Master Thesis

at the University Hamburg

Sub-/Millimeter Profile Retrieval with EOF and Neural Net

Author:
Maximilian Schaper

Supervisors:
Prof. Dr. Stefan Buehler
Dr. Manfred Brath

March 13, 2018

Topic: ISMAR profile retrieval with EOF and neural net

Abstract

This thesis examines the use of empirical orthogonal functions and neural nets to retrieve atmospheric profiles based on radiometer measurements. The retrieval is developed around millimeter and submillimeter measurements from two passive airborne radiometers (ISMAR and MARSS). As data source atmospheric profiles from the ICON model are used, matching the conditions of flight B897 of the FAAM BAe-146 Atmospheric Research Aircraft. The model data is complemented by the results of a radiative transfer simulation, utilizing the Atmospheric Radiative Transfer Simulator. This dataset is used to assess the retrieval's performance based on the bias and the fractional error. The retrieved quantities of interest are cloud snow water content, cloud ice water content, cloud liquid water content and rain water content. Also water vapour content is retrieved and used to demonstrate fundamental retrieval properties. Due to its lower variability throughout different atmospheric states, it can be retrieved more precisely. Additional error measures are introduced to estimate the retrieval performance solely based on measurements. At last, retrieval results for the B897 flight are compared to ICON model data. They imply that the retrieval works perfectly for frozen hydrometeor contents above $10^{-4}\,\mathrm{kg/m^3}$. At the lower edge of dense clouds a masking effect results in an underestimation of the hydrometeor content, while the smoothing error results in an overestimation beneath thin clouds with a sharp lower edge.

Kurzfassung

Die vorliegende Studie beschreibt die Anwedung von empirischen orthogonalen Funktionen und neuronalen Netzen auf Radiometermessungen, um aus ihnen atmosphärische Profile zu rekonstruieren. Das entwickelte Retrieval basiert auf Millimeter- und Submillimetermessungen zweier passiver, luftgestützter Radiometer (ISMAR und MARSS). Als Datengrundlage werden atmosphärische Profile aus dem ICON Modell verwendet, die den Flugbedingungen des Fluges B897 des FAAM BAe-146 Atmospheric Research Aircraft entsprechen. Die Modelldaten werden mit den Ergebnissen einer Strahlungstransfersimulation des 'Atmospheric Radiative Transfer Simulator' ergänzt, um den zugrunde liegenden Datensatz zu vervollständigen. Dieser Datensatz wird verwendet, um das Retrieval basierend auf dem Bias und 'fractional error' zu untersuchen. Bei den abgeleiteten Größen handelt es sich um Schnee, Eis, Wolkenwasser und Regen. Auch der Wasserdampfgehalt der Atmosphäre wird errechnet und zum Veranschaulichen der Retrieval-Methode herangezogen, da die geringere Variabilität von Wasserdampf den Retrievalprozess vereinfacht. Es werden zusätzliche Fehlermaße eingeführt, um die Retrieval Genauigkeit nur auf Grundlage von Messungen abzuschätzen. Zum Abschluss werden die Retrieval-Ergebnisse von Flug B897 mit entsprechenden ICON Modelldaten verglichen. Die Ergebnisse implizieren, dass sich die untersuchte Retrieval-Methode perfekt für das rekonstruieren von Eiswolken mit einem Hydrometeorgehalt von über $10^{-4}\,\mathrm{kg/m^3}$ eignet. An der Unterkante von dichten Wolken unterschätzt das Retrieval den Hydrometeorgehalt, aufgrund eines maskirenden effekts der dichten Wolken. Dem entgegen steht der 'smoothing error', der ein überschätzten des Hydrometeorgehalts an einer scharfen Unterkante von Wolken mit einem geringen Hydrometeorgehalt verursacht.

Contents

1 Introduction **1**

2 The Sensor and Data Sources **3**

3 Retrieval Overview **7**

4 Empirical Orthogonal Functions **11**

4.1 Basic principle 11

4.2 Choosing the number of EOF modes 17

5 Operating a Neural Net **23**

5.1 How does it work? 23

5.2 Training with noise 26

6 Retrieval Insights **29**

6.1 Case study 29

6.2 Analysing the retrieved data 32

6.3 Applying error measures 37

6.4 Estimating the retrieval error 39

6.4.1 Minimum content boundary 39

6.4.2 Noise ensemble standard deviation 44

7 Applying the Retrieval **47**

7.1 Retrieving a simulated flight 47

7.2 Applying the retrieval to measurements 50

8 Conclusion **53**

List of Figures

1 Jacobian sketch . . . 9
2 Retrieval conceptual flowcharts . . . 10

3 Generic EOF example . . . 12
4 WVC EOF modes . . . 13
5 Cumulative eigenvalues for WVC . . . 14
6 Covariancematrix reconstruction precision . . . 15
7 AKM of WV . . . 18
8 Retrieved information content few modes . . . 19
9 DOF with maximum . . . 21
10 WVC retrieved information content using 56 modes . . . 22

11 Neural net sketch . . . 23
12 Effect of noise on the NN . . . 26

13 Chosen example profile . . . 29
14 Retrieval of the selected example profile . . . 31
15 Retrieval results for the testing database . . . 33
16 Error covariance for WVC . . . 34
17 Retrieved hydrometeor paths . . . 36
18 Retrieval analysis for the testing dataset . . . 38
19 Calculating the minimum content boundary . . . 40
20 MFE and minimum content boundary for WVC . . . 41
21 MFE and minimum content boundary for hydrometeors . . . 43
22 Noise ensemble analysis . . . 45

23 Simulated flight retrieval of SWC . . . 48
24 B897 flight retrieval of hydrometeors . . . 51

Chapter 1

Introduction

Based on the retrieval method examined by Brath et al. (2018) the study at hand targets to deepen the understanding of its potential applications. This work goes one step further by actually retrieving atmospheric profiles of the hydrometeor content instead of solely the integrated hydrometeor content as in Brath et al. (2018). Currently, active measurement systems like cloud radars (e.g. CloudSat) are predominantly utilized for such applications. However, in contrast to their active counterparts, passive sensor systems, like ISMAR[1] and MARSS[2] used in this study, have a wider field of view. Most retrieval products based on passive systems are up until today limited to integrated quantities like the water vapour path or ice water path.

The proposed retrieval approach is special in the sense that the computational time, after the framework is set up, is rather small compared to alternative methods. Some other methods involve a full radiative transfer simulation to execute the retrieval or need a big database to be consulted at runtime. An example for those are the optimal estimation method (OEM) or the Bayesian Monte Carlo Integration (BMCI). The approach used within Brath et al. (2018) is able to get around this by utilizing neural networks (NNs) for the retrieval.

NNs are used to close the gap between the 'measurement space' (actual brightness temperatures measured by the satellite sensor) and the 'state space' (referring to the corresponding atmospheric profiles). Using this method reduces the retrieval to a few matrix multiplications and it can also be applied to many measurements at once. More details and explanations on the utilized NNs can be found in chapter 5.

Furthermore, the complexity of training the NNs is reduced by transforming both spaces using an empirical orthogonal function (EOF) transformation. The EOF transformation provides a way of separating the noise within a dataset from the signals of interest for the retrieval. Thus, it is possible to reduce the complexity of the retrieval problem, introducing less but better matching variables in respect to the existing degrees of freedom. Thereby, the connection the NN has to achieve between measurement space and state space is simplified. A detailed explanation on the fundamental properties and parameters of interest can be found within chapter 4.

The same dataset used by Brath et al. (2018) is utilized to train and test the EOFs and NNs. For this study it is partitioned into a training and a testing dataset. The training dataset

[1] International Submillimetre Airborne Radiometer
[2] Microwave Airborne Radiometer Scanning System

contains 40 % of the profiles and is used to train both EOFs and NNs. The resulting matrices and weights for both parts are stored and subsequently applied to the testing dataset. As a result independence between training and testing datasets is achieved, which is deemed to produce representative results for datasets with coinciding statistical properties as the testing dataset. Chapter 2 covers the data sources as well as the characteristics of the sensors, utilized in this study.

An overview on the entire retrieval setup is given in chapter 3. The thesis further examines what insights can be gained by analysing the retrieval results (sections 6.2). This is done by using only statistical error measures and not an analytical propagation of errors. Furthermore, two additional error measures are introduced to estimate whether the retrieval produces reasonable results at a specific point in the profile or if the result is uncertain (section 6.4). The last part of the study examines the retrieval properties with respect to measurements (chapter 7). Currently, only airborne measurements exist, while the corresponding sensor is expected to be in space within the early 2020s on board of the 'MetOp-SG B' satellite.

Chapter 2

The Sensor and Data Sources

Within this chapter the used satellite sensor and data sources are presented. The properties of the sensor as well as some additional details about the used measurements are outlined. The information about the sensors is based on Fox et al. (2017) and McGrath and Hewison (2001). Next, the database used for training the NN is described. The flight on which the actual measurement data used within this study was taken is discussed at the end of the chapter. The same database and flight measurements were used within Brath et al. (2018) and the descriptions follow his work.

As the goal of this thesis is to retrieve an atmospheric profile of hydrometeor content, the sensor of choice would be the 'Ice Cloud Imager' (ICI). The sensor is designed to measure cloud ice and will be on board of the MetOp second-generation satellites (Kangas et al., 2012). While the MetOp second-generation is proposed to be launched within the 2020s, the sensor does not currently exist. But the 'International Submillimetre Airborne Radiometer' (ISMAR), which is designed as a predecessor for the ICI sensor, does exist.

