38.5	-18.4

Table 4.3: Transition times of mixing state.

Despite the sparse spatial coverage and the irregular spacing between matches, the simulated satellite observations yield basically the same dynamics. The transition rates would suggest around 13.5 and 18.9 days to reach an M of 0.5 depending on the mixing scenario and an additional 12.4 and 16.7 days to become visually indistinguishable from the background, in good agreement with the results from the regularly gridded data. Satellite observations are therefore capable of studying the dynamics and the mixing processes in the middle world. But how likely is it to actually revisit a particular filament?

4.5.2 Revisit times

Generally, trajectories depend on local winds. A satellite, however, follows a predetermined orbit and cannot be adjusted to accommodate a particular region of interest. Whether a particular trajectory is revisited within a specified period is, therefore, up to chance. The statistic of such matches must be derived numerically and implicitly depends on the individual physical situation. Foremost, it depends on the *swath width* of the instrument, i.e. its horizontal coverage, which is

therefore an essential parameter when designing a mission for the purpose of studying mixing. Such a statistic is visualized in Fig. 4.12:

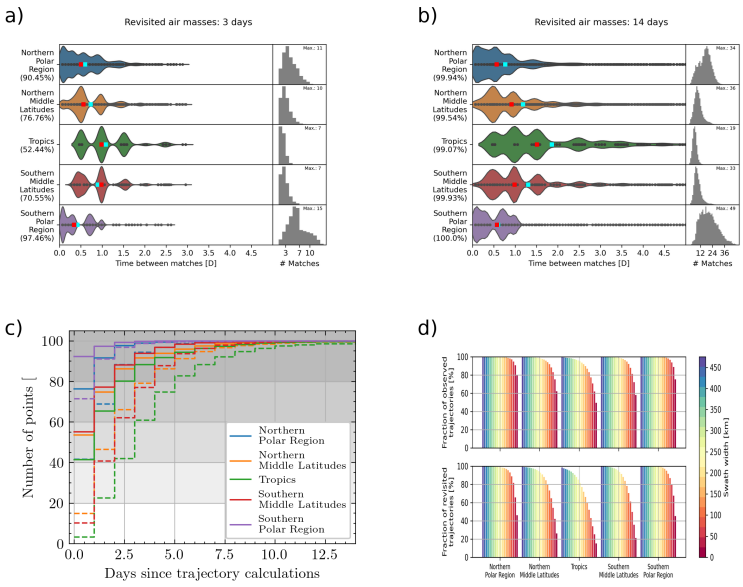

Figure 4.12: Summary of matching statistics. Panels a, b: Distribution of times between matches (revisit time) over a total period of three days and 14 days, respectively. Mean and median are indicated as red and cyan markers, respectively. Distributions are split into five climatological regions (cf. Tab. 4.4). The percentage of revisited points in each region is shown on the left–hand side. The numbers of matches along a single trajectory are summarized by histograms on the right–hand side. Panel c: Percentage of seen (solid lines) and revisited (dashed lines) air parcels over the span of 14 days. The climatological regions are shown in colour–code. Panel d: Effect of swath width on the number of revisited air parcels. A period of 72 hours is considered. The track width is shown in colour–code.

Climatological zone	Northern polar	Northern mid-lat	Tropics	Southern mid-lat	Southern polar
Latitude range	82.6°N–66.6°N	66.6°N–23.4°N	23.4°N–23.4°S	23.4°S–66.6°S	66.6°S–82.6°S

Table 4.4: Definition of the climatological zones used in Fig. 4.12. The zonal boundaries towards the poles follow the highest latitudes retrievable by CAIRT.

Panels a, b show the revisit times, i.e. the time between two consecutive matches. For this study we assume that we have three days to find and revisit a short-lived event and 14 days to find a long-lived event before it dissolves. All observations/matches in this time frame are counted. The total likelihood hence increases with the lifetime. The revisit statistics for short–lived (panel a) and long-lived events (panel b) are binned into the climatological zones (cf. Tab. 4.4) according to the latitude of the initial positions. The fractions of the observed air parcels that were also revisited

are indicated in percentages on the left–hand side. How many times an air parcel was observed is indicated in the plot on the right–hand side. The fraction of observed/revisited air parcels with respect to the observation time is shown in Fig. 4.12c. The revisit times are roughly periodic in multiples of 12 h due to the measurement geometry of an orbiting satellite. For short–lived events, the fraction of revisited air parcels depends on the latitude and lies between 50% in the tropics and 97% in the polar regions. The bulk of revisits happens within the first 48 h, with average revisit times of around 12 h in the polar regions and 36 h in the tropics. After three days, between 70 and 90% of all original air parcels were observed, with between 2 and 6 revisits on average. For a long-lived event the revisit statistics change only marginally, with average revisit times shifting slightly towards longer times and revisits after longer periods (> 3 days) becoming more prevalent, particularly in the tropics. The coverage of matched air parcels approaches 100% outside of the tropics after 8 days, with 100% of revisited air parcels likely reached after 10 days. On average an air parcel is revisited 12 to 24 times. The simulated CAIRT instrument has a swath width of around 475 km consisting of 20 tracks. The effect of the swath width on the observed/revisited air parcels is illustrated in Fig. 4.12d for a short-lived event. Given the matching parameters used, a global matching coverage of 50% requires a swath width of at least 350 km, which translates to only 18% global revisit coverage. An effectively single–track instrument, i.e. an instrument with poor horizontal resolution, would yield global matching and revisiting coverage of 20% and < 3%, respectively. Good horizontal coverage is therefore key for measuring mixing processes. Of course, vertical resolution is equally important for resolving trajectories. We have found that while satellite observations are spatially and temporally sparse – and revisiting air parcels is difficult – they yield similar dynamics and dependencies on M as the regularly gridded data. With that in mind, one may consider whether it is possible to learn the general dynamics of the atmosphere from these observations alone, despite their sparsity. Neural networks would be ideal candidates, as they do not require any prior parametrization of the mixing processes and could learn them from the satellite observations instead.

4.6 Prediction of filament evolution

Since every change in the composition of an air parcel due to mixing is governed by the underlying mixing processes, information about these processes is in turn contained within the timeseries of an air parcel and thus can be captured by a suitable neural network. Given a time series $\mathbf{s}(t)$ of some values s_i known at m discrete times $\{t_0, t_1, ..t_m\}$, the task is to predict some s_k at some future time t_k, with $k > m$. Usually the time intervals $\Delta t_n := t_{n+1} - t_n$ are regularly spaced, as is the case for the trajectory matches of the regularly gridded data, such that the time dependency of $\mathbf{s}(t)$ can be covered implicitly. Most common architectures rely on this assumption and are therefore well suited out–of–the–box for the regular trajectories. For the irregular trajectories certain adaptations must be applied. The simplest way to address irregularly spaced time steps is to include the time steps as additional features to the network, which is the choice for this section. For future prediction models two sets of parameters are essential: The first set covers how many time steps are provided as model input, i.e. the *lookback window*, and how many time steps are predicted, i.e. the *forecast horizon*. In this thesis, we simplify these formulations to past and future, respectively, which should allow for more fluid, more intuitive reading. Models with a future of one are referred to as single–cast prediction models, whereas models with a larger future are referred to as multiple–cast prediction models. The size of the past should be chosen with respect to the inertia of the dynamical system as well as its continuity. A slowly developing system may not necessarily benefit from a long past along which changes are minor. Likewise, if no sudden changes are to be expected, shorter pasts also suffice. In the case of atmospheric filaments, whose lifetimes we estimated to be around 14 days, long pasts are not necessary, as a longer past generally does not guarantee better predictions. As a general rule of thumb, the future should not exceed the past to keep a balance between known information and the unknown prediction.

The second set covers what values s_i are included in the past and which values s_j are actually predicted. Typically s_j are included in s_i, i.e. past values of s_i are used to predict future values of s_i, but this is not required. The s_i with $i \neq j$ are called *covariates*, because they provide information through their covariances to the s_j. A model for which $i = j$ is referred to as an *autoregressive* model, although the term is often understood to mean *linear* autoregressive models. In this study, the term is not limited to linear models. For irregularly spaced time series the time intervals Δt are natural covariates to the other s_j. The goal of the following neural networks is to learn the underlying mixing processes and predict the correct dissipation rates.

The basis for the neural networks are the matches between trajectories and the available back-

ground data, resulting in regularly spaced time series for the regularly gridded data and irregularly spaced time series for the matches on the synthetic observations. The s_j of these time series then are normalized to equalize their numerical weight and divided into suitable sequences. Different scales along different trajectories suggest per-trajectory normalization, which allows one to better capture dissipation rates of significantly differing magnitude. For regularly spaced time series this form of normalization is applied. For irregularly spaced time series this form of normalization is not feasible because of potentially strongly irregular matches, which can skew the normalization. Instead, the time series are normalized globally, which equalizes the numerical weight of all s_j.

The time series are divided into sequences of length past+future, of which the first past values are used as inputs and the last future values are used as prediction targets. In order to validate the models against data which was not seen during training, only the first week's worth of time series is used to create the training sequences, leaving the last week for validation. Dividing irregularly spaced time series into sequences is consequently more difficult, as not all time series may contain at least past+future entries. Matches with large revisit times are also to be avoided and may be filtered out. To generate the necessary volume of sequences required for training, the training domain may need to be extended beyond the first week for some choice of past and future.

For predicting dissipation rates we use temporal CNN, see Sect. A.4), i.e. a neural network architecture which determines temporal development through the structure of a time series. In the limit of very short sequences, these models become similar to Multi-Layer Perceptron (neural network) (MLP), and are thus suitable for a wide range of sequence lengths. The architecture of such a CNN and how its design relates to the parameters is illustrated in Fig. 4.13: Panel a shows a hypothetical architecture for regularly gridded data; panel b shows the same architecture adapted to work with irregularly spaced data. This architecture predicts a single quantity s_0 **two** steps into the future from the past **three** values of s_0 and **one** additional covariate s_1. Therefore, its input layer consists of $3 \cdot (1 + 1) = 6$ neurons. The inputs are flattened into a single 1–D vector, which is passed to a series of convolution layers with 32 channels each. (A more detailed overview of convolution layers is given in App. A.4.) These convolution layers use kernels of size 3 and increasing dilation rates to learn progressively larger–scale structures. The convolution layers are regularized via L_2–regularization, which suppresses large weights. This processed input is passed to an averaging layer, which distills the derived information, followed by a linear, transformative layer to apply the correct numerical scale to the prediction. Finally, a reshaping

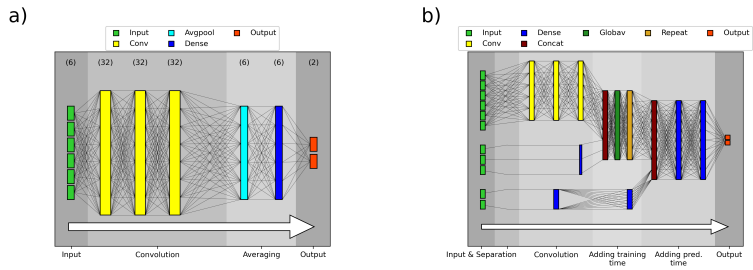

Figure 4.13: Architectures of different CNNs. Panel a: predictive CNN for regularly gridded data. The model input assumes 3 values into the past including 1 covariate for a total 6-D input vector. The input is passed into convolution layers, which are L_2–regularized. The final convolution is averaged across its channels and reshaped into the output shape (here: 2 values into the future). Panel b: The same model as in panel a, but adjusted to accept irregularly spaced time series. The input is split into the s_i values (upper input), which are passed through L_2–regularized convolution layers, and into the input time steps (middle input) and the prediction time steps (lower input). Input time steps are adequately merged back after the convolution, after which the prediction time steps are added in as well. The recombined and processed inputs are then passed through an MLP-like section and returned as the output shape (here: 2 values into the future).

layer returns the correct shape of the data, here 2 time steps with 1 variable channel each.

The irregular variant differs from the regular CNN by including both the time steps of the input and the time steps of the prediction, adding three additional neurons (one per past) for the input and two additional neurons (one per future) for the prediction to the input layer. This is necessary, as the prediction times can no longer be assumed to be equal across sequences. The irregular variant therefore requires $6 + 3 + 2 = 11$ input parameters. In Fig. 4.13b, the input layer is shown split into these types of inputs. Since two identical time steps may correspond to two completely different changes in values s_i, the time steps of the input are separated from the sequences of values s_i. The values are passed through a set of L_2–regularized convolution layers and are merged afterwards with the input time steps. Since for irregular time steps the times at which the predictions are made are not determined by default, these time steps need to be provided to the model as well. The prediction time steps are merged with the convolved values and then passed through a set of dense layers. These layers determine the mapping of the final prediction between the temporal structure of the values at the given input times and the provided prediction times.
Since the mixing state M is artificially clipped for values beyond either tropical/stratospheric cluster, it may introduce issues in future predictions. Instead, the model AoA is used as the primary value to predict. The AoA acts as a perfectly inert tracer and ultimately conveys the same information about the dynamics. Since these CNNs should learn the correct dissipation rates, they should be capable of accurately predicting future s_i, which given the slow dynamics, is likely easy.

However, the dissipation rates, for example derived as the linear trend of a trajectory, require consistently accurate predictions. A model which accurately predicts future s_i, but has a bias towards over– or underestimation cannot accurately predict these dissipation rates. These linear trends further require multiple predictions, such that the prediction horizon of these models must be investigated.

In short, we investigate the predictive capabilities of multiple configurations regarding the following questions:

- What influence do the choice of *past* and *future* have on the predictions?
- Can covariates be used to increase performance?
- Do irregularly spaced time series have different performance?
- How far into the future can we make reliable predictions?

To answer the first question, different configurations are considered. First, a set of single prediction models (with a future of one) is considered, consisting of differently sized pasts (minimal past (1) with or without covariates) and adapted to either regularly or irregularly spaced time series. The performance of each model is summarized in their Mean-Absolute-Error (MAE)s ,see Eq. 2.29, in Fig. 4.14a: A suitable baseline to compare the performance of the models against is persistence, which assumes that no changes occur between time steps, since the dissipation rates are fairly slow. Since all models are regularized, a baseline which utilizes perfect knowledge of the atmosphere is unrealistic. Instead, Gaussian noise between zero and five percent is added to the model values, resulting in the performance MAE values indicated as the gray background in Fig. 4.14a–c. The models with past and future of one trained on regularly spaced time series (Fig. 4.14a) achieve better performance (lower MAE) than a persistence with perfect knowledge (darkest shade). The equivalent models trained on irregularly spaced time series show comparatively lower performance, but still exceed a baseline of 1% noise. Increasing the past value to three (orange) or five (green) does not consistently improve performance, but rather shows a slight negative trend. Models trained on irregularly spaced time series also tend to perform slightly worse. However, all models achieve better performance than a 1% noise baseline, i.e. assuming no changes if the original state is known within 1% accuracy. The predicted value deviates from the true model value on average by around one month. Including covariates (light colours) can

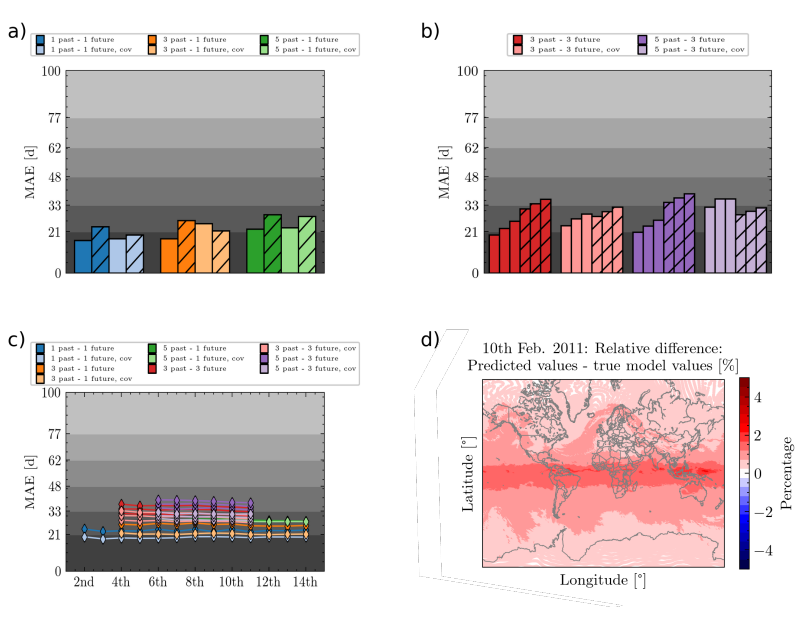

Figure 4.14: Performance of different CNN architectures. Panel a: Performance of the single prediction models. Shown are models with past values of 1, 3 and 5 in blue, orange and green, respectively. Models which use irregularly spaced time series are shown hatched, models considering covariates are shown in light colours. Persistence is shown in the background for different percentages of noise. Panel b: Performance of multiple prediction models. Shown are models with a future value of three and past values of 3, 5 in red and violet, respectively. Irregularly spaced time series and covariates are indicated similarly to panel a. Persistence is shown in the background for different percentages of noise. Performance is derived for each of the three predicted values individually. Panel c: Performance of models trained on regularly spaced time steps resolved by the predicted day. Performance is calculated for those days where true model values are available for each predicted value. Predictions beyond the 7th lie outside of the training domain. Persistence is shown in the background for different percentages of noise. In panels a – c, the performance is indicated by the mean absolute error (MAE). Panel d: Relative difference between predicted and true model values expressed in percentages exemplarily for 10 February 2011.

lead to improvements, but does not guarantee improvements.

Since multiple predictions are required to estimate dissipation rates, a second set of models is trained following the same setup as the single prediction models, but with futures of three and five, respectively. These models simultaneously predict multiple values into the future, and therefore require evaluations separated by the predicted time step. The multiple prediction models (Fig. 4.14b) show an interesting detail: since these models predict multiple time steps into the future at once, they consistently perform worse than their single prediction model counterparts, with the benefit of providing predictions further into the future. Their performance decreases with the number of days predicted into the future, with models trained on irregularly spaced time series again performing worse. An outlier is the model with a past of five trained on regularly spaced time

intervals including covariates, as this model performs significantly worse than its peers. While all models perform better than a baseline of 2% noise, corresponding to average deviations from the model values of around 1.5 months, it becomes clear that these kinds of models are ultimately limited in their prediction horizon. Predicting far into the future ultimately comes at the price of decreasing performance. This is a general issue, as is illustrated in Fig. 4.14c, which shows the performance of the models trained on regularly spaced time series resolved by the predicted day. Each model has consistent performance regardless of which day is predicted, i.e. the performance of predicting in-domain is equal to the performance of predicting out–of–domain. Depending on the goal, such predictions are still very accurate. Fig. 4.14d shows the spatial distribution of the deviation between predicted and model values as percentages of the model values for 10th February 2011 for a model with past and future of one. The Age-of-Air is systematically overestimated in the tropics, with accuracy increasing towards the poles. All predictions are accurate within 5% of the true model value, which is as mentioned before more accurate than even very precise baselines.

CNNs are well suited for accurately predicting Age–of–Air from time series. Particularly, they retain good performance when relying on irregularly spaced time series as could be derived from matching trajectories on satellite observations. The issue for deriving dissipation rates, however, is the need for many predictions into the future. More future predictions would provide a more stable estimate for the corresponding dissipation rate, but on the other hand will naturally yield lower prediction accuracy. Multiple prediction models therefore do not scale well in this regard. One may wonder: Is there a different architecture which could mitigate this limitation and still produce accurate dissipation rates?

4.6.1 Recursive prediction models

Instead of predicting multiple steps into the future at once, one may predict one step into the future and then use this prediction as a new input to preidct the next step. A model which operates this way is referred to as a *recursive network*. A simple recursive network requires a prediction output which is suitable to act as a new input, i.e. requires $s_i = s_j$. Thus, covariates cannot be utilized in recursive predictions, but can be incorporated into the training process. A form of recursive model which is capable of handling covariates is the *Non-linear autoregressive exogenous* (NARX) model. Details on this particular architecture can be found in App. A.5. Differing past and future values can be addressed by introducing a backwards shift when iterating through the data.

In the following, NARX models are considered if for a particular configuration the covariates led to improvements in performance. Otherwise, the regular models are treated as recursive models. Beginning with an initial state of the first one or three days of model values, predictions for these initial states are performed by the different models and their predictions are treated as the new initial state for the next prediction until the final day is predicted. Each prediction is compared to the corresponding model value. The performances for recursive predictions of different recursive models are shown in Fig. 4.15:

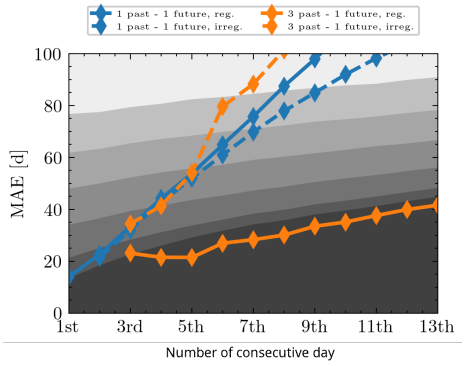

Figure 4.15: Performance of the recursive models. Based on the initial state on 1 February, each following day is predicted from the prediction of the last day. Different baselines based on different levels of noise are indicated in the background similar to Fig. 4.14a. These baselines naturally change with the predicted day due to error propagation. Shown are recursive models with past values of one (blue) and three (orange). Models trained on irregularly spaced time steps are shown as dashed lines. These models are replaced by a NARX variant with the same parameters.

With a growing number of recursions the models show rapidly decreasing performance. This is expected, as the error in the last prediction is fed back into the next one. The prediction error grows for models with a past and future of one by roughly 10 days per iteration, being outperformed by a baseline with 3% noise after the 5th iteration. The model trained on irregularly spaced time series with a past of three shows worse behaviour, losing performance much faster than the other models. The model trained on regularly spaced time series and a past of three stands out as consistently exceeding even the perfect baseline, with an average deviation from the true model values of around one month after 13 iterations. The models, after training, only require a single initial state to make predictions far into the future.

If these models are not only sufficiently accurate in their predicted values, but also accurately capture the correct rate by which the s_j change, it should be possible for linear trends to be derived from such recursive predictions. Fig. 4.16 shows those linear trends:

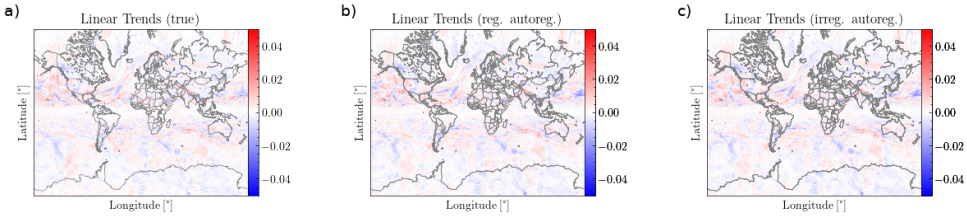

Figure 4.16: Comparison of linear trends for an initial state at 1 February. Panel a: Linear trends derived from true (model) values. Panel b: Linear trends derived from recursive predictions based on regularly spaced time series. Panel c: Linear trends derived from recursive NARX model predictions based on irregularly spaced time series.

The linear trends derived from the full regularly gridded data (without predictions) are shown in Fig. 4.16a as a reference. Panels b, c show the linear trends based on recursive NARX model predictions trained on regularly gridded and irregular data, respectively. The fine structure of the dynamics is captured by both models in great detail. Individual predictions may be occasionally imprecise, but the ensembles required for linear trends are robust. There are two notable qualities in these results: first, despite the sparse spatial coverage and therefore highly irregular spacing between matches, it is possible for a CNN to accurately learn the dissipation rates from satellite observations given some trajectories of air parcels. Second, the dissipation rates are learned based on the initial state only, implying that the initial state alone conveys sufficient information to predict whether an air parcel will undergo rapid mixing or not. This is actually not as surprising as it may seem, because if the mixing can be described by an ordinary differential equation, the Picard–Lindelöf theorem does guarantee the uniqueness of its solution given only the initial state. We have ample reason to believe that this is indeed the case, further supported by the results of this study.

4.7 Chapter conclusion

In this chapter, we have analysed filaments of tropical air leaking from the tropical pipe at altitudes beyond the reach of most aircraft. Observations of such filaments require satellite instruments, which we have substituted by model simulations. The longevity of filaments at this altitude requires focusing on equally long-lived trace gases. We have derived a parametrization M of the atmospheric state with respect to these LLTs, which describes the mixing between the tropical troposphere and the extratropical stratosphere. This parametrization is well suited for both spatial analysis (detection) of filaments and dynamical analysis, since it also reflects the chemical changes over time of the system. Using this parametrization, we constructed definitions of the tropical pipe, which allow us to define filaments as spatial anomalies of tropical air. Trajectory calculations, combined with matching to modelled/observed tracers, allowed us to determine both the spatial evolution and the dissipation rates of the mixing processes. We found mixing to be strongest at the edges and the 'tip' of the filaments. A similar analysis based on simulated satellite observations was performed, and the results found for regularly gridded data could be replicated from these spatially limited synthetic observations as well. Using convolutional neural networks, we have explored to what extent the dynamics of the AoA in this compartment of the atmosphere can be predicted. We found that these neural networks can learn the mixing processes from the information contained in matched trajectories and that filaments as dynamical anomalies can accurately be identified as such from recursive predictions. This means that given real observations, the spatial sparsity of satellite observations can be effectively bridged (with respect to the mixing processes) by learning mixing processes from those air parcels that are observed multiple times. This information can then be applied to any other air parcel within the given time frame and accurate predictions can be made up to around one week by utilizing recursive predictions.

Chapter 5

Beyond filaments – mixing in the general circulation

In the previous two chapters, mixing processes within individual, meso- to synoptic-scale filaments were investigated explicitly. Such localized analyses are possible because mixing at these scales occurs rapidly enough to track individual air parcels via trajectory calculations and resolve their changes in chemical composition. However, on a global scale, the net transport and mixing of trace gases are much less localized, and are instead driven by the Brewer-Dobson Circulation (BDC). In the zonally averaged framework, the BDC can be functionally divided into the advective *residual mean circulation*, which describes the net overturning mass flux, and the large-scale *eddy mixing* induced by breaking planetary waves. Both are intertwined and cannot be described independently from each other. The BDC can be outlined as follows: It consists of upward vertical transport of air in the tropics and horizontal, meridional transport and downwards vertical transport in the extra-tropics (see Fig. 5.1). The general transport is commonly described by a shallow branch, extending approximately to 20 km, and a deep branch, reaching up to the stratopause (around 50 km). In a broader sense, the meridional circulation in the mesosphere can be considered part of the deep branch. The air descends in the mid to high latitudes, where the circulation closes. It is one of the most influential circulations for distributing trace gas species, such as ozone or tropospheric species, and is important for the composition and, consequently, radiative properties of the atmosphere (Diallo et al., 2021; Poshyvailo-Strube et al., 2022). The timescale of the BDC is typically of the order of 2 years or less in the shallow branch (Lin and Fu, 2013) and up to 6 years for the deep branch (Poshyvailo et al., 2018). On these timescales we are no longer

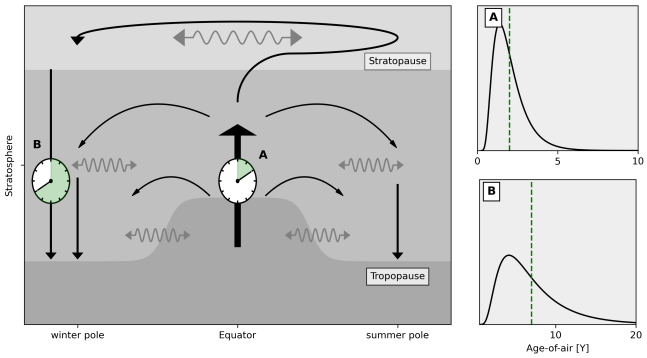

Figure 5.1: Illustration of the mean Age–of–Air concept. Air is uplifted from the troposphere in the tropics and transported polewards via either branch of the BDC (plain arrows). The two-sided wavy arrows indicate two-way mixing. Air may also be transported upwards as far as into the mesosphere, where it may reside for around one year (Ray et al., 2017). The age spectra for both scenarios are sketched in the insets A and B, respectively. Adapted from ESA (2025).

tracking individual air parcels, but rather air masses which have been mixed together on their pathway. This makes identifying individual transport pathways difficult. Nevertheless, the underlying mixing is not strictly affected by such a limitation. Instead, the concept of *Age-of-Air* (AoA) is generally used to characterize the residual circulation. AoA indicates the mean lag/transport time since an air parcel left the troposphere (Hall and Plumb, 1994; Garny et al., 2024a). Different pathways therefore lead to different AoA values and the gradients in AoA consequently also reflect the mixing between air masses. An air parcel is assigned a spectrum of ages (Johnson et al., 1999; Ploeger and Birner, 2016; Ray et al., 2024) reflecting the contribution of mixing with differently aged air parcels (see insets A and B in Fig. 5.1). Generally, in the literature the term AoA refers to the first moment of this distribution (mean Age-of-Air, indicated by the green lines in Fig. 5.1), as it is the effective AoA of an air parcel and the closest to an observable quantity. In climate models, AoA can be easily tracked by using an ideal clock tracer, which increases linearly at the boundary and has no sources or sinks elsewhere. The AoA of an air parcel can then be determined exactly from the stratospheric mixing ratio of this ideal clock tracer by the difference between the observation time and the matching boundary time (Hall and Plumb, 1994).

Global climate models today do not give decisive answers to the question of how the BDC changes under increased greenhouse gas forcing. Butchart et al. (2006) and McLandress and Shepherd (2009) predicted an acceleration of the BDC, manifested as enhanced tropical upwelling and faster transport. This would lead to significant changes in the distribution of trace gas species.

Observations of trace gas mixing ratios, however, show no evidence of such long-term strengthening (e.g. Waugh, 2009; Engel et al., 2009a). A more recent study by Ray et al. (2026), which analysed observations again, did find negative AoA trends, albeit much stronger than what the climate models predict. Atmospheric reanalysis datasets – designed to provide optimal estimates of atmospheric state – exhibit different BDC trends over recent decades (Ploeger et al., 2019a, 2021a). The future development of the BDC is currently unclear. It is therefore necessary to improve upon our understanding of the BDC and the wider circulation, particularly with the support of real measurements. The wind velocities of the residual circulation are too slow to be directly observed (on the order of 0.2 – 0.5 mm/s), such that the resulting distribution of AoA is generally used to quantify the dynamics of the BDC.

In the real world, no ideal clock tracers exist, aside from highly undesirable radioactive material, which no longer exists in the atmosphere in 'sufficient' quantities. Long-lived trace gases (LLTss), such as SF_6, can act as a proxy for an ideal clock tracer, and have been used to derive AoA (Maiss et al., 1996; Garny et al., 2024b). These LLTs have no sources in the stratosphere and stratospheric lifetimes which are long compared to the timescale of the BDC. While the different LLTs in this study have varying lifetimes, their individual lifetimes are ultimately unimportant for the results.

While being chemically inert, real LLTs do have sinks in the form of, for example, slow decomposition via electron attachment in the mesosphere (Loeffel et al., 2022), introducing a substantial positive bias in AoA estimation (the air appears older or contains less LLTs molecules), which needs to be addressed in the form of a mesospheric sink correction (Garny et al., 2024b). The exact quantification of these sinks is a major challenge for accurate determination of AoA. Volk et al. (1997) developed a method to determine AoA from the LLTs SF_6 and CO_2 by convolution of tropospheric time series of these LLTs with an idealized spectrum, and estimated the lifetimes of other LLTs through a second–degree polynomial fit. Voet et al. (2025) expanded the number of LLTs to the six LLTs used in this study to derive AoA estimates from the correlations of said LLTs. The same six LLTs were used in Chap. 4 to classify the mixing in the stratospheric surf zone and were used as covariates for predicting the temporal evolution of AoA in distinct air parcels, successfully. One may wonder whether a neural network is capable of similarly predicting AoA from these LLTs, learning the intricate chemical sinks, mixing rates and transport pathways from the chemical signature of these LLTs. In this chapter, different architectures and configurations of neural networks are presented and analysed with respect to their predictive capabilities. The

insight gained at the beginning of Chap. 4 that these LLTs form a 1-D manifold in phase space raises the suspicion that there may exist an alternative, model-agnostic representation of the BDC strongly related to AoA.