ISMAR is currently operated on the FAAM[1] BAe-146 Atmospheric Research Aircraft. It provides a similar set of channels as the proposed ICI sensor. Five channels are centred on the 118.75 GHz oxygen absorption line, three channels on each water vapour absorption line at 325.15 GHz and 448 GHz, and two channels (horizontal and vertical polarisation) for each of the two atmospheric quasi-windows at 243.2 GHz and 664 GHz. The calibration process for ISMAR is described in Fox et al. (2017). Within this study, the provided measurements of the sensors are taken without further processing and thus resemble Rayleigh-Jeans equivalent brightness temperatures.

To match the full set of channels of the ICI sensor, the 'Microwave Airborne Radiometer Scanning System' (MARSS) is operated along side of ISMAR. It provides additional atmospheric window channels at 89 GHz and 157 GHz, and three further channels at the water vapour absorption line at 183 GHz. Advanced details on the MARSS sensor can be found in McGrath and Hewison (2001) and following the citations within. In total, the entire setup consists of 20 separate channels.

For the scope of this thesis only the nadir view of the sensors is used. Therefore, the

[1] Facility for Airborne Atmospheric Measurements

polarization for the window channels of the ISMAR sensor is assumed to make no difference for the measurement. The reason for this is that, within the nadir view, both horizontal and vertical polarization are assumed to be orthogonal to the surface and hydrometeors within clouds are assumed to be randomly oriented. Thus, the polarized ISMAR channels are respectively averaged for both frequencies, leaving 18 separate channels. Given the examined retrieval setup it would also be possible to run a retrieval for other viewing angle than nadir if the training dataset is adjusted accordingly.

A detailed description on the generation of the database is given in Brath et al. (2018). At first atmospheric profiles are needed to generate the database. Those are obtained by randomly selecting profiles from a regional ICON forecast run with a grid resolution of 10 km. The time (March 2015) and region are selected to match the ensuing database statistics to the chosen flight measurements. Most importantly the profiles within the dataset store information about, the pressure grid with 90 levels ranging from 0.02 hPa down to surface pressure, the hydrometeor content, and the atmospheric humidity converted to volume mixing ratio (vmr).

To obtain the brightness temperatures corresponding to the profiles a radiative transfer simulation was performed by using the "Atmospheric Radiative Transfer Simulator" (ARTS). For further details on ARTS Buehler et al. (2017) can be consulted. ARTS can be used for polarized radiative transfer calculations including scattering, making it ideal to be used for the chosen profiles as they contain several hydrometeors. In order to execute the radiative transfer calculation explicit assumptions have to be made for each hydrometeor type. Those assumptions alongside with a detailed description of the ARTS setup can be found within Brath et al. (2018).

The four hydrometeor types contained within the ICON model are based on a 1-moment microphysics scheme. Explicitly these types are cloud ice water (IW), snow water (SW), cloud liquid water (LW) and rain water (RW). Further on if a 'C' is added to these abbreviations, it refers to the hydrometeor content. Following the work of Brath et al. (2018) the content is retrieved as the decadic logarithm with the help of a respective unity content as shown in equation 2.1. Taking SWC as an example, x is the original SWC, x_0 equals $1\,\mathrm{kg/m^3}$ and $\tilde{x}$ resembles the retrieval quantity.

$$\tilde{x} = \log_{10}\left(\frac{x}{x_0}\right) \tag{2.1}$$

The reason for this choice is the wide dynamic range of the hydrometeor content and the fact that the content is always greater or equal to zero. An evident drawback of this approach is that the decadic logarithm is not defined for zero. Therefore, a minimum value has to be chosen. For the retrieval setup within this study $10^{-12}\,\mathrm{kg/m^3}$ and respectively 10^{-12} vmr for WVC are the best values to use as minimum. All values smaller than the minimum value are replaced by it. The reasoning for this choice can be found within section 6.4.1. Basically, the chosen minimum value fits to the fact that the values, which are smaller than the minimum, are retrieved with a consistently different precision than the values above the minimum. Therefore, one can assume those to be clear sky cases. The conducted analyses are always based on the decadic values as it is the retrieval quantity.

The dataset is further on referred to as x and respectively $\widetilde{x}$ for the content of the retrieval quantity. x is structured as shown in equation 2.2, where i denotes the different profiles and z denotes the height level. The corresponding brightness temperatures in the dataset are referred to as y and are structured in the same way as x, while i corresponds to the profile and c describes the different channels.

$$x = \begin{bmatrix} x_{0,0} & \cdots & x_{0,i} \\ \vdots & \ddots & \vdots \\ x_{z,0} & \cdots & x_{z,i} \end{bmatrix}, \qquad y = \begin{bmatrix} y_{0,0} & \cdots & y_{0,i} \\ \vdots & \ddots & \vdots \\ y_{c,0} & \cdots & y_{c,i} \end{bmatrix} \tag{2.2}$$

The flight measurements used in this study were conducted on March 18th, 2015 by the FAAM aircraft. The aircraft measured a precipitating frontal system west of Iceland. A more detailed description on the flight (called B897) can be found in Brath et al. (2018). Brath et al. (2018) used the measurement data to compare the conducted hydrometeor path retrieval to ICON model data. In the outlook of this thesis (chapter 7) the same will be done, with the difference that the hydrometeor content within each level of the profile will be compared. The important point to note about the measurement data of flight B897 is that all measurements were taken over water, as are the profiles for the training and testing dataset.

Further limitations to the measurement data were applied, following the work of Brath et al. (2018). Only measurements with an altitude above 5000 m, a pitch angle of the aircraft between 4° and 6° and a roll angle below 5° were taken into account. Additionally, the altitude of the aircraft is not allowed to change more than 7 m between two measurements in time for the measurements to be valid. These limitations ensure only measurements with a stable aircraft at a fixed altitude are used.

Chapter 3

Retrieval Overview

The aim of this chapter is to present fundamental questions and concepts concerning a retrieval. Also the most important and fundamental retrieval errors are outlined. Finally, the retrieval setup utilized in this study is explained in detail.

Retrievals are necessary tools to extract information about the atmospheric state out of satellite measurements. So far there is no model that is able to reconstruct the exact atmospheric state out of the current space borne measurements. This is due to several reasons, addressed later on in this chapter. In order to go deeper into these problems it is necessary to understand the basic ideas on which a retrieval is build upon. Detailed information on this can be found in Rodgers (2000), which also provides the explanations used in the following.

$$\mathbf{y} = \mathbf{f}(\mathbf{x}) + \boldsymbol{\varepsilon} \tag{3.1}$$

Equation (3.1) can be viewed as the most fundamental explanation of the retrieval problem. Where $\mathbf{y}$ describes the measurement, $\mathbf{x}$ the atmospheric state and $\boldsymbol{\varepsilon}$ the measurement error. $\mathbf{f}$ can be understood as the forward model to describe the relation between the atmospheric state and the radiation that will be emitted and subsequently measured. The goal of a retrieval is to find $\mathbf{x}$ for a given $\mathbf{y}$. The equation can be solved to find $\mathbf{y}$ for a given $\mathbf{x}$ by using a radiative transfer model as $\mathbf{f}$. Thus, one could state that the goal of a retrieval is to find an inverse radiative transfer model $\mathbf{f}^{-1}$.

Finding $\mathbf{f}^{-1}$ would be trivial if $\mathbf{f}$ would be linear. However, radiative transfer models are highly nonlinear and can also contain iterative processes, which further increase the difficulty of inverting such a model. This is such a difficult task, that constructing a new model (the retrieval), which serves as $\mathbf{f}^{-1}$, is deemed to be easier.

The question that arises now is 'how to construct such a retrieval?'. Rodgers (2000) gives a detailed explanation on how to do so in general. For the scope of this thesis only a few of those concepts are of essence. The core problem is that the connection between the measurement space and the state space is ill-conditioned. This means there can be multiple atmospheric states that would result in the exact same measurement. Vice versa there can be multiple measurements that correspond to the same atmospheric state. This can happen if a specific

part of the atmospheric state might not influence the measurement and vice versa, resulting in an independency of those components.

The first idea to simplify the problem at hand is to reduce the complexity of both the measurement space and state space. In the scope of this study this is done by applying an EOF transformation to both spaces. A detailed explanation on this can be found in chapter 4. The result gained by the transformation is twofold. A separation of signal and noise contained within the dataset is achieved which reduces the complexity of the two spaces and, more importantly, the independency of the most dominant components within each space is achieved. After the transformation all retrieval quantities are linear independent of each other. Thus, the retrieval of a specific quantity won't be influenced by retrieving another one. This enables the use of the next tool introduced to bridge the gap between the two spaces.

The next step towards a simple retrieval approach is to find a model to bridge the gap between measurement space and state space. This model needs to be able to describe the nonlinear relation between these two spaces and should furthermore be simple to apply. A NN fulfils these criteria. While not necessarily requiring the above mentioned EOF transformation, the training strategy for the NN would have to be more sophisticated without it. More details on the NN can be found in chapter 5.

After the fundamental questions on how to construct a retrieval are adressed, another question arises. What kind of errors are to be expected by using such a retrieval? A detailed mathematical explanation of these errors can be found in Rodgers (2000). The following paragraphs are going to outline some of those errors and basic approaches on how to assess them. However, due to the fact that this thesis is done using a NN, all the applied error measures are only based on statistical analyses and no analytical propagation of uncertainty is calculated.

The most pressing issue arises from the so called null space. The null space describes the part of the state space that is not covered by any measurement. Thus, it is impossible to perform a retrieval on that part of the atmospheric state. For example, if there is no part of the measurement that is influenced by rain, it is impossible to retrieve rain by solely taking said measurement into account. Thus, resulting in an error that can't even be estimated based on the measurement. But, by using some a priori information it is possible to make an educated guess on the properties of such a component. For example: rain is likely to appear beneath dense[1] clouds.

Another problem is based on the measurement and retrieval technique: the so called smoothing error. The smoothing error is already introduced by the measurement itself. A passive satellite sensor does not receive information about the radiation within every layer of the atmosphere. Instead, the received signal is a result of the complex radiative transfer processes that take place within the atmosphere. The contribution of each layer to the final signal can be calculated if the atmospheric state is known. This contribution can be condensed into a weighting function, which is called a Jacobian. An example for such a Jacobian is given in figure 1. It shows the typical shape for a Jacobian without ground influence. The sensitivity is highest at a certain height level and decreases around it. Each measurement channel has its own Jacobian and the

[1] the density of a cloud further on refers to the hydrometeor content within it

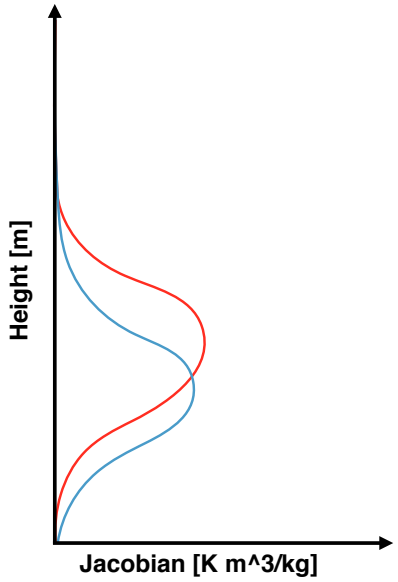

Figure 1: The sketch shows idealized Jacobians for two channel responsive to hydrometeor content (blue and red line). Both contain no ground influence.

bandwidth of each channel influences the shape of the Jacobian. Therefore, the bandwidth influences the height level to which it is most sensitive. Thus, the same measurement frequency is used for a number of channels with different bandwidth to gain information about different layers of the atmosphere. These measurement properties are described by the smoothing error. Another part of the smoothing error lies within the applied retrieval method. Using the EOF transformation does not change the data or information contained within it, but reducing the complexity by cutting off some noise does. A more detailed explanation on this can be found in chapter 4.

Up until now yet another error has not been addressed, the one contained within the measurements itself. Measurements aren't perfectly accurate. They contain some noise based on the utilized measurement system. Most satellite sensors have a very high sensitivity so that minor temperature changes of the measurement system itself can change their behaviour. Thus, noise is introduced to every conducted measurement.

The retrieval setup used within this thesis is supposed to meet all mentioned requirements and should provide good results regarding the fundamental retrieval errors. Figure 2 contains two flowcharts describing the retrieval procedure utilized within this thesis. On the left the training concept is visualized, while the testing concept is on the right. The training process consists of two steps. At first the EOFs are generated based on the training dataset (see chapter 4). The 'in/out' within the flowchart refers to the fact that two individual EOF transformations are necessary. One for the input data to the NN, the $\widetilde{\mathrm{x}}_{\mathrm{truth}}$ and one for the output data $\widetilde{\mathrm{y}}_{\mathrm{truth}}$. Where $\widetilde{\mathrm{x}}_{\mathrm{truth}}$ refers to the brightness temperatures, generated by the radiative transfer simulations with ARTS and $\widetilde{\mathrm{y}}_{\mathrm{truth}}$ refers to the corresponding atmospheric profiles, on which the radiative transfer simulations are based. After both parts of the dataset are converted using the EOF transformation, they are fed into the training algorithm of the NN (see chapter 5).

To test the performance of the retrieval the testing data $\widetilde{\mathrm{x}}_{\mathrm{truth}}$ is run through the retrieval

and compared to the truth $\widetilde{y}_{truth}$. Therefore, $\widetilde{x}_{truth}$ is run through the EOF transformation 'in', through the NN and lastly reconstructed using the EOF transformation 'out'. The result is $\widetilde{y}_{retrieved}$, which can be directly compared to the truth. $x_{retrieved}$ is compared against the corresponding x_{truth}.

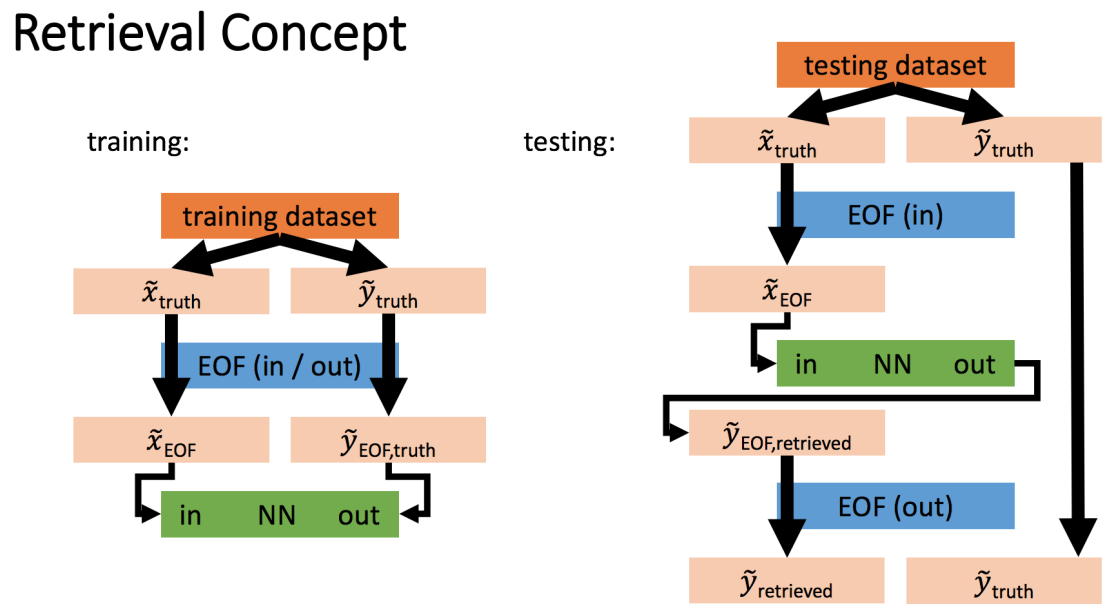

Figure 2: The testing (right) and training (left) concept of the retrieval is visualized as a flowchart. $\widetilde{x}$ refers to the brightness temperatures, while $\widetilde{y}$ refers to the retrieval quantity. The 'in/out' for the EOF transformation refers to the respective transformation for $\widetilde{x}$ and $\widetilde{y}$. Arrows describe the processing chain.

Chapter 4

Empirical Orthogonal Functions

This chapter examines the properties and reasoning behind a transformation with empirical orthogonal functions (EOFs). As mentioned above, the transformation is used to gain a deeper understanding of the data at hand. Transforming the data by using an EOF transformation yields a simple separation of actual signals and noise within the data. Hence, the complexity of the data is decreased, as some information is discarded for the analyses to follow.

4.1 Basic principle

Transforming a dataset using EOFs yields a change of the underlying coordinate system. A visual example is given in figure 3. The blue dots represent data scattered around a centre point marked by the red dot. The dashed lines represent the original basis vectors of the 'normal-space'. In order to change the coordinate system a new one has to be defined first. This is done by identifying the axis alongside which the data contains the strongest variability. After identifying it the procedure is repeated for additional axes. Every axis has to be orthogonal to the already used ones and proceed along the direction of the strongest variability left within the uncovered data. For the given example one can only select up to two axis. The number of axis can be as high as the number of existing dimensions. Green and orange lines in figure 3 represent the selected axis. The green axis contains the strongest variability and orange the rest of the variability within the data. Centre point for the new coordinate system is the centre point of the data (the red dot). The selected axis and the centre point are used to define the 'eof-space'.

A detailed mathematical description of the entire process can be found in Richards and Jia (2006). But the core of the transformation consists of basic mathematical equations, which are well described by Strang (2014). The explanations about eigenvalues and eigenvectors given later on are based on this book.

Calculating the axis with the strongest variability is done by using the covariance matrix of the data. A covariance matrix contains the variance for a specific variable on the diagonal and the covariances between different variables fill up the rest of the square matrix. Equation 4.1 is used to calculate the entries of a covariance matrix.

$$A_{a,b} = \left\langle \left(\widetilde{x}_a - \overline{\widetilde{x}_a}\right)\left(\widetilde{x}_b - \overline{\widetilde{x}_b}\right)\right\rangle \tag{4.1}$$

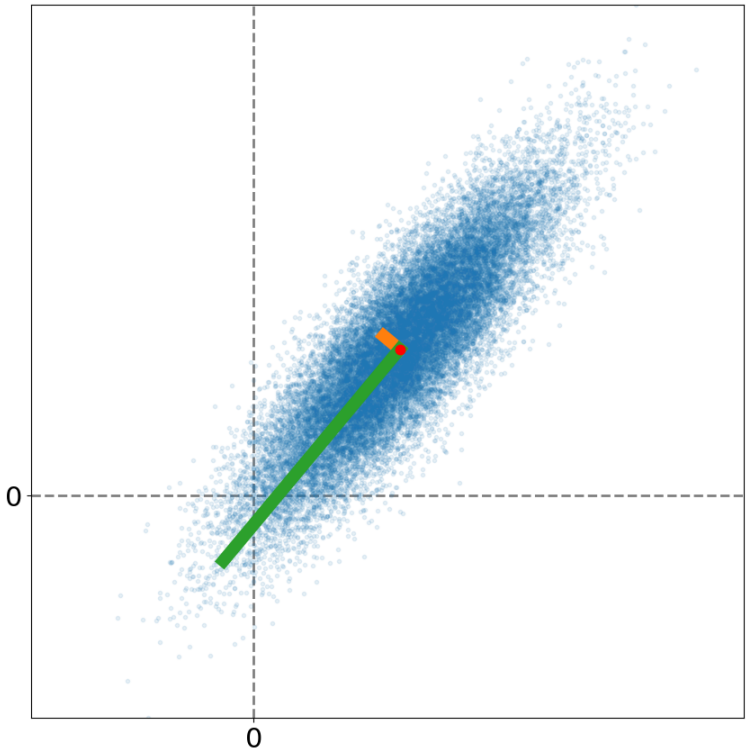

Figure 3: Blue dots represent the data. The green and orange lines represent scaled eigenvectors of the EOF transformation. The red dot marks the origin for the eof-space. Dashed lines symbolise the axis of the original coordinate system.