This chapter aims to answer the following questions:

- How accurately can AoA be derived using neural networks from only LLTs?
- What effect does measurement uncertainty have on the prediction, and how to best mitigate it?
- How well does a neural network predict AoA outside of its training domain, and are there covariates which can improve the predictions?
- If the LLTs lie in a 1-D manifold, can this manifold be determined by a neural network, and what is its relationship to AoA?

5.1 Dataset preparation

The data used in this study come from the same simulations used for the analysis in Chap. 4. However, since the posed problem is *in principle* limited to the chemical phase space, there are no special demands on the data such that the raw, unmodified output of CLaMS can be used. No form of interpolation is needed, and to an extent, interpolation is even discouraged. The data used cover the year 2011, which will be the main focus, and the following one, five and ten years (2012, 2016, 2021) for validation purposes. We consider the first of every month of each year to get an adequate coverage of potential trends and seasonal effects. We select all data points with potential temperature between 420 K and 900 K, corresponding to an altitude range between 14 km and 35 km. Similar methods for AoA determination are considered reliable from the tropopause to around 25 km (Kouznetsov et al., 2020; Voet et al., 2025), and marginally reliable between 25 km and 35 km. Conventional methods generally cover a similar altitude range. We aim to predict AoA across the maximum vertical range for which reliable data are available. We create datasets for each of the dates considered. All sets contain the LLTs $\mathrm{CH_4}$, $\mathrm{CFC{-}11}$, $\mathrm{CFC{-}12}$, $\mathrm{HCFC{-}22}$, $\mathrm{N_2O}$ and $\mathrm{SF_6}$, the CLaMS model AoA and the variables latitude, time and potential temperature. The time variable indicates the month of the dataset. A cyclic time variable, for which values for January and December converge, has been considered and found to not be beneficial. The selected sets are then normalized (per variable) to the range [0, 1] to apply

equal numerical weight to each variable. From each normalized set a subset of 50,000 random samples is created. The values in each dataset therefore are uniform, random samples across different altitudes, latitudes, and longitudes. These subsets are split into a training and a validation set, which contain 40,000 (80%) and 10,000 (20%) points, respectively. Sets containing several months may be created by merging their respective training and validation data sets, combining for instance the subset for January 2011 and the subset for February 2011 to create a subset for both months, with 80,000 points in the training set and 20,000 points in the validation set. The features of the following neural networks always consist of the previously mentioned LLTs, but may include any of the additional variables except for CLaMS model AoA, which is exclusively used as a label. The *training* data are used exclusively for training of the neural networks; the validation of these networks is performed exclusively using the validation data. This ensures that the performance of a model is judged by the prediction of unseen data only. Validation between different months (or years) is particularly interesting, since the simulation state for these cases is very different from the states used for training. Successful predictions beyond the training domain indicate that overarching structures and relationships were learned. The ideal model data have very clear relationships between LLTs and the CLaMS model AoA, since all quantities are driven by the same wind data. The real relationships observed in real measurements may deviate from them. In addition to the random samples of the simulation, we emulate observations by adding relative (per sample) Gaussian white noise to the features, which are saved as new variables in the datasets. We consider noise levels of 1%, which roughly corresponds to the measurement precision of the in situ measurement technique AirCores, which is used to measure LLTs (Laube et al., 2020), and 2%, 3% and 5%, which is the order of magnitude a limb–viewing satellite instrument could achieve (Voet et al., 2025). While the white noise does capture the statistical uncertainty of an instrument, real instruments additionally produce systematic uncertainties. These systematic uncertainties may be more difficult for a neural network to learn and compensate for if not yet seen during training. Noisy inputs therefore may not be a 100% accurate substitute for real observations. The effects of systematic uncertainties are covered at the end of Sect. 5.2.1.

5.2 Supervised MLPs

The first task to be investigated is: if the correct model AoA is known, can this AoA be accurately predicted using only our set of LLTs as features? For this a simple, regressive Multi-Layer Perceptron (MLP) architecture is considered. An MLP is a simple, feed-forward neural network consisting of at least one input layer, one output layer and at least one hidden layer in between. In this section, a total of three hidden layers with 300 neurons each are considered. Fig. 5.2 shows the architecture in two variations: Panel a shows a regular MLP which takes inputs with six

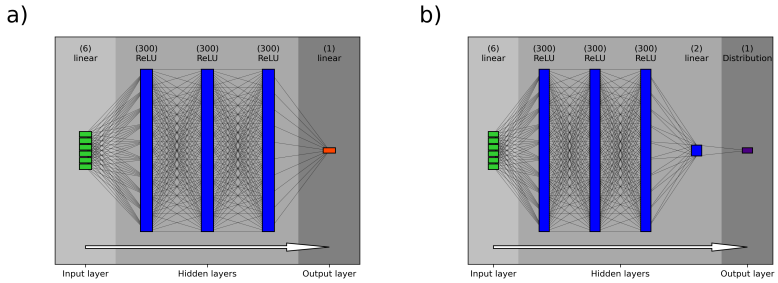

Figure 5.2: Sketch of a typical (supervised) MLP architecture. Panel a shows a regular MLP, panel b shows a probabilistic MLP. Shown is the architecture for the six LLT training scenario as an example. This architecture consists of 3 hidden layers containing 300 neurons each and an output layer of 1 neuron. The sizes shown are reduced compared to the true sizes for visualization purposes.

features and predicts a single label. Such a model learns a mapping between scalar input values and a scalar output. The predicted values are then compared against the true model values via the mean squared error (squared RMSE, see Eq. 2.28). The MLP learns which values of AoA best match with the true CLaMS model AoA given some input.

Alternatively, the output layer can be replaced by a Gaussian distribution layer with a suitable prior dense layer to get the correct number of inputs (a single Gaussian distribution has two free parameters). Rather than predicting a single scalar value, this *probabilistic* MLP predicts a Gaussian distribution. The model checks the likelihood that the true CLaMS model AoA x would have been sampled from a Gaussian distribution given the predicted distribution parameters μ, σ via the negative log-likelihood (NLL) loss:

$$\text{NLL}(x|\mu,\sigma) = \frac{1}{2}\log(2\pi\sigma^2) + \frac{1}{2\sigma^2}(x-\mu)^2 \tag{5.1}$$

Since these parameters must match across many similar inputs, they reflect (technically: can

reflect) the sensitivity of the prediction.

We are interested in the answers to the following three questions:

- How accurately can AoA be predicted based on LLTs? In other words: what is the aleatoric uncertainty of such a model?
- How robust are these predictions with regard to measurement noise? In other words: what is the epistemic uncertainty of such a model?
- What is the contribution of either type of uncertainty to the final prediction?

5.2.1 Prediction performance and influence of noise

We begin with a simple, naive setup that trains a regular MLP on ideal model LLTs of a single dataset of August 2011. This model is validated on the validation data of the same dataset. There are many metrics to consider in order to contextualize the predictive capabilities of this model, see Sect. 2.6. We will discuss them in greater detail in the following: An overview of some essential metrics is given in Fig. 5.3: Panels a, b show the predicted AoA values for the regular and probabilistic MLP, respectively, against the true values of the CLaMS model. The predictions in Fig. 5.3a were performed based on input data without additional noise. A perfect model would predict the exact model values, such that all points would lie on the identity line (shown in grey). The latitude associated with each prediction (i.e. the latitude of the point from which the LLT values are taken) is indicated in colour-code. The regular MLP is capable of predicting AoA from these LLT values within a margin of just under two weeks. There is also no clearly visible dependence on the latitude of the predictions. The performance drops towards the higher AoA values, but stays within an accuracy of around one month. Comparing this result to the probabilistic MLP (Fig. 5.3b), we see a very similar picture for the predictions based on unmodified inputs (0% noise): The probabilistic model achieves a performance of the same magnitude (RMSE of 11 days), but simultaneously provides a measure of uncertainty in the form of the width of the predicted distributions, shown as error bars. Comparing the accuracy (difference between predicted and true values) to these error bars we see that the vast majority of differences lie within one standard deviation. The model is capable of recognizing its prediction precision.

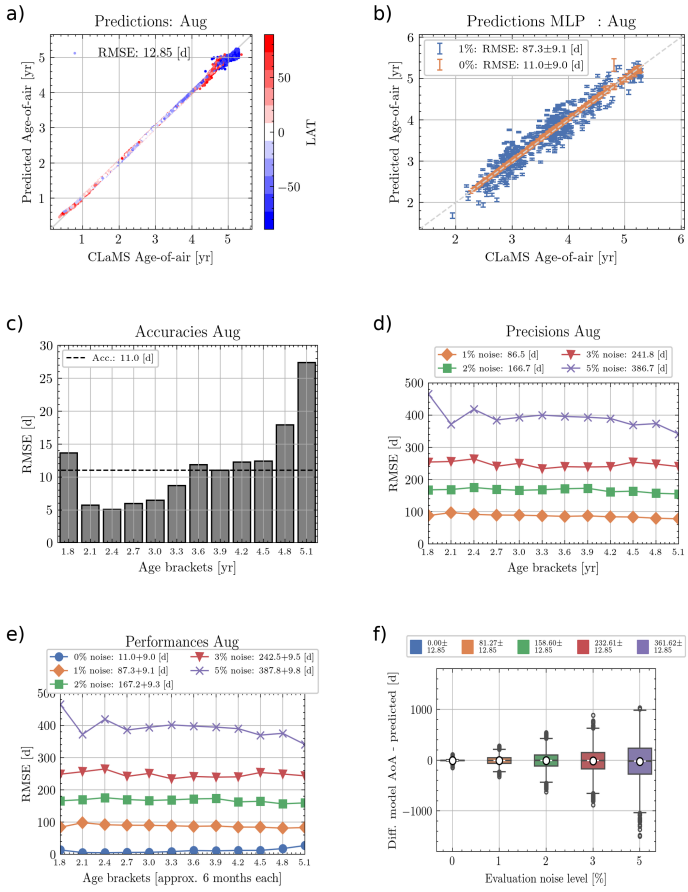

Figure 5.3: Diagnostics of the regular and probabilistic MLP. Quantities that are shown only once are almost identical for each MLP. Panel a: Correlation between the predicted AoA based on noise-free features and the true CLaMS model AoA. The grey line indicates a perfect prediction (identity). The corresponding latitude of each point is indicated in colour-code. The performance is summarized by the RMSE(Y_p, Y_r) value. Panel b): The same correlation, but for values predicted by a probabilistic MLP. Shown is the correlation between the expectation values of the predicted distributions and the true CLaMS model AoA. The standard deviations of each predicted distribution are indicated as error bars. The performance is summarized by the RMSE(Y_p, Y_r). Panel c: Distribution of the model accuracy split by predicted AoA value. The average accuracy is indicated by the black vertical line. Panel d): Precision of the predictions, split by the predicted AoA values. The precision is expressed as the RMSE(Y_p, Y_δ), i.e. the difference between optimal prediction and noisy prediction, with higher RMSE values indicating worse precision. The distortion added to the inputs is indicated in colour-code and varies between 0% and 5% of the true value. The average precision RMSE is shown in the legend above. Panel e): Same as panel d, but for the model performance, indicated by the RMSE(Y_δ, Y_r), i.e. the difference between true CLaMS model values and noisy predictions. Panel f): Distribution of the difference between true CLaMS model values (Y_r) and noisy predictions (Y_δ). The distributions for the different noise levels are shown as boxplots, with 50% of all values contained within the box, 85% of all values within the whiskers and all other values indicated by the outliers.

These error bars are not uniform and may vary between predictions, thus providing a potentially realistic measure of the prediction uncertainty. However, if we consider a perturbed input (1% noise, Fig. 5.3b blue), a problem becomes obvious: A slight variation in the inputs has increased the RMSE almost tenfold, while the average predicted uncertainty has stayed constant. The precision of the predictions decreases so strongly that it is no longer reflected by the estimated uncertainties. To understand how the performance can decrease this strongly, the full diagnostics of the models should be considered; since both models have nearly identical diagnostics, it suffices to discuss the probabilistic MLP: Fig. 5.3c shows the accuracy represented by the RMSE(Y_r, Y_p) resolved by the predicted AoA, with the average RMSE indicated by the black dashed line. While we see a drop in accuracy (higher RMSE) for the highest predicted AoA values (see also Fig. 5.3a), the accuracy of the model stays well below one month of predicted AoA. Since accuracy reflects the aleatoric uncertainty of the model, we can deduce that in principle AoA can be highly accurately predicted from LLTs alone. The response of the model for inputs with noise added is interesting, and is illustrated in Fig. 5.3d through the precision represented by the RMSE(Y_δ, Y_p). The errors for 1, 2, 3 and 5 percent input noise are shown as a function of different values of AoA. While the model shows a minor increase in precision (lower RMSE) towards higher predicted AoA, each percent of noise added to the input increases the mean RMSE by around 100 days. The slight decrease in RMSE at high AoA for 5% noise (purple line, 'x'-markers) is negligible relative to the year-scale mean error ($\approx$364 days). The model shows the characteristic signs of overfitting, as small deviations in the inputs lead to progressive loss of precision. The poor precision indicates a very high epistemic uncertainty, which is the limiting factor of this model. The predictive capabilities of the model can be summarized in the performance of the model, which is shown in Fig. 5.3e. In terms of RMSE, the performance of a model is given by the quadratic sum of accuracy and precision. With very high epistemic uncertainty the performance is dominated by the poor precision. Lastly, rather than only considering the average deviations, the distribution of the deviations should be considered as well, summarized in Fig. 5.3f via boxplots. We see that with increasing input noise the spread of the differences becomes disproportionately larger (wider whiskers), with a visible asymmetry in outliers for the highest input noise towards overestimation. The main issue with the models is that no form of regularization was performed, which led to extreme overfitting. While such an overfitted model can achieve very high accuracy, its lack of robustness towards input noise is the limiting factor. While the model is capable of predicting AoA based on inputs similar to its training data, it fails to learn overarching structures.

Next, we train new MLPs (both regular and probabilistic) with the same architectures on the dataset of the same day (1 August 2011). Each model is now trained on noisy inputs (1%, 2%, 3% and 5% noise, respectively), for a total of eight individual models. Each model is validated in the same way as the models before. Fig. 5.4 shows the distribution of the differences for the probabilistic MLPs as an example.

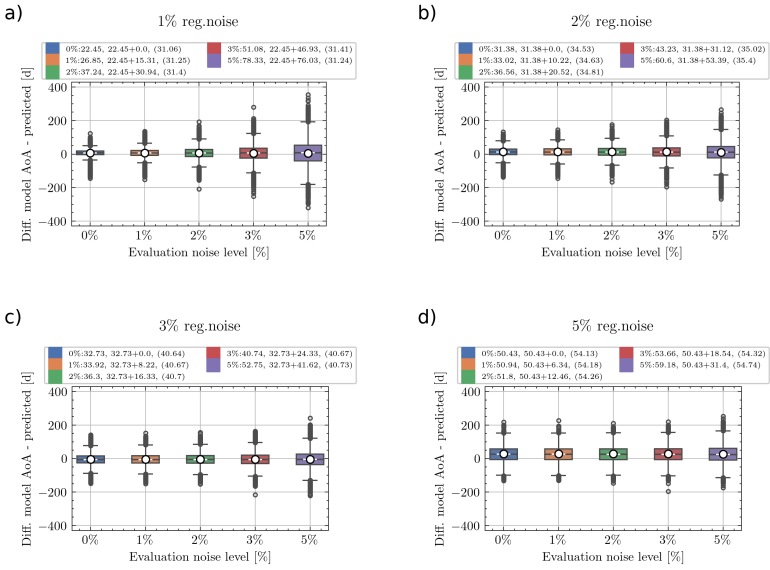

Figure 5.4: Difference diagnostic for probabilistic MLPs trained on different levels of feature–noise. Panels a – d show the distribution of the differences for 1%, 2%, 3% and 5% regularization noise, respectively. The corresponding precision and accuracy of each validation is shown in the legends. The models show fairly constant performances for the different validations, since the accuracy is determined during training, and while it is not impossible for a model to adjust its accuracy for different predictions, the models fail to do so. The statistics are functionally identical to Fig. 5.3f .

The legend summarizes (from left to right) the prediction RMSE (performance), the accuracy RMSE (aleatoric uncertainty), the precision RMSE (precision) and the average standard deviation of the predicted distributions (estimated accuracy). As expected, the introduction of regularization noise has reduced the model's accuracies significantly. However, the predictions of the models have become much more robust against input noise, with precisions rarely exceeding RMSE values of two months. The resulting performance is therefore decreased for low noise levels, but vastly improved for higher noise levels. However, the accuracy remains constant across different validation noise levels, indicating that the model accuracy is learned during training and does not

change during validation. Instead, the loss in performance is attributed completely to the epistemic uncertainty of the model. With increasing regularization noise the difference distributions become much narrower with fewer extreme outliers. As seen in Fig. 5.4d, too high regularization can introduce systematic patterns such as consistent underestimation of AoA.

A special pattern stands out if one considers the validations on noise levels similar to the regularization noise: if a model is validated on noise levels similar to the noise used for regularization we see that the performance is dominated by the aleatoric uncertainty. This makes sense: if the model has seen similarly noisy values during training, it can accurately recognize its prediction limits. If we consider the average standard deviations of the predicted distributions we see that these match up very well with the accuracies of the model for these cases. While the full diagnostic requires knowledge of the unperturbed inputs for the accuracy determination, the predicted standard deviations do not. Since they line up well with the analytically calculated accuracies, for these cases they can be considered a good proxy for the model accuracy. Since the performance is dominated by the epistemic uncertainty for these cases, they are equally a good measure for the overall prediction performance. This is the main advantage of a probabilistic MLP.

Regularization can be performed in many ways, such as L_1 or L_2 regularization or dropout. However, using regularization noise has the unique advantage that it allows us to link the expected model performance to the predicted standard deviations. For real applications, as a rule of thumb, the regularization noise should be of the same magnitude (and shape) as the expected measurement noise, as this optimizes both performance and provides realistic performance estimates. This effect is further visualized in Appendix A.7.

One may ask: if the model learns its aleatoric uncertainty from the variability in the training data, could a more robust model be trained by including differing regularization noise levels, such that the epistemic uncertainty is correctly estimated? The very constant accuracy across different levels of epistemic uncertainty is a known issue in data science, and subject to current debate and research (e.g. Ovadia et al., 2019; Foong et al., 2020; Kirsch, 2025). It is sometimes known as the 'epistemic collapse', where a model fails to recognize changes in epistemic uncertainty. To date there exists no universal, clear way to avoid it.

Regularization with Gaussian white noise can capture the effects of measurement uncertainties well, but such noise is generally considered rather 'harmless' since it does not change expectation values compared to the more problematic systematic uncertainties, which are inherent to the individual instrument. The most common forms of systematic uncertainties are calibration

offsets, i.e. a fixed offset between measured and real values, and scaling errors, i.e. changes in observed values proportional to the observed signal. How robust are MLPs with respect to such errors? Generally we do not anticipate any MLP to perform well with such data, since these systematic errors fundamentally change the mapping the network learns. Fig. 5.5 shows the performance of a set of MLPs following the aforementioned architecture, regularized with 2% noise and evaluated within the training domain. We train one instance of MLP based on no systematics,

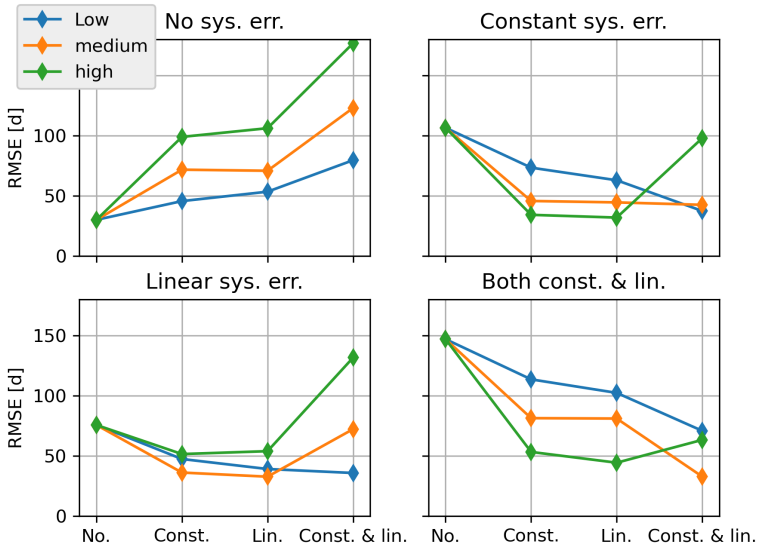

Figure 5.5: Effect of different kinds of systematic uncertainties on the prediction performance, expressed by the RMSE. A model regularized on 2% Gaussian noise is tasked to predict AoA based on inputs subject to a constant offset (Const.), a linear scaling factor (Lin.) or both (Const. & lin.), see panel a. Three degrees of severity (low, medium, high) of each form of systematic uncertainty are considered. Panels b – d show the same evaluations, but for a modified model which was trained on data containing the same combinations of systematic uncertainties of medium severity, respectively.

a systematic offset, a linear scaling error and both offset and scaling error. We consider a random offset, which differs in scale across all features (medium), and two variations of said offset, which are half as large (low) and 1.5 times as large (high). For the linear scaling we consider systematic, linear increases of 1% (low), 2% (medium) and 5% (high). In a combination of both systematic effects, the scaling is applied first and consequently the offset. During training, medium systematic effects are considered. Every design is evaluated against data with any of the systematic errors of differing magnitude. The performance of each validation, represented by the RMSE, is shown for

each configuration of systematic error in colour-code. If no form of systematic error was included in the training data performance, the error can be doubled for each systematic error present in the data. Including an offset in the training data helps mitigate performance loss significantly for both offset and linear systematic error, up to the point when the true systematics are no longer effectively compensated. A similar effect is observed when including a systematic linear scaling in the training data, which improves the performance for either offset or linear systematic error, but struggles to compensate both for pronounced effects. Including both types of systematic errors works best to mitigate these effects if present in the data, but generally leads to significant loss of performance if they are not present.

Systematic offsets and linear scaling errors can be compensated by including either in the training data. In a sense, these systematic effects can be approximated by each other. Even a broad guess about the scale of these effects can increase performance by around 25 days in this example. Generally, handling systematic uncertainties is a difficult task, and similarly to the noise regularization, an accurate tuning is required. Performance of course is best if the correct systematics are included in the training data, but a performance loss from one month to $2-3$ months is likely.

5.2.2 Feature selection and training domain

Up to this point we have only considered the performance of the models *in-domain*, i.e. validation of data which is explicitly similar to that in the training domain. A good model should learn overarching patterns which extend far beyond its training domain.

With expansion beyond that training domain, systematic shortcomings of a model may become visible in the spatial distribution of the predicted values. Fig. 5.6 shows the spatial distribution of the differences between true CLaMS model and predicted AoA and the performances for the other months of 2011. A significant bias in the predictions is visible in Fig. 5.6a. The model systematically underestimates AoA in the southern hemisphere and systematically overestimates it in the northern hemisphere. Asymmetries in AoA between the hemispheres are a known feature (Konopka et al., 2015; Ploeger and Garny, 2022; Dubé et al., 2025), which this model cannot recognize from its LLT features alone. If the model had explicit spatial knowledge, such as the latitude of the input points, the model could potentially learn these patterns. Another type of bias is illustrated in Fig. 5.6b, which shows the performance of the model for all months of 2011, not only for the training month August. The performance RMSE is reasonably low for the late summer

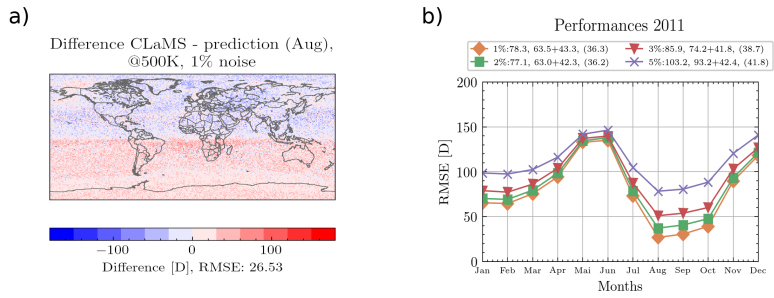

Figure 5.6: Diagnostics for particular model artefacts. Panel a: Spatial distribution of the differences between true CLaMS model and predicted AoA for a probabilistic MLP trained with 1% regularization noise. Shown is an isentrope of 500 K potential temperature, corresponding to an altitude range between 19 km and 22 km. Predicted are inputs with 1% simulated noise. Red regions indicate underestimation of AoA, blue regions indicate overestimation. The blur in the distribution stems from the inputs with added noise used to predict AoA. Panel b): Performance of the same model for different datasets of 2011. The performances are expressed in terms of RMSE(Y_δ, Y_r) for different validation noises. The precision and accuracy of each validation is shown in the legend.

and autumn months (August – October), but increases significantly for May and June, as well as December. The model performs well in-domain, but has issues performing out-of-domain. The processes determining the relationships between LLTs and AoA, such as mixing, chemistry and transport, are known to have seasonal effects, which a model cannot learn if no such seasonal patterns are included in the training data. Information about these seasonal patterns can be provided by including more training data from more months, i.e. by expanding the training domain.

To address these issues, we apply the following adaptations to the current training setup: the latitude coordinate can be included in the features, which transfers information about the meridional distribution of AoA. Instead of a single monthly dataset, multiple datasets may be used during training. Including all months during training is possible in principle, but would ultimately lead back to the original problem that the model cannot be trained on *all* possible data. This of course would improve prediction performance, but the aim should be to learn overarching patterns from limited amounts of data. For these reasons we include the months February, May, August and November in the training data, i.e. every third month and thus one month per season. We refer to such a dataset as a seasonal dataset. In addition to using a seasonal dataset, time can be included in the features as well by numbering the months. In addition to the LLTs, such a dataset contains a feature whose value indicates the month of any particular data point. This partitions the training data into seasons, whose influence on the prediction is then determined by the model during training. A cyclic time coordinate was also tested, but we found that this yields worse

performance. This might be due to the limited number of months considered and may improve once larger, more complete datasets are considered. Since both latitude and time are known without significant uncertainty, we do not apply noise to these features.

The performances for the different configurations are shown in Fig. 5.7: Training an MLP with

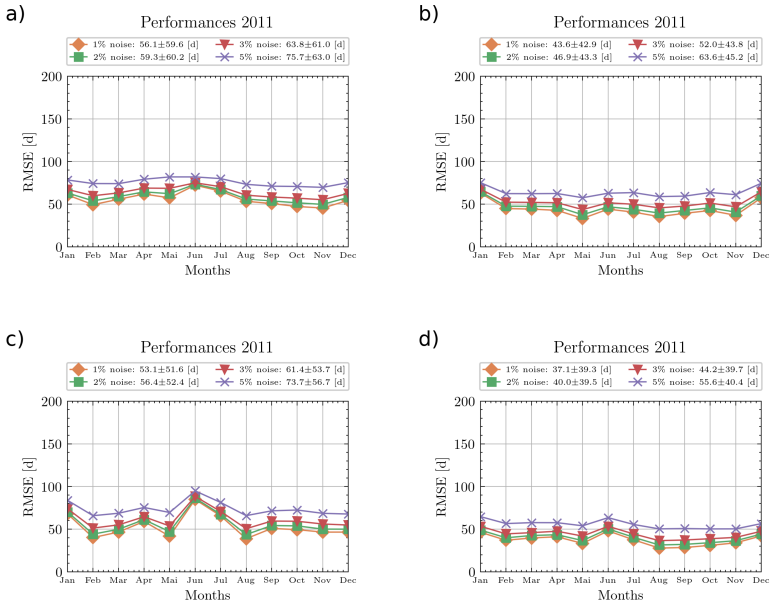

Figure 5.7: Annual performances of probabilistic MLPs trained on different features. All models were regularized with different noise levels where applicable, and therefore show reduced variability between validation noise levels. Panel a: Previous design as shown in Fig. 5.6, but trained on a seasonal dataset with implicit time dependence only. Panel b: MLP trained on a seasonal dataset which includes latitude as a feature. Panel c: MLP trained on a seasonal dataset which includes time explicitly as feature. Panel d: MLP trained on a seasonal dataset which includes both latitude and time as features.

a seasonal dataset (Fig. 5.7a improves the performance by around 20%, and shows only small variations across the seasons. As expected, including more training data (or volume) leads to higher performance. The model trained with explicit spatial knowledge (Fig. 5.7b) shows greatly improved performance by around 40% compared to the single-month model, but retains some systematic performance issues, particularly in December. The model trained with time as an explicit parameter shows the poorest overall performance of any of the adjusted designs (Fig. 5.7c), with an exceptionally poor performance in June. A potential reason for this performance loss from simply including more information is as follows: If there is only a weak correlation between AoA

and the time coordinate, the dataset including the time coordinate naturally has higher variability (with respect to AoA) than a dataset without it. Each feature also contributes through its correlation to other features; it acts as a covariate. Such covariates may show complex, incidental correlations to other features, which the neural network can learn but ultimately do not contribute to accurate predictions. If the time coordinate alone does not provide valuable information, it can lead to reductions in performance. The exact reasons why the time coordinate decreases performance are currently unknown. However, combining both time and latitude in a seasonal training dataset yields a model with the highest performance so far, see Fig. 5.7d. While the systematic issues remain, particularly for June and December, the performance only exceeds 60 days for the highest validation noise. Whatever negative effect was introduced by the inclusion of the time coordinate is seemingly resolved by the inclusion of the latitude. It would seem as if the time coordinate only acts as a meaningful covariate to the latitude. For seasonal structures, which are also localized, this explanation appears plausible. For example, the time variable may tell us if the polar vortex is strong, but cannot tell us where it is located, and the latitude can indicate where the vortex is located, given that it is indeed present at the given time. Both variables in tandem provide sufficient information to learn the complete pattern. If we remove those months with poor performance from the total annual performance, we would consistently get an average performance of around 1.5 months. From these experiments it becomes clear that seasonality of AoA is only poorly captured by a simple MLP. While the performance is still high, one must recognize these variations in performance. One may consider training an additional, dedicated model for a time period where a more general model shows poor performance. The equivalent, regular MLPs yield very similar performance and are not shown here.