$A_{a,b}$ referring to the covariance of variable a and b within the dataset $\widetilde{x}$. If a is equal to b it describes the variance. $\overline{\widetilde{x}_a}$ describes the mean value for variable a within the data and the same goes for $\overline{\widetilde{x}_b}$. This is done for the training part of the dataset.

$$Av = \lambda v \tag{4.2}$$

Eigenvectors v are defined by the property, that they solve equation 4.2. Solving the equation for v with A as the training covariance matrix results in the axis necessary for the transformation. Every combination of λ and v which solve this equation yields another axis. With λ (the eigenvalue) describing the amount of variability on the axis defined by v (the eigenvector). These axis v are also called 'EOF modes' and sorted by decreasing eigenvalue.

After constructing the EOF transformation based on a specific dataset one gains a fixed set of eigenvectors and eigenvalues. These fixed eigenvectors can then be used to transform further data into eof-space. Equation 4.4 describes this transformation. First the eigenvectors v have to be put together as described in equation 4.3. The resulting matrix V consists of all the eigenvectors that should be used. Indices m and i describe the number of the EOF modes and the size of the eigenvectors. Next up, the matrix V is inserted into equation 4.4. $\widetilde{x}$ refers back to the normed dataset (see equation 2.1). $\overline{\widetilde{x}_{train}}$ is the mean profile of the training dataset. If the training dataset is meant to be referenced, the respective variable always has the '$_{train}$'

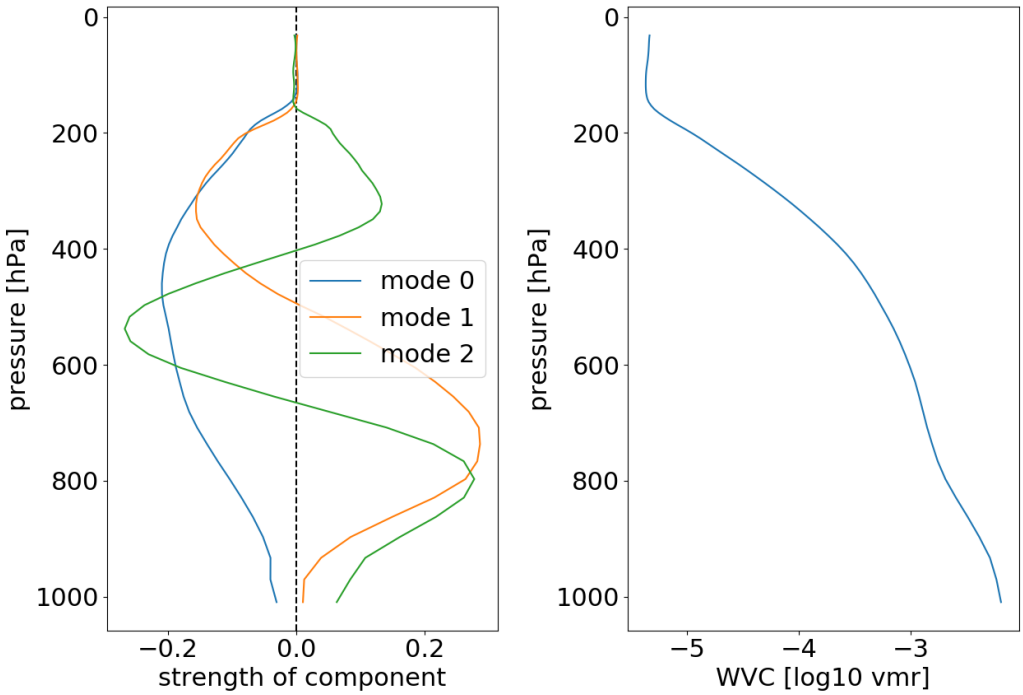

Figure 4: The three most dominant EOF modes for water vapor content (WVC) are shown on the left and the corresponding mean profile of WVC on the right.

annotation. After processing the equation the data is transformed into the eof-space ($\widetilde{x}_{EOF}$). The reconstruction of the data to normal-space is done using equation 4.5.

$$V = \begin{bmatrix} v_{0,0} & \cdots & v_{0,i} \\ \vdots & \ddots & \vdots \\ v_{m,0} & \cdots & v_{m,i} \end{bmatrix} \tag{4.3}$$

$$\widetilde{x}_{EOF} = V \times \left(\widetilde{x} - \overline{\widetilde{x}_{train}}\right) \tag{4.4}$$

$$\widetilde{x} = \left(\widetilde{x}_{EOF} \times V^{T}\right) + \overline{\widetilde{x}_{train}} \tag{4.5}$$

A visual example for the eigenvectors and the mean profile is given within figure 4. The eigenvectors (on the left side) are resulting from the database used for training the NN. The numbers '0' through '2' within the legend correspond to the order of the eigenvectors (EOF modes). The most prominent feature within the training dataset is the overall change of WVC. This change is described by the '0' EOF mode. It is continuously negative and possesses a maximum at 420 hPa. The second mode ('1') contains one zero point pass close to the maximum of the first EOF mode. Thus, resulting in a negative coupling of the upper and lower troposphere. If the WVC within the upper troposphere increases, the WVC in the lower troposphere decreases. Following the scheme, the third mode contains an additional zero point pass which results in an

even more complex change of WVC within the troposphere.

On the right side of figure 4 the mean profile of WVC is shown. In order to reconstruct the data after it was transformed to eof-space the mean profile is needed in addition to the eigenvectors. Essentially, the mean profile is used to shift the origin of the coordinate system back to its original position. The eigenvectors solely contain information about the strength of the variability around the mean profile. The transformed data ($\widetilde{\mathrm{x}}_{\mathrm{EOF}}$) describes the strength of the EOF modes which have to be applied to the mean profile to reconstruct the original data.

Further EOF modes for the WVC are not shown within this figure, but they follow the same scheme as the first three and increase in complexity. So, the possible 'information content' existing within the eof-space is increased, until the entire variability can be represented.

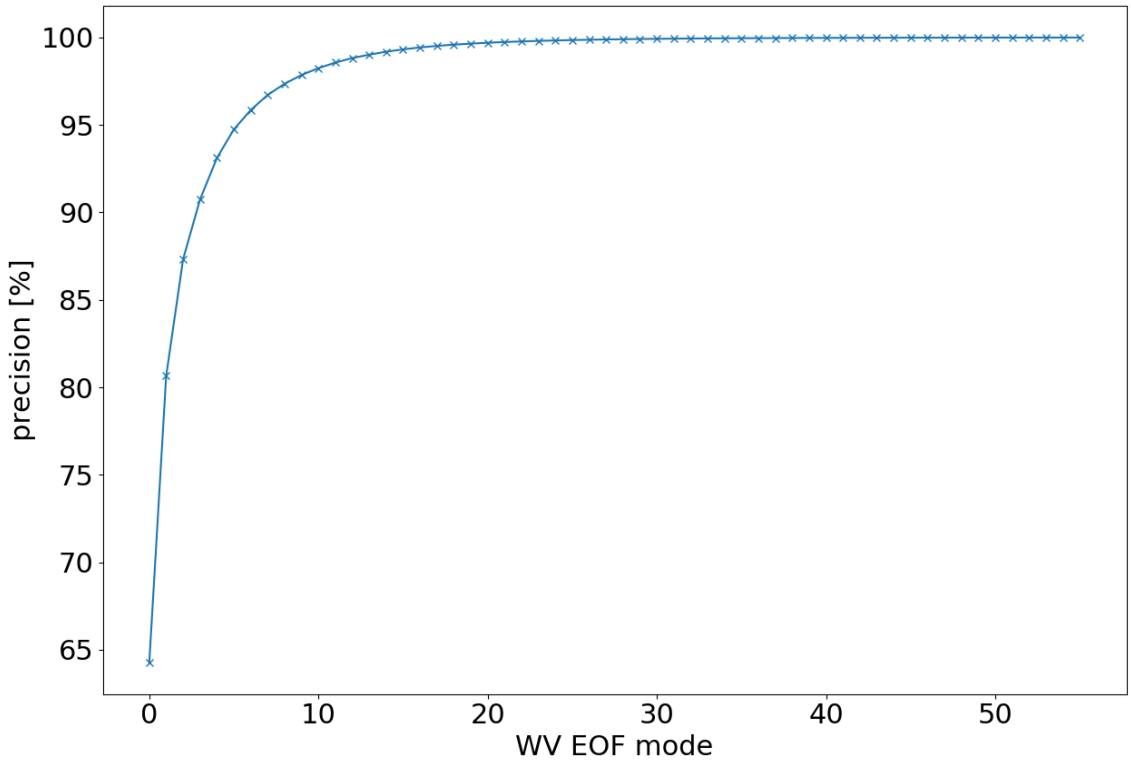

Figure 5: Cumulative eigenvalues for WVC are shown as percentages with respect to the contained variability of the training dataset.