Similar to conventional methods, the supervised neural networks exhibit exceptional performance. The model with optimized features and training domain is capable of accurately predicting AoA over an entire year within 1.5 months performance. The model retains high performance even at high altitudes, where the mixing ratios of many LLTs begin to vanish. It would be worthwhile to investigate the lower limit of the LLTs for usage in MLPs. The main caveat of neural networks is their nature as a 'black box'. Generally, the exact function (or mapping) that is learned by a neural network is either inaccessible or incomprehensible (usually due to sheer size) to the researcher. Ultimately, a model should be in line with previously established knowledge of the modelled physical processes. There exist, however, various methods by which one may attempt to understand the function learned by the model. Fig. 5.8 shows multiple analyses of the optimized model: The

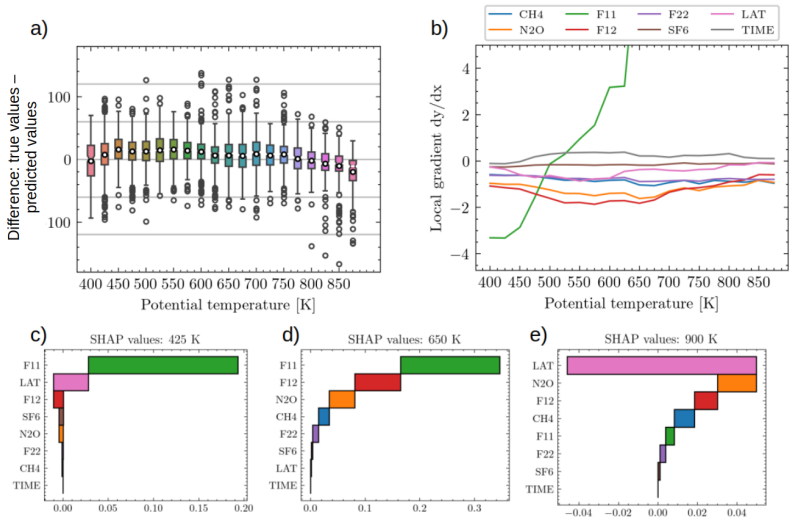

Figure 5.8: Analysis metrics used to describe the function of a neural network: Panel a: Distribution of the differences between predicted and true CLaMS model AoA similarly to Fig. 5.4, but resolved by potential temperature. Panel b): Gradients of the neural network, taken at representative inputs for each of the potential temperature brackets used in panel a. Panels c – e: (SHapley Additive exPlanations (SHAP) values for representative inputs corresponding to potential temperatures of 425, 650 and 900 K, respectively. The SHAP values indicate the importance of each input to the final prediction.

figure shows the difference statistic similarly to Fig. 5.4, but resolved by the potential temperature of the predicted values. The distributions are very stably centred around zero and show no significant systematic over- or underestimation. The width of the distributions are consistent across different potential temperature levels, but shows a slight tendency towards overestimation for outliers towards the highest potential temperature. The model retains its performance over the entire considered vertical range. Insights into the geometry of the learned function are difficult to visualize, since the function is generally defined on a high-dimensional domain. However, by using backpropagation (see Eq. 2.33) we can determine an effective 'local gradient' for particular inputs, which is shown in Fig. 5.8b. These gradients describe how much a small change in the particular input parameter would translate into a change in predicted value. Steep gradients indicate a particularly high sensitivity of that input to the prediction, which implies that this input is particularly important, whereas low gradients suggest the opposite. Since these gradients vary across the entire domain, they must be derived locally for specific input vectors. Here we choose the averaged feature values of 100 random samples from within the potential temperature brackets used in Fig. 5.8a. The gradient for $\mathrm{CFC}{-}11$ stands out as anomalously steep. The most likely explana-

tion is that $\mathrm{CFC}-11$ in isentropes below 650 K shows very pronounced variations and structures, which allow one to capture the fine structure of AoA. A high gradient, however, does not necessitate that $\mathrm{CFC}-11$ is actually the most important input for all potential temperatures, since the corresponding weights could vanish. This will be made apparent shortly. Since gradients point in the direction of change, most gradients are negative as expected. The positive gradient for the seasonal 'TIME' variable is interesting. The model introduces a small correction towards higher TIME-values (the boreal winter months), which is likely responsible for the improvements due to the inclusion of the TIME-parameter. Latitude (LAT) shows more pronounced adjustments through its gradients, indicating that the inclusion of latitudes did indeed introduce small corrections reflecting the spatial distribution of AoA. $\mathrm{CH_4}$ and $\mathrm{SF_6}$ show constant, weak gradients, mostly due to their strong correlation with AoA. $\mathrm{N_2O}$ and $\mathrm{CFC}-12$ show the largest (not anomalous) gradients, indicating that the content of these LLTs is essential for the fine structure of AoA. This reflects the different sinks of these LLTs depending on altitude, latitude and solar radiation, leading to more complicated relationships.

While the gradients tell us how *sensitive* the model is to particular inputs, they do not tell us how important these inputs are for the predicted value. For this the SHAP values are shown in Fig. 5.8c–e for three vertical regions. The SHAP values are based on the Shapley values (a Nobel Prize-winning concept in game theory), which measure how much a particular input (player) contributes to a particular prediction (game state). SHAP values are local and additive and compare each prediction against an averaging baseline. Positive SHAP values move the prediction in the positive direction (higher predicted AoA values), and negative SHAP values in the negative direction (lower predicted AoA values). At lower altitudes (425 K potential temperature) the model prediction is dominated by $\mathrm{CFC}-11$, with a significant positive contribution by latitude and only minor negative contributions by the other LLTs. A steep gradient in combination with high SHAP values does indicate the aforementioned importance of $\mathrm{CFC}-11$ for the predicted AoA. However, it could also mean that the model overestimates AoA primarily based on $\mathrm{CFC}-11$ and then reduces that value according to the other LLTs. The absolute SHAP values indicate only a relatively small deviation from the baseline (< 0.2). At middle altitudes (650 K potential temperature) all features contribute positively to the predicted AoA, with $\mathrm{CFC}-11$, $\mathrm{CFC}-12$ and $\mathrm{N_2O}$ being the largest contributors. Latitude appears to have very little significance in this region. The larger absolute SHAP values indicate much stronger deviation from the baseline (< 0.3). At the highest altitudes (900 K potential temperature), the contribution of the LLTs is similar to

before, but now latitude has a significant contribution towards lowering the predicted AoA. This is likely due to the polar vortex, which is located at the highest latitudes, where AoA is often largely overestimated by conventional methods. The MLP has automatically learned a suitable correction. Despite its large gradient, the SHAP values suggest that $\mathrm{CFC}-11$ is not actually very relevant for the determination of AoA in this potential temperature bracket.

5.2.3 Predicting the future

So far the performance of MLPs for predicting a complete year from limited training input was discussed. The best design yielded an average performance of around 1.5 months. However, we have only considered data sampled from one year so far. A useful prediction method should be transferable to other data significantly outside of its training domain. It is a well-established result from conventional approaches that models cannot, in general, be directly applied to data from substantially different distributions. If that were not the case, we would have access to a universal formula or algorithm which determines AoA for all cases. Since a neural network only learns a function as implicitly encoded within the training data, predictions of inputs from outside of that training domain are only possible if the function conveys an overarching structure for many months – or even years – into the future. Any changes in that function in the future McLandress and Shepherd (e.g. as predicted by 2009); Ploeger et al. (e.g. as predicted by 2021a), could lead to significant performance losses.

So far no conventional method is capable of long–term prediction of AoA without proper adjustments, i.e. relying on new observations or existing time series. Following the results of the previous sections, we take the probabilistic MLP trained on a seasonal dataset of 2011 with all previously considered input features (see Fig. 5.7d) and evaluate it on the corresponding datasets created from data of the years 2012, 2016 and 2021. The performances are shown in Fig. 5.9: As expected, the performance for each year is reduced compared to the training year. However, the performance is much less reduced than one might have expected. For 2012 in Fig. 5.9b, pronounced systematic differences in performance are visible, with poorer performance in the boreal spring months. Average performance lies below 3 months, with (estimated) epistemic uncertainty of around 40 days (down from 60 days for 2011), indicating that the model actually does not recognize its shortcomings. Five years later (Fig. 5.9c), the situation is comparable, with slightly decreased performance of just over 3 months, similar seasonal patterns and the same epistemic uncertainty. In Fig. 5.9d, the performances for 2021, i.e. one decade after the training data, are

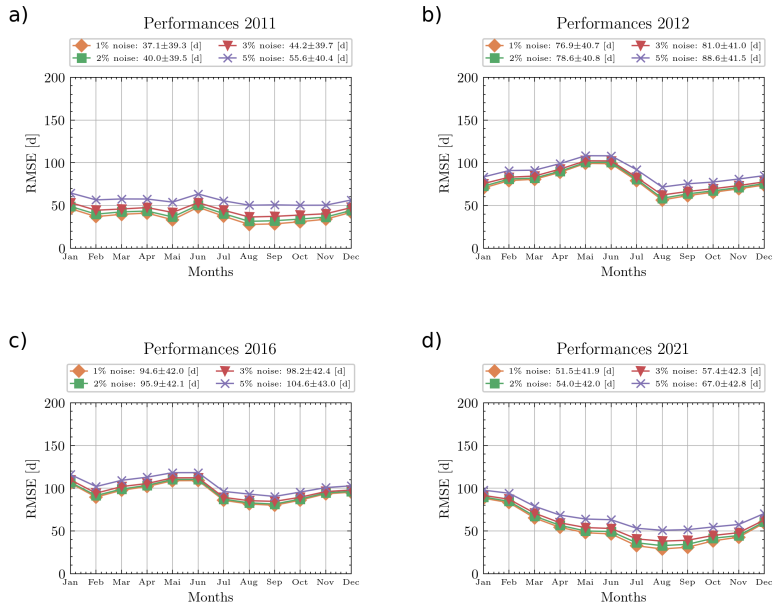

Figure 5.9: Performance evaluation of datasets from different years. The model is a probabilistic MLP trained on a seasonal dataset of 2011 with full features and various feature noise regularization. Shown are the years 2011, 2012, 2016 and 2021 in panels a, b, c, and d, respectively.

shown. The model performs better for this year than the other two, with average performance comparable in size to the training year 2011, but with a new, seasonal pattern in its performance with high–performing months in boreal autumn and low–performing months in boreal winter.

We are unaware of any reason why this should be. Ultimately we must attribute this high performance to chance. Still, the model retains a high level of performance over many years far outside its training domain. Across all validated years the (estimated) epistemic uncertainty of the model remains close to 40 days, such that this estimate is likely inaccurate and most of the uncertainty is aleatoric in nature. The model cannot intrinsically reconcile these issues. There is ultimately a limit to how different a dataset can be to be accurately predicted, even if the precision is surprisingly high. The regular MLP shows similar predictions and is not shown here.

Conventional methods require current data in order to determine the tropospheric mixing ratios. Thus, data for 2021 would be needed if one wanted to determine AoA for that year. The MLPs show that they do not rely on this and retain high performance even for those years which were not known during training. In this study, a model trained on data from 2011 can still accurately predict AoA from 'measurements' of 2021. One reason why an MLP appears not to depend on

this and is inherently capable of retaining its performance across many years lies in the preprocessing of the input data. The MLP is trained to map certain feature constellations to certain AoA values, or in the case of a probabilistic MLP to express the training inputs as distributions from which AoA is predicted. During preprocessing every dataset is normalized to the range [0, 1] such that the absolute scale is lost, but the correlations are retained. While absolute values may change significantly over many years, the correlations between the normalized features stay approximately constant. The scaling of input data back to physically relevant quantities such as AoA is learned by the MLP from the training output. Any linear transformation that translates the mixing ratios observed in one year to those observed ten years later is implicitly accounted for during preprocessing. Thus, the scaling of input data remains valid for future data. It must be noted that it is assumed that the principal mechanisms governing the distribution of LLTs and their correlation with AoA do not undergo strong changes. AoA in a changed climate cannot be accurately predicted by such an MLP. While an MLP is more robust than conventional methods with respect to the application to new datasets, it cannot be used to learn new mechanisms that were not included in the training data. For this, a different approach would be needed, which can learn such new (or unknown) mechanisms from new data. The next section introduces such a concept.

In summary, MLPs are a powerful tool for predicting AoA from LLTs. If suitably regularized, an MLP can accurately and precisely predict AoA within 1.5 months on average. In order for the model to correctly recognize the epistemic uncertainty, i.e. its prediction confidence, regularization via input modification can be applied. If the regularization noise is comparable in scale to the additional noise contained in the prediction inputs, the model becomes well-calibrated. A probabilistic, well-calibrated MLP can provide reasonable estimates for its epistemic uncertainty via the widths of the predicted distributions. By training an MLP on data which implicitly include seasonal effects the model retains its high performance for arbitrary months of the year. Including latitude and time as explicit features further improves the model's performance, which is well retained for predicting data from years up to a decade after the training year. No conventional method so far can achieve this performance. By applying various methods from the field of explainable AI, it was demonstrated how the different features contribute to the predictions, and that the model implicitly learns an equivalent of the modern mesospheric sink correction at high altitudes near the polar vortex.

5.3 Unsupervised autoencoders – an alternative approach

The success of the previous supervised MLPs has demonstrated that AoA can be accurately and precisely predicted from a set of LLTs only. This implies the existence of some function $F(\boldsymbol{X}) = x$, where $\boldsymbol{X}$ denotes a vector of LLTs. Indeed, in Chap. 4 it was already shown that a 1-D embedding of this 6-D $\boldsymbol{X}$ exists. We may therefore assume that this embedding can be used to find a parametrization which directly correlates to AoA, similarly to the mixing state M in Chap. 4.

It is a long-standing goal to become independent of climate models when deriving AoA, as individual models themselves carry their own inherent biases, be it through the data used to initialize trace gases, the data used to drive the simulation or the chemistry / mixing implemented in the model (e.g. Dietmüller et al., 2018; Abalos et al., 2021, 2019). While there are well-established standard methods to infer AoA from measurements of approximate clock tracers, most commonly SF_6 (e.g. Volk et al., 1997; Garny et al., 2024b)), no method is completely free of any model information (e.g. correction of mesospheric SF_6–sink, knowledge about the ratio of moments). The dependence on climate models becomes particularly relevant if one considers the anticipated climate change.

In order to find this 1-D embedding, we employ autoencoders (, see Sect. 2.5.2). This form of dimensionality reduction enables highly non-linear transformations which are likely required to accurately determine the underlying manifold. Since none of the LLTs strongly interact with other species, their change over time (including mixing and decay) should approximately be only a function of time. If this function $F(\boldsymbol{X}(t))$ can be inverted, it would be possible to directly infer t (representing AoA) from $\boldsymbol{X}$. Contrary to the supervised MLPs, an AE does not require any form of model AoA as a reference.

A simple approach would be to mimic the architecture of a successful MLP, taking only the LLTs as features. An AE consists of two neural networks: an *encoder*, which maps the features to a latent space of dimension n (here: one), and a *decoder*, which attempts to reconstruct the original features from the encoding in latent space. Since an AE only works with data which can be represented by an invertible function, additional features such as latitude cannot be considered. The mapping between the embedding and its features must be at least approximately injective, which is not the case for latitude, where many different values for AoA could be mapped to the same latitude. The architecture of a simple and a probabilistic AE is shown in Fig 5.10.

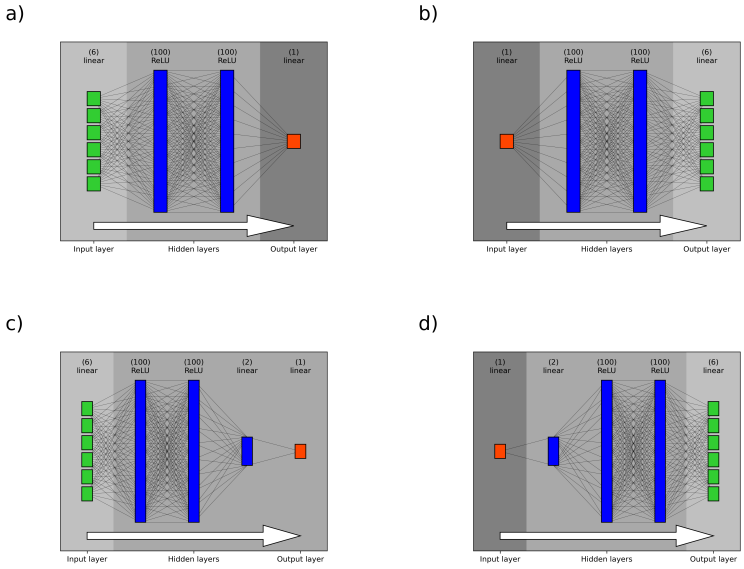

Figure 5.10: Architecture of the AEs. Panel a, c: Architecture of the regular and probabilistic encoder, respectively. The encoders take the six LLTs as features and return a 1-D output. Panel b, d: Architecture of the corresponding decoder. The decoders take the 1-D output of the encoder and return a 6-D output. During training, the decoder output is compared against the encoder input. In all panels, the number of neurons shown in the hidden layers is reduced to 25 for visibility reasons.

Due to the symmetry of the problem, the encoder and decoder are mirror-symmetrical. During training we found that both regular and probabilistic AEs perform similarly, such that in the following only the probabilistic AE is discussed. The AE is trained on data for August 2011 without regularization noise to explore the capabilities of the concept. Regularized AEs are considered at a later point. Since an AE is an unsupervised neural network, there is no direct quantity to compare the encoding against. To get a better understanding of the embedding Γ and its relation to the real, physical quantities, it is beneficial to focus on a small dataset first. For this, the 500 K potential temperature isentrope and its spatial structure (see e.g. Fig. 5.6a) are considered.

The spatial structure of the embedding is shown in Fig. 5.11. The encoding and the decoding (reconstructed original features) of this AE are shown in Fig. 5.12a, b, respectively:

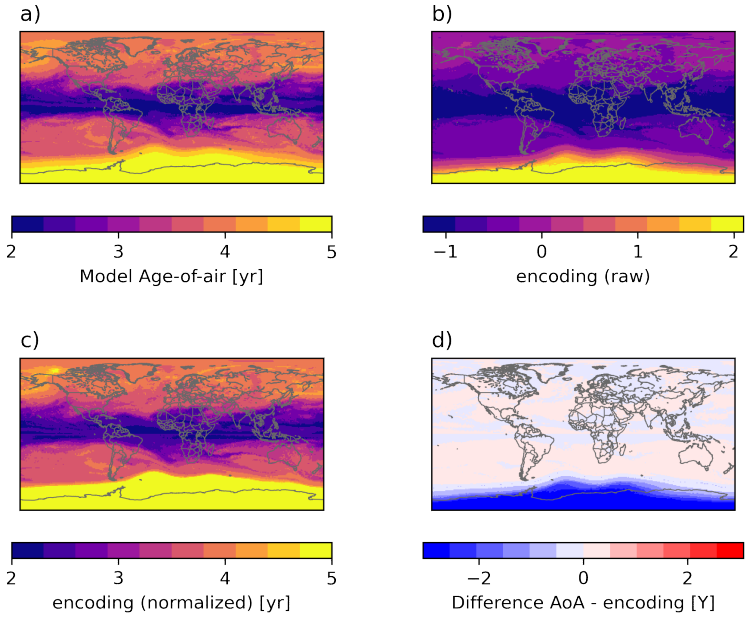

Figure 5.11: Spatial distribution of a probabilistic AE 1-D embedding of LLTs. Panel a: CLaMS model AoA values. Panel b: 1-D embedding, as provided by the encoder. The scale of the encoding follows the scale of the inputs. Panel c: 1-D embedding, linearly scaled to match the model value range based on the linear regime in Fig. 5.12a. The colour-scale is limited to the value range found in the model AoA. Panel d: Difference between CLaMS model AoA and linearly scaled embedding values.

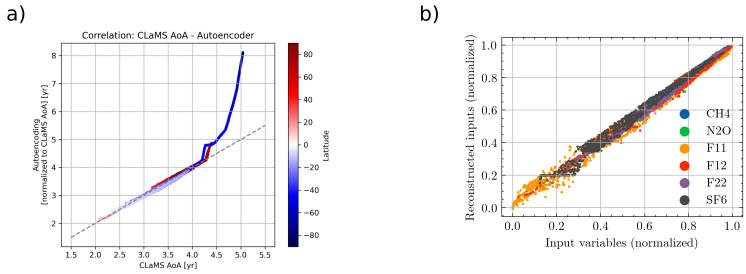

Figure 5.12: Relations between AE embedding and model values. Panel a: Correlation between CLaMS model AoA and linearly scaled 1-D embedding. The latitude of the encoded values is indicated in colour-code. The embedding is scaled such that it corresponds to the CLaMS model AoA in the linear regime (< 4 years). Panel b): Correlation between the model LLTs used as features and the decoding of the embedding. The decoded values need no calibration.

In Fig. 5.12a, the latitude of the input values is shown in colour-code. We see that the correlation between Γ and AoA is remarkably compact and that the encoding Γ is roughly linear as a

function of AoA for AoA < 4 years. Since there are no constraints on the numerical magnitude of Γ, it tends to be expressed in units of similar scale as the features. In Fig. 5.12a, this linear relationship has been used to roughly match the scale of Γ to AoA. Beyond four years the linear relationship begins to break down, with Γ now significantly exceeding AoA, particularly towards high latitudes. The AE learns that Γ should increase significantly inside the polar vortex, similarly to how a model without mesospheric sink correction would. One may suspect that the mixing ratios in the polar vortex become too small such that they can no longer be accurately reconstructed by the decoder. Considering Fig. 5.12b, this is not the case. While the decoding is less accurate for lower values (higher AoA), it cannot explain the discrepancy between Γ and AoA.

However, this primitive Γ is already reasonably suitable for describing the physical situation. Compared to the CLaMS model AoA (Fig. 5.11a), the raw Γ reflects the different physical features such as the tropical pipe, filaments and the polar vortex well (Fig. 5.11b). If Γ is scaled to the linear regime in Fig. 5.12a, it also accurately reflects the correct magnitude of these features (Fig. 5.11c), with the exception of the polar vortex. In fact, considering Fig. 5.11d we see that in the polar vortex the discrepancy between CLaMS model AoA and the primitive Γ exceeds two years.

Rather than the CLaMS model AoA, Γ appears to represent the apparent AoA τ. This makes sense when considering that the AE only has knowledge of the different LLTs, which are subject to different chemical sinks. Since the AE has no additional domain knowledge, it cannot tell if certain values are likely subject to strong sinks such that Γ needs to be adjusted accordingly. The question is: Can the embedding be improved by providing this domain knowledge, and can this information be extracted from the LLT input natively?

The answer to the first question, outlined in Kaumanns et al. (2026), is as follows: If we consider Γ as a function of the true model CLaMS AoA in Fig. 5.12a, we can see that Γ could be divided into two functions: a linear function f_1 for AoA < 4 years, and some non-linear function f_2 for AoA > 4 years. We know that this linear behaviour points to weak chemical sinks, such that the LLTs change linearly over time. This part represents the relationship between Γ and the LLTs in the absence of chemical sinks. The non–linear deviation of f_2 from f_1 is due to the effect of the chemical sinks. If f_2 could be expressed as a function of f_1, then this new function f_3 would only be a function of the chemical sinks, if roughly constant chemical sinks can be assumed. By training two AEs – one on the linear, one on the non-linear regime – we can create these functions f_1, f_2, and if f_3 is a monotonic function (which it is), a calibrated Γ can be derived, which

accurately reflects the true AoA. This approach is discussed in Kaumanns et al. (2026) using a hemispheric split and the domain knowledge of where the linear regime ends. This minimizes the dependence on model AoA, but does not completely remove it. It was also applied only to data up to 600 K potential temperature.

To answer the second question, we must understand how an AE learns: Provided some training domain, an AE learns an encoding function E, which maps the training inputs X to a manifold Γ such that there exists a decoding function D with $D(D(X)) = X$. The decoder D then maps Γ back to X. The AE judges its performance through its loss function L by comparing $D(E(X))$ to X, here:

$$L = \mathbb{E}_{X\sim p(X)}\left[\|X - D(E(X))\|^2\right] \tag{5.2}$$

where $p(X)$ denotes the (probability) distribution of the training data X. We already know that Γ is in principle only a function of time. We can exploit that the depletion of an LLT via a chemical sink is functionally identical to the depletion due to long transport times. We introduce an apparent velocity v_{apparent}, which describes the rate at which an LLT depletes with respect to Γ. In this apparent velocity, chemical sinks appear as if the system sped up compared to the regimes with weak chemical sinks. In an abstract phase space, chemical sinks increase the local curvature of the manifold. Herein lies the explanation of the problem: For a dynamical system, $p(X)$ is inversely proportional to the local (apparent) velocity of the system:

$$p(X) \sim \frac{1}{\|v_{\text{apparent}}\|} = \left\|\frac{dX}{d\tau}\right\|^{-1} = \left\|\frac{dX}{dt}\right\|^{-1} \cdot \left|\frac{dt}{d\tau}\right|^{-1} \tag{5.3}$$

In the last step the chain rule was used. Thus, regions with strong chemical sinks barely contribute to the shape of Γ, such that the AE performs poorly in these regions, or rather expresses them as a function of the true time t it learned from the other regions. The previous approach works because each AE is trained in physically coherent regions, where $\frac{dt}{d\tau}$ is fairly constant. Both time scales are then unified by expressing one in terms of the other. The process is illustrated in App. A.13 roughly through a 3-D representation of the data used.

In order to divide the training domain X natively into regions of approximately constant $\frac{dt}{d\tau}$, a *Mixture Of Experts* (MOE) model can be used. A MOE is a neural network architecture which consists of a number of (functionally) identical neural networks (here: AEs). These AEs are coupled to a router network, which usually takes the form of an MLP. Given an input X, the router network passes X either to one of the experts (hard assignment) or to all of the experts (soft assignment).

If an expert performs well for a specific X, the router network learns to pass this kind of X to this particular expert in the future, and vice versa. Since regions with strongly different $\frac{dt}{d\tau}$ cannot perform well on the same expert, the router network natively sorts X by different $\frac{dt}{d\tau}$, i.e. divides the training domain based on the local curvature of the manifold. A MOE also natively solves the problem of expressing the Γs of the many experts in a unified basis, which can then be calibrated in the linear regime using CLaMS model AoA. Details of this particular architecture and the different parameters that can be chosen to achieve the best performance are described in App. A.8. Here we use the LLTs $\mathrm{SF}_6, \mathrm{CH}_4$ and $\mathrm{CFC}{-}12$ as the main inputs, with the LLTs $\mathrm{CFC}{-}11, \mathrm{HCFC}{-}22$ and $\mathrm{N}_2\mathrm{O}$ acting as covariates for the embedding. The routing network is provided information about the season (time), latitude, potential temperature and potential vorticity. This configuration should allow the MOE AEs to determine a physical mapping of the atmosphere and find a compact embedding while optimizing the information yield of all available LLTs. Contrary to a simple AE, MOE AEs allow training on a seasonal dataset, since they can resolve the seasonal structures. The MOE AEs presented are trained on 1% measurement noise, which directly influences the embedding. While embeddings without measurement noise are visually more compact, they show the same structure as the embeddings presented here. Since the regularized MOE AEs are used for validation, the general concept is demonstrated using these noisy embeddings.

The result of the MOE AEs are illustrated in Fig. 5.13:

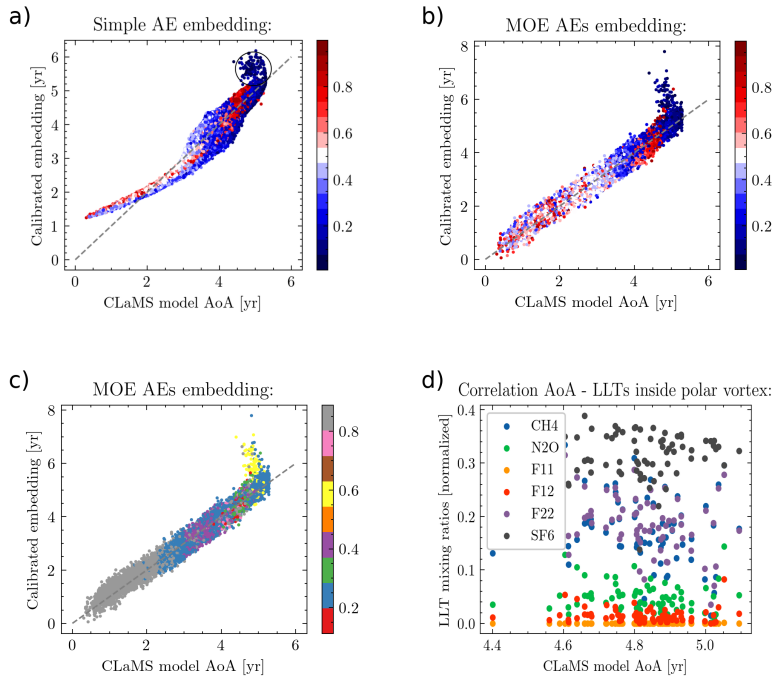

Figure 5.13: Overview of MOE AEs performance. Panel a: Correlation between the true CLaMS model AoA and the embedding of a simple AE. The corresponding latitude values are shown in colour-code. The embedding is linearly calibrated. The prominent feature (located at CLaMS model AoA > 5 years and calibrated embedding > 5 years) is marked with a black circle. Panel b: The same correlation as in panel a, but for the embedding of MOE AEs. The latitude is shown in colour-code. Panel c: The same correlation, but shown in colour-code is the ID of the expert used predominantly for the embedding instead. Panel d: Correlation between the true CLaMS model AoA and the different LLTs for the values contained within the circle in panel a.

As a reference case, despite the manifold spanned by the LLTs being 1-dimensional, a single autoencoder (see panel a) is incapable of determining an embedding Γ which reflects the true CLaMS model AoA. The strong deviations of this Γ from the CLaMS model AoA negatively affect the calibration as well, such that this embedding does not represent the CLaMS model AoA. We do see that this embedding is rather linear in the regime AoA < 3 years, since according to Eq. 5.3 the region with only weak chemical sinks dominates AE's loss function. The determined Γ does reflect the curvature of the manifold relative to its curvature for low AoA. The MOE AEs, being similar AEs, but instead trained on dedicated subsections of the manifold, yield a Γ which

resembles the true CLaMS model AoA much more closely (see panel b). A small section located close to the polar vortex (latitude $< -75°$) stands out, where Γ significantly deviates from the CLaMS model AoA. If we consider the ID of the expert used predominantly for each data point, we see that this section was assigned a dedicated expert (yellow), with some contribution of the blue expert. Identifying this yellow expert in the primitive embedding in Fig. 5.13a, we find that this encoder is active in the uppermost, almost circular feature indicated by the circle. Identifying this section in the spatial distribution of the data reveals that this section corresponds to the polar vortex. The polar vortex is historically very difficult for AoA determination, and the failure of the MOE AEs here is unfortunate. It raises the question: can MOE AEs be improved to also work well inside the vortex? The answer may be disappointing: panel d shows the correlation of each LLT to the CLaMS model AoA inside the selected section. Inside the polar vortex, there is no longer a compact correlation between any of the LLTs (and consequently Γ). Indeed, calculating the intrinsic dimensionality for this section yields a clear value of 2.02. It is physically impossible with this data to accurately predict AoA inside the polar vortex. Nonetheless, the MOE AEs do provide a proxy for AoA anywhere else by only exploiting the curvature of the compact manifold, thereby not relying on any model AoA whatsoever.

Fig. 5.14 shows the spatial distribution of difference between the CLaMS model AoA and the MOE AEs' Γ for select isentropes:

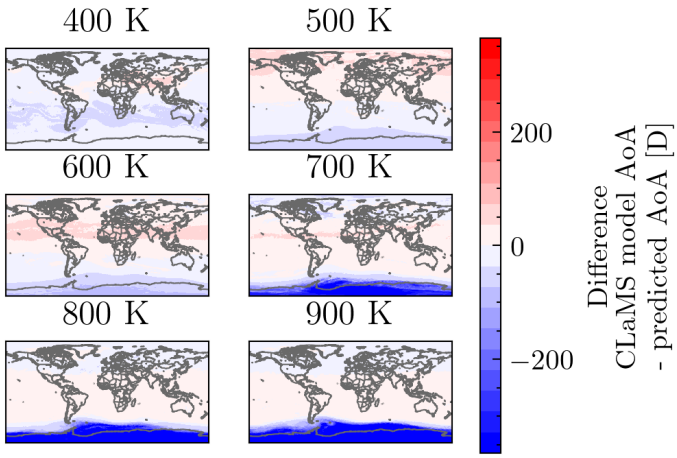

Figure 5.14: Evaluation of the physics-informed MOE AEs. Shown are the spatial distributions of the differences between the true CLaMS model AoA and the calibrated embedding on isentropes for 1 August 2011.

The MOE AEs are capable of accurately predicting an embedding strongly related to the CLaMS model AoA within two months for potential temperatures between 400 K and 600 K, as shown in Fig. 5.14a. While the similarity between this embedding and the CLaMS model AoA is still remarkable almost everywhere for potential temperatures as high as 900 K (see Fig. 5.14d), the embedding fails to accurately reflect the CLaMS model AoA as mentioned before inside the Antarctic polar vortex, with deviations up to many years. Knowing this, the MOE AEs are validated in the following only for potential temperatures up to 600 K, where the influence of the polar vortex is minimal.

Fig. 5.15 shows the performances of the MOE AEs resolved by month and year similarly to Fig. 5.9. Compared to a supervised MLP, the MOE AEs achieve an overall lower performance,

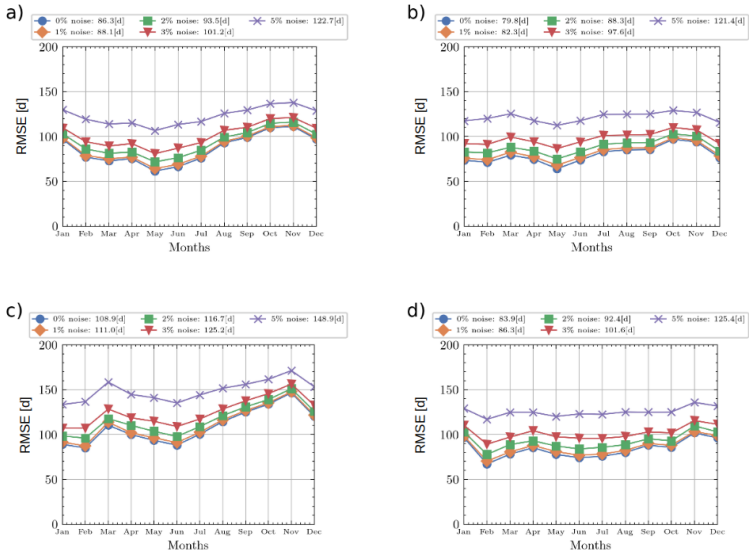

Figure 5.15: Performance evaluation of datasets from different years. The models here are calibrated Mixture of Experts autoencoders trained on 1% measurement noise using only data from August 2011. Validation data is sampled between 420 K and 600 K potential temperature. Shown are the years 2011, 2012, 2016 and 2021 in panel a, b, c, and d, respectively.

with around one month higher RMSE values. Some seasonal effects are visible, since MOE AEs cannot be trivially expanded to seasonal datasets and require a dedicated design, for reasons discussed in App. A.8. Still, similarly to the MLPs, the MOE AEs retain high performance for data far outside their training domain, and are only relatively weakly affected by measurement noise.

In summary: AEs are unsupervised neural networks which can generate a 1-D representation of LLTs. During training no knowledge of the CLaMS model AoA is provided. The structure of this embedding is similar to the apparent AoA and related to the true AoA. Calibration of this embedding is linear for lower values of AoA. By dividing the encoding domain into suitable segments, it is possible to create multiple embeddings, which are piecewise compactly correlated to the CLaMS model AoA. By combining these embeddings, a continuous proxy for the CLaMS model

AoA can be constructed, which has no explicit knowledge of the CLaMS model AoA, and thus is independent of the model itself, if the influence of the model on the used LLT values is neglected. A mixture of experts (MOE) setup can be used to allow the neural network to figure out the best partitioning of the domain itself, as well as making each embedding continuous with respect to one another. While MOE AEs still fail to accurately predict the AoA inside the polar vortices, their embedding has proven to be an accurate proxy for AoA.

5.4 Chapter conclusion

In this chapter, we have explored methods to derive AoA from LLT features using supervised MLPs and unsupervised AutoEncoder (neural network) (AE)s. Since AoA is an important diagnostic quantity to describe and understand the Brewer-Dobson Circulation (BDC), particularly in a changing climate, it is important to obtain suitable estimates or proxies for AoA based on real observations.

We find that given model values, an MLP can predict AoA with high accuracy and precision. These MLPs retain high performance even when predicting inputs with large added noise similar to real–life measurements. Using probabilistic MLPs allows us to obtain predictive uncertainties for individual predictions. If the MLP is calibrated with noise similar to the measurement noise, the uncertainty estimates are in good agreement with the observed prediction errors, indicating that the model provides well-calibrated estimates of the aleatoric uncertainty. A regularization of 1–2% on its training inputs improves the robustness of an MLP when predicting data outside of its training domain, such as data from different months or years.

The approach of using LLTs can be further improved upon if other features such as spatial information like latitude or temporal information such as a monthly index are included. The monthly index is found to contribute positively only if used in conjunction with latitude. Since MLPs are trained on normalized inputs, they retain high performance even when predicting values a decade after their training data. Seasonal variability can be adequately accounted for if the training domain is expanded conservatively. Such an MLP which is trained extensively on a large amount of data is in principle feasible and could result in a powerful standard tool for AoA prediction. We found that a few distinct datasets could not be accurately predicted, hinting at physical irregularities within these datasets. Further investigation into the reasons is required.

To become (partially) independent of model values, we utilize MOE AEs to find a 1-D embedding

of the manifold spanned by the LLTs. This embedding acts as a proxy for AoA and does not rely on the CLaMS model AoA. It accurately reflects the spatial structure and the magnitude of AoA outside of the polar vortex, but deviates inside of it due to the breakdown of the correlations between AoA and the LLTs. Neural networks are a powerful tool to determine AoA from observable quantities, with high performance and robustness for supervised MLPs utilizing model AoA, and for alternative, purely data-driven approaches utilizing only observable quantities.

Chapter 6

Summary and outlook

The stratospheric circulation is fundamentally driven by a wave–induced residual mean circulation, which is deeply connected to various mixing processes. While the residual mean transport establishes the global overturning structure, localized mixing processes further shape the spatial distribution of chemical species. These species, in turn, drive climate feedbacks, influence the stratospheric radiative budget, and feed back onto the global circulation itself.

Despite their critical importance, resolving these mixing processes remains an important bottleneck in state–of–the–art climate modelling, forcing global circulation models (GCMs) to rely on sub–grid parametrsations of various kinds. The description of a mixing process relies on a well–defined initial state, such that changes of this state may be attributed to mixing. Capturing and validating these parametrsations through observations presents an equally severe challenge, requiring dedicated, high–resolution instruments capable of tracking and revisiting unique air parcels along trajectories to derive changes in their chemical composition over time. Atmospheric filaments, which are induced by breaking Rossby waves, are attractive phenomena on which to study mixing, as they undergo localized and rapid mixing. In this thesis, mixing processes and suitable measurement concepts are discussed for the different compartments of the stratosphere, from the UTLS over the stratospheric surf zone to the more general BDC:

- Mesoscale (Filaments in the UTLS)

  In the first part of this work, high–resolution, three–dimensional tomographic measurements from the "Gimballed Limb Observer for Radiance Imaging of the Atmosphere" (GLORIA) were analysed. Using 3–D tomographic retrievals of multiple trace gas species obtained during a dedicated self–matching experiment performed during the "Probing High Latitude

Export of air from the Asian summer Monsoon" (PHILEAS) campaign in 2023, the mixing within a filament of Asian summer monsoon outflow near a tropopause fold was studied. By exploiting the internal structure of the retrievals using Bayesian Gaussian Mixture Models, we derive a classification that describes the number and the composition of all air types involved, of which none need to be known beforehand. This approach allows one to utilize observations of any number of available trace gases to maximize the information yield. Despite not utilizing the spatial information available, this classification accurately reflects the spatial structure of each air type and the observed mixing processes. The classification allows one to identify mixing processes in the matched air parcels as changes in classification even for complex three–way mixing, which is otherwise very difficult to identify. The probabilistic nature of the classification allows one to identify ongoing mixing via gradients of classification probabilities. The approach was further validated by applying it to the other 2–D data products obtained during the experiment, showing similar success.

- Synoptic scale (Filaments in the surf zone)

  The second part of this work focuses on the stratospheric surf zone, in which filaments drawn from the tropical pipe into the extra–tropical stratosphere are studied. Synthetic observations following the design specifications of the "Changing–Atmosphere Infra–Red Tomography Explorer" (CAIRT) instrument, simulated by the "Chemical Lagrangian Model of the Stratosphere" (CLaMS) are used. Adapting the previous classification approach, we derive a parametrsation which allows one to identify filaments as anomalous spatial structures as well as describe the mixing processes as the transition from one air type to another. Using this parametrsation the entire spatial and dynamical situation can be described using a single, scalar quantity. By calculating trajectories and efficiently matching these trajectories with the synthetic observations we obtain irregularly spaced time series, from which explicit mixing processes are derived. These mixing processes are analysed with respect to the geometry of the filaments and to the mixing strength implemented in the CLaMS simulations. The likelihood of achieving such a match between air masses at subsequent times was derived with respect to geographical location and the track width of the instrument. We demonstrate that the mixing rates obtained from these synthetic CAIRT observations are nearly identical to those obtained by using fully spatially resolved, ideal data. We also demonstrate how the information about the mixing processes can be distilled from these time series using Convolutional Neural Networks (CNNs), such that the evolution of an air

parcel can be accurately predicted.

- Planetary scale (Age–of–Air)

  The third part of this work focuses on the BDC, and the net effect of large–scale eddy mixing beyond individual filaments. An important diagnostic for the BDC is stratospheric Age–of–Air (AoA), which acts as a macroscopic indicator of the circulation strength. Deep neural networks are trained to resolve the complex relationship between long–lived trace gases (LLTs) and AoA, providing highly accurate and precise predictions of AoA beyond the capabilities of conventional methods. The best model can predict AoA up to one month of the CLaMS model AoA, which significantly exceeds conventional methods. This high performance is retained even for high AoA, which is usually the most difficult to accurately predict. The models are optimized with respect to anticipated measurement noise, input features and robustness outside their training domain. As a culmination of the insight gained during this work, this thesis introduces a purely data–driven proxy for AoA based on physics–informed Mixture of Experts (MOE) autoencoders, which map long–lived tracers onto a suitable manifold directly correlated to the true AoA. Since this proxy only relies on directly observable quantities, it can serve as a validation tool for climate models given real observations.

This dissertation demonstrates that a multi–species classification of air masses provides a powerful, physically consistent framework to identify and quantify implicit mixing processes embedded within complex atmospheric datasets. By looking at traditional transport challenges through the lens of non–parametric machine learning, the main challenges when considering mixing – the definition of the initial state and the possible mixing pathways that lead to change of this initial state – are solved elegantly. The methodology established in this thesis – particularly in Chap. 4 – opens the way for future studies of mixing processes through satellite data. Using a similar multi–species approach supported by neural networks, this dissertation shows how AoA can be accurately derived from such satellite data with high precision.

Future research should investigate how the data–driven AoA proxy performs with respect to real observations. If the proxy conforms with our physical understanding of the atmosphere it could be considered an additional tool for climate model validation. While the CAIRT instrument concept will not be realized in its originally envisioned form, the conceptually similar "Stratosphere Troposphere Response using Infrared Vertically–resolved light Explorer" (STRIVE) mission is currently considered for deployment in 2030 by the American National Aeronautics and Space Administra-

tion (NASA), potentially satisfying the demand for highly resolved tomographic observations. If simultaneous observations of many trace gas species are available, the multi–species classification presented in this thesis can describe even complex physical situations in simple, meaningful terms.

As a general remark – this thesis showed how machine learning provides valuable synergy in atmospheric science. While machine learning has yet to be broadly applied in atmospheric science, it is a promising approach for future research and will help us better understand our atmosphere.

Chapter 7

Danksagung

Zuallererst möchte ich mich bei meinem Doktorvater und Mentor Prof. Dr. Felix Plöger bedanken für diese großartige Gelegenheit, an diesem Projekt arbeiten zu dürfen und seiner wunderbaren Unterstützung. Gleichermaßen möchte ich mich bei Prof. Dr. Michaela Imelda Hegglin bedanken für ihre Unterstützung und das Vertrauen, das sie mir entgegengebracht hat, sowie bei Prof. Dr. Martin Riese, ohne den ich diese Gelegenheit niemals bekommen hätte. Zudem möchte ich mich herzlich bei meinen Betreuern Dr. Peter Preuße und Dr. Jörn Ungermann für die massive Unterstützung während meines Promotionsprojektes bedanken, sowohl auf professioneller als auch auf menschlicher Ebene. Des Weiteren möchte ich mich bei den Mitgliedern unserer Arbeitsgruppe sowie den vielen Mitarbeitern und Doktoranden unseres Instituts bedanken, die eine wunderbare und einzigartige Arbeitsatmosphäre geschaffen haben. Ich werde auf meine Zeit in diesem Promotionsprojekt als eine meiner besten Zeiten zurückblicken.

Appendix

A.1 Numerical calculation of trajectories

In practice, the trajectory calculation of an air parcel needs to be performed numerically by solving the equation of motion Eq. (2.6) iteratively. To ensure the numerical stability of the trajectory, a Runge Kutta scheme is generally used. If Eq. 2.6 is simply solved naively by calculating the local derivatives, the trajectory always overshoots along the tangent of the wind field. Instead, the (fourth order) Runge Kutta scheme averages multiple different derivatives to better approximate the final derivative. These derivatives are:

$$\mathbf{k}_1 = \mathbf{v}(\mathbf{x}_n, t_n) \tag{A.1}$$

$$\mathbf{k}_2 = \mathbf{v}\left(\mathbf{x}_n + \frac{\Delta t}{2}\mathbf{k}_1, t_n + \frac{\Delta t}{2}\right) \tag{A.2}$$

$$\mathbf{k}_3 = \mathbf{v}\left(\mathbf{x}_n + \frac{\Delta t}{2}\mathbf{k}_2, t_n + \frac{\Delta t}{2}\right) \tag{A.3}$$

$$\mathbf{k}_4 = \mathbf{v}(\mathbf{x}_n + \Delta t\,\mathbf{k}_3, t_n + \Delta t) \tag{A.4}$$

The solution for the time step t_n is:

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \frac{\Delta t}{6}\left(\mathbf{k}_1 + 2\mathbf{k}_2 + 2\mathbf{k}_3 + \mathbf{k}_4\right) \tag{A.5}$$

The Runge Kutta scheme allows one to calculate stable trajectories over long periods of time (many iterations for solving the trajectories). To ensure that the scheme is stable, the approximation error must be at most linear: Assuming we want to solve the equation:

$$\frac{d\mathbf{x}}{dt} = \mathbf{J}\mathbf{x} \tag{A.6}$$

We can consider the stability function $R(z)$, which satisfies:

$$\mathbf{x}_{n+1} = R(z) \cdot \mathbf{x}_n \tag{A.7}$$

For a fourth order Runge Kutta R is:

$$R(z) = 1 + z + \frac{z^2}{2} + \frac{z^3}{6} + \frac{z^4}{24}, \quad \text{with } z = \lambda \cdot \Delta t \tag{A.8}$$

with λ being an eigenvalue of **J**. R represents the error that is made during each time step by approximating the true solution. If this error is too large, it compounds and the trajectory is no longer correctly calculated. Formally, if

$$|R(z)| \leq 1 \tag{A.9}$$

then the scheme is stable and the approximation error can be neglected. This does constrain the time step used in the scheme. Practically, the limiting factor is the largest wind gradient anticipated for the trajectory.

A.2 Matching of trajectories

The following algorithm is designed to determine all matches between a trajectory $\boldsymbol{x}_{traj}(t)$ and a discrete set of observations $\boldsymbol{x}_{obs}(t)$. We understand a match between $\boldsymbol{x}_{traj}(t^{'})$ and $\boldsymbol{x}_{obs}(t^{'})$ at a time $t^{'}$ as $\boldsymbol{x}_{traj}(t^{'})$ being at most ϵ away to $\boldsymbol{x}_{obs}(t^{'})$ for some choice of ϵ. Mathematically, this is described by Eq. A.10:

$$\|\boldsymbol{x}_{obj}(t^{'}) - \boldsymbol{x}_{traj}(t^{'})\| < \epsilon \tag{A.10}$$

where $\|\cdot\|$ denotes some metric. If a quantity is continuous and changes only slightly (and slowly) near $\boldsymbol{x}_{traj}(t^{'})$, Eq. A.10 can be relaxed to allow a matching time period δ with:

$$\|\boldsymbol{x}_{obj}(t_1^{'}) - \boldsymbol{x}_{traj}(t_2^{'})\| < \epsilon \tag{A.11}$$

with $t_1^{'}, t_2^{'} \in [t^{'} - \frac{\delta}{2}, t^{'} + \frac{\delta}{2}]$ and $t^{'}$ satisfying Eq. A.10. The task then is to find all times $\hat{t}$ for which Eq. A.11 holds for some choice of ϵ, δ. Generally, trajectories are given as a set of discrete points at discrete times, whose resolution may be less than either ϵ or δ. In such a case, the trajectories can be linearly interpolated between points to a higher resolution. How fine this interpolation

needs to be depends on the choice of ϵ, δ. Theoretically, a simple brute–force approach is possible, but generally not feasible due to the high number of $\boldsymbol{x}_{obs}(t)$ and the interpolated $\boldsymbol{x}_{traj}(t)$. The first step is to choose a suitable metric $\|\cdot\|$ which ideally can cover both spatial and temporal closeness. We would want the spatial components (governed by ϵ) to be independent of the temporal component (governed by δ). The Chebychev metric $d_\infty(x, y)$ between two points x, y, given by:

$$d_\infty(x, y) := \max_i |x_i - y_i| \tag{A.12}$$

does not correlate individual components and is therefore a suitable choice. By scaling (dividing) the spatial components of x by $\epsilon/2$ and the temporal components by $\delta/2$, all points which satisfy Eq. A.12 lie within a hyper–sphere of radius 1/2 (with respect to d_∞). The points of $\boldsymbol{x}_{traj}(t)$ and $\boldsymbol{x}_{obs}(t)$ scaled this way can then be binned into bins with width 1/2. Two points in neighbouring bins then are at most 1 unit apart from each other, i.e. fulfill Eq. 4.7. Such a discretization not only allows one to determine matches efficiently, but also provides the optimal interpolation resolution of $\boldsymbol{x}_{traj}(t)$. Given the binned starting point $\boldsymbol{x}_1$ and the binned end point $\boldsymbol{x}_2$ of a linear segment of $\boldsymbol{x}_{traj}(t)$, all points which lie on a line between them can be calculated by Bresenham's line algorithm (Bresenham, 1965) in terms of bins. Therefore, any point $\boldsymbol{x}_3$ which lies close to the linear segment between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ in terms of d_∞ also lies close to at least one of these derived points. A higher interpolation resolution yields no advantage. This yields the optimal number of points to check. Bresenham's line algorithm can also be easily adapted to work in higher dimensions. The other advantage of the Chebychev metric is that it is one of the metrics for which a k–d tree (Bentley, 1975) can be constructed. A k-d tree is a binary space-partitioning data structure that recursively divides a set of k–dimensional points along coordinate axes, which yields a hierarchical index that enables efficient nearest–neighbour and range searches. It is an efficient way to perform the necessary operations as follows: First, a k–d tree is created from the binned $\boldsymbol{x}_{obs}(t)$. Then a ball search, i.e. all points which lie in a hyper–sphere (or ball) of radius 1 around a point, is performed for all interpolated $\boldsymbol{x}_{traj}(t)$. Those $\boldsymbol{x}_{traj}(t)$ for which the search does not yield any points are therefore not matched. To find the $\boldsymbol{x}_{obs}(t)$ which best match a particular $\boldsymbol{x}_{traj}(t)$, a nearest–neighbour search is performed for the matched $\boldsymbol{x}_{traj}(t)$. By construction these nearest neighbours are at most 1 unit apart from the $\boldsymbol{x}_{traj}(t)$. The desired quantity of these $\boldsymbol{x}_{obs}(t)$ can then be transferred to the $\boldsymbol{x}_{traj}(t)$ by either identification with the nearest neighbour or by averaging of all suitable neighbours. The main advantage of this algorithm, besides the aforementioned fit for this particular problem, is its high performance. On a regular computer we found that this

setup can perform around 8.3 million checks per second without explicit parallelization. 24 h of observations (228,580 points) can be matched to 65,160 trajectories with 2 minutes temporal and 1° spatial resolution in around 2 minutes. A caveat for this approach is a varying real distance between two points $\boldsymbol{x}_1, \boldsymbol{x}_2$, whose positions are generally given by longitude and latitude. The spatial resolution is therefore given as an angular precision. The algorithm assumes that for small distances the curvature of the Earth can be neglected. There are minor differences between matched points depending on their latitude, such that matching in the tropics is less strict than at the poles. We do not deem this significant for matching satellite observations with trajectories, particularly with respect to actual satellite resolution.

A.3 Modified DBSCAN algorithm

The following is a modification of the DBSCAN algorithm for feature detection with respect to data barriers. The core problem is to find coherent, connected structures in data. Generally, the DBSCAN algorithm partitions data into ϵ–connected sets of sufficient size. It only considers the data points which it is provided with. For feature detection we may have a larger set of points $\boldsymbol{x}$, which we may divide into valid points $\boldsymbol{x}_{\text{valid}}$ and invalid points $\boldsymbol{x}_{\text{invalid}}$, here into points with valid mixing state $M < 0.425$ and points with invalid mixing state $M > 0.425$. The clustering is normally performed only on $\boldsymbol{x}_{\text{valid}}$. Connectivity can therefore be broken by data gaps or by large sections of invalid data, which appear as data gaps as well. Let us assume we have two coherent sets of data $\boldsymbol{X}_1, \boldsymbol{X}_2$, which are ϵ–connected with respect to some small ϵ. We would like to merge these sets if they are separated via data gaps, and not merge them if they are separated via invalid data. Distance alone is an insufficient criterion, as it does not account for invalid data. If we imagine invalid data as deep holes that we cannot pass and data gaps as shallow pits which are traversable, the distance between edges does not indicate whether the space between them is shallow or deep. In this image, the gap between sets is 'traversable' if there exists a path from a point on $\boldsymbol{X}_1$ to a point on $\boldsymbol{X}_2$ that leads only through shallow pits. Such a problem can be solved iteratively by searching for nearby shallow pits and approaching the other side. Mathematically this is best solved by using directed graphs. Alternatively, the algorithm can be simplified by considering only straight connections between $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$. This neglects connectivity through curved data gaps, which is not an issue if the data gaps are at least radially convex. By binning the data (both $\boldsymbol{x}_{\text{valid}}$ and $\boldsymbol{x}_{\text{invalid}}$) we can determine those regions which contain invalid data and those

which are simply data gaps (whose bins do not contain any points). The bins making up the linear connections can be calculated using Bresenham's line algorithm (Bresenham, 1965). These linear connections between pairs of points in $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are then checked for bins containing invalid data until a valid connection is found, in which case the search can be terminated and $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are merged. If no such valid connection is found the sets are considered separated by invalid data and not merged. An example where this algorithm is important is indicated in Fig. 4.10b in the southern hemisphere, where many fragments appear close, but remain separated through invalid regions (not explicitly shown). Unreasonable merging of fragments exacerbates the problem of keeping structures distinct, easily leading to the formation of a single superstructure which no longer resolves individual filaments.

A.4 Convolution

A convolutional neural network (CNN) uses convolution to learn information about the structure of given data. A general overview of different aspects of convolution is given in Fig. A.1: The convolution of two functions f and g is defined as

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau)\, g(t-\tau)\, \mathrm{d}\tau. \tag{A.13}$$

Intuitively, convolution is an operation that "slides" one function over another, multiplies them pointwise where they overlap, and integrates (or sums, in the discrete case) this product, producing a third function that describes how the two originals overlap as a function of shift. In Fig. A.1, the input data (representing f) is shown in blue. The function g with which f is convolved is a simple kernel function, i.e. a function which 'selects' an interval (or set, in the discrete case). In this example, the kernel has a fixed size of three. The resulting convolution is shown in Fig. A.1 in orange. A simple convolution (with stride and dilation of one), shown in the first row, creates a mapping based on the previous three inputs. Each input therefore contributes to (up to) three entries of the convolution. The *dilation* parameter determines which elements within the convolution kernel are actually used for the convolution. A dilation of two (second row) ignores every second entry of the kernel, resulting in only the first and the third element in the kernel being used in the convolution. The *stride* parameter determines how the kernel is 'slid' across the input. A stride of two (and dilation of one, third row) moves the kernel two units after each convolution, resulting in fewer convolution entries. Every input now contributes to between one and two convolution en-

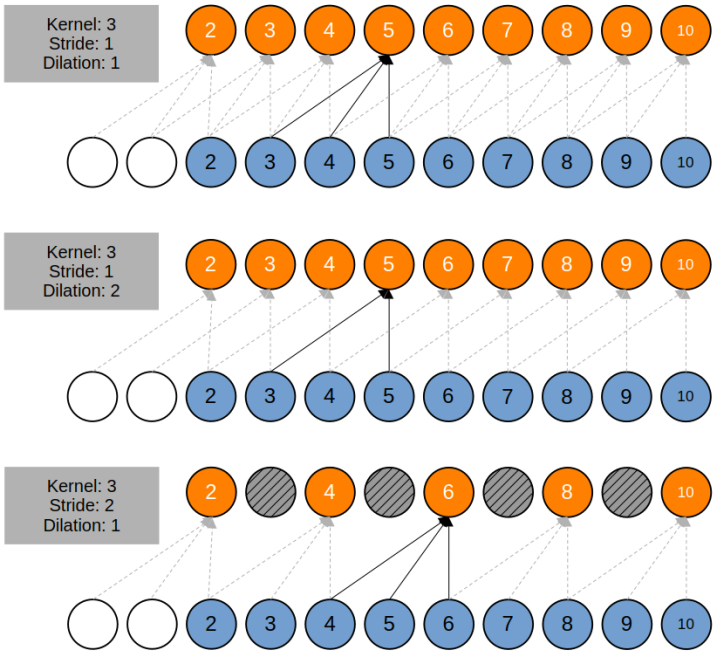

Figure A.1: Illustration of different kernel parameters: First row: A basic convolution with a convolution kernel of size three and stride/dilation of one. The input is indicated in blue, the resulting convolution is indicated in orange. The contribution of each input to the convolution is indicated by grey arrows. Second row: Same setup, but with dilation of two. Third row: Same setup, but with stride of two.

tries. A convolutional layer of a CNN operates on the same principle. However, different instances of convolution using different convolution weights are performed simultaneously, yielding multiple (as many as the layer size) convolutions. Every convolutional layer therefore increases the complexity of its output compared to its input. It is therefore advisable to increase dilation across consecutive instances of convolutional layers, and ultimately condense the many convolutions into a more feasible output.

A.5 NARX models

A *Non–linear AutoRegressive eXogenous* (NARX) model is a neural network architecture defined by exogenous inputs. It is used to predict future values of a quantity s_0 based on past values of s_0 and past values of external (exogenous) quantities $s_1, ..., s_n$, which act as covariates to s_0. In this study, we use a particular design of NARX model which utilizes *information distillation*. In principle, such a NARX model consists of two different architectures: The first one is a prediction model that accepts the full input (both endogenous s_0 and exogenous inputs $s_1, ..., s_n$) to predict s_0. This model is trained on the provided training data. The second architecture is a second autoregressive model, which simply predicts future s_0 from past s_0, such that autoregressive predictions are possible. This model, however, is not trained on the actual training data, but on the predictions of the first model. It therefore inherits the knowledge learned from the exogenous model and 'distills' it down to be applicable to the endogenous inputs only. It is trained to reproduce the predictions of the exogenous model based on the endogenous inputs only. The NARX model provides an additional uncertainty term in the form of the average performance of the exogenous model, which ultimately limits the performance of the autoregressive predictions. There exist different ways to design a NARX model, which cannot be covered in this thesis, but are definitely worth a read.

A.6 Time series of trajectories used for filament detection validation

The following figures show the full trajectories used in Chap. 4. The left–hand side shows the position of those air parcels identified as filaments, the right–hand side shows all air parcels, with their mixing state M shown in colour–code. Each row shows the positions and the updated mixing states of the air parcels after further 24 h.

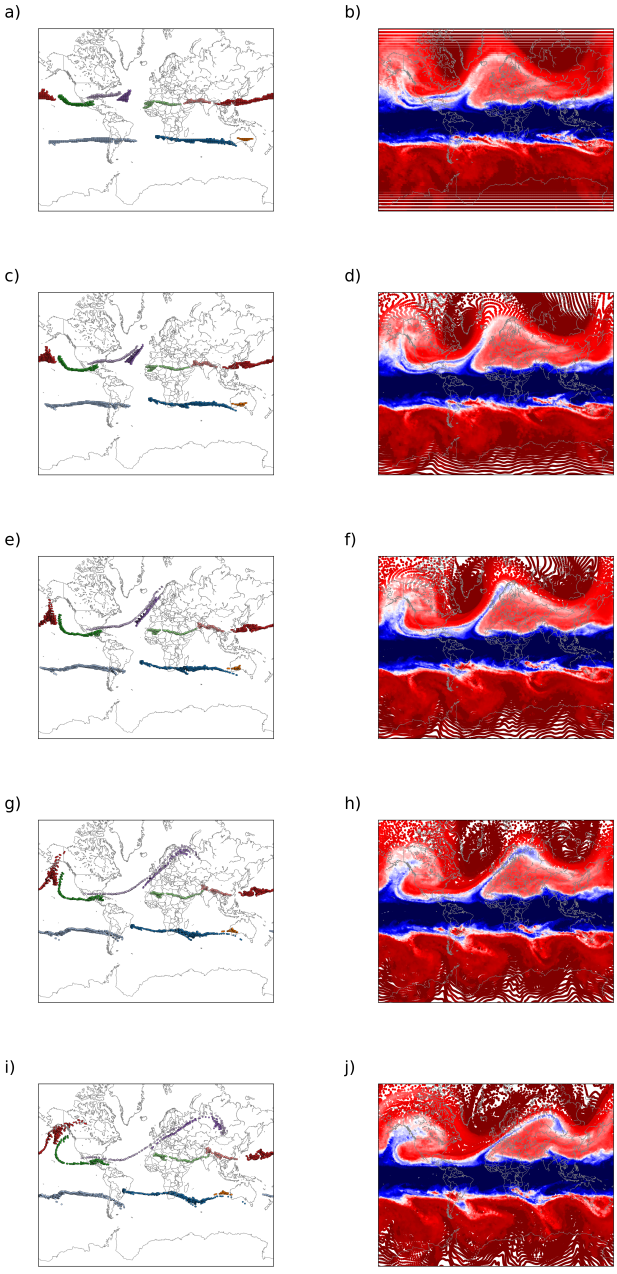

Figure A.2: First part of the trajectories (1 February 2011 – 5 February 2011)

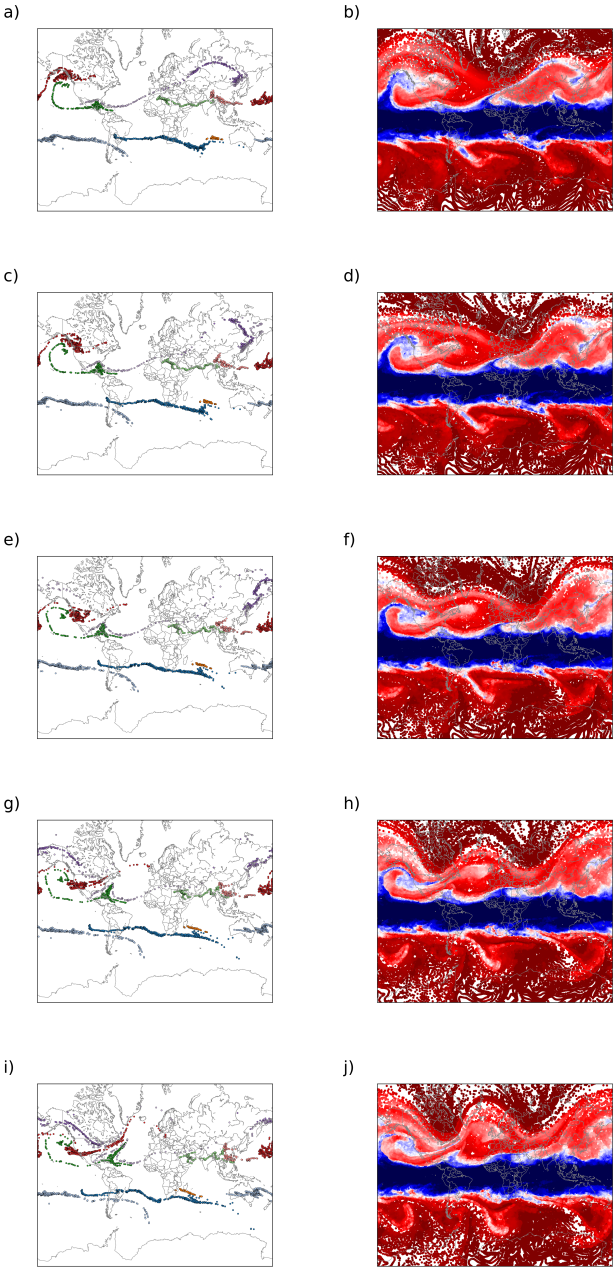

Figure A.3: Second part of the trajectories (6 February 2011 – 10 February 2011)