Finally, the question remains which information is still valuable for the retrieval and which information is impossible to retrieve. Answers can be provided by further analysing the eigenvalues of the transformation. Eigenvalues of the same WVC transformation as before are shown in figure 5. The eigenvalues are given as cumulative percentage values describing the explained variability of the original dataset up to a specific mode. Thus, the given percentage also describes the precision to which the transformed data can be reconstructed. As figure 5 shows, the first five modes are sufficient for reconstructing over 90% of the variability within the original data after the transformation. Adding the last few modes yields next to no additional gain of precision. Reason for this is the separation of signal and noise throughout the EOF modes. The modes of a higher order (last few modes) are mostly influenced by noise within the dataset,

which is the least dominant variability within the original dataset. To achieve a more visible representation of the achieved precision, covariance matrices can be used.

By having a look at the covariance of the original data as well as the same data transformed to eof-space and back, we can see which part of the variability is lost by the transformation. Now there is one important parameter to look at for the entire EOF transformation. The precision to which the variability of the data should be reconstructible after the transformation. This precision is determined by the number of EOF modes chosen for the transformation and will further be discussed in section 4.2. For this example the number of modes are based on arbitrarily chosen precisions. The resulting covariance matrices can be seen in figure 6.

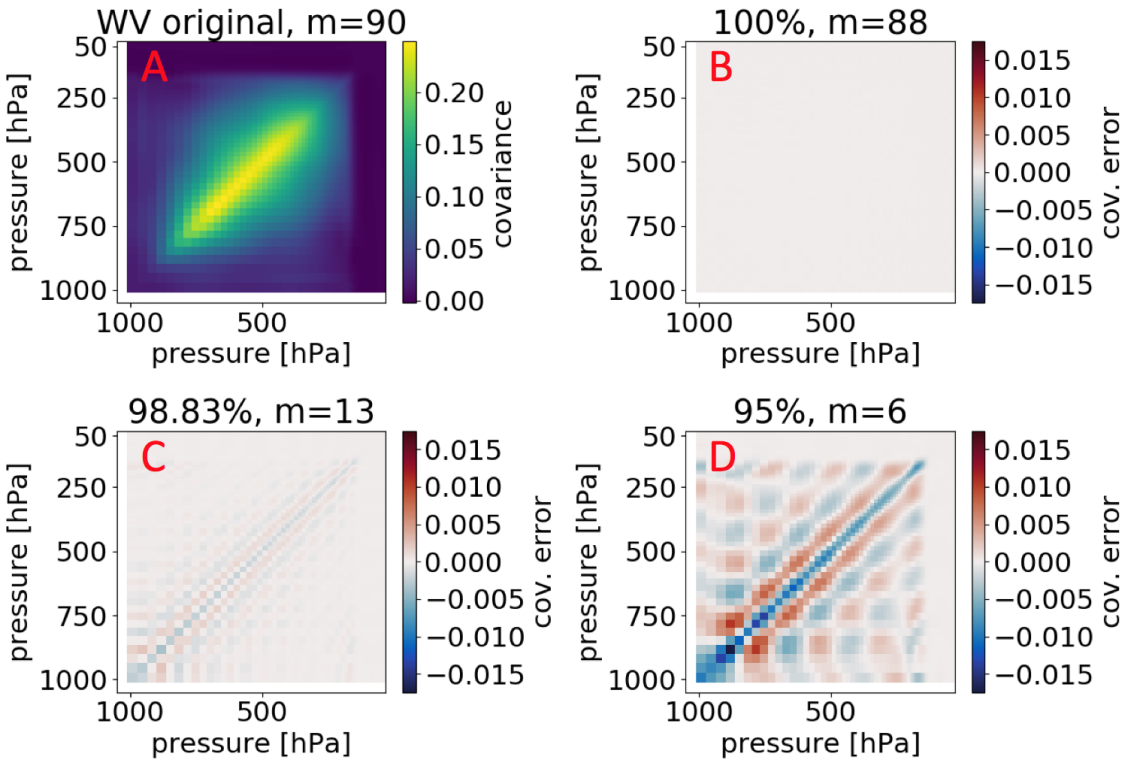

Figure 6: The panel A shows the original covariance matrix of WVC, followed by the error covariance matrices (B-D) gained through EOF transformation and reconstruction. The respective EOF transformation is set to a specific cutoff N_{EOF} denoted as 'm', which is shown together with the reconstruction precision on top of each panel.

Panel A shows the covariance matrix for the training dataset. Variability is mainly found within the troposphere, while the variability within the upper atmosphere is zero. The number of dimensions m within the original dataset relates to the existing height levels of the model data, which are 90. Panel B through D show the error resulting out of the EOF transformation and reconstruction of the training dataset. The EOF transformation is executed using the number of EOF modes denoted by the m. The reconstruction precision is denoted by the percentage value above each panel. Therefore, one can see that the original dataset can be fully reconstructed using only 88 EOF modes instead of the original 90 height levels. As it is to be expected the resulting error increases for a decreasing number of EOF modes. However, the most interesting feature is the change between the C and D panel. The unresolved patterns are smaller within C

as they are in D. Demonstrating the effect additional EOF modes have on the retrieval. Higher order modes resolve smaller features within the training dataset. Thus, the resulting error pattern decrease not only in strength, but also in size.

4.2 Choosing the number of EOF modes

This section starts by explaining the use of the averaging kernel matrix (AKM). The AKM is used to evaluate which number of EOF modes is reasonable to run the retrieval. Afterwards, the results of this analysis will be discussed and a decision on the number of used modes for this study is made. Following this decision the corresponding errors will be discussed.

The AKM is calculated using equation 4.7. As before x refers to the content of the testing dataset, while $\mathbb{1}$ is the identity matrix. To calculate the AKM the covariance matrix $A_{\mathrm{EOF,train}}$ (see equation 4.1) has to be inverted. Then, the covariance matrix of the difference between retrieved and true data $\mathrm{E_{EOF}}$ has to be calculated. Multiplying those two matrices yields the AKM $\mathrm{K_{EOF}}$. All of these calculations have to be done in eof-space, because it is (based on the training database) not possible to invert the original covariance matrix in normal-space. The reason for this is that in the upper atmosphere the training dataset does not contain any variability. Thus, the covariance matrix in normal-space contains columns where all entries are zero. This holds true for all hydrometeors and the WVC, leading to the fact that the covariance matrix can't be inverted in normal-space.

$$\mathrm{E_{EOF}} = \mathrm{cov}\left(\widetilde{\mathrm{x}}_{\mathrm{EOF,retrieved}} - \widetilde{\mathrm{x}}_{\mathrm{EOF,truth}}\right) \tag{4.6}$$

$$\mathrm{K_{EOF}} = \mathbb{1} - \mathrm{E} \times A^{-1}_{\mathrm{EOF,train}} \tag{4.7}$$

In order to fully understand the resulting matrix $\mathrm{K_{EOF}}$ one has to analyze the steps necessary to process equation 4.7. First, one has to calculate the error covariance matrix of the retrieved testing dataset $\mathrm{E_{EOF}}$. This yields a matrix describing the variability of the overall retrieval error. The overall error is afterwards multiplied with the inverse of the covariance matrix of the original training dataset. Most importantly, one has to note that this covariance matrix $A_{\mathrm{EOF,train}}$ is diagonal by design and the values one the diagonal are decreasing in strength. After inverting this covariance matrix it resembles a weighting for the error covariance matrix $\mathrm{E_{EOF}}$ (equation 4.6). While the variability within the lower orders of the transformed dataset (corresponding to first few EOF modes) is high, higher orders have a lower variability. This accumulates to the fact that the inverted covariance matrix punishes relatively small errors within higher order modes, while lower order modes are allowed to have bigger errors. The matrix resulting out of the described weighting of $\mathrm{E_{EOF}}$ describes the information that is lost relative to the information that was available within the covariance of the original dataset. At last this resulting matrix is subtracted from the identity matrix to describe the information preserved throughout the retrieval. So in conclusion, the AKM $\mathrm{K_{EOF}}$ describes the information content, which is preserved within a specific mode relative to the variability of the original training dataset. Thus, the information content refers to the resolved variability of the testing dataset with respect to the variability in the training dataset.

Figure 7 shows the resulting AKM for WV. Based on this data it is possible to estimate the optimal number of EOF modes to be used and define a justifiable cutoff. Tests showed that

the information content within one specific mode is invariant in respect to the number of EOF modes used for the retrieval. Meaning the AKM remains unchanged regardless of the number of chosen EOF modes. Only the size of the AKM itself changes. Subsequently, it is valid to look at different cutoffs without recalculating the AKM.