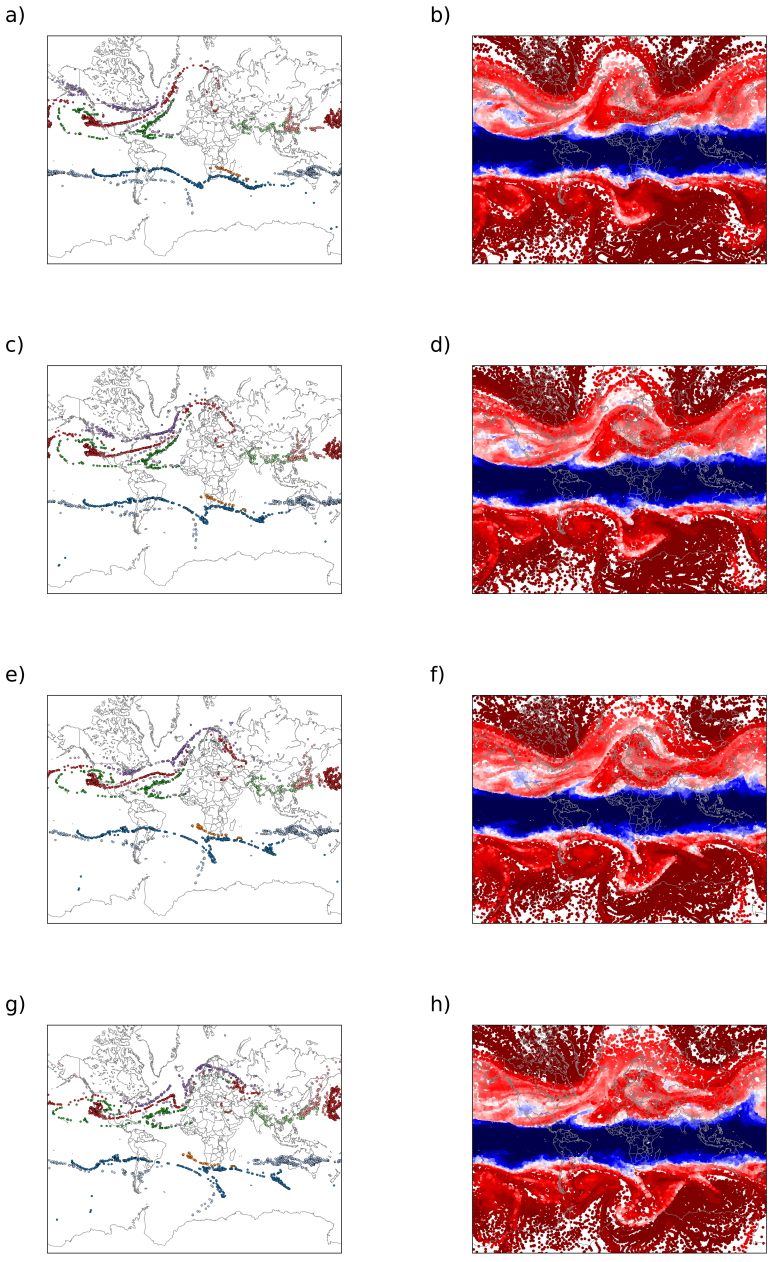

Figure A.4: Third part of the trajectories (11 February 2011 – 14 February 2011)

A.7 Regularization scheme

As stated beforehand, regularizing a probabilistic MLP with noise comparable to the noise inherent in the validation data yields accurate estimates of the aleatoric uncertainty. Let us assume that the real AoA is some function $f(x)$ plus some random, Gaussian distributed noise $\epsilon(x)$, where x denotes the LLT measurements.

$$y = f(x) + \epsilon(x), \quad \text{with } \epsilon(x) \sim \mathcal{N}\left(0, \sigma^2_{\text{true}}(x)\right) \tag{A.14}$$

We denote the real spread of these measurements by σ_{true}. The true, conditional probability of the real AoA is then given by:

$$p(y \mid x) = \mathcal{N}\left(f(x), \sigma^2_{\text{true}}(x)\right) \tag{A.15}$$

The probabilistic model predicts both the expectation value μ_θ and σ_θ. Let us assume we have an infinite number of values to predict, such that the epistemic uncertainty vanishes. If we optimize the model using the negative log–likelihood (see Eq. 5.1) we get the following expectation values:

$$\mathbb{E}_{y\sim p(y|x)}[L(\theta)] = \mathbb{E}_{y\sim p(y|x)}\left[\frac{1}{2}\ln\left(2\pi\sigma^2_\theta(x)\right) + \frac{\left(y-\mu_\theta(x)\right)^2}{2\sigma^2_\theta(x)}\right] \tag{A.16}$$

Since the logarithmic term is independent of y, we can rewrite:

$$\mathbb{E}_{y\sim p(y|x)}[L(\theta)] = \frac{1}{2}\ln\left(2\pi\sigma^2_\theta(x)\right) + \frac{1}{2\sigma^2_\theta(x)}\mathbb{E}_{y\sim p(y|x)}\left[\left(y-\mu_\theta(x)\right)^2\right] \tag{A.17}$$

We assume that the model is sufficiently trained such that it minimizes the loss function:

$$\frac{\partial}{\partial\mu_\theta(x)}\mathbb{E}_{y\sim p(y|x)}[L(\theta)] = \frac{1}{2\sigma^2_\theta(x)}\mathbb{E}_{y\sim p(y|x)}\left[-2\left(y-\mu_\theta(x)\right)\right] = 0 \tag{A.18}$$

$$\implies \mathbb{E}_{y\sim p(y|x)}[y] - \mu_\theta(x) = 0 \tag{A.19}$$

$$\implies \mu^*_\theta(x) = \mathbb{E}_{y\sim p(y|x)}[y] = f(x) \tag{A.20}$$

Thus, the expectation value predicted by the probabilistic model corresponds to the real AoA value. Replacing μ_θ by $f(x)$ and using $\mathbb{E}[(y-f(x))^2] = \sigma^2_{true}$ we similarly get:

$$\frac{\partial}{\partial\left(\sigma_{\theta}^{2}(x)\right)}\mathbb{E}_{y\sim p(y|x)}[L(\theta)]=\frac{1}{2\sigma_{\theta}^{2}(x)}-\frac{\sigma_{\text{true}}^{2}(x)}{2\left(\sigma_{\theta}^{2}(x)\right)^{2}}=0 \tag{A.21}$$

$$\implies \frac{1}{2\sigma_{\theta}^{2}(x)}\left(1-\frac{\sigma_{\text{true}}^{2}(x)}{\sigma_{\theta}^{2}(x)}\right)=0 \tag{A.22}$$

$$\implies \sigma_{\theta}^{2*}(x)=\sigma_{\text{true}}^{2}(x) \tag{A.23}$$

Thus, the predicted variance corresponds to the true variance. The model learns the aleatoric uncertainty of the prediction. During training the model sees some $\epsilon_{\text{train}} \sim \mathcal{N}\left(0, \sigma_{\text{train}}^{2}(x)\right)$, and thus learns σ_{train}. If $\sigma_{\text{train}} \neq \sigma_{\text{true}}$, then naturally this aleatoric uncertainty is no longer accurately reflected. A more intuitive view is presented in Fig. A.5, which shows the residuals of the predictions of a model regularized with 1% regularization noise and the predicted σ_{train} as uncertainties: Since the model converged, all predictions are centred around zero. The model is considered

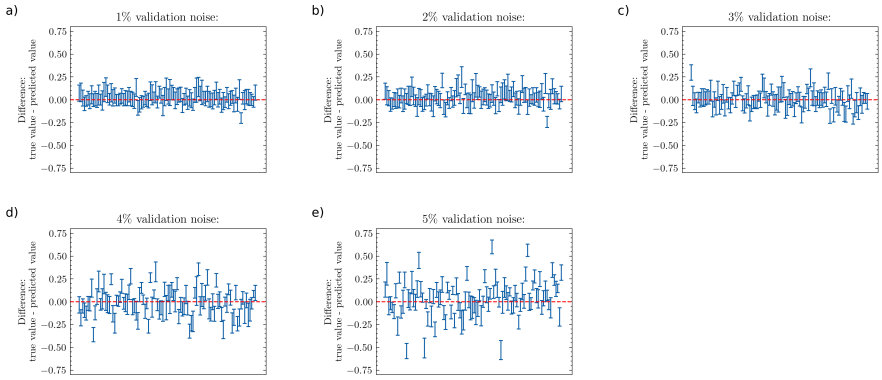

Figure A.5: Illustration of the relationship between estimated and true aleatoric uncertainty of probabilistic MLPs. Shown are the residuals between true and predicted AoA, with the estimated aleatoric uncertainty (in the form of the standard deviation of the predicted distribution) indicated by error bars. The true aleatoric uncertainty is well estimated if the residuals are distributed within around one standard deviation.

calibrated (low epistemic uncertainty) if the aleatoric uncertainty reflects the spread of the predictions. We see good agreement for 1% and 2% validation noise, beyond which it becomes clear that the model no longer captures the spread of its predictions precisely.

A.8 MOE AE setup and configuration

The MOE AE network (in the following simply MOE) allows one to perform many local embeddings simultaneously to capture a hypothetical, underlying parametrization of the training domain. In this section, the architecture of the MOE is discussed in detail, as well as the technical measures taken to provide a stable convergence.

An MOE consists of essentially two parts: the first part is the *router network*, which often takes on the form of a simple MLP, as is the case here. The second part is the autoencoders, whose number is determined by the MOE. During training each input $\boldsymbol{x}$ is fed into each autoencoder, and the reconstruction (decoder) loss is calculated. Depending on how well each autoencoder performed, the router network predicts weights of each AE, with high performance yielding a high weight. The predicted value is then the weighted mean of the embeddings of each AE. Details on this process are described at a later point. If a particular AE has performed well encoding some particular $\boldsymbol{x}$, the router network anticipates that this AE will perform similarly well for a similar input $\hat{\boldsymbol{x}}$, and thus puts more weight on this AE's encoding. Each AE can therefore try to specialize in a particular part of the training domain, i.e. become an expert. An expert is allowed to perform badly for some x if there is another expert which performs well with it. The performance of a MOE is therefore the best performance achieved by any of its expert.

How does such an architecture allow us to learn AoA from a set of LLTs? Let us consider the theoretical foundation of an autoencoder in more detail, and discuss the implications for a MOE: Let us assume that all LLTs $\boldsymbol{x}$ are only a function of time t, i.e. their mixing ratio inside the stratosphere is governed by some function f satisfying:

$$\frac{d\boldsymbol{x}}{dt} = f(\boldsymbol{x}) \tag{A.24}$$

In the absence of sources, f is always negative, and thus injective. An AE consists of an encoding function E and a decoding function D, which in an ideal scenario satisfy:

$$D(E(\boldsymbol{x})) = D(b) = \boldsymbol{x} \tag{A.25}$$

where $b = E(\boldsymbol{x})$ is the embedding of the AE. We can thus express Eq. A.24 in the latent space b as:

$$\frac{db}{dt} = \nabla_x E(\boldsymbol{x}) \cdot \frac{d\boldsymbol{x}}{dt} = \nabla_x E(\boldsymbol{x}) \cdot f(\boldsymbol{x}) \tag{A.26}$$

Since the decay is irreversible, $\frac{d\boldsymbol{x}}{dt}$ is never zero. If the AE always performs well, i.e. if Eq. A.25 always holds, then $\nabla_x E(\boldsymbol{x})$ is never zero. Therefore, $\frac{db}{dt}$ is always positive, and thus monotonic. If a parametrization $b^{'}(t)$ exists, then an AE will be able to find an embedding $b(\boldsymbol{x})$ which is a monotonic function of $b^{'}(t)$. The time t and the embedding b are diffeomorphic. The idea behind a MOE is that while this monotonic function does not trivially relate to t, it is locally linear in t, since it is a continuous function. One may think of the Taylor expansion. A finely resolved (local) embedding b thus reflects the curvature of the manifold with respect to some parametrization. The goal now is to ensure that all local embeddings are continuous with respect to x. In order to achieve this, the MOE AEs are trained on the same domain, but with different weights indicating different sections of the domain. Since the embeddings are continuous, their sum (weighted mean) is also continuous. Since the domains of each AE overlap, there are no discontinuities in the embeddings between domains. If we force non–negative weights in the output layer of each encoder, each embedding is positive, and thus avoids averaging of positive and negative values that would counter each other.

An additional constraint can be introduced by adding a penalty to the AEs if the resulting embedding b is not monotonic in time. We can use SF_6 as a monotonic function in time to validate whether b is in turn monotonic with respect to SF_6. Adding this constraint stabilizes the training process. Note that the MOE does not learn anything from this comparison to SF_6; it only constrains the embedding.

Instead of performing a full dimensionality reduction on all LLTs, some LLTs can be used as covariates to the manifold. The relationship between CFC−11 and AoA for example is not very compact. The inclusion of CFC−11 introduces (unwanted) degrees of freedom to the AEs, such that the decoder even with an otherwise perfect embedding cannot reconstruct the original LLTs well. However, CFC−11 does convey information and provides a meaningful contribution to the manifold. We therefore modify the training loss of the decoder to only consider the reconstruction of a few LLTs, thus treating the rest as covariates.

The main part of a MOE is the distribution of the available experts to different parts of the domain. While only those features that are quasi–bijective contribute to the final embedding, additional features can be used to inform the MOE how to distribute the different experts. We provide the router network of the MOE with information about latitude and potential temperature, which an ordinary AE cannot utilize. The training domain of each expert is thus determined by the spatial location of the training data. To avoid the so–called *expert collapse*, where the domain is

split among very few experts, we add an additional entropy penalty to the training loss to penalize the MOE for using too few experts.

The final architecture is a physics–informed (due to the monotonicity loss and gating inputs), covariate (due to the existence of covariates for the embedding) Mixture–of–Experts of Autoencoders network.

A.9 Gaussian mixture model

A statistical mixture model is a probabilistic model which describes a statistical population as the combination of many, statistically distinct sub–populations. In simple terms, given an N–dimensional dataset **X**, a mixture model assumes that the dataset **X** was generated by sampling from k random variables with N–dimensional probability densities $(\boldsymbol{D_1, D_2, ..., D_k})$ with different (mixing) weights $(w_0, ..., w_k)$ (Baxter, 2010). The statistical properties of each component distribution $(\boldsymbol{D_k})$ are inferred directly from the dataset $(\mathbf{X})$. A *Gaussian* mixture model assumes each component distribution to be a Gaussian distribution $(\boldsymbol{\mathcal{N}})$. The mixture distribution from which **X** was generated is then given by:

$$\begin{aligned} D(\mathbf{X} \mid \boldsymbol{\mu_1}, ..., \boldsymbol{\mu_k}, \boldsymbol{\Sigma_1}, ..., \boldsymbol{\Sigma_k}, w_1, ..., w_k) &= \sum_{i=1}^{k} w_i \cdot (2\pi)^{-\frac{N}{2}} \cdot \det(\boldsymbol{\Sigma_i})^{-\frac{1}{2}} \cdot \exp\left(-\frac{1}{2}(\mathbf{X} - \boldsymbol{\mu_i})^T \boldsymbol{\Sigma_i^{-1}} (\mathbf{X} - \boldsymbol{\mu_i})\right) \\ &= \sum_{i=1}^{k} w_i \, \boldsymbol{\mathcal{N}}(\mathbf{X} | \boldsymbol{\mu_i}, \boldsymbol{\Sigma}_i) \end{aligned} \tag{A.27}$$

where $\boldsymbol{\mu_i}, \boldsymbol{\Sigma_i}, w_i$ are the mean vector, covariance matrix and mixing weight of each component, with $\sum_{i=0}^{k} w_i = 1$. The number of components k is the free parameter of the statistical mixture model. In contrast to a standard Gaussian mixture model, which determines its parameters using expectation–maximization (Dempster et al., 1977), a Bayesian Gaussian mixture model (BGMM) determines its parameters via variational inference. The necessary a priori values for this Bayesian approach can be provided by a Dirichlet process. A trained BGMM provides probabilities for each data point $\boldsymbol{x}$ according to Eq. (A.27). Since it is assumed that a point must have been sampled from one of the distributions, these probabilities can be normalized:

$$p_i(\boldsymbol{x}) = \frac{w_i \, \boldsymbol{\mathcal{N}}(\boldsymbol{x} | \boldsymbol{\mu_i}, \boldsymbol{\Sigma}_i)}{\sum_{j=1}^{k} w_j \, \boldsymbol{\mathcal{N}}(\boldsymbol{x} | \boldsymbol{\mu_j}, \boldsymbol{\Sigma}_j)} \tag{A.28}$$

A BGMM can also be used as a classifier by assigning each point $\boldsymbol{x}$ a label according to the most likely component the point was sampled from. The label is assigned according to:

$$l(\boldsymbol{x}) = \mathrm{argmax}(\log(w_k) + \log(\mathcal{N}_k)) \tag{A.29}$$

The maximum value of Eq. A.28 then corresponds to the probability that the point was accurately labelled. For disjoint distributions Eq. A.28 typically takes values between 1 (high confidence) and 0.5 (ambiguous labelling between two classes). In this thesis, this quantity is referred to as the class alignment.

The classifier works well for all points which have suitably high probabilities to be sampled from any one of the component distributions and thus are well represented by the model. If points are not well represented by the model, each probability $\mathcal{N}_i$ vanishes and Eq. A.29 is dominated by its second term. Labels are then determined by which distribution $\mathcal{N}_i$ vanishes most slowly, regardless of the mixture weights w_i, which leads to sudden changes in labelling on the edge of phase space. Fig. A.6 illustrates the effect for a GMM in one dimension and two components.

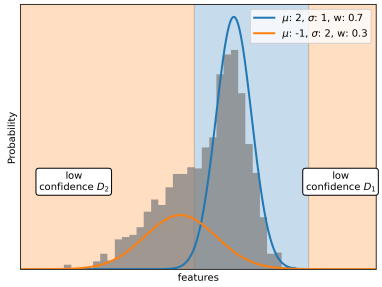

Figure A.6: Illustration of unphysical labelling artefacts. Shown are random samples of two Gaussian distributions with different parameters (see legend). The labels derived from Eq. A.29 are indicated in the background in colour–code. For low label probability for either distribution (right side) the broader distribution determines the label, resulting in a jump in labelling.

Since each Gaussian distribution is compact in phase space, the labels should be compact likewise. To account for this behaviour we determine labels using Eq. A.29 only for points with at least one sufficiently high probability and determine labels for all other points using a nearest–neighbour classifier. Only few points require such a correction and the resulting labels are compact in phase space.

A.10 DBSCAN

The Density–Based Spatial Clustering of Applications with Noise (DBSCAN) is a density–based clustering algorithm, which partitions data into an arbitrary number of coherent structures. The DBSCAN algorithm has two parameters: a maximum distance ϵ and the minimum number of neighbours m. The algorithm starts by finding points which have at least m neighbouring points within a distance ϵ, which are labelled as *core points*. Core points that lie within a distance ϵ (i.e. are ϵ–reachable) of other core points are then merged into clusters. Points that are ϵ–reachable, but lack the minimum number of ϵ–reachable neighbours m, are added to the clusters but cannot be used to merge clusters further. Points that are not ϵ–reachable by any cluster are considered outliers. Fig. A.7 illustrates the DBSCAN algorithm for a simple example of $\epsilon = 1$ and $m = 3$:

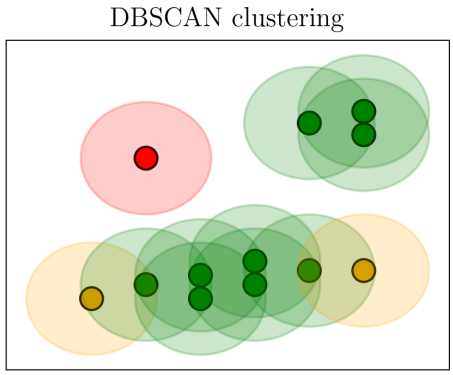

Figure A.7: Illustration of the DBSCAN algorithm. Data points are indicated by dots. An ϵ–environment is indicated by the transparent circles around each dot. The type of classification of each dot is indicated in colour–code: Points, which are ϵ–reachable by at least m neighbours are considered core points (green). Points, which are ϵ–reachable by at least one core point, but are not core points themselves, are considered border points (yellow). Points, which are ϵ–reachable by fewer than m points, of which none are core points, are considered outliers (red).

DBSCAN determines all points within the data that form coherent structures. The main applications for DBSCAN are outlier detection, i.e. finding isolated points inside the data, and partitioning of data. In particular, it can be used to find coherent structures in irregularly spaced data, including data with irregular temporal sampling, e.g. in satellite observation data.

A.11 Properties of the PCA with respect to clustering

In Sect. 3.2, it was stated that certain statements about the original data can be derived from the structure of its PCA. These statements were mostly derived from experience and have clear limitations, which is why these limitations are discussed in the following in more detail. Formally, we consider only such PCAs which retain the majority of the data variance as described in Jolliffe and Cadima (2016a). Let $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ be centred data with covariance Σ, and let PCA choose an orthogonal basis $\mathbf{U}= [\boldsymbol{u}_1, ..., \boldsymbol{u}_p]$ with eigenvalues $\lambda_1 \geq ... \geq \lambda_p$. Let the cut–off of the PCA be k. The $(p-k)$–dimensional manifold z is then given by:

$$\boldsymbol{z} = \boldsymbol{U}_k^T \boldsymbol{x} \tag{A.30}$$

where $\boldsymbol{U}_k$ denotes the first k principal components. Since k is chosen to retain the majority of the variance, we get:

$$\hat{\boldsymbol{x}} = \boldsymbol{U}_k \boldsymbol{z} \approx \boldsymbol{x} \tag{A.31}$$

and

$$\sum_{j=1}^{k} \lambda_j / \sum_{j=1}^{p} \lambda_j \approx 1 \tag{A.32}$$

From these relations we can infer: 1. The distances and inner products of $\boldsymbol{X}$ are well retained in $\boldsymbol{z}$:

$$\|\boldsymbol{x}_i - \boldsymbol{x}_j\| = \|\boldsymbol{z}_i - \boldsymbol{z}_j\| + \delta \tag{A.33}$$

where δ is some small residual. Thus, structure similarity in X is retained in z.

2. If $\boldsymbol{z}$ can be separated into two orthogonal manifolds with centers $\boldsymbol{z}_0^{(1)}, \boldsymbol{z}_0^{(2)}$, i.e. can be expressed by a 1–D affine subspace:

$$\boldsymbol{S}_1 = \{\boldsymbol{z}_0^{(1)} + \alpha \boldsymbol{v}_1 : \alpha \in \mathbb{R}\} \tag{A.34}$$

$$\boldsymbol{S}_2 = \{\boldsymbol{z}_0^{(2)} + \beta \boldsymbol{v}_2 : \beta \in \mathbb{R}\} \tag{A.35}$$

with $\boldsymbol{v}_1 \perp \boldsymbol{v}_2$, then these manifolds are also well described by similar 1–D affine subspaces in $\boldsymbol{X}$:

$$\boldsymbol{L}_1 = \{\boldsymbol{x}_0^{(1)} + \alpha \boldsymbol{U}_k \boldsymbol{v}_1\} \tag{A.36}$$

$$\boldsymbol{L}_2 = \{\boldsymbol{x}_0^{(2)} + \beta \boldsymbol{U}_k \boldsymbol{v}_2\} \tag{A.37}$$

A similar argument follows for non–1–D subspaces. If the bulk of variance is contained in these subspaces, we can infer that there exist two nearly orthogonal linear modes of variation in $\boldsymbol{X}$. However, this does not imply that these modes are strictly orthogonal in $\boldsymbol{X}$. The discarded variation contained in δ may also contribute in some way. 3. Clusters in $\boldsymbol{X}$ are retained in $\boldsymbol{z}$. Let $\boldsymbol{C}_1, ..., \boldsymbol{C}_m$ be subsets in $\boldsymbol{z}$ with pairwise distances large relative to their diameters (formal cluster description). Then the preimages $\boldsymbol{U}_k\boldsymbol{C}_j$ are also clusters, since $\boldsymbol{U}_k$ is an isometry from $\mathbb{R}^k$ into its image. Overlap along discarded components δ cannot be ruled out. This is also limited to Euclidean metrics, and generally does not need to hold with respect to nonlinear class boundaries. 4. A strong structure also implies that $\boldsymbol{\Sigma}$ is close to low rank, i.e. for those structures are few eigenvalues dominate. Formally:

$$\boldsymbol{\Sigma} \approx \sum_{j=1}^{k} \lambda_j \boldsymbol{u}_j \boldsymbol{u}_j^T \tag{A.38}$$

These dominating eigenvalues indicate a 'direction' in $\boldsymbol{X}$ along which the structure is positioned (the combination of the $\boldsymbol{x}_i$ spanning the $\boldsymbol{u}_j$). For multiple such structures we can infer that these modes are mostly uncorrelated, which implies that the data contain a mixture of these modes.
There are strict limitations for deriving statements regarding $\boldsymbol{X}$ from $\boldsymbol{z}$. Geometry does not strictly necessitate causal or physical mechanisms. While for observational data this is plausible, geometry in $\boldsymbol{z}$ can be introduced as artefacts. Such artefacts are prevalent if a linear manifold does not actually describe the underlying structure well. Non–linear manifolds may be poorly represented by a PCA, for example 'Swiss roll' data, which is in essence a 2–D manifold 'rolled up' in space. Finally, the orthogonality of the principal components $\boldsymbol{u}_j$ does not imply independence. Even orthogonal components may be strongly dependent on each other. This discussion was adapted from various sources. See Jolliffe and Cadima (2016a) for a general overview of PCA. Limitations were adapted from Björklund (2019) and Jolliffe and Cadima (2016b).

A.12 Effect of topology on clustering

The topology of a space is generally not considered in the vast majority of problems. Whenever a mathematical (or physical) problem is stated, it is generally implied that the underlying topology is flat. An exception is astronomy, where general relativity becomes relevant. Deep Embedding Clustering (DEC) networks utilize transformations of the topology to better structure data into clusters. The concept is visualized in Fig. A.8 based on a simple example:

Fig. A.8a shows the samples from two Gaussian distributions, indicated in colour–code. These

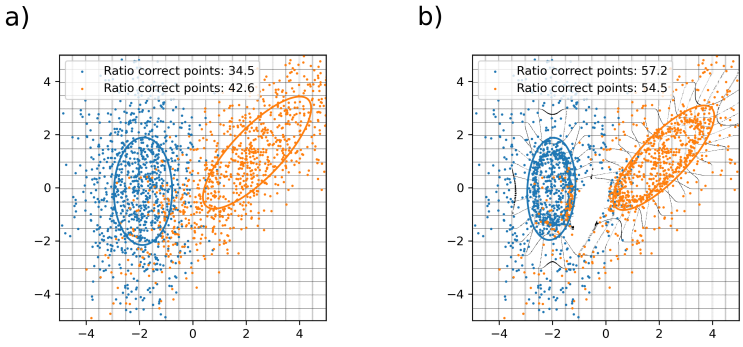

Figure A.8: Illustration of the effect of topology on clustering. Panel a) shows two slightly intertwined Gaussian distributions. A 1–σ environment is indicated by the coloured ellipses. The underlying topology (flat space) is indicated by the grid. The percentage of accurately classified points inside each ellipse is shown in the legend. Panel b) shows the same data, but with two topological lenses applied around the estimated shape of each blob. The topological distortions are visible in the grid. Another Gaussian fit is performed, and the resulting 1–σ environment as well as the percentage of accurately classified points are indicated similarly to panel b).

distributions show some overlap and are not strictly separable. Training a GMM on this data yields distributions whose covariances are indicated in the ellipses. As a simple concept we consider a point classified as belonging to either cluster if it is located within these ellipses, yielding 42.6% (close to the theoretically expected value) and 34.5% (a little lower, but reasonable given the deviation between fitted and true distribution). A simple topological transformation suitable for this problem is a *topological lens*, which 'contracts' space locally. Two lenses are assumed at the location of the fitted distributions and are set to focus the regions between the 1–σ environment and the 3–σ environment close to the former. The resulting distributions are shown in Fig. A.8b. The changes in topology are visible in the underlying grid, which is contorted. Therefore, the position of the data points is similarly changed, resulting in a new distribution, on which a new instance of GMM is trained. The 1–σ environments of this new fit are indicated similarly to Fig. A.8a. The points are now much more clustered compared to a flat topology, resulting in more compact clusters, which are easier to detect and contain a much higher fraction of the points (between 10 and 20 percentage points). In particular, there are now sections close to the edges of the topological lenses which are 'drawn' into the cluster. The border between the two clusters (white space in Fig. A.8b around (0, -1)) is now distorted such that a sharp (topological) border between the clusters exists. A DEC optimizes both topological transformation and cluster detection simultaneously. The shape of a cluster due to the clustering algorithm used (Gaussian for a GMM or

polygon for k–means clustering) becomes mostly irrelevant if the underlying topology is malleable. Two concentric circles on a sheet of paper may become clusters if we crumble the paper such that the edges meet in an almost drop–like shape. A DEC is a very potent tool for cluster detection, whose capabilities go beyond many traditional clustering algorithms. The caveat is much less transparency and potentially non–deterministic results.

A.13 Visualization of the manifold formed by Long–Lived Tracers (LLTs)

The methodology described in Sect. 5.3 to derive a representation of Age–of–Air (AoA) relies on the fact that the curvature of this approximately 1–D manifold can be exploited to identify regions with strong chemical sinks, indicated by regions of strong topological curvature. Sect. 5.3 describes the process and the reasoning in abstract terms. To obtain a more intuitive understanding, it is beneficial to visualize this manifold. As in Chap. 4, the full phase space is too high–dimensional for any conventional visualization. However, since the manifold is effectively 1–D, there exist higher–dimensional representations, which in essence represent the same manifold. By training an autoencoder with a latent dimension of three, such a 3–D manifold representation of the original 6–D data can be created, which includes all relevant information, since little to no information is lost during the projection to a lower dimensional subspace, as described by the Whitney Embedding theorem. Fig. A.9 shows this 3–D embedding from different angles. The physical 1–D embedding found by the Mixture–of–Experts autoencoders in Sect. 5.3 is shown in colour–code, the polar vortex (where the methodology fails) is marked in red:

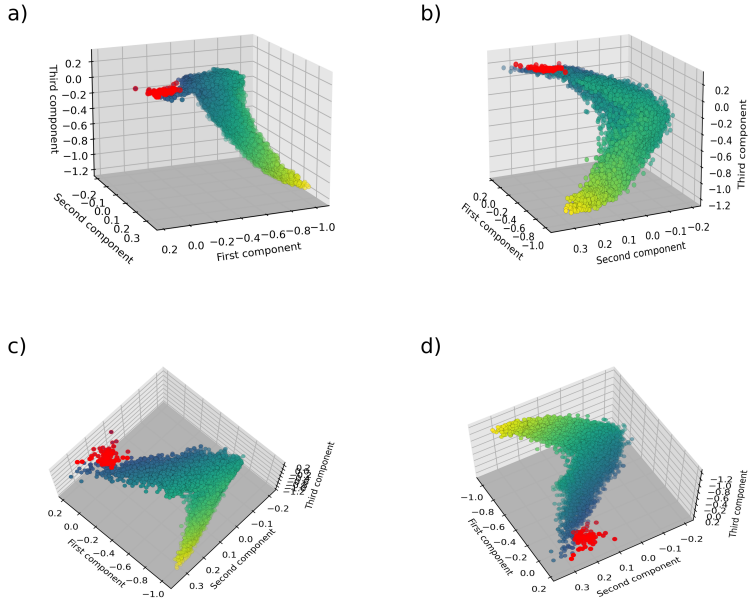

Figure A.9: 3–D visualization of the 6–D phase space spanned by the LLTs. Shown is a 3–D representation of the 6–D phase space (with effective dimensionality of one) obtained through a dedicated autoencoder with latent dimension of three. The three components do not natively correspond to any of the original LLTs. The embedding obtained through the MOE AEs is shown in colour–code. The values inside the polar vortex are marked in red.

List of Abbreviations

Abbreviation	Full Meaning
A	Asian summer monsoon outflow (classification)
AE	AutoEncoder (neural network)
AoA	Age-of-Air
ASM	Asian Summer Monsoon
ASMO	Asian summer monsoon outflow (characterization)
B	Blended Air
BDC	Brewer-Dobson circulation
BoB	Bay of Bengal
CAIRT	Changing Atmosphere Infra-Red Tomography Explorer
CAMS	Copernicus Atmosphere Monitoring Service
CLaMS	Chemical Lagrangian Model of the Stratosphere
CNN	Convolutional Neural Network
CT	Continental, Tropospheric Air
DBSCAN	Density-Based Spatial Clustering of Applications with Noise (algorithm)

Abbreviation	**Full Meaning**
DEC	Deep Embedding Clustering (neural network)
ECH	Eastern China
ECMWF	European Centre for Medium-Range Weather Forecasts
ESA	European Space Agency
GLORIA	Gimballed Limb Observer for Radiance Imaging of the Atmosphere
GMM	Gaussian Mixture Model
GOES West	Geostationary Operational Environmental Satellite
HALO	High Altitude and Long Range research aircraft
ICON	ICOsahedral Nonhydrostatic (modelling framework)
IND	Indian Subcontinent
LLT	Long-Lived Tracer
MAE	Mean-Absolute-Error
MLP	Multi-Layer Perceptron (neural network)
MOE	Mixture Of Experts (neural network)
MT	Maritime, Tropospheric Air
NARX	Non-linear autoregressive exogenous (neural network)
NASA	National Aeronautics and Space Administration
NIN	Northern INdian subcontinent
NIO	Northern Indian Ocean
NWP	North-West Pacific

Abbreviation	Full Meaning
PCA	Principal Component Analysis
PH13	Research flight 13 of the PHILEAS campaign
PH14	Research flight 14 of the PHILEAS campaign
PHILEAS	Probing High Latitude Export of air from the Asian Summer Monsoon
PV	Potential Vorticity
S	Stratopheric Air
SEA	South East Asia
SHAP	(SHapley Additive exPlanations
STRIVE	Stratosphere Troposphere Response using Infrared Vertically-resolved light Explore
TIB	Tibetan Plateau
TWP	Tropical Western Pacific
UTLS	Upper Troposphere – Lower Stratosphere

Symbols

Symbol	Description	Unit	Page
$\boldsymbol{a}$	Activation vector (output vector) of a neural network layer	—	24
A	Area enclosed by a tracer contour	m^2	13
$a(i)$	Mean distance from the i-th data element to all other points in the same cluster	—	20
τ	Apparent time, corresponds to the transport time as derived from long-lived tracers subject to chemical sinks	s	116
b	Bias (offset) of a neuron	—	24
b	Latent space embedding (representation vector) of an autoencoder	—	142
$b(i)$	Mean distance from the i-th data element to all points in the nearest neighboring cluster	—	20
β	Rossby parameter (meridional gradient of the Coriolis parameter)	$m^{-1}\ s^{-1}$	15
C_k	The k-th cluster (subset of data elements)	—	19

Symbol	Description	Unit	Page
$\boldsymbol{c}(t)$	Curve (trajectory) connecting key states in the phase space	variable	66
c_{gx}	Zonal component of the group velocity	m s^{-1}	17
c_{gy}	Meridional component of the group velocity	m s^{-1}	17
c_{gz}	Vertical component of the group velocity	m s^{-1}	17
c	Similarity threshold	—	18
C	Conservative tracer field concentration	—	13, 14
c_x	Zonal component of the phase velocity	m s^{-1}	16
c_y	Meridional component of the phase velocity	m s^{-1}	16
c_z	Vertical component of the phase velocity	m s^{-1}	16
$\boldsymbol{x}_i^j$	jth chemical state that can be reached by	variable	65
c_p	Specific heat capacity at constant pressure	J kg^{-1} K^{-1}	12
D	Decoding function of an autoencoder	—	117, 142
$\boldsymbol{D}$	Deformation tensor (symmetric part of the velocity gradient tensor)	s^{-1}	14
D	Probability density function of a statistical mixture distribution	—	144
$d_\infty(x, y)$	Chebychev distance metric (maximum norm)	variable	132
d_r	Relative distance of a point to the filament boundary (alpha shape) versus its skeleton	—	74

Symbol	Description	Unit	Page
$\boldsymbol{\delta}$	Local gradient vector (error vector) in a neural network layer	—	24
δ	Small geometric residue or reconstruction error magnitude	—	148
δ	Small temporal tolerance window for trajectory matching	s	76, 131
$\dim_{\mathrm{H}}(X)$	Hausdorff dimension of a metric space X	—	21
E	Encoding function of an autoencoder	—	117, 142
ϵ	Connectivity parameter of the DBSCAN algorithm	variable	133, 146
ϵ	Model error or deviation metric	—	22
ϵ	Maximum spatial distance threshold for trajectory matching	variable	76, 131
η	Learning rate for gradient descent optimization	—	25
f_0	Reference Coriolis parameter on a constant latitude	s^{-1}	16
f	First general mathematical function in a convolution	—	134
f	Correlation function (compactly supported)	—	63
$\boldsymbol{F}$	Frictional force per unit mass (or non-conservative forces)	$\mathrm{m\,s}^{-2}$	13
F	Mathematical mapping function (or model approximation)	—	113

Symbol	Description	Unit	Page
$f(x)$	True underlying deterministic mapping function	—	140
g	Second general mathematical function in a convolution	—	134
$\mathcal{H}^d_\delta(S)$	Hausdorff pre-measure of a subset S at scale δ for dimension d	—	21
$\mathcal{H}^d(S)$	The d-dimensional Hausdorff measure of a subset S	—	21
$\boldsymbol{J}_{\text{total}}$	Total flux density vector (advective and diffusive)	kg m^{-2} s^{-1}	14
K	Total number of clusters	—	19
k	Cut-off threshold representing the number of retained principal components	—	147
K	Magnitude of the wave vector	m^{-1}	16
k	Zonal wavenumber	m^{-1}	16
κ	Diffusivity (diffusion coefficient)	m^2 s^{-1}	15
κ_{eff}	Effective diffusivity due to flow deformation	m^2 s^{-1}	15
L	Cut-off radius (correlation length threshold)	m	65
L	Loss function of the neural network	—	24, 117, 140
L_{min}	Minimum possible contour length (undisturbed latitude circle)	m	15
L	Loadings (correlation coefficients) of a principal component analysis	—	41

Symbol	Description	Unit	Page
l_r	Relative length coordinate of a point projected onto the filament skeleton	—	74
l	Meridional wavenumber	m^{-1}	16
λ	Exponential growth rate (Lyapunov exponent)	s^{-1}	15
λ	Eigenvalue of some matrix	s^{-1}	131, 147
n	dimension of latent space	—	113
Γ	Projection of $\boldsymbol{X}$ to a latent space of dimension 1	variable	114
m	Minimum size for clusters in the DBSCAN algorithm	variable	146
M	Mixing state parameter (clipped phase space coordinate)	—	66, 113, 133, 136
$\mathcal{M}$	Predictive model or mapping function	—	22
m	Total number of discrete historical time steps in a sequence	—	83
m	Vertical wavenumber	m^{-1}	16
$\boldsymbol{\mu}_k$	Centroid (mean vector) of the k-th cluster	—	19
$\boldsymbol{\mu}_{\text{global}}$	Global centroid (mean vector) of the entire dataset	—	19
μ	Arithmetic mean of a dataset or tracer field	—	32, 144
μ	Predicted mean parameter of the Gaussian distribution	—	97, 140
N	Brunt-Väisälä frequency (buoyancy frequency)	s^{-1}	16

Symbol	Description	Unit	Page
N	Total number of elements in a set	—	18
$\|\| \cdot \|\|$	Distance metric or norm	—	19
ω	Angular frequency of the wave	rad s^{-1}	16
p	Pressure	Pa	12
p_0	Reference pressure (or surface pressure)	Pa	12
p	Original dimensionality (number of features) of the dataset	—	147
PV	Ertel's potential vorticity	K m^2 kg^{-1} s^{-1}	13
ϕ	Geographic latitude	$^\circ$ or rad	15
ϕ	Activation function of a neural network	—	24
ϕ_e	Equivalent latitude	$^\circ$ or rad	13
ψ	Geostrophic stream function	m^2 s^{-1}	16
Q	Specific diabatic heating rate	J kg^{-1} s^{-1}	12
q	Quasi-geostrophic potential vorticity	s^{-1}	16
R	Specific gas constant (for dry air)	J kg^{-1} K^{-1}	12
r	Spatial distance between two points or air parcels	m	65
R_E	Radius of the Earth	m	13
R	Stability function (amplification factor) of the numerical scheme	—	131

Symbol	Description	Unit	Page
R	Coefficient of determination (R-squared) from regressing a variable on all other predictors	—	50
ρ	Air density	kg m^{-3}	15
$\rho(r)$	Spatial covariance (or correlation coefficient) as a function of distance	—	65
$s(i)$	Silhouette coefficient of the i-th data element	—	20
$\boldsymbol{s}(t)$	Time series vector of a sequence of observed or modeled values	variable	83, 136
σ	Standard deviation of a dataset or tracer field	—	32, 144, 148
$\boldsymbol{\Sigma}_i$	Covariance matrix of the i-th component in a mixture model	—	144
$\boldsymbol{\Sigma}$	Covariance matrix of a dataset	—	147
σ	Predicted standard deviation parameter of the Gaussian distribution	—	97, 140
SS_{between}	Inter-cluster variability (between-cluster sum of squares)	—	19
SS_{within}	Inner-cluster variability (within-cluster sum of squares)	—	19
T	Absolute temperature	K	12
t	Time	s	76, 142
Δt	Numerical time step	s	130

Symbol	Description	Unit	Page
t^*	Normalized coordinate (position) along the phase space curve	—	66
θ	Potential temperature	K	12
θ	Parameter vector (weights and biases) of the neural network	—	140
U_c	Critical zonal wind speed for vertical wave propagation	m s^{-1}	17
$\mathbf{U}$	Orthogonal basis matrix containing the eigenvectors (principal components) from PCA	—	147
$\bar{u}$	Mean zonal velocity component	m s^{-1}	16
v	Wind velocity (or velocity component)	m s^{-1}	14
$\boldsymbol{W}$	Vorticity tensor (asymmetric part of the velocity gradient tensor)	s^{-1}	14
$\boldsymbol{w}$	Weight vector of a neuron	—	24
X	Abstract placeholder for an initial air mass type (or class label)	—	44
x	True mean age of air (AoA) from the CLaMS model	yr	97, 113
x	General data element (or data point in a feature space)	—	18
X	Set of all data elements	—	19, 133, 144, 147
$\hat{x}$	Standardized data element (zero mean, unit variance)	—	32

Symbol	Description	Unit	Page
X	Input variable (predictor) of the predictive model	—	22
$\boldsymbol{x}$	Position vector of an air parcel	m	14, 76, 131, 133
$\boldsymbol{X}$	Input feature vector consisting of long-lived tracer concentrations	—	113, 142, 159
$\boldsymbol{X}$	Data vector (or sample) within the lower-dimensional subspace	—	22
Y	Abstract placeholder for a resulting air mass type (or class label)	—	44
Y	Output variable (predictand) of the predictive model	—	22
$\boldsymbol{Y}$	Data vector (or sample) within the higher-dimensional space	—	22
z	Dimensionless numerical parameter ($z = \lambda \Delta t$)	—	131
z	Pre-activation input (weighted sum) of a neuron	—	24
ζ_a	Absolute vorticity	s^{-1}	13

Chemical Compounds

Symbol	Description	Unit	Page
C_2H_4	Ethylene (Ethene)		50
C_2H_6	Ethane		50
CCl_4	Carbon tetrachloride (Tetrachromethane)		50
$CFC-11$	Trichlorofluoromethane (Fluorocarbon-11)		50, 61, 63, 95, 108, 171
$CFC-12$	Dichlorodifluoromethane (Fluorocarbon-12)		35, 63, 95, 109
CH_3OH	Methanol		50
CH_4	Methane		63, 95, 109
$ClONO_2$	Chlorine nitrate		50
H_2O	Water vapour		35
$HCFC-22$	Chlorodifluoromethane (Hydrofluorocarbon-22)		50, 63, 95
$HCOOH$	Formic acid		50
HNO_3	Nitric acid		35
N_2O	Nitrous oxide (Laughing gas)		63, 64, 95, 109, 171
O_3	Ozone		35

Symbol	Description	Unit	Page
PAN	Peroxyacetyl nitrate		30, 35
$\mathrm{SF_6}$	Sulfur hexafluoride		63, 95, 109

List of Figures

1.1 The vertical structure of the atmosphere and the atmospheric circulation with respect to mixing. Panel a: Vertical temperature profile. Beginning in the troposphere (0 km), temperature decreases at an average rate of 6.5 K/km until the tropopause ($\approx 15\,km$), beyond which the lapse rate becomes positive due to absorption of radiation by ozone. Temperature increases with height until the stratopause ($\approx 55\,km$) is reached, and the lapse rate becomes negative again in the mesosphere. At the mesopause ($\approx 85\,km$) the lapse rate becomes positive again due to a thin atmosphere and absorbed high-energy radiation. Panel b: Time series of zonal mean water vapour in the tropics (from -15° to 15° latitude), also known as the tape-recorder. Due to the asymmetry in wave activity between the hemispheres, the tropopause has lower water vapour content in boreal winter. This freeze-dried air can be seen in the zonal mean signatures as they propagate upwards like an imprint on a tape. The vertical wind speeds can be estimated from the slope of the water vapour signatures and amount to 0.2-0.5 mm/s. Based on data by Davis et al. (2016). Panel c: Summary of the global circulation. The heart of the circulation is the Brewer-Dobson Circulation (BDC), which is driven by Rossby waves (purple arrows). The momentum deposited by breaking of these waves induces mixing (blue arrows), which drives upwelling of tropical air (grey arrow) and horizontal transport in the extratropics (red arrows). The tropical air is confined to the tropical pipe (red shading), from where air is transported dominantly towards the winter pole and the polar vortex (blue shading). Gravity waves (yellow arrows) drive the overturning mesospheric circulation (green arrows), which transports tropical air across the mesosphere towards the polar vortex. 3

3.1 Map of surface-origin tracers used in the global 3-D CLaMS simulation for 2023. . 29

3.2 Summary of the double flight experiment. Panels a, b: CLaMS South Asia tracer (= 'ECH' + 'TIB' + 'NIN' + 'IND' + 'BoB' + 'NIO', see Fig.3.1 at 360 K potential temperature. PV values are shown as gray isolines. Panel a: Initial flight on 10 September 2023 UTC. A large filament with increased mixing ratios of South Asian air is projected to be in range of HALO research flights from the base in Anchorage. The first hexagon is located towards the end of the flight near the edge of the filament and is marked in red. Panel b: Follow-up flight on 11 September 2023 UTC. The filament 'rolls over' southeastwards. The second hexagon (marked in green) is located inside the outflow from the first hexagon at that time. Panel c: Forward trajectories of hexagon 1. The trajectories are shown as gray lines, beginning inside hexagon 1 at 01:35 UTC and ending at 01:50 UTC on the next day. The trajectories were started at every point on the retrieval grid and were driven by ERA5 (1 ° x 1 °) reanalysis. Potential temperature of end points is shown in colour-code. 31

3.3 Schematic outline of the data pipeline. The workflow goes from left to right. Model data and their data products are indicated in green. Measurement data and their data products are indicated in blue and red depending on which measurement. Transformations and classifiers that are not explicitly dependent on a dataset are marked in black. Transformations and classifiers that explicitly depend on a specific dataset are outlined in that dataset's colour. Transformations and classifiers which are applied to multiple datasets individually are outlined in brown. 33

3.4 Geometry of a 3-D tomographic retrieval. All data points are contained within a hexagonal prism, whose foundation is given by the flight path. These data points consist of those points within the retrieval volume with high measurement influence (black dots). For further analysis, the cross-sections indicated in red, blue and green are presented to represent the total volume. 34

3.5 PAN mixing ratios inside the first hexagon on 10 September 2023 (upper row: GLORIA retrievals; lower row: CAMS model values). Shown are the cross-sections indicated in Fig. 3.4. PV isolines are plotted as white dashed lines, potential temperature isolines are given as gray dashed lines. The dynamical tropopause adapted from Kunz et al. (2015) is indicated as a black line in all panels. 35

3.6 **(a-e)**: Retrieved trace gas species inside the first hexagon on 10 September 2023 UTC. Extraordinary regions, such as the maximum signature in PAN and HNO_3, which contain apparent inconsistencies, are marked by the bright and dark gray arrows, respectively. **f)**: Potential vorticity derived from ERA5. Shown is the green cross-section in Fig. 3.4. The dynamical tropopause and PV-isolines are indicated as black and white lines, respectively. The potential temperature is indicated by the gray dashed lines in each panel. 36

3.7 GMM classification based on the first three principal components. **(a, c)**: Full manifold of hexagon 1 and hexagon 2 consisting of all three principal components. **(b, d)**: First two principal components of the manifold of hexagon 1 and 2, respectively, for better visualization. All panels show the labels in colour-code. For each Gaussian distribution a 1-σ environment is indicated by ellipses. The PCA of the mean of each measured trace gas weighted by the relevant surface-origin tracers of the CLaMS simulations is shown in red markers. The red markers for 'SEA' and 'TWP' in panel b) lie close to each other and basically overlap (both in 3-D and 2-D). The percentages shown are the maximum percentage of each respective surface-origin tracer. In all panels, the classes found by the GMM are shown in colour. Air parcels that were observed in both hexagons are indicated by a black outline. For each class the cluster centre is indicated by white dots. 38

3.8 Panel a): Chemical composition of each cluster centre as obtained by the inverse PCA transformation. The chemical composition of each cluster is shown as percentiles of the total distribution, i.e. a value of 20 indicates that 20% of all observed values are lower than this value. Panel b): Loadings of each principal component. The length of each bar indicates the importance of that tracer to that principal component. 40

3.9 Class distribution inside **(a, b, c)** hexagon 1 and inside **(d, e, f)** hexagon 2. Cross-sections are identical to Fig. 3.5a-c. The dynamical tropopause is indicated as black line. PV isolines are shown in white. Measurement points within each cross-section are shown as solid markers. Matched points are indicated by black borders. Class assignment is indicated in colour-code, with high class alignment shown in more opaque colours. In hexagon 2, the background of both tropospheric classes is displayed in blue as if they were a single class. For clarity gray labelled points are still shown in gray. The background shows the classification applied to the trace gases interpolated in each cross-section. Spaces containing few measurement points are more uncertain. 42

3.10 Confusion matrix of class assignment of matched air parcels. The number of each transition is shown in each matrix element. The columns indicate the class assignment of all air parcels in hexagon 1; the rows indicate the class assignment of all air parcels in hexagon 2. The air types are indicated by the abbreviations (T, A, B, S). Both tropospheric classes (CT and MT) are combined into one (T). Each possible transition has a unique colour variation of the colour used to encode its class assignment in hexagon 1. The total number of parcels in each class in hexagon 1 and 2 is indicated on each row/columns edges, respectively. 44

3.11 Visualization of transition states. Left: trajectory starting points inside hexagon 1. Transition state is indicated in colour-code, see legend. Right: trajectory end points inside hexagon 2. Transition state is indicated in colour-code consistent with Fig. 3.10. Rare and missing transitions are grayed out for visibility. 46

3.12 **Panels a) - e)**: Overview of tracer changes between the two hexagons. The identity (no changes) is shown in the background as a line. The different transitions are indicated in colour-code in accordance with the confusion matrix shown again in panel h). Uncertainties due to retrieval parameters are shown as well (small). Histograms of measurements are shown on the edges. A Gaussian kernel density estimation is shown as shading. **Panel a)**: The value range for $\mathrm{H_2O}$ is artificially limited to focus on low values. **Panel f)**: Changes in potential vorticity (ERA5 data) during mixing. **Panel g)**: Changes in potential temperature (ERA5 data) during mixing. 47

3.13 Map of all lightning events detected by GOES-18 between the time hexagon 1 and hexagon 2 were measured. The inset focuses on the relevant section including both measurements. Both hexagons are shown in colour, as well as the trajectories with anomalously increased HNO_3. 49

3.14 Classification of the curtains using GMMs. Panels a, c, e, g (left-hand side): Distribution of PAN, O_3 , labels and probabilities for PH13, respectively. Panels b, d, f, h (right-hand side): Distribution of PAN, O_3, labels and probabilities for PH14, respectively. The labels are shown in colour-code consistent with previous classifications such that similar air types share similar colours. The probabilities are colour-coded such that they indicate label ambiguity (green: unambiguous, yellow: two competing classes, orange: three competing classes ...). The thermal tropopause is shown in gray markers; each hexagon is indicated by the white, vertical lines. 52

3.15 Technical information for the GMM classifications of the two 2-D curtains. Panels a,c show the chemical composition of each cluster centre after inverse PCA transformation for PH13 and PH14, respectively; values are shown as percentiles of the respective total distribution. Panels b, d show the corresponding PCA loadings, with the associated trace gas indicated by colour. For cluster-related quantities in panels a, c, colours encode the physically interpreted air-mass category consistently across the 3-D retrievals and both curtains: blue denotes tropospheric air, orange ASM outflow, green blended air, purple stratospheric air, and red strongly polluted air. Different shades denote distinct classes within a category. 53

3.16 Summary of the DEC classifications of the 2-D curtains. The left and right columns show PH13 and PH14, respectively. The upper row shows the spatial distribution of the DEC-derived classes; the lower row shows their chemical composition. Colours are assigned by matching each DEC cluster to the chemically most similar air-mass category identified by the GMM classifications: blue denotes tropospheric air, orange ASM outflow, green blended air, purple stratospheric air, and red strongly polluted air. Different shades distinguish separate but chemically similar classes. . 57

4.1 Visualization of the different CLaMS simulations and data products. All panels show the spatial distribution of $\mathrm{CFC}-11$ at 21 km altitude for the 1 February 2011. Regularly gridded data for 12:00 UTC are shown in the first column, the synthetic observations are shown in the second column. The synthetic observations are the cumulative observations over 24 h at individual times. Panels a, b show the control mixing simulation, panels c, d show the high mixing simulation. 61

4.2 Correlation between $\mathrm{CFC}-11$ and $\mathrm{N_2O}$. Shown is data for the 1 February 2011 for the control mixing scenario. On the left-hand side correlations for the isentropes of 500, 550 and 600 K are shown, respectively. The latitude is shown in colour-code. On the right-hand side the point density for 550 K is indicated in the vertical axis, with outstanding peaks (clusters) marked by the numbers 0 – 5. 64

4.3 Visualization of the mixing state. Panel a): Visualization of the definition according to Eq. 4.3. The mixing state is shown in colour-code. The chain (no filter) is shown in cyan, cluster centres are indicated as white dots. Panel b): Spatial distribution of the mixing state for 1 February 2011 on a level of 21 km altitude. 67

4.4 Visualization of the mixing dynamics for 1 February 2011. The first row shows values derived from the high mixing scenario, the second row shows values derived from the control mixing scenario. First column: linear trend of the static mixing states. Second column: linear trend of the dynamic mixing states. Third column: linear trend of Age–of–Air. 69

4.5 Panel a: Zonal means of the linear trends in the static mixing state for the high mixing scenario. A section of anomalous negative trend is indicated by the red dashed line. Panel b - c: Initial positions and final positions of the air parcels associated with the anomalous negative trend, respectively. The current mixing state is shown in colour-code in either panel. 70

4.6 Average linear trend divided by initial M. Shown are the high mixing scenario (dashed line) and the control mixing scenario (dotted line) for February 2011. The error bars indicate the standard deviation of the average linear trend of each ensemble. The initial mixing state is also indicated in colour-code. 70

4.7 Definition of the tropical pipe and the filaments. Panel a): Zonal mean mixing states. The black lines show the zonal means over all longitudes (thick line) and longitude slices (thin lines). The latitudes where the threshold 0.35 is met are indicated by the red vertical lines and the black markers, respectively. Panel b): Spatial distribution of the boundary latitudes. The zonally resolved latitudes are fitted by a high–degree polynomial (green). M is indicated in the background. Panel c): Resulting filaments defined as tropical air outside the tropical pipe. M is indicated in the background. . 73

4.8 Visualization of the geometrical properties of sets of points on the example of a banana. Panel a: All points making up the total banana shape. Panel b: Different geometrical properties. Shown are: convex hull (orange, dashed line), alpha shape (red line) and the skeleton (green line). Panel c: Relative distance (does a point lie close to the shape's edge?). Panel d: Relative length (does a point lie close to the start of the shape?). 74

4.9 Pearson correlation coefficient between mixing state M and geometrical properties of the filaments. The x–axis shows correlation to the position along the filament, the vertical direction shows correlation to the position with respect to the filaments edges. The error bars indicate the limits forming a 95% confidence interval. The p–values of the different regression of d_r, l_r corresponding to a 95% confidence interval are indicated in the legend. 75

4.10 The different steps of the filament detection. For a description of the algorithm see the text below. Panel a: Spatially coherent structures (fragments) found among all satellite data. The colour–code indicates the ID of each of the ~ 130 fragments. Panel b: Spatially and temporally coherent structures, formed from the fragments, bridge data gaps. The colour–code indicates the ID of these fragments. Panel c: Initial positions and initial mixing states (shown in colour-code) of all original trajectories matched with the fragment points. Panel d: Positions of the originally detected filaments as a reference. 78

4.11 Panel a: Linear trends derived from matches on satellite observations. Panel b: Statistical dissipation rates of trajectories sorted by initial mixing state similarly to Fig. 4.6. The dissipation rates are represented by the average linear trend of each group. Positive numbers indicate an increase in mixing state, negative ones indicate decreases. Zero (no change) is indicated as a black, dashed line. Trends are shown for the high mixing scenario (dashed lines) and for the control mixing scenario (dotted lines). 80

4.12 Summary of matching statistics. Panels a, b: Distribution of times between matches (revisit time) over a total period of three days and 14 days, respectively. Mean and median are indicated as red and cyan markers, respectively. Distributions are split into five climatological regions (cf. Tab. 4.4). The percentage of revisited points in each region is shown on the left–hand side. The numbers of matches along a single trajectory are summarized by histograms on the right–hand side. Panel c: Percentage of seen (solid lines) and revisited (dashed lines) air parcels over the span of 14 days. The climatological regions are shown in colour–code. Panel d: Effect of swath width on the number of revisited air parcels. A period of 72 hours is considered. The track width is shown in colour–code. 81

4.13 Architectures of different CNNs. Panel a: predictive CNN for regularly gridded data. The model input assumes 3 values into the past including 1 covariate for a total 6-D input vector. The input is passed into convolution layers, which are L_2–regularized. The final convolution is averaged across its channels and reshaped into the output shape (here: 2 values into the future). Panel b: The same model as in panel a, but adjusted to accept irregularly spaced time series. The input is split into the s_i values (upper input), which are passed through L_2–regularized convolution layers, and into the input time steps (middle input) and the prediction time steps (lower input). Input time steps are adequately merged back after the convolution, after which the prediction time steps are added in as well. The recombined and processed inputs are then passed through an MLP-like section and returned as the output shape (here: 2 values into the future). 85

4.14 Performance of different CNN architectures. Panel a: Performance of the single prediction models. Shown are models with past values of 1, 3 and 5 in blue, orange and green, respectively. Models which use irregularly spaced time series are shown hatched, models considering covariates are shown in light colours. Persistence is shown in the background for different percentages of noise. Panel b: Performance of multiple prediction models. Shown are models with a future value of three and past values of 3, 5 in red and violet, respectively. Irregularly spaced time series and covariates are indicated similarly to panel a. Persistence is shown in the background for different percentages of noise. Performance is derived for each of the three predicted values individually. Panel c: Performance of models trained on regularly spaced time steps resolved by the predicted day. Performance is calculated for those days where true model values are available for each predicted value. Predictions beyond the 7th lie outside of the training domain. Persistence is shown in the background for different percentages of noise. In panels a – c, the performance is indicated by the mean absolute error (MAE). Panel d: Relative difference between predicted and true model values expressed in percentages exemplarily for 10 February 2011. 87

4.15 Performance of the recursive models. Based on the initial state on 1 February, each following day is predicted from the prediction of the last day. Different baselines based on different levels of noise are indicated in the background similar to Fig. 4.14a. These baselines naturally change with the predicted day due to error propagation. Shown are recursive models with past values of one (blue) and three (orange). Models trained on irregularly spaced time steps are shown as dashed lines. These models are replaced by a NARX variant with the same parameters. . 89

4.16 Comparison of linear trends for an initial state at 1 February. Panel a: Linear trends derived from true (model) values. Panel b: Linear trends derived from recursive predictions based on regularly spaced time series. Panel c: Linear trends derived from recursive NARX model predictions based on irregularly spaced time series. . 90

5.1 Illustration of the mean Age–of–Air concept. Air is uplifted from the troposphere in the tropics and transported polewards via either branch of the BDC (plain arrows). The two-sided wavy arrows indicate two-way mixing. Air may also be transported upwards as far as into the mesosphere, where it may reside for around one year (Ray et al., 2017). The age spectra for both scenarios are sketched in the insets A and B, respectively. Adapted from ESA (2025). 93

5.2 Sketch of a typical (supervised) MLP architecture. Panel a shows a regular MLP, panel b shows a probabilistic MLP. Shown is the architecture for the six LLT training scenario as an example. This architecture consists of 3 hidden layers containing 300 neurons each and an output layer of 1 neuron. The sizes shown are reduced compared to the true sizes for visualization purposes. 97

5.3 Diagnostics of the regular and probabilistic MLP. Quantities that are shown only once are almost identical for each MLP. Panel a: Correlation between the predicted AoA based on noise-free features and the true CLaMS model AoA. The grey line indicates a perfect prediction (identity). The corresponding latitude of each point is indicated in colour-code. The performance is summarized by the RMSE(Y_p, Y_r) value. Panel b): The same correlation, but for values predicted by a probabilistic MLP. Shown is the correlation between the expectation values of the predicted distributions and the true CLaMS model AoA. The standard deviations of each predicted distribution are indicated as error bars. The performance is summarized by the RMSE(Y_p, Y_r). Panel c: Distribution of the model accuracy split by predicted AoA value. The average accuracy is indicated by the black vertical line. Panel d): Precision of the predictions, split by the predicted AoA values. The precision is expressed as the RMSE(Y_p, Y_δ), i.e. the difference between optimal prediction and noisy prediction, with higher RMSE values indicating worse precision. The distortion added to the inputs is indicated in colour-code and varies between 0% and 5% of the true value. The average precision RMSE is shown in the legend above. Panel e): Same as panel d, but for the model performance, indicated by the RMSE(Y_δ, Y_r), i.e. the difference between true CLaMS model values and noisy predictions. Panel f): Distribution of the difference between true CLaMS model values (Y_r) and noisy predictions (Y_δ). The distributions for the different noise levels are shown as boxplots, with 50% of all values contained within the box, 85% of all values within the whiskers and all other values indicated by the outliers. 99

5.4 Difference diagnostic for probabilistic MLPs trained on different levels of feature–noise. Panels a – d show the distribution of the differences for 1%, 2%, 3% and 5% regularization noise, respectively. The corresponding precision and accuracy of each validation is shown in the legends. The models show fairly constant performances for the different validations, since the accuracy is determined during training, and while it is not impossible for a model to adjust its accuracy for different predictions, the models fail to do so. The statistics are functionally identical to Fig. 5.3f 101

5.5 Effect of different kinds of systematic uncertainties on the prediction performance, expressed by the RMSE. A model regularized on 2% Gaussian noise is tasked to predict AoA based on inputs subject to a constant offset (Const.), a linear scaling factor (Lin.) or both (Const. & lin.), see panel a. Three degrees of severity (low, medium, high) of each form of systematic uncertainty are considered. Panels b – d show the same evaluations, but for a modified model which was trained on data containing the same combinations of systematic uncertainties of medium severity, respectively. 103

5.6 Diagnostics for particular model artefacts. Panel a: Spatial distribution of the differences between true CLaMS model and predicted AoA for a probabilistic MLP trained with 1% regularization noise. Shown is an isentrope of 500 K potential temperature, corresponding to an altitude range between 19 km and 22 km. Predicted are inputs with 1$\%$ simulated noise. Red regions indicate underestimation of AoA, blue regions indicate overestimation. The blur in the distribution stems from the inputs with added noise used to predict AoA. Panel b): Performance of the same model for different datasets of 2011. The performances are expressed in terms of RMSE(Y_δ, Y_r) for different validation noises. The precision and accuracy of each validation is shown in the legend. 105

5.7 Annual performances of probabilistic MLPs trained on different features. All models were regularized with different noise levels where applicable, and therefore show reduced variability between validation noise levels. Panel a: Previous design as shown in Fig. 5.6, but trained on a seasonal dataset with implicit time dependence only. Panel b: MLP trained on a seasonal dataset which includes latitude as a feature. Panel c: MLP trained on a seasonal dataset which includes time explicitly as feature. Panel d: MLP trained on a seasonal dataset which includes both latitude and time as features. 106