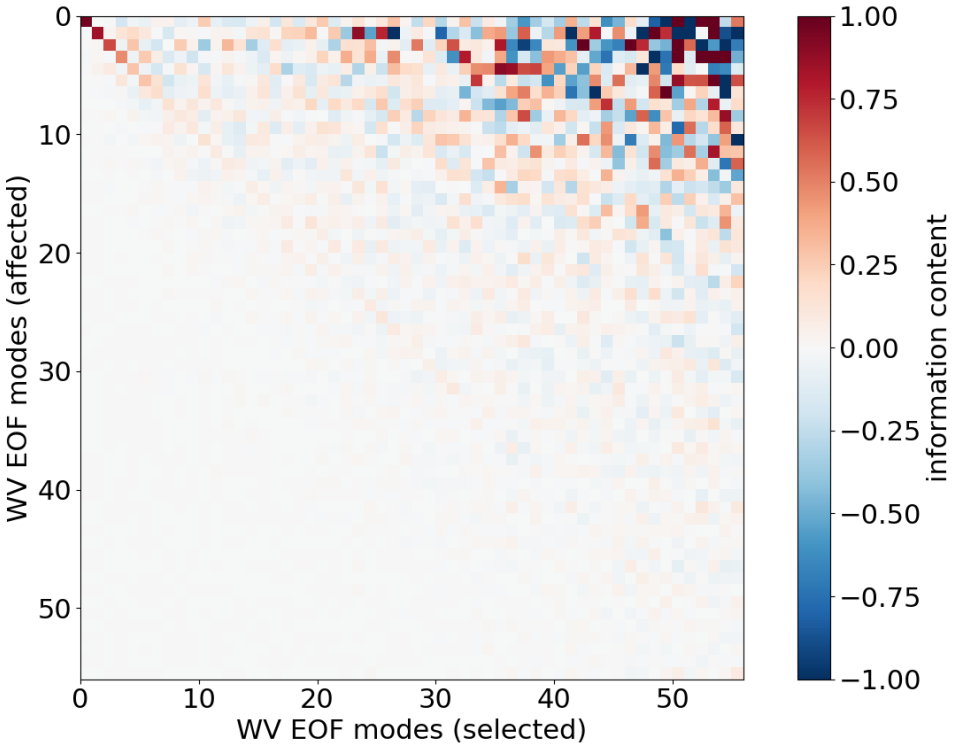

Figure 7: AKM for the WV retrieval. The information content (colors) describes the variability of the testing dataset that the retrieval is able to capture within a specific EOF mode. The dimension 'affected' contains the effect a specific 'selected' EOF mode has on the distribution of the retrieved information content I*. For a specific 'affected' mode* I *is calculated by summing up the information content of all 'selected' modes. The trace of this matrix describes the* DOF *that the retrieval results achieves.*

To further condense the information contained within the AKM there are two possibilities. One is to have a look at the 'retrieved information content' defined by equation 4.8. The number of used EOF modes $\mathrm{N_{EOF}}$ extents the sum over the columns of the AKM, while m denotes the row of the matrix $\mathrm{K_{EOF}}$ and therefore the order of a specific mode. Thus, I (the retrieved information content) describes the variability represented in a specific mode m if a certain number of modes $\mathrm{N_{EOF}}$ is used for the retrieval in relation to the variability contained within the training dataset.

$$\mathrm{I_m} = \sum_{\mathrm{i=0}}^{\mathrm{N_{EOF}}} \mathrm{K_{m,i}} \tag{4.8}$$

A visual example of I is given in figure 8. Each black line (the retrieved information content) is a sum over all the colored lines for the specific EOF mode. A retrieved information content of 1 denotes the best possible result, in respect to the used training dataset. In other words, based

on the variability of the training dataset, the retrieval could not have resolved the feature within the testing dataset better than it did. Opposite to that case, a retrieved information content of 0 corresponds to the fact that the feature is not contained within the result at all. The next paragraph further examines this, using figure 8 as an example.

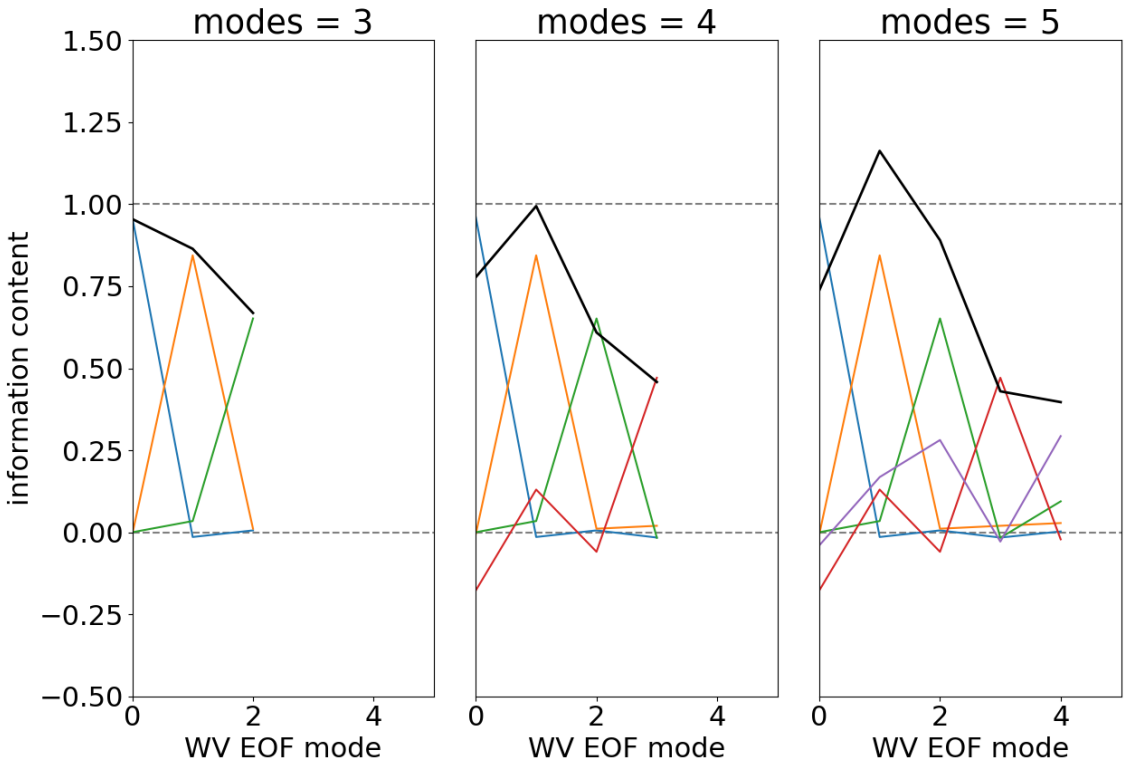

Figure 8: The retrieved information content (black lines) for WV is shown for different N_{EOF} *cutoffs. The coloured lines are based on the entries of the AKM. Each line shows the effect of one specific mode has one the retrieved information content ('0' is blue, '1' is orange, '2' is green, '3' is red, '4' is purple). The used* N_{EOF} *cutoff is denoted above each graph. Two dashed lines enclose the expected range for the retrieved information content.*

The figure shows the retrieved information content for some arbitrarily chosen cutoff values. For each graph (from left to right) an additional mode is added. With only 3 modes (left graph) each mode is only influenced by itself. This means that the retrieved information content within the mode results only out of the feature covered by the mode itself. For example the retrieved information content (black line) of the second mode (order '1'), results from the feature that is covered by the second mode itself. This can be seen by having a look at the orange line within the left graph of figure 8. The orange line corresponds to the information content gained by the retrieval of the second EOF mode. It has the same magnitude as the retrieved information content (black line), while every other colored line is zero at this point. Due to the fact that the retrieved information content is not 1, the feature is not perfectly retrieved in comparison to the variability of the testing data. Because, the information content is scaled based on the covariance of the training data.

Moving one graph to the right, a fourth mode is introduced. This fourth mode is the first one that does not only influence itself, but also affects the second mode. It causes a decrease

in the retrieved information content within the second mode. The origin of this is the error covariance matrix of the retrieval E_{EOF}. If there is a relation between two modes within this error covariance matrix, the EOF modes are not independent throughout the retrieval. While the modes are independent of one another in eof-space, within the retrieval process the different modes can be related to one another. This change in the retrieved information content results out of a different distribution of information contained within the given brightness temperatures (y). In other words, while adding more modes, the information contained within y is distributed differently over all of them.

The rightmost panel in figure 8 showcases that the added information content can be arbitrarily distributed over other modes. The additional information content added by the fifth mode (order '4') is distributed evenly over the fifth and third mode (order '2'). If the retrieved information content exceeds 1, it suggests that such a precision should not have been possible to reach with the given retrieval setup. This phenomenon appears due to the applied scaling of E_{EOF} with the covariance matrix $A_{EOF,train}$. If it falls below 0, the opposite occurs. The retrieved information content is below the expected minimum.

The other possibility to condense the information of the AKM is given by equation 4.9. It describes the degrees of freedom (DOF), which the retrieval is able to extract out of the given y in order to describe the resulting $x_{retrieved}$. To obtain the DOF, the product of all diagonal elements of the AKM is taken up to a given N_{EOF}. In other words, the trace of the AKM is calculated based on the cutoff mode.

$$DOF = \sum_{i=0}^{N_{EOF}} K_{i,i} \tag{4.9}$$

Due to the fact that the EOF modes are independent of one another, the number of modes (N_{EOF}) describes the degrees of freedom existing within the profile x. Comparing the N_{EOF} to the achieved DOF yields information about the degrees of freedom (not DOF) that were contained within the testing dataset x, but could not be extracted from the testing dataset y. Figure 9 shows those two variables plotted against each other for every retrieval quantity considered within this study. The dots mark the maximal DOF for the given retrieval quantity. The number above the dot denotes the maximum DOF that the retrieval can extract out of y. The number below the dot denotes the degrees of freedom (not DOF) that were used for x to run the retrieval (N_{EOF}).