5.8 Analysis metrics used to describe the function of a neural network: Panel a: Distribution of the differences between predicted and true CLaMS model AoA similarly to Fig. 5.4, but resolved by potential temperature. Panel b): Gradients of the neural network, taken at representative inputs for each of the potential temperature brackets used in panel a. Panels c – e: SHAP values for representative inputs corresponding to potential temperatures of 425, 650 and 900 K, respectively. The SHAP values indicate the importance of each input to the final prediction. 108

5.9 Performance evaluation of datasets from different years. The model is a probabilistic MLP trained on a seasonal dataset of 2011 with full features and various feature noise regularization. Shown are the years 2011, 2012, 2016 and 2021 in panels a, b, c, and d, respectively. 111

5.10 Architecture of the AEs. Panel a, c: Architecture of the regular and probabilistic encoder, respectively. The encoders take the six LLTs as features and return a 1-D output. Panel b, d: Architecture of the corresponding decoder. The decoders take the 1-D output of the encoder and return a 6-D output. During training, the decoder output is compared against the encoder input. In all panels, the number of neurons shown in the hidden layers is reduced to 25 for visibility reasons. 114

5.11 Spatial distribution of a probabilistic AE 1-D embedding of LLTs. Panel a: CLaMS model AoA values. Panel b: 1-D embedding, as provided by the encoder. The scale of the encoding follows the scale of the inputs. Panel c: 1-D embedding, linearly scaled to match the model value range based on the linear regime in Fig. 5.12a. The colour-scale is limited to the value range found in the model AoA. Panel d: Difference between CLaMS model AoA and linearly scaled embedding values. . . 115

5.12 Relations between AE embedding and model values. Panel a: Correlation between CLaMS model AoA and linearly scaled 1-D embedding. The latitude of the encoded values is indicated in colour-code. The embedding is scaled such that it corresponds to the CLaMS model AoA in the linear regime (< 4 years). Panel b): Correlation between the model LLTs used as features and the decoding of the embedding. The decoded values need no calibration. 115

5.13 Overview of MOE AEs performance. Panel a: Correlation between the true CLaMS model AoA and the embedding of a simple AE. The corresponding latitude values are shown in colour-code. The embedding is linearly calibrated. The prominent feature (located at CLaMS model AoA > 5 years and calibrated embedding > 5 years) is marked with a black circle. Panel b: The same correlation as in panel a, but for the embedding of MOE AEs. The latitude is shown in colour-code. Panel c: The same correlation, but shown in colour-code is the ID of the expert used predominantly for the embedding instead. Panel d: Correlation between the true CLaMS model AoA and the different LLTs for the values contained within the circle in panel a. 119

5.14 Evaluation of the physics-informed MOE AEs. Shown are the spatial distributions of the differences between the true CLaMS model AoA and the calibrated embedding on isentropes for 1 August 2011. 121

5.15 Performance evaluation of datasets from different years. The models here are calibrated Mixture of Experts autoencoders trained on 1% measurement noise using only data from August 2011. Validation data is sampled between 420 K and 600 K potential temperature. Shown are the years 2011, 2012, 2016 and 2021 in panel a, b, c, and d, respectively. 122

A.1 Illustration of different kernel parameters: First row: A basic convolution with a convolution kernel of size three and stride/dilation of one. The input is indicated in blue, the resulting convolution is indicated in orange. The contribution of each input to the convolution is indicated by grey arrows. Second row: Same setup, but with dilation of two. Third row: Same setup, but with stride of two. 135

A.2 First part of the trajectories (1 February 2011 – 5 February 2011) 137

A.3 Second part of the trajectories (6 February 2011 – 10 February 2011) 138

A.4 Third part of the trajectories (11 February 2011 – 14 February 2011) 139

A.5 Illustration of the relationship between estimated and true aleatoric uncertainty of probabilistic MLPs. Shown are the residuals between true and predicted AoA, with the estimated aleatoric uncertainty (in the form of the standard deviation of the predicted distribution) indicated by error bars. The true aleatoric uncertainty is well estimated if the residuals are distributed within around one standard deviation. . . 141

A.6 Illustration of unphysical labelling artefacts. Shown are random samples of two Gaussian distributions with different parameters (see legend). The labels derived from Eq. A.29 are indicated in the background in colour–code. For low label probability for either distribution (right side) the broader distribution determines the label, resulting in a jump in labelling. 145

A.7 Illustration of the DBSCAN algorithm. Data points are indicated by dots. An ϵ–environment is indicated by the transparent circles around each dot. The type of classification of each dot is indicated in colour–code: Points, which are ϵ–reachable by at least m neighbours are considered *core points* (green). Points, which are ϵ–reachable by at least one core point, but are not core points themselves, are considered *border points* (yellow). Points, which are ϵ–reachable by fewer than m points, of which none are core points, are considered *outliers* (red). 146

A.8 Illustration of the effect of topology on clustering. Panel a) shows two slightly intertwined Gaussian distributions. A 1–σ environment is indicated by the coloured ellipses. The underlying topology (flat space) is indicated by the grid. The percentage of accurately classified points inside each ellipse is shown in the legend. Panel b) shows the same data, but with two topological lenses applied around the estimated shape of each blob. The topological distortions are visible in the grid. Another Gaussian fit is performed, and the resulting 1–σ environment as well as the percentage of accurately classified points are indicated similarly to panel b). . 149

A.9 3–D visualization of the 6–D phase space spanned by the LLTs. Shown is a 3–D representation of the 6–D phase space (with effective dimensionality of one) obtained through a dedicated autoencoder with latent dimension of three. The three components do not natively correspond to any of the original LLTs. The embedding obtained through the MOE AEs is shown in colour–code. The values inside the polar vortex are marked in red. 151

List of Tables

4.1 Average transition times between mixing states. Shown are the times in days an air parcel takes on global average to undergo a change in mixing state of 0.1. Positive numbers indicate increases, negative numbers indicate decreases. These estimates are uncertain particularly for a mixing state of zero, as very little change is anticipated. 71

4.2 Percentages of trajectories belonging to the originally detected filaments that were matched with points identified as filaments in the observational data. 78

4.3 Transition times of mixing state. 80

4.4 Definition of the climatological zones used in Fig. 4.12. The zonal boundaries towards the poles follow the highest latitudes retrievable by CAIRT. 81

Bibliography

Abalos, M., Polvani, L., Calvo, N., Kinnison, D., Ploeger, F., Randel, W., and Solomon, S.: New Insights on the Impact of Ozone-Depleting Substances on the Brewer-Dobson Circulation, J. Geophys. Res., 124, 2435–2451, doi: https://doi.org/10.1029/2018JD029301, 2019.

Abalos, M., Calvo, N., Benito-Barca, S., Garny, H., Hardiman, S. C., Lin, P., Andrews, M. B., Butchart, N., Garcia, R., Orbe, C., Saint-Martin, D., Watanabe, S., and Yoshida, K.: The Brewer–Dobson circulation in CMIP6, Atmos. Chem. Phys., 21, 13 571–13 591, doi: 10.5194/acp-21-13571-2021, 2021.

Albergante, L., Bac, J., and Zinovyev, A.: Estimating the effective dimension of large biological datasets using Fisher separability analysis, in: Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), pp. 1–8, IEEE, doi: 10.1109/IJCNN.2019.8852156, 2019.

Austin, J. and Li, F.: On the relationship between the strength of the Brower-Dobson circulation and the age of stratospheric air, Geophys. Res. Lett., 33, L17807, doi: 10.1029/2006GL026867, 2006.

Bacmeister, J. T., Küll, V., Offermann, D., Riese, M., and Elkins, J. W.: Intercomparison of satellite and aircraft observations of ozone, CFC-11, and NO_y using trajectory mapping, J. Geophys. Res., 104, 16 379–16 390, doi: 10.1029/1999JD900173, 1999.

Baikhadzhaev, R., Ploeger, F., Preusse, P., Ern, M., and Birner, T.: A dynamics-based separation of deep and shallow stratospheric circulation branches, Atmos. Chem. Phys., 25, 12 753–12 777, doi: 10.5194/acp-25-12753-2025, 2025.

Bak, J., Kim, J., Koo, J.-H., Pan, L. L., Ryoo, J.-M., Newman, P. A., Franchin, A., Liu, X., Abad, G. G., Lee, H.-J., Jeon, W., and Kim, C.-H.: An integrated analysis of Ozone and

Carbon Monoxide Over the Western Pacific Using Satellite and Aircraft Measurements during the ACCLIP Summer Campaign 2022, J. Geophys. Res. Atmos., 130, e2024JD042 771, doi: https://doi.org/10.1029/2024JD042771, 2025.

Baker, A. K., Schuck, T. J., Slemr, F., van Velthoven, P., Zahn, A., and Brenninkmeijer, C. A. M.: Characterization of non-methane hydrocarbons in Asian summer monsoon outflow observed by the CARIBIC aircraft, Atmos. Chem. Phys., 11, 503–518, doi: 10.5194/acp-11-503-2011, 2011.

Baxter, R. A.: Mixture Model, pp. 680–682, Springer US, URL "https://doi.org/10.1007/978-0-387-30164-8_547", 2010.

Bentley, J. L.: Multidimensional Binary Search Trees Used for Associative Searching, Communications of the ACM, 18, 509–517, doi: 10.1145/361002.361007, 1975.

Bergman, J. W., Fierli, F., Jensen, E. J., Honomichl, S., and Pan, L. L.: Boundary layer sources for the Asian anticyclone: Regional contributions to a vertical conduit, J. Geophys. Res., 118, 2560–2575, doi: 10.1002/jgrd.50142, 2013.

Bian, J., Pan, L. L., Paulik, L., Vömel, H., and Chen, H.: In situ water vapor and ozone measurements in Lhasa and Kunmin during the Asian summer monsoon, Geophys. Res. Lett., 39, L19808, doi: 10.1029/2012GL052996, 2012.

Bian, J., Li, D., Bai, Z., Li, Q., Lyu, D., and Zhou, X.: Transport of Asian surface pollutants to the global stratosphere from the Tibetan Plateau region during the Asian summer monsoon, Natl. Sci. Rev., 7, 516–533, doi: 10.1093/nsr/nwaa005, 2020.

Birner, T. and Boenisch, H.: Residual circulation trajectories and transit times into the extratropical lowermost stratosphere, Atmos. Chem. Phys., 11, 817–827, doi: 10.5194/acp-11-817-2011, 2011.

Björklund, M.: Be careful with your principal components, Evolution, 73, 2151–2158, doi: 10.1111/evo.13835, 2019.

Bönisch, H., Engel, A., Birner, T., Hoor, P., Tarasick, D. W., and Ray, E. A.: On the structural changes in the Brewer-Dobson circulation after 2000, Atmos. Chem. Phys., 11, 3937–3948, doi: 10.5194/acp-11-3937-2011, 2011.

Bourassa, A. E., Robock, A., Randel, W. J., Deshler, T., Rieger, L. A., Lloyd, N. D., Llewellyn, E. J. T., and Degenstein, D. A.: Large Volcanic Aerosol Load in the Stratosphere Linked to Asian Monsoon Transport, Science, 337, 78–81, doi: 10.1126/science.1219371, 2012.

Bresenham, J. E.: Algorithm for Computer Control of a Digital Plotter, IBM Syst. J., 4, 25–30, doi: 10.1147/sj.41.0025, 1965.

Brewer, A. W.: Evidence for a World Circulation Provided by the Measurements of Helium and Water Vapour in the Stratosphere, Quart. J. Roy. Meteorol. Soc., 75, 351–363, 1949.

Brown, A., Volk, C. M., Schoeberl, M. R., Boone, C. D., and Bernath, P. F.: Stratospheric lifetimes of CFC-12, CCl_4, CH_4, CH_3Cl and N_2O from measurements made by the Atmospheric Chemistry Experiment-Fourier Transform Spectrometer (ACE-FTS), Atmos. Chem. Phys., 13, 6921–6950, doi: 10.5194/acp-13-6921-2013, 2013.

Butchart, N., Scaife, A. A., Bourqui, M., de Grandpre, J., Hare, S. H. E., Kettleborough, J., Langematz, U., Manzini, E., Sassi, F., Shibata, K., Shindell, D., and Sigmond, M.: Simulations of anthropogenic change in the strength of the Brewer-Dobson circulation, Clim. Dyn., 27, 727–741, 2006.

Butchart, N., Cionni, I., Eyring, V., Shepherd, T. G., Waugh, D. W., Akiyoshi, H., Austin, J., Brühl, C., Chipperfield, M. P., Cordero, E., Dameris, M., Deckert, R., Dhomse, S., Frith, S. M., Garcia, R. R., Gettelman, A., Giorgetta, M. A., Kinnison, D. E., Li, F., Mancini, E., McLandress, C., Pawson, S., Pitari, G., Plummer, D. A., Rozanov, E., Sassi, F., Scinocca, J. F., Shibata, K., Steil, B., and Tian, W.: Chemistry-climate model simulations of twenty-first century stratospheric climate and circulation changes, J. Clim., 23, 5349–5374, 2010.

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

Charlesworth, E. J., Dugstad, A.-K., Fritsch, F., Jöckel, P., and Plöger, F.: Impact of Lagrangian Transport on Lower-Stratospheric Transport Time Scales in a Climate Model, Atmos. Chem. Phys., 20, 15 227–15 245, doi: 10.5194/acp-20-15227-2020, 2020.

Charney, J. G. and Drazin, P. G.: Propagation of planetary-scale disturbances from the lower into the upper atmosphere, J. Geophys. Res., 66, 83–109, doi: 10.1029/JZ066i001p00083, 1961.

Chen, B., Xu, X. D., Yang, S., and Zhao, T. L.: Climatological perspectives of air transport from atmospheric boundary layer to tropopause layer over Asian monsoon regions during boreal summer inferred from Lagrangian approach, Atmos. Chem. Phys., pp. 5827–5839, doi: 10.5194/acp-12-5827, 2012.

Curbelo, J. and Linz, M.: Lagrangian coherent structures to examine mixing in the stratosphere, Atmos. Chem. Phys., 25, 7941–7957, doi: 10.5194/acp-25-7941-2025, 2025.

Davis, S. M., Rosenlof, K. H., Hassler, B., Hurst, D. F., Read, W. G., Vömel, H., Selkirk, H., Fujiwara, M., and Damadeo, R.: The Stratospheric Water and Ozone Satellite Homogenized (SWOOSH) database: a long-term database for climate studies, Earth Syst. Sci. Dat., 8, 461–490, doi: 10.5194/essd-8-461-2016, 2016.

Dempster, A. P., Laird, N. M., and Rubin, D. B.: Maximum Likelihood from Incomplete Data via the EM Algorithm, J. Roy. Stat. Soc. B Met., 39, 1–22, doi: 10.1111/j.2517-6161.1977.tb01600.x, 1977.

Diallo, M., Ern, M., and Ploeger, F.: The advective Brewer–Dobson circulation in the ERA5 reanalysis: climatology, variability, and trends, Atmos. Chem. Phys., 21, 7515–7544, doi: 10.5194/acp-21-7515-2021, 2021.

Dietmüller, S., Eichinger, R., Garny, H., Birner, T., Boenisch, H., Pitari, G., Mancini, E., Visioni, D., Stenke, A., Revell, L., Rozanov, E., Plummer, D. A., Scinocca, J., Jöckel, P., Oman, L., Deushi, M., Kiyotaka, S., Kinnison, D. E., Garcia, R., Morgenstern, O., Zeng, G., Stone, K. A., and Schofield, R.: Quantifying the effect of mixing on the mean age of air in CCMVal-2 and CCMI-1 models, Atmos. Chem. Phys., 18, 6699–6720, doi: 10.5194/acp-18-6699-2018, 2018.

Dijkstra, E. W.: A Note on Two Problems in Connexion with Graphs, Numer. Math., 1, 269–271, doi: 10.1007/BF01386390, 1959.

D'Isidoro, M., Maurizi, A., and Tampieri, F.: Effects of resolution on the relative importance of numerical and physical horizontal diffusion in atmospheric composition modelling, Atmos. Chem. Phys., 10, 2737–2743, doi: 10.5194/acp-10-2737-2010, 2010.

Dubé, K., Tegtmeier, S., Ploeger, F., and Walker, K. A.: Hemispheric asymmetry in recent stratospheric age of air changes, Atmos. Chem. Phys., 25, 1433–1447, doi: 10.5194/acp-25-1433-2025, 2025.

Engdahl, N. B., Benson, D. A., and Bolster, D.: Lagrangian simulation of mixing and reactions in complex geochemical systems, Water Resources Research, 53, 3513–3522, doi: 10.1002/2017WR020362, 2017.

Engel, A., Möbius, T., Bönisch, H., Schmidt, U., Heinz, R., Levin, I., Atlas, E., Aoki, S., Nakazawa, T., Sugawara, S., Moore, F., Hurst, D., Elkins, J., Schauffler, S., Andrews, A., and Boering, K.: Age of stratospheric air unchanged within uncertainties over the past 30 years, Nature Geoscience, 2, 28–31, doi: 10.1038/ngeo388, 2009a.

Engel, A., Möbius, T., Bönisch, H., Schmidt, U., Heinz, R., Levin, I., Atlas, E., Aoki, S., Nakazawa, T., Sugawara, S., Moore, F., Hurst, D., Elkins, J., Schauffler, S., Andrews, A., and Boering, K.: Age of stratospheric air unchanged within uncertainties over the past 30 years, Nature Geosci., 2, 28–31, doi: 10.1038/ngeo388, 2009b.

Engel, A., Bönisch, H., Ullrich, M., Sitals, R., Membrive, O., Danis, F., and Crevoisier, C.: Mean age of stratospheric air derived from AirCore observations, Atmos. Chem. Phys., 17, 6825–6838, doi: 10.5194/acp-17-6825-2017, 2017.

ESA: Earth Explorer 11 candidate mission CAIRT: report for assessment, Tech. rep., European Space Agency, ESTEC, Keplerlaan 1, 2200 AG Noordwijk, The Netherlands, 2023.

ESA: Earth Explorer 11 candidate mission CAIRT: report for mission selection, Tech. rep., European Space Agency, ESTEC, Keplerlaan 1, 2200 AG Noordwijk, The Netherlands, doi: 10.5281/zenodo.15606819, 2025.

Farman, J. C., Gardiner, B. G., and Shanklin, J. D.: Large losses of total ozone in Antarctica reveal seasonal ClOx/NOx interaction, Nature, 315, 207–210, doi: 10.1038/315207a0, 1985.

Feely, H. W. and Spar, J.: Tungsten-185 from nuclear bomb tests as a tracer for stratospheric transport, Nature, 188, 1062–1064, doi: 10.1038/1881062a0, 1960.

Foong, A. Y., Burt, D. R., Li, Y., and Turner, R. E.: On the expressiveness of approximate inference in Bayesian neural networks, in: Advances in Neural Information Processing Systems (NeurIPS), vol. 33, pp. 15 897–15 908, 2020.

Friedl-Vallon, F., Gulde, T., Hase, F., Kleinert, A., Kulessa, T., Maucher, G., Neubert, T., Olschewski, F., Piesch, C., Preusse, P., Rongen, H., Sartorius, C., Schneider, H., Schönfeld, A., Tan, V., Bayer, N., Blank, J., Dapp, R., Ebersoldt, A., Fischer, H., Graf, F., Guggenmoser,

T., Höpfner, M., Kaufmann, M., Kretschmer, E., Latzko, T., Nordmeyer, H., Oelhaf, H., Orphal, J., Riese, M., Schardt, G., Schillings, J., Sha, M. K., Sumińska-Ebersoldt, O., and Ungermann, J.: Instrument concept of the imaging Fourier transform spectrometer GLORIA, Atmos. Meas. Tech., 7, 3565–3577, doi: 10.5194/amt-7-3565-2014, 2014.

Garfinkel, C. I., Aquila, V., Waugh, D. W., and Oman, L. D.: Time-varying changes in the simulated structure of the Brewer–Dobson Circulation, Atmos. Chem. Phys., 17, 1313–1327, doi: 10.5194/acp-17-1313-2017, 2017.

Garny, H. and Randel, W. J.: Dynamic variability of the Asian monsoon anticyclone observed in potential vorticity and correlations with tracer distributions, J. Geophys. Res., 118, 13,421–13,433, doi: 10.1002/2013JD020908, 2013.

Garny, H., Dameris, M., Randel, W., Bodeker, G. E., and Deckert, R.: Dynamically forced increase of tropical upwelling in the lower stratosphere, J. Atmos. Sci., 68, 1214–1233, doi: 10.1175/2011JAS3701.1, 2011.

Garny, H., Birner, T., Bönisch, H., and Bunzel, F.: The effects of mixing on Age of Air, J. Geophys. Res., 119, doi: 10.1002/2013JD021417, 2014.

Garny, H., Eichinger, R., Laube, J. C., Ray, E. A., Stiller, G. P., Bönisch, H., Saunders, L., and Linz, M.: Correction of stratospheric age of air (AoA) derived from sulfur hexafluoride (SF_6) for the effect of chemical sinks, Atmos. Chem. Phys., 24, 4193–4215, doi: 10.5194/acp-24-4193-2024, 2024a.

Garny, H., Ploeger, F., Abalos, M., Bönisch, H., Castillo, A. E., von Clarmann, T., Diallo, M., Engel, A., Laube, J. C., Linz, M., Neu, J. L., Podglajen, A., Ray, E., Rivoire, L., Saunders, L. N., Stiller, G., Voet, F., Wagenhäuser, T., and Walker, K. A.: Age of stratospheric air: progress on processes, observations, and long-term trends, Rev. Geophys., 62, e2023RG000 832, doi: https://doi.org/10.1029/2023RG000832, 2024b.

Gettelman, A.: Rainbows and climate change: a tutorial on climate model diagnostics and parameterization, Geosci. Model Dev., 16, 4937–4956, doi: 10.5194/gmd-16-4937-2023, 2023.

Gettelman, A., Hoor, P., Pan, L. L., Randel, W. J., Hegglin, M. I., and Birner, T.: The extra tropical upper troposphere and lower stratosphere, Rev. Geophys., 49, RG3003, doi: 10.1029/2011RG000355, 2011.

Gupta, A., Linz, M., Curbelo, J., Pauluis, O., Gerber, E. P., and Kinnison, D. E.: Estimating the Meridional Extent of Adiabatic Mixing in the Stratosphere Using Age-Of-Air, J. Geophys. Res. Atmos., 128, doi: 10.1029/2022JD037712, 2023.

Hall, T. M. and Plumb, R. A.: Age as a diagnostic of stratospheric transport, J. Geophys. Res., 99, 1059–1070, 1994.

Haynes, P. and Shuckburgh, E.: Effective diffusivity as a diagnostic of atmospheric transport: 1. Stratosphere, J. Geophys. Res., 105, 22 777–22 794, doi: https://doi.org/10.1029/2000JD900093, 2000a.

Haynes, P. and Shuckburgh, E.: Effective diffusivity as a diagnostic of atmospheric transport: 2. Troposphere and lower stratosphere, J. Geophys. Res., 105, 22 795–22 810, doi: https://doi.org/10.1029/2000JD900092, 2000b.

Hegglin, M. I., Brunner, D., Wernli, H., Schwierz, C., Martius, O., Hoor, P., Fischer, H., Spelten, N., Schiller, C., Krebsbach, M., Parchatka, U., Weers, U., Staehelin, J., and Peter, T.: Tracing troposphere-to-stratosphere transport above a mid-latitude deep convective system, Atmos. Chem. Phys., 4, 169–206, 2004.

Highwood, E. J. and Hoskins, B. J.: The tropical tropopause, Q. J. R. Meteorol. Soc., 124, 1579 – 1604, 1998.

Holton, J. R., Haynes, P. H., McIntyre, M. E., Douglass, A. R., Rood, R. B., and Pfister, L.: Stratosphere-Troposphere Exchange, Rev. Geophys., 33, 403–439, 1995.

Hoppe, C. M., Hoffmann, L., Konopka, P., Grooß, J.-U., Ploeger, F., Günther, G., Jöckel, P., and Müller, R.: The implementation of the CLaMS Lagrangian transport core into the chemistry climate model EMAC 2.40.1: application on age of air and transport of long-lived trace species, Geosci. Model Dev., 7, 2639–2651, doi: 10.5194/gmd-7-2639-2014, 2014.

Hotelling, H.: Analysis of a complex of statistical variables into principal components, Journal of Educational Psychology, 24, 417–441, doi: 10.1037/h0071325, 1933.

Hsu, C. J. and Plumb, R. A.: Nonaxisymmetric Thermally Driven Circulations and Upper-Tropospheric Monsoon Dynamics, J. Atmos. Sci., 57, 1255–1276, 2001.

Ivaniha, O., Abalos, M., Calvo, N., Stiller, G. P., Shah, K., and Davis, S.: Stratospheric Subtropical Transport Barriers in CESM1-WACCM and Observations: Climatology, Variability, and Trends, J. Geophys. Res. Atmos., 131, doi: 10.1029/2025JD045069, 2026.

James, R. and Legras, B.: Mixing processes and exchanges in the tropical and the subtropical UT/LS, Atmos. Chem. Phys., 9, 25–38, 2009.

Johnson, D. G., Jucks, K. W., Traub, W. A., Chance, K. V., Toon, G. C., Russell III, J. M., and McCormick, M. P.: Stratospheric age spectra derived from observations of water vapor and methane, J. Geophys. Res., 104, 21 595, doi: 10.1029/1999JD900363, 1999.

Jolliffe, I. and Cadima, J.: Principal component analysis: a review and recent developments, Philosophical transactons, Series A, Mathematical, physical and engineering sciences, 374, doi: 10.1098/rsta.2015.0202., 2016a.

Jolliffe, I. T. and Cadima, J.: Principal component analysis: a review and recent developments, Nature Methods, 13, doi: 10.1038/nmeth.4346, 2016b.

Kaas, E., Lauritzen, P. H., Sørensen, J. H., and Christensen, J. H.: A Hybrid Eulerian–Lagrangian Numerical Scheme for Solving the Tracer Advection Equation, Geosci. Model Dev., 6, 2023–2047, doi: 10.5194/gmd-6-2023-2013, 2013.

Kaumanns, J., Voet, F., Ploeger, F., Preuße, P., Ungermann, J., and Hegglin, M. I.: On using neural networks to predict mean Age-of-Air from long-lived tracers, EGUsphere, 2026, 1–36, doi: 10.5194/egusphere-2026-4183, 2026.

Keene, W. C., Galloway, J. N., and Holden, J. D.: Measurement of HCl in clean marine air, J. Geophys. Res. Oceans, 88, 5122–5130, doi: 10.1029/JC088iC09p05122, 1983.

Kirsch, A.: (Implicit) Ensembles of Ensembles: Epistemic Uncertainty Collapse in Large Models, Trans. Mach. Learn. Res., URL `https://arxiv.org/abs/2409.02628`, 2025.

Konopka, P., Spang, R., Günther, G., Müller, R., McKenna, D. S., Offermann, D., and Riese, M.: How homogeneous and isotropic is stratospheric mixing? Comparison of CRISTA-1 observations with transport studies based on the Chemical Lagrangian Model of the Stratosphere (CLaMS), Quart. J. Roy. Meteorol. Soc., 131, 565–579, 2005.

Konopka, P., Grooß, J.-U., Günther, G., Ploeger, F., Pommrich, R., Müller, R., and Livesey, N.: Annual cycle of ozone at and above the tropical tropopause: observations versus simulations

with the Chemical Lagrangian Model of the Stratosphere (CLaMS), Atmos. Chem. Phys., 10, 121–132, doi: 10.5194/acp-10-121-2010, 2010.

Konopka, P., Ploeger, F., Tao, M., Birner, T., and Riese, M.: Hemispheric asymmetries and seasonality of mean age of air in the lower stratosphere: Deep versus shallow branch of the Brewer-Dobson circulation, J. Geophys. Res., 120, 2053–2066, doi: 10.1002/2014JD022429, 2015.

Konopka, P., Tao, M., von Hobe, M., Hoffmann, L., Kloss, C., Ravegnani, F., Volk, C. M., Lauther, V., Zahn, A., Hoor, P., and Ploeger, F.: Tropospheric transport and unresolved convection: numerical experiments with CLaMS-2.0/MESSy, Geosci. Model Dev., 15, 7471–7487, doi: 10.5194/gmd-15-7471-2022, 2022.

Konopka, P., Ploeger, F., D'Amato, F., Campos, T., von Hobe, M., Honomichl, S. B., Hoor, P., Pan, L. L., Santee, M. L., Viciani, S., Walker, K. A., and Hegglin, M. I.: Isentropic Mixing vs. Convection in CLaMS-3.0/MESSy: Evaluation Using Satellite Climatologies and In Situ Carbon Monoxide Observations, EGUsphere, 2025, 1–32, doi: 10.5194/egusphere-2025-1155, 2025.

Konopka, P. and Tao, M. and Ploeger, F. and Diallo, M. and Riese, M.: Tropospheric mixing and parametrization of unresolved convective updrafts as implemented in the Chemical Lagrangian Model of the Stratosphere (CLaMS v2.0), Geosci. Model Dev., 12, 2441–2462, doi: 10.5194/gmd-12-2441-2019, 2019.

Kouznetsov, R., Sofiev, M., Vira, J., and Stiller, G.: Simulating age of air and the distribution of SF6 in the stratosphere with the SILAM model, Atmos. Chem. Phys., 20, 5837–5859, doi: 10.5194/acp-20-5837-2020, 2020.

Krasauskas, L., Ungermann, J., Ensmann, S., Krisch, I., Kretschmer, E., Preusse, P., and Riese, M.: 3-D tomographic limb sounder retrieval techniques: irregular grids and Laplacian regularisation, Atmos. Meas. Tech., 12, 853–872, doi: 10.5194/amt-12-853-2019, 2019.

Krishnamurti, T. N. and Bhalme, H. N.: Oscillations of a monsoon system. Part I. Observational aspects, J. Atmos. Sci., 33, 1937—-1954, doi: 10.1175/1520-0469(1976)033<1937:OOAMSP>2.0.CO;2, 1976.

Krüger, K., Langematz, U., Grenfell, J. L., and Labitzke, K.: Climatological features of stratospheric streamers in the FUB-CMAM with increased horizontal resolution, Atmos. Chem. Phys., 5, 547–562, doi: 10.5194/acp-5-547-2005, 2005.

Kunz, A., Konopka, P., Müller, R., and Pan, L. L.: Dynamical tropopause based on isentropic potential vorticity gradients, J. Geophys. Res., 116, D01 110, doi: 10.1029/2010JD014343, 2011.

Kunz, A., Müller, R., Homonnai, V., M. Jánosi, I., Hurst, D., Rap, A., M. Forster, P., Rohrer, F., Spelten, N., and Riese, M.: Extending water vapor trend observations over Boulder into the tropopause region: Trend uncertainties and resulting radiative forcing, J. Geophys. Res., 118, 11,269–11,284, doi: 10.1002/jgrd.50831, 2013.

Kunz, A., Sprenger, M., and Wernli, H.: Climatology of potential vorticity streamers and associated isentropic transport pathways across PV gradient barriers, J. Geophys. Res., 120, 3802–3821, doi: 10.1002/2014JD022615, 2014JD022615, 2015.

Laube, J. C., Elvidge, E. C. L., Adcock, K. E., Baier, B., Brenninkmeijer, C. A. M., Chen, H., Droste, E. S., Grooß, J.-U., Heikkinen, P., Hind, A. J., Kivi, R., Lojko, A., Montzka, S. A., Oram, D. E., Randall, S., Röckmann, T., Sturges, W. T., Sweeney, C., Thomas, M., Tuffnell, E., and Ploeger, F.: Investigating stratospheric changes between 2009 and 2018 with halogenated trace gas data from aircraft, AirCores, and a global model focusing on CFC-11, Atmos. Chem. Phys., 20, 9771–9782, doi: 10.5194/acp-20-9771-2020, 2020.

Li, Q., Jiang, J. H., Wu, D. L., Read, W. G., Livesey, N. J., Waters, J. W., Zhang, Y., Wang, B., Filipiak, M. J., Davis, C. P., Turquety, S., Wu, S., Park, R. J., Yantosca, R. M., and Jacob, D. J.: Convective outflow of South Asian pollution: A global CTM simulation compared with EOS MLS observations, Geophys. Res. Lett., 32, L14 826, doi: 10.1029/2005GL022762, 2005.

Lin, P. and Fu, Q.: Changes in various branches of the Brewer–Dobson circulation from an ensemble of chemistry climate models, J. Geophys. Res., 118, 73–84, doi: https://doi.org/10.1029/2012JD018813, 2013.

Loeffel, S., Eichinger, R., Garny, H., Reddmann, T., Fritsch, F., Versick, S., Stiller, G., and Haenel, F.: The impact of sulfur hexafluoride (SF6) sinks on age of air climatologies and trends, Atmos. Chem. Phys., 22, 1175–1193, doi: 10.5194/acp-22-1175-2022, 2022.

Machta, L.: Transport in the Stratosphere and Through the Tropopause, vol. 6, Academic Press, doi: 10.1016/S0065-2687(08)60114-4, 1959.

Maiss, M., Steele, L., Francey, R. J., Fraser, P. J., Langenfelds, R. L., Trivett, N. B., and Levin, I.: Sulfur hexafluoride — A powerful new atmospheric tracer, Atmos. Environment, 30, 1621–1629, doi: https://doi.org/10.1016/1352-2310(95)00425-4, 1996.

Manney, G. L., Daffer, W. H., Lawrence, B. N., Hegglin, M. I., Schwartz, M. J., Fishbein, R. E. S., Pawson, S., McCormack, D. P., and Zawodny, J. R.: Reanalysis comparisons of upper tropospheric/lower stratospheric jets and multiple tropopauses, Atmos. Chem. Phys., 17, 11 541–11 566, doi: 10.5194/acp-17-11541-2017, 2017.

Marro, M., Nironi, C., Salizzoni, P., and Soulhac, L.: Dispersion of a Passive Scalar Fluctuating Plume in a Turbulent Boundary Layer. Part II: Analytical Modelling, Bound. Lay. Meteor., 156, 447–469, doi: 10.1007/s10546-015-0041-9, 2015.

Mason, R. B. and Anderson, C. E.: The development and decay of the 100-mb. summertime anticyclone over southern Asia, Monthly Weather Review, 91, 3–12, doi: 10.1175/1520-0493(1963)091<0003:TDADOT>2.3.CO;2, 1963.

McFarlane, N. A.: The effect of orographically excited gravity wave drag on the general circulation of the lower stratosphere and troposphere, J. Atmos. Sci., 44, 1775–1800, 1987.

McKenna, D. S., Grooß, J.-U., Günther, G., Konopka, P., Müller, R., Carver, G., and Sasano, Y.: A new Chemical Lagrangian Model of the Stratosphere (CLaMS) 2. Formulation of chemistry scheme and initialization, J. Geophys. Res., 107, doi: 10.1029/2000JD000113, 2002a.

McKenna, D. S., Konopka, P., Grooß, J.-U., Günther, G., Müller, R., Spang, R., Offermann, D., and Orsolini, Y.: A new Chemical Lagrangian Model of the Stratosphere (CLaMS) 1. Formulation of advection and mixing, J. Geophys. Res., 107, doi: 10.1029/2000JD000114, 2002b.

McLandress, C. and Shepherd, T. G.: Simulated anthropogenic changes in the Brewer-Dobson circulation, including its extension to high latitudes, J. Clim., 22, 1516–1540, doi: 10.1175/2008JCLI2679.1, 2009.

Minschwaner, K., Hoffmann, L., Brown, A., Riese, M., Müller, R., and Bernath, P. F.: Stratospheric loss and atmospheric lifetimes of CFC-11 and CFC-12 derived from satellite observations, Atmos. Chem. Phys., 13, 4253–4263, doi: 10.5194/acp-13-4253-2013, 2013.

Molina, M. J. and Rowland, F. S.: Stratospheric sink for chlorofluoromethanes: chlorine atom-catalyzed destruction of ozone, Nature, 249, 810–812, 1974.

Mote, P. W., Rosenlof, K. H., McIntyre, M. E., Carr, E. S., Gille, J. C., Holton, J. R., Kinnersley, J. S., Pumphrey, H. C., Russell, J. M., and Waters, J. W.: An atmospheric tape recorder: The imprint of tropical tropopause temperatures on stratospheric water vapor, J. Geophys. Res., 101, 3989–4006, doi: 10.1029/95JD03422, 1996.

Orbe, C., Takacs, L. L., El Akkraoui, A., Wargan, K., Molod, A., and Pawson, S.: Changes in Stratospheric Climate and Age-Of-Air in Recent GEOS Systems Since MERRA-2, J. Adv. Model. Earth Syst., 17, e2024MS004 442, doi: 10.1029/2024MS004442, 2025.

O'Shea, S. J., Allen, G., Gallagher, M. W., Bauguitte, S. J.-B., Illingworth, S. M., Le Breton, M., Muller, J. B. A., Percival, C. J., Archibald, A. T., Oram, D. E., Parrington, M., Palmer, P. I., and Lewis, A. C.: Airborne observations of trace gases over boreal Canada during BORTAS: campaign climatology, air mass analysis and enhancement ratios, Atmos. Chem. Phys., 13, 12 451–12 467, doi: 10.5194/acp-13-12451-2013, 2013.

Ovadia, Y., Fertig, E., Ren, J., Lakshminarayanan, B., Nowozin, S., Sculley, D., and Snoek, J.: Can you trust your model's uncertainty? Evaluating predictive uncertainty under dataset shift, in: Advances in Neural Information Processing Systems (NeurIPS), vol. 32, 2019.

Palmer, T. N., Shutts, G. J., and Swinbank, R.: Alleviation of a systematic weterly bias in general circulation and numerical weather prediction models trough an orographic gravity wave drag parameterization, Quart. J. Roy. Meteorol. Soc., 112, 1001–1093, 1986.

Pan, L. L., Kinnison, D., Liang, Q., Chin, M., Santee, M. L., Flemming, J., Smith, W. P., Honomichl, S. B., Bresch, J. F., Lait, L. R., Zhu, Y., Tilmes, S., Colarco, P. R., Warner, J., Vuvan, A., Clerbaux, C., Atlas, E. L., Newman, P. A., Thornberry, T., Randel, W. J., and Toon, O. B.: A Multimodel Investigation of Asian Summer Monsoon UTLS Transport Over the Western Pacific, J. Geophys. Res. Atmos., 127, e2022JD037 511, doi: https://doi.org/10.1029/2022JD037511, 2022.

Park, M., Randel, W. J., Kinnison, D. E., Garcia, R. R., and Choi, W.: Seasonal variation of methane, water vapor, and nitrogen oxides near the tropopause: Satellite observations and model simulations, J. Geophys. Res. Atmos., 109, D03 302, doi: 10.1029/2003JD003706, 2004.

Park, M., Randel, W. J., Gettelman, A., Massie, S. T., and Jiang, J. H.: Transport above the Asian summer monsoon anticyclone inferred from Aura Microwave Limb Sounder tracers, J. Geophys. Res., 112, D16 309, doi: 10.1029/2006JD008294, 2007.

Pearson, K.: LIII. On lines and planes of closest fit to systems of points in space, The London, Edinburgh, and Dublin philosophical magazine and journal of science, 2, 559–572, doi: 10.1080/14786440109462720, 1901.

Ploeger, F. and Birner, T.: Seasonal and inter-annual variability of lower stratospheric age of air spectra, Atmos. Chem. Phys., 2016, 10 195–10 213, doi: 10.5194/acp-16-10195-2016, 2016.

Ploeger, F. and Garny, H.: Hemispheric asymmetries in recent changes in the stratospheric circulation, Atmos. Chem. Phys., 22, 5559–5576, doi: 10.5194/acp-22-5559-2022, 2022.

Ploeger, F., Günther, G., Konopka, P., Fueglistaler, S., Müller, R., Hoppe, C., Kunz, A., Spang, R., Grooß, J.-U., and Riese, M.: Horizontal water vapor transport in the lower stratosphere from subtropics to high latitudes during boreal summer, J. Geophys. Res., 118, 8111–8127, doi: 10.1002/jgrd.50636, 2013.

Ploeger, F., Abalos, M., Birner, T., Konopka, P., Legras, B., Müller, R., and Riese, M.: Quantifying the effects of mixing and residual circulation on trends of stratospheric mean age of air, Geophys. Res. Lett., 42, 2047–2054, doi: 10.1002/2014GL062927, 2015.

Ploeger, F., Konopka, P., Walker, K., and Riese, M.: Quantifying pollution transport from the Asian monsoon anticyclone into the lower stratosphere, Atmos. Chem. Phys., 17, 7055–7066, doi: 10.5194/acp-17-7055-2017, 2017.

Ploeger, F., Legras, B., Charlesworth, E., Yan, X., Diallo, M., Konopka, P., Birner, T., Tao, M., Engel, A., and Riese, M.: How robust are stratospheric age of air trends from different reanalyses?, Atmos. Chem. Phys., 19, 6085–6105, doi: 10.5194/gmd-12-2441-2019, 2019a.

Ploeger, F., Legras, B., Charlesworth, E., Yan, X., Diallo, M., Konopka, P., Birner, T., Tao, M., Engel, A., and Riese, M.: How robust are stratospheric age of air trends from different reanalyses?, Atmos. Chem. Phys., 19, 6085–6105, doi: 10.5194/gmd-12-2441-2019, 2019b.

Ploeger, F., Diallo, M., Charlesworth, E., Konopka, P., Legras, B., Laube, J. C., Grooß, J.-U., Günther, G., Engel, A., and Riese, M.: The stratospheric Brewer–Dobson circulation inferred from age of air in the ERA5 reanalysis, Atmos. Chem. Phys., 21, 8393–8412, doi: 10.5194/acp-21-8393-2021, 2021a.

Ploeger, F., Diallo, M., Charlesworth, E., Konopka, P., Legras, B., Laube, J. C., Grooß, J.-U., Günther, G., Engel, A., and Riese, M.: The stratospheric Brewer–Dobson circulation inferred from age of air in the ERA5 reanalysis, Atmos. Chem. Phys., 21, 8393–8412, doi: 10.5194/acp-21-8393-2021, 2021b.

Plumb, R. A.: A 'tropical pipe' model of stratospheric transport, J. Geophys. Res., 101, 3957–3972, 1996.

Plumb, R. A.: Tracer interrelationships in the stratosphere, Rev. Geophys., 45, RG4005, doi: https://doi.org/10.1029/2005RG000179, 2007.

Plumb, R. A. and Eluszkiewicz, J.: The Brewer-Dobson Circulation: Dynamics of the Tropical Upwelling, J. Atmos. Sci., 56, 868–890, 1999.

Plumb, R. A. and Ko, M. K. W.: Interrelationships between mixing ratios of long-lived stratospheric constituents, J. Geophys. Res., 97, 10 145–10 156, 1992.

Polvani, L. M. and Plumb, R. A.: Rossby Wave Breaking, Microbreaking, Filamentation, and Secondary Vortex Formation: The Dynamics of a Perturbed Vortex, J. Atmos. Sci., 49, 462–476, 1992.

Polvani, L. M., Waugh, D., and Plumb, R. A.: On the subtropical edge of the stratospheric surf zone, J. Atmos. Sci., 52, 1288–1309, 1995.

Polvani, L. M., Wang, L., Abalos, M., Butchart, N., Chipperfield, M. P., Dameris, M., Deushi, M., Dhomse, S. S., Jöckel, P., Kinnison, D., Michou, M., Morgenstern, O., Oman, L. D., Plummer, D. A., and Stone, K. A.: Large Impacts, Past and Future, of Ozone-Depleting Substances on Brewer-Dobson Circulation Trends: A Multimodel Assessment, J. Geophys. Res. Atmos., 124, 6669–6680, doi: 10.1029/2018JD029516, 2019.

Pommrich, R., Müller, R., Grooß, J.-U., Konopka, P., Ploeger, F., Vogel, B., Tao, M., Hoppe, C. M., Günther, G., Spelten, N., Hoffmann, L., Pumphrey, H.-C., Viciani, S., D'Amato, F., Volk, C. M., Hoor, P., Schlager, H., and Riese, M.: Tropical troposphere to stratosphere transport of carbon monoxide and long-lived trace species in the Chemical Lagrangian Model of the Stratosphere (CLaMS), Geosci. Model Dev., 7, 2895–2916, doi: 10.5194/gmd-7-2895-2014, 2014.

Popovic, J. M. and Plumb, R. A.: Eddy shedding from the upper tropospheric Asian monsoon anticyclone, J. Atmos. Sci., 58, 93–104, doi: 10.1175/1520-0469(2001)058<0093:ESFTUT>2.0.CO;2, 2001.

Poshyvailo, L., Müller, R., Konopka, P., Günther, G., Riese, M., Podglajen, A., and Ploeger, F.: Sensitivities of modelled water vapour in the lower stratosphere: temperature uncertainty, effects of horizontal transport and small-scale mixing, Atmos. Chem. Phys., 18, 8505–8527, doi: 10.5194/acp-18-8505-2018, 2018.

Poshyvailo-Strube, L., Müller, R., Fueglistaler, S., Hegglin, M. I., Laube, J. C., Volk, C. M., and Ploeger, F.: How can Brewer–Dobson circulation trends be estimated from changes in stratospheric water vapour and methane?, Atmos. Chem. Phys., 22, 9895–9914, doi: 10.5194/acp-22-9895-2022, 2022.

Randel, W. J. and Park, M.: Deep convective influence on the Asian summer monsoon anticyclone and associated tracer variability observed with Atmospheric Infrared Sounder (AIRS), J. Geophys. Res., 111, D12 314, doi: 10.1029/2005JD006490, 2006.

Ray, E. A., Moore, F. L., Elkins, J. W., Rosenlof, K. H., Laube, J. C., Röckmann, T., Marsh, D. R., and Andrews, A. E.: Quantification of the SF_6 lifetime based on mesospheric loss measured in the stratospheric polar vortex, J. Geophys. Res. Atmos., 122, 4626–4638, doi: 10.1002/2016JD026198, 2017.

Ray, E. A., Moore, F. L., Garny, H., Hintsa, E. J., Hall, B. D., Dutton, G. S., Nance, D., Elkins, J. W., Wofsy, S. C., Pittman, J., Daube, B., Baier, B. C., Li, J., and Sweeney, C.: Age of air from in situ trace gas measurements: insights from a new technique, Atmos. Chem. Phys., 24, 12 425–12 445, doi: 10.5194/acp-24-12425-2024, 2024.

Ray, E. A., Baier, B. C., Moore, F. L., Garny, H., Jesswein, M., Li, J., Hintsa, E. J., Hall, B. D., Dutton, G. S., Nance, D., and Plummer, D.: Observed stratospheric mean age decrease consistent with circulation acceleration, Nature Geoscience, 19, 612–618, doi: 10.1038/s41561-026-02011-3, 2026.

Rex, M., Harris, N. R. P., von der Gathen, P., Lehmann, R., Braathen, G. O., Reimer, E., Beck, A., Chipperfield, M. P., Alfier, R., Allaart, M., O'Connor, F., Dier, H., Dorokhov, V., Fast, H., Gil, M., Kyrö, E., Litynska, Z., Mikkelsen, I. S., Molyneux, M. G., Nakane, H., Notholt, J., Rummukainen, M., Viatte, P., and Wenger, J.: Prolonged stratospheric ozone loss in the 1995–96 Arctic winter, Nature, 389, 835–838, doi: 10.1038/39849, 1997.

Riese, M., Grooß, J.-U., Feck, T., and Rohs, S.: Long-term changes of hydrogen-containing species in the stratosphere, J. Atmos. Solar Terr. Phys., 68, 1973–1979, 2006.

Riese, M., Ploeger, F., Rap, A., Vogel, B., Konopka, P., Dameris, M., and Forster, P.: Impact of uncertainties in atmospheric mixing on simulated UTLS composition and related radiative effects, J. Geophys. Res. Atmos., 117, n/a–n/a, doi: 10.1029/2012JD017751, 2012.

Riese, M., Oelhaf, H., Preusse, P., Blank, J., Ern, M., Friedl-Vallon, F., Fischer, H., Guggenmoser, T., Höpfner, M., Hoor, P., Kaufmann, M., Orphal, J., Plöger, F., Spang, R., Suminska-Ebersoldt, O., Ungermann, J., Vogel, B., and Woiwode, W.: Gimballed Limb Observer for Radiance Imaging of the Atmosphere (GLORIA) scientific objectives, Atmos. Meas. Tech., 7, 1915–1928, doi: 10.5194/amt-7-1915-2014, 2014.

Riese, M., Hoor, P., Rolf, C., Kunkel, D., Vogel, B., Köllner, F., Pöhlker, M., Ploeger, F., Ungermann, J., Woiwode, W., Johansson, S., Bauer, R., Barmpounis, K., Borrmann, S., Brauner, P., Clemens, J., Dragoneas, A., Ekinci, F., Emig, N., Engel, A., Eppers, O., Fadnavis, S., Friedl-Vallon, F., Geldenhuys, M., Günther, G., Grooß, J., Hegglin, M., Höpfner, M., Jesswein, M., Joppe, P., Kaumanns, J., Kachula, O., Keber, T., Kretschmer, E., H.C., L., Lauther, V., Lloyd, P., Molleker, S., Müller, R., Neubert, T., Ort, L., Pöschl, U., Pöhlker, C., Rapp, M., Retzlaff, M., Rhode, S., Schneider, J., Schuck, T., Sinnhuber, B.-M., Spelten, N., Strobel, J., Tomsche, L., Turhal, K., van Luijt, R., Versick, S., Voigt, C., Volk, M., von Hobe, M., Weyland, F., Zahn, A., Ziereis, H., and Zlotos, L.: Long-range transport of polluted Asian summer monsoon air to high latitudes during the PHILEAS campaign in the boreal summer 2023, Bull. Amer. Meteor. Soc., pp. BAMS–D–24–0232.1, doi: 10.1175/BAMS-D-24-0232.1, 2025.

Rohs, S., Schiller, C., Riese, M., Engel, A., Schmidt, U., Wetter, T., Levin, I., Nakazawa, T., and Aoki, S.: Long-term changes of methane and hydrogen in the stratosphere in the period 1978 – 2003 and their impact on the abundance of stratospheric water vapor, J. Geophys. Res., 111, D14315, doi: 10.1029/2005JD006877, 2006.

Rosenlof, K. H., Tuck, A. F., Kelly, K. K., Russell, J. M., and McCormick, M. P.: Hemispheric asymmetries in water vapor and inferences about transport in the lower stratosphere, J. Geophys. Res., 102, 13 213–13 234, 1997.

Rousseeuw, P. J.: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, jcam, 20, 53–65, doi: 10.1016/0377-0427(87)90125-7, 1987.

Rowland, F. S., Spencer, J. E., and Molina, M. J.: Stratospheric formation and photolysis of chlorine nitrate, J. Phys. Chem., 80, 2711–2713, doi: 10.1021/j100565a002, 1976.

Seguel, R. J., Opazo, C., Cohen, Y., Cooper, O. R., Gallardo, L., Sinnhuber, B.-M., Obersteiner, F., Zahn, A., Hoor, P., Rohs, S., and Marsing, A.: Hemispheric differences in ozone across

the stratosphere–troposphere exchange region, Atmos. Chem. Phys., 25, 8553–8573, doi: 10.5194/acp-25-8553-2025, 2025.

Solomon, S., Rosenlof, K. H., Portmann, R. W., Daniel, J. S., Davis, S. M., Sanford, T. J., and Plattner, G.-K.: Contributions of stratospheric water vapor to decadal changes in the rate of global warming, Science, 327, 1219–1223, doi: 10.1126/science.1182488, 2010.

Stiller, G. P., von Clarmann, T., Haenel, F., Funke, B., Glatthor, N., Grabowski, U., Kellmann, S., Kiefer, M., Linden, A., Lossow, S., and López-Puertas, M.: Observed temporal evolution of global mean age of stratospheric air for the 2002 to 2010 period, Atmos. Chem. Phys., 12, 3311–3331, doi: 10.5194/acp-12-3311-2012, 2012.

Stiller, G. P., Fierli, F., Ploeger, F., Cagnazzo, C., Funke, B., Haenel, F., Reddmann, T., Riese, M., and von Clarmann, T.: Shift of subtropical transport barriers explains observed hemispheric asymmetry of decadal trends of age of air, Atmos. Chem. Phys., 17, 11 177–11 192, 2017.

Stohl, A., Forster, C., Frank, A., Seibert, P., and Wotawa, G.: Technical note: The Lagrangian particle dispersion model FLEXPART version 6.2, Atmos. Chem. Phys., 5, 2461–2474, doi: 10.5194/acp-5-2461-2005, 2005.

Stolarski, R. S. and Cicerone, R. J.: Stratospheric Chlorine: A possible sink for ozone, Can. J. Chem., 52, 1610–1615, 1974.

Strube, C., Ern, M., Preusse, P., and Riese, M.: Removing spurious inertial instability signals from gravity wave temperature perturbations using spectral filtering methods, Atmos. Meas. Tech., 13, 4927–4945, doi: 10.5194/amt-13-4927-2020, 2020.

Tie, X., Zhang, R., Brasseur, G., Emmons, L., and Lei, W.: Effects of lightning on reactive nitrogen and nitrogen reservoir species in the troposphere, J. Geophys. Res. Atmos., 106, 3167–3178, doi: https://doi.org/10.1029/2000JD900565, 2001.

Ungermann, J.: Improving retrieval quality for airborne limb-sounders by horizontal regularisation, Atmos. Meas. Tech., 6, 15–32, doi: 10.5194/amt-6-15-2013, 2013.

Ungermann, J., Hoffmann, L., Preusse, P., Kaufmann, M., and Riese, M.: Tomographic retrieval approach for mesoscale gravity wave observations by the PREMIER Infrared Limb-Sounder, Atmos. Meas. Tech., 3, 339–354, doi: 10.5194/amt-3-339-2010, 2010a.

Ungermann, J., Kaufmann, M., Hoffmann, L., Preusse, P., Oelhaf, H., Friedl-Vallon, F., and Riese, M.: Towards a 3-D tomographic retrieval for the air-borne limb-imager GLORIA, Atmos. Meas. Tech., 3, 1647–1665, doi: 10.5194/amt-3-1647-2010, 2010b.

Ungermann, J., Blank, J., Lotz, J., Leppkes, K., Hoffmann, L., Guggenmoser, T., Kaufmann, M., Preusse, P., Naumann, U., and Riese, M.: A 3-D tomographic retrieval approach with advection compensation for the air-borne limb-imager GLORIA, Atmos. Meas. Tech., 4, 2509–2529, doi: 10.5194/amt-4-2509-2011, 2011.

Ungermann, J., Blank, J., Dick, M., Ebersoldt, A., Friedl-Vallon, F., Giez, A., Guggenmoser, T., Höpfner, M., Jurkat, T., Kaufmann, M., Kaufmann, S., Kleinert, A., Krämer, M., Latzko, T., Oelhaf, H., Olchewski, F., Preusse, P., Rolf, C., Schillings, J., Suminska-Ebersoldt, O., Tan, V., Thomas, N., Voigt, C., Zahn, A., Zöger, M., and Riese, M.: Level 2 processing for the imaging Fourier transform spectrometer GLORIA: derivation and validation of temperature and trace gas volume mixing ratios from calibrated dynamics mode spectra, Atmos. Meas. Tech., 8, 2473–2489, doi: 10.5194/amt-8-2473-2015, 2015.

Ungermann, J., Ern, M., Kaufmann, M., Müller, R., Spang, R., Ploeger, F., Vogel, B., and Riese, M.: Observations of PAN and its confinement in the Asian summer monsoon anticyclone in high spatial resolution, Atmos. Chem. Phys., 16, 8389–8403, doi: 10.5194/acp-16-8389-2016, 2016.

Ungermann, J., Kleinert, A., Maucher, G., Bartolomé, I., Friedl-Vallon, F., Johansson, S., Krasauskas, L., and Neubert, T.: Quantification and mitigation of the instrument effects and uncertainties of the airborne limb imaging FTIR GLORIA, Atmos. Meas. Tech., 15, 2503–2530, doi: 10.5194/amt-15-2503-2022, 2022.

Voet, F., Ploeger, F., Laube, J., Preusse, P., Konopka, P., Grooß, J.-U., Ungermann, J., Sinnhuber, B.-M., Höpfner, M., Funke, B., Wetzel, G., Johansson, S., Stiller, G., Ray, E., and Hegglin, M. I.: On the estimation of stratospheric age of air from correlations of multiple trace gases, Atmos. Chem. Phys., 25, 3541–3565, doi: 10.5194/acp-25-3541-2025, 2025.

Vogel, B., Pan, L. L., Konopka, P., Günther, G., Müller, R., Hall, W., Campos, T., Pollack, I., Weinheimer, A., Wei, J., Atlas, E. L., and Bowman, K. P.: Transport pathways and signatures of mixing in the extratropical tropopause region derived from Lagrangian model simulations, J. Geophys. Res., 116, doi: 10.1029/2010JD014876, 2011.

Vogel, B., Günther, G., Müller, R., Grooß, J.-U., Afchine, A., Bozem, H., Hoor, P., Krämer, M., Müller, S., Riese, M., Rolf, C., Spelten, N., Stiller, G. P., Ungermann, J., and Zahn, A.: Long-range transport pathways of tropospheric source gases originating in Asia into the northern lower stratosphere during the Asian monsoon season 2012, Atmos. Chem. Phys., 16, 15 301–15 325, doi: 10.5194/acp-16-15301-2016, 2016.

Vogel, B., Lauther, V., Köllner, F., Ekinci, F., Rolf, C., Strobel, J., van Luijt, R., Volk, C. M., Borrmann, S., Dragoneas, A., Eppers, O., Molleker, S., Hoor, P., Ort, L., Weyland, F., Zahn, A., Clemens, J., Günther, G., Kachula, O., Müller, R., Ploeger, F., and Riese, M.: Continental and marine source regions contributing to the outflow of the Asian summer monsoon anticyclone during the PHILEAS campaign in summer 2023, Atmos. Chem. Phys., 26, 6283–6319, doi: 10.5194/acp-26-6283-2026, 2026.

Volk, C. M., Elkins, J. W., Fahey, D. W., Dutton, G. S., Gilligan, J. M., Loewenstein, M., Podolske, J. R., Chan, K. R., and Gunson, M. R.: Evaluation of source gas lifetimes from stratospheric observations, J. Geophys. Res., 102, 25 543–25 564, 1997.

von der Gathen, P., Rex, M., Harris, N. R. P., Lucic, D., Knudsen, B. M., Braathen, G. O., De Backer, H., Fabian, R., Fast, H., Gil, M., Kyrö, E., Mikkelsen, I. S., Rummukainen, M., Stähelin, J., and Varotsos, C.: Observational evidence for chemical ozone depletion over the Arctic in winter 1991–92, Nature, 375, 131–134, doi: 10.1038/375131a0, 1995.

von Neumann, J. and Richtmyer, R. D.: A Method for the Numerical Calculation of Hydrodynamic Shocks, J. Appl. Phys., 21, 232–237, doi: 10.1063/1.1699639, 1950.

Waugh, D. W.: The age of stratospheric air, Nature Geoscience, 2, 14–16, 2009.

Weigel, K., Hoffmann, L., Guenther, G., Khosrawi, F., Olschewski, F., Preusse, P., Spang, R., Stroh, F., and Riese, M.: A stratospheric intrusion at the subtropical jet over the Mediterranean Sea: air-borne remote sensing observations and model results, Atmos. Chem. Phys., 12, 8423–8438, doi: 10.5194/acp-12-8423-2012, 2012.

WMO: Meteorology – a three-dimensional science: Second session for the commission for aerology, WMO Bull., 6, 134–138, 1957.

Yan, X., Konopka, P., Ploeger, F., Podglajen, A., Wright, J. S., Müller, R., and Riese, M.: The efficiency of transport into the stratosphere via the Asian and North American summer monsoon circulations, Atmos. Chem. Phys., 19, 15 629–15 649, doi: 10.5194/acp-19-15629-2019, 2019.

Zhang, J., Li, D., Bian, J., and Bai, Z.: Deep stratospheric intrusion and Russian wildfire induce enhanced tropospheric ozone pollution over the northern Tibetan Plateau, Atmospheric Research, 259, 105 662, doi: https://doi.org/10.1016/j.atmosres.2021.105662, 2021.

Zhang, T. Y. and Suen, C. Y.: A Fast Parallel Algorithm for Thinning Digital Patterns, Communications of the ACM, 27, 236–239, doi: 10.1145/357994.358023, 1984.

Zuiderweg, A., Kaiser, J., Laube, J. C., Röckmann, T., and Holzinger, R.: Stable carbon isotope fractionation in the UV photolysis of CFC-11 and CFC-12, Atmos. Chem. Phys., 12, 4379–4385, doi: 10.5194/acp-12-4379-2012, 2012.

Band / Volume 735
Dynamical drivers of deep and shallow branches in the stratospheric circulation and transport
R.- A. Baikhadzhaev (2026), 120 pp
ISBN 978-3-95806-996-1 (Print)
ISBN 978-3-95806-997-8 (E-Book)

Band / Volume 736
Silicon heterojunction solar cells on thin wafers – prospects in different utilization scenarios
U. K. Chime (2026), 151 pp
ISBN 978-3-95806-998-5 (Print)
ISBN 978-3-95806-999-2 (E-Book)

Band / Volume 737
Advancing the understanding of actual evapotranspiration and critical soil moisture thresholds using lysimeter networks
X. Lu (2026), XXX, 159 pp
ISBN 978-3-69367-000-2 (Print)
ISBN 978-3-69367-001-9 (E-Book)

Band / Volume 738
Einfluss des automatisierten Fahrens auf die technoökonomische Performance von Verkehrssystemen
E. Jungblut (2026), VIII, 304 pp
ISBN 978-3-69367-006-4 (Print)
ISBN 978-3-69367-007-1 (E-Book)

Band / Volume 739
Photon-assisted Degradation and Recovery of Transparent Passivating Contact Silicon Solar Cells
B. Xu (2026), xii, 189 pp
ISBN 978-3-69367-008-8 (Print)
ISBN 978-3-69367-009-5 (E-Book)

Band / Volume 740
Effects of Microplastics on Carbon and Nitrogen Turnover Processes in the Rhizosphere of Crops
K. J. Neubert (2026), xxii, 193 pp
ISBN 978-3-69367-010-1 (Print)
ISBN 978-3-69367-011-8 (E-Book)

Band / Volume 741
Comprehensive Understanding and Optimization of the Tape Casting Process for the Fabrication of LLZO-Based Solid-State Battery Components
K. N. El K. Touidjine (2026), XV, 170 pp
ISBN 978-3-69367-016-3 (Print)
ISBN 978-3-69367-017-0 (E-Book)

Band / Volume 742
Microstructure and Functionality of Heterojunction contacts on crystalline Silicon solar cells
S. Alkhereibi (2026); 170 pp
ISBN 978-3-69367-018-7 (Print)
ISBN 978-3-69367-019-4 (E-Book)

Band / Volume 743
Deployment of Fuel Cell Vehicles in Road Transport and the Expansion of the Hydrogen Refueling Station Network – 2026 Update
E. Jungblut, T. Grube, J. Neußer (2026), iii, 28 pp
ISBN 978-3-69367-020-0 (Print)
ISBN 978-3-69367-021-7 (E-Book)

Band / Volume 744
Electrochemical CO_2 Reduction for Photovoltaic Energy Storage under Implementation-relevant Conditions
T. E. Cibaka (2026), getr. Paginierung
ISBN 978-3-69367-022-4 (Print)
ISBN 978-3-69367-023-1 (E-Book)

Band / Volume 745
Regional Feasibility of the German Heating Transition
K. Dabrock (2026), xiv, 250 pp
ISBN 978-3-69367-024-8 (Print)
ISBN 978-3-69367-025-5 (E-Book)

Band / Volume 746
In Bearbeitung

Band / Volume 747
Identifying and quantifying mixing processes in filaments using computer vision and Artificial Intelligence
J. Kaumanns (2026), 201 pp
ISBN 978-3-69367-028-6 (Print)
ISBN 978-3-69367-029-3 (E-Book)

Weitere ***Schriften des Verlags im Forschungszentrum Jülich*** unter
http://wwwzb1.fz-juelich.de/verlagextern1/index.asp

Mitglied der Helmholtz-Gemeinschaft