Focusing on the line for WV, at first the DOF increase rapidly, while after adding about 10 EOF modes to the retrieval, the increase in DOF reduces and holds at a N_{EOF} of 56. Thereafter, the DOF starts to decrease. This final decrease in DOF is due to the fact that x_{EOF} becomes more and more complex, until the point where additional EOF modes introduce too much noise. The noise contained within the higher order modes can't be resolved by the information retrieved out of y, distorting the distribution of the information content. As an example for this distortion, figure 10 shows the retrieved information content for WV when N_{EOF} is set to 56. One can see that the retrieved information content is out of bounds, especially for the lower order modes,

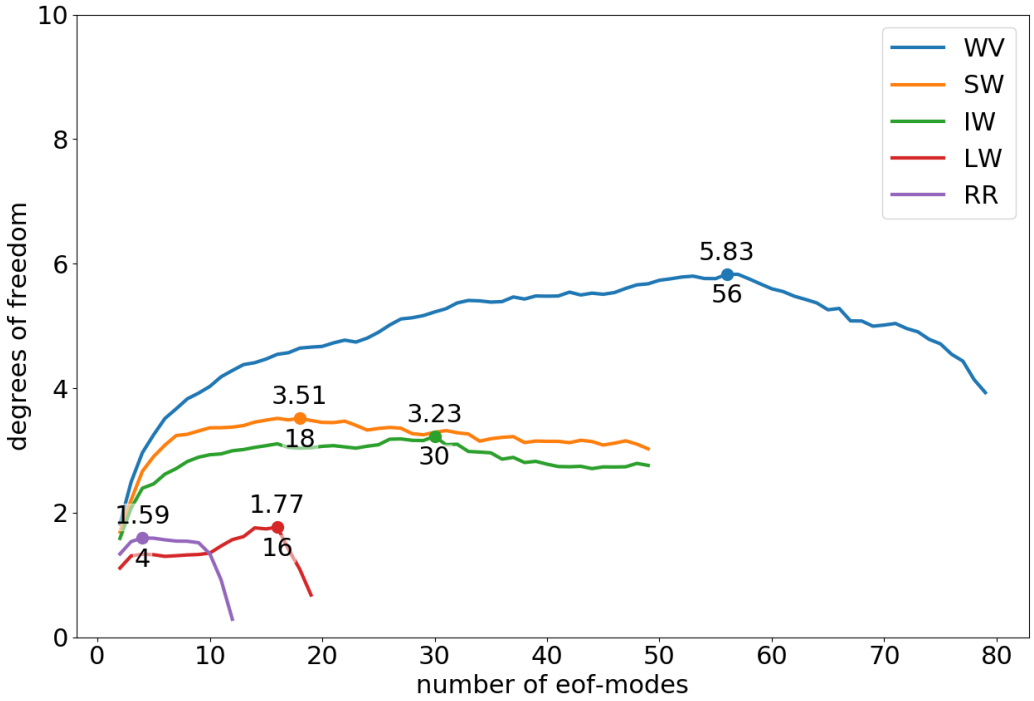

Figure 9: Shown are degrees of freedom (DOF*) for a specific number of EOF modes that are used for the retrieval. The indicators give information about the maximum* DOF *for every retrieved quantity and the corresponding number of selected EOF modes.*

which describe most of the existing variability within the training dataset.

Using all the knowledge gathered within this chapter, it is time to chose the $\mathrm{N_{EOF}}$ and therefore the precision of the EOF transformation that shall be used for this thesis. The selection is done by combining the precision estimate discussed within the prior section 4.1 and the information about the maximum DOF achievable. The first criteria is that the explained variability (calculated by the eigenvalues as in figure 5) should be at least 99 %, while the upper limit of $\mathrm{N_{EOF}}$ is set to the point at which the DOF reach the maximum. Table 1 shows the cutoffs based on these criteria as well as the resulting errors. For every quantity the lowest $\mathrm{N_{EOF}}$ is from then on used for all following calculations within this thesis. The transformation and reconstruction error is calculated by running only the EOF transformation on the testing dataset, using the cutoffs and transforming it back to normal-space. Then, the residual between truth x and reconstructed data $\mathrm{x_{recon}}$ is taken as the error measure.

For the remaining EOF transformation on the brightness temperature y all EOF modes are used. Doing so adds additional noise to the retrieval. However, the effect can be considered negligible considering the results in section 5.2.

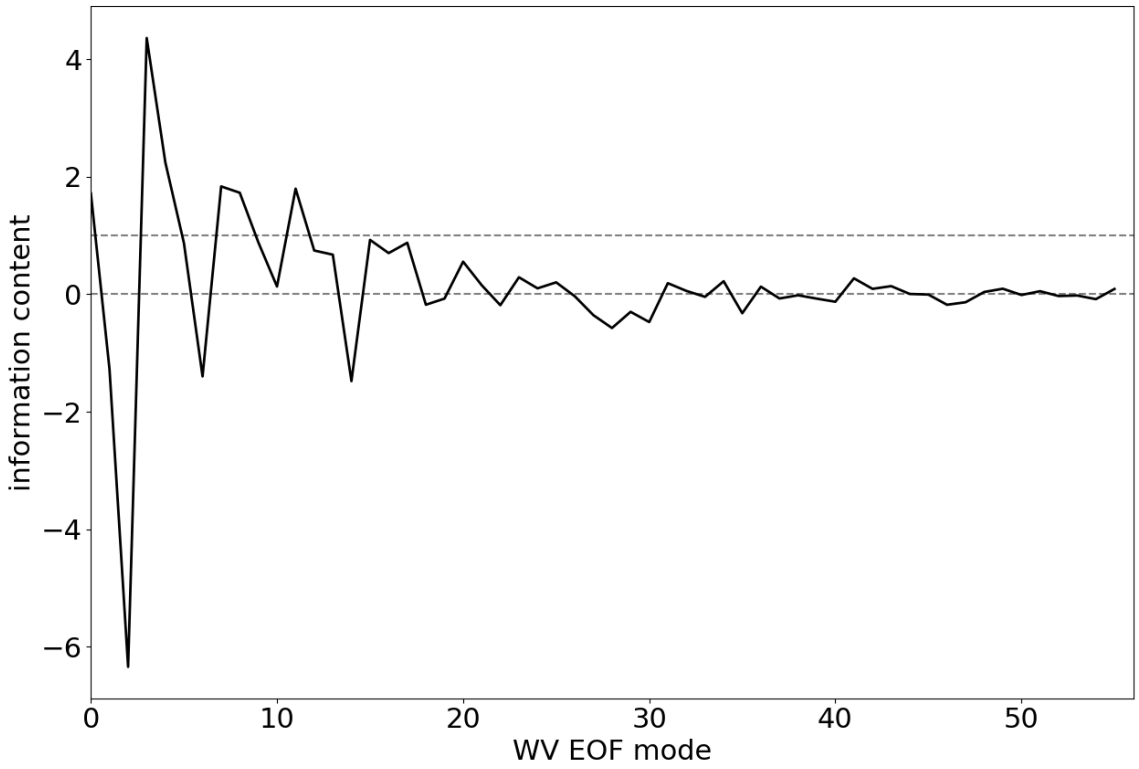

Figure 10: The retrieved information content for WV is shown for a N_{EOF} cutoff of 56. The two dashed lines enclose the expected range for the retrieved information content.

Table 1: For all retrieval quantities the EOF settings are listed as well as the corresponding transformation errors. Columns 'precision' and 'max DOF' show the N_{EOF} based on the corresponding cutoff. 'used N_{EOF}' lists the N_{EOF} cutoffs used in this study. The next column 'resulting DOF' contains the DOF based on the set cutoffs. The last columns show the mean value for the testing dataset and both error columns describe the specified percentiles of the EOF transformation error.

quantity	WV	SW	IW	LW	RR
precision	13	23	33	12	7
max DOF	56	18	30	16	4
used N_{EOF}	13	18	30	12	4
resulting DOF	4.19	3.43	3.16	1.58	1.56
mean value	$6.5\,e^{-4}$	$7.5\,e^{-6}$	$1.4\,e^{-6}$	$2.3\,e^{-6}$	$0.6\,e^{-6}$
90% error	$1.1\,e^{-4}$	$1.1\,e^{-6}$	$3.8\,e^{-7}$	$9.0\,e^{-15}$	$3.0\,e^{-14}$
75% error	$1.1\,e^{-5}$	$1.2\,e^{-11}$	$4.1\,e^{-13}$	0	0

Chapter 5

Operating a Neural Net

Despite the relative simplicity of the used NN, it still requires some mindful tuning. Consequently, a through examination of the training dataset and the implications of noise are evaluated in this chapter. Fundamental concepts and properties of NNs are given in the first, while the introduction of noise is further examined in the second section.

5.1 How does it work?

This section explains the structure and properties of NNs. Regarding the enormous amount of NN set ups and their fields of application, solely the fundamentals are covered here. A detailed description about NNs can be found in Bishop (2006).

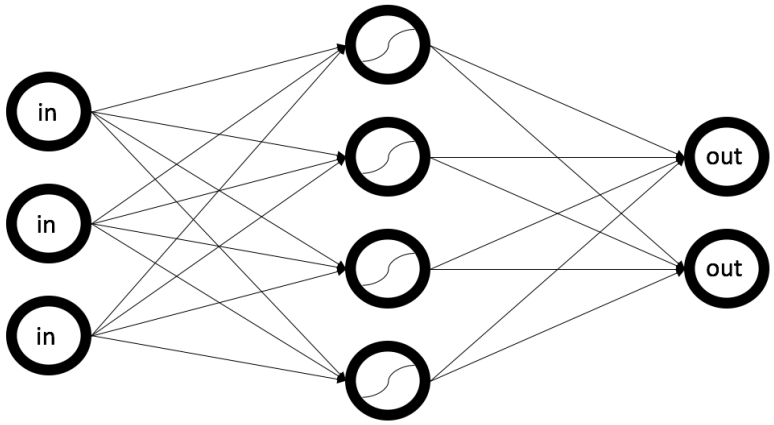

Figure 11: A sketch of a neural net. The circles symbolise the neurons. On the left are three input neurons, in the middle are four hidden layer neurons with activation function and on the right are two output neurons. The arrows symbolise the weighted connections between the neurons.

For this study the main motivation for using the NN is that it provides a functional connection

between the two retrieval spaces. However, this is not the only asset, as it is also relatively simple to configure and to understand. Figure 11 shows the structure of the utilized NN. It is a so called 'feed-forward' NN. Information is only passed directly through the net without recursions. The circles are called neurons and can be thought of as storage units. Starting on the left, the brightness temperatures of all channels are fed into the NN. Those values are stored in the leftmost neurons. As soon as those neurons receive an input they are triggered, forwarding the given signal to the next neuron. This communication is illustrated by the black lines between the neurons. In the end the final signals received within the rightmost neurons are the results.

As one would expect, reality proves to be more difficult than the basic explanation given above suggests. First, the signals fed into the NN should be normalized. The reason for that is, that within the training process the NN attributes a weight to every connection between to neurons. Those weights determine the strength of the signal that will be passed onto the next neuron. If multiple signals end up at the same neuron, the sum over all these incoming signals is taken to determine the value stored within the receiving neuron. Thus, it is beneficial for the training process if all incoming signals are normalized and independent of one another. So that the NN can connect the given informations without an implicit connection within it.

For the study at hand this normalisation is only partly done. The values fed into the NN are transformed using the EOF transformation. Therefore, the strength of the resulting components is not normalized between each other. However, the strength directly relates to the strength of a certain feature within the original atmospheric profile and the most dominant features result in the strongest signals in eof-space. Thus, a normalisation of the input signal isn't necessary as the information contained within the input values is deemed to be valuable. Although the EOF transformation does not normalise the input data, the point of origin is afterwards by design of the transformation always zero. This plays a major role for the upcoming step.

The neurons in the middle of the sketch are the so called 'hidden layer neurons'. The hidden layer can contain an arbitrarily chosen number of neurons, while the input and output layer depend on the dimensions of the respective data. Also the hidden layer is special in the processing of the received signals from the input layer. Instead of just storing the sum of the received signals, the signals are processed by a so called 'activation function'. The reason for this is to ensure, that all the values stored within the hidden layer neurons are representative for certain features within the data. The activation function utilized within this study is the commonly used sigmoid function. Its limit approaches 1 for values greater zero and -1 for values smaller zero. For an input signal of 0 the sigmoid function returns 0. Thus, it is best to apply the sigmoid function to data centred around zero. The feature described by the hidden layer neurons is determined by the weights of the connections between the neurons. As those weights can either be negative or positive the input signal does not determine the resulting sign as the training progresses. However, once the NN is trained the weights are kept unchanged and therefore the input signal will from then on change the sign of the resulting signal.

The concept of the connections between the hidden and the last layer, containing the results, is the same as between the input and the hidden layer. During the training process all of the connections are trained based on an algorithm called 'backpropagation'. At first, the weights for all connections are randomly set. One set of input values is used from the training dataset

and processed through the net. The resulting output is compared to the truth, which is also part of the training dataset. Thus, an error can be calculated. This error is fed back through the NN, contributing an error to every neuron. The error is dependent on the error fraction that was caused by the specific neuron. After this propagation of the error, the weights of the connections between the neurons are adjusted to better match the input to the given output data. This adjustment involves further settings which are not discussed within this thesis but can be found in Bishop (2006). The details and a mathematical explanation on all steps can also be found within the same book.

Despite numerous other methods, a basic configuration of training and structuring a NN was chosen for this study, as it solves the task at hand and also follows the work done by Brath et al. (2018). As for the training algorithm the 'Adam' solver was used, which is described within Kingma and Ba (2014). Thus, there are only three important parameters to be considered in the scope of this study. First the number of hidden neurons has to be chosen. It is set to be 25, because the number of input neurons is 18, while the number of output neurons varies between 4 for RW and 30 for IW. Thus, 25 hidden neurons are deemed to capture all features within the input data and further translate them to the output neurons. A proof that this works sufficiently is given in figure 8. The information content within one specific EOF mode does not change based on the number of retrieved modes. Therefore, one can assume that all available information content is captured in either case. Which leaves two parameters to be set.

This study uses not one NN, but an ensemble of NNs. The reason for this can be found in the well known 'bootstrap aggregating' method also called bagging. It is commonly used in informatics to stabilise the results of an estimator and achieve less overfitting. Meaning that by using more than one estimator the result will vary less, based on the given input data and will become more accurate in general. Details on this can be found within James et al. (2014). For the scope of this study bagging is used to disregard any further issues with overfitting or a suboptimal training procedure. Therefore, a NN ensemble is trained, where every NN is only trained on 10 % of the entire training dataset, chosen on a random basis. These 10 % represent one of the arbitrarily chosen parameters, while the other one is the use of 100 NNs as the ensemble.

5.2 Training with noise

This section examines the effect of noise on the brightness temperatures y, which are used as input data for the retrieval. As it is the goal to retrieve atmospheric profiles from real measurements, the purpose of this section is to find the setup, yielding the best results.

To analyse the problem of noise on the y dataset, four cases are considered. These four consist of a full permutation of training and testing with or without noise. The measurement noise is calculated based on the sensor accuracy for the respective channel. Originally the dataset calculated by the radiative transfer model ARTS does not include any measurement noise. For each atmospheric profile within y a randomly generated noise is added, resulting in two variations of the same dataset y and $\mathrm{y}_{\mathrm{noise}}$. Whereas $\mathrm{y}_{\mathrm{noise}}$ represents a realistic measurement, y represents an idealised measurement without any noise.

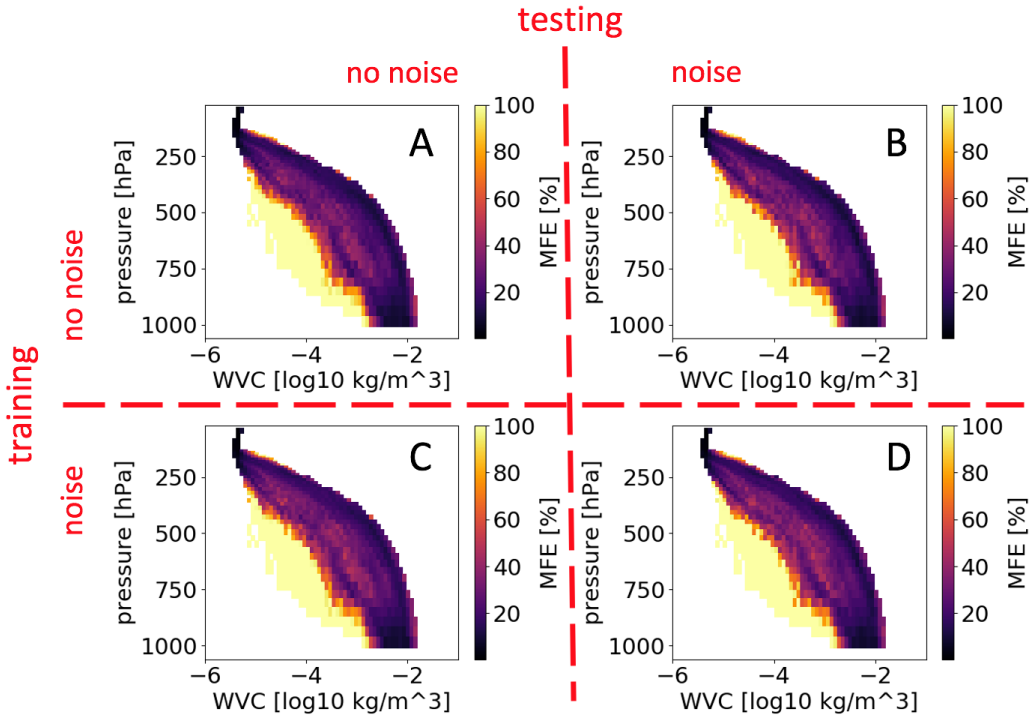

Figure 12: Four different cases are presented. A and B are trained without, while C and D are trained with noise. A and C are tested without noise, while B and D are tested with noise. The resulting MFE is shown for the retrieval of WV.

Figure 12 shows the resulting effect based on the MFE[1] of the WV retrieval. The MFE is explained in section 6.1. It is a measure for relative errors. For the results shown in A and B the retrieval is trained using the ideal data. For C and D the NN is trained using the noisy training dataset. The graphs on the left (A and C) are tested without noise, while one the right (B and D) they are tested with the noisy testing dataset.

One can see that the differences between those approaches are negligible. However, the highest agreement is achieved by training and testing without noise (panel A). Training with

[1] median fractional error

noise results in an increase in errors, while adding noise to the testing dataset yields the highest errors regardless of the training procedure. These results showcase the benefit of NNs and the chosen setup. Given the retrieval setup the impact of noise on the training dataset is close to none. The bigger issue is the noise introduced with the testing dataset, which can not be ignored, because the process of measuring does not come without noise. However, the effect is also rather small and therefore the measurement error is not further analysed in this study. Since, ultimately measurement data should be retrieved, the retrieval is trained and tested against data with noise.

Another information can be drawn from figure 12. Splitting the dataset into a training and a testing one did not change the statistical properties of either one. If the statistical properties of the training dataset are different to those of the testing dataset the retrieval would not work. This is due to the null space error. As explained before, features that are not covered by the training dataset can not be retrieved. Therefore, the training dataset has to contain the same features as the testing dataset. Now, one might notice distinct areas where the retrieval fails to perform well. However, as shown in section 6.3, those areas can be related to the sparseness of training data.

Chapter 6

Retrieval Insights

Within this chapter all results obtained by running the retrieval on the testing dataset are explained. Moreover, this chapter explores the statistical performance of the retrieval, using the testing dataset. Additionally, two error measures are introduced, which can be used to estimate the retrieval performance even if the true atmospheric state is unknown.

6.1 Case study

At first one has to understand the fundamental problems and properties accompanying the used retrieval. In the following one arbitrarily chosen profile will be used as an example for showcasing these problems.

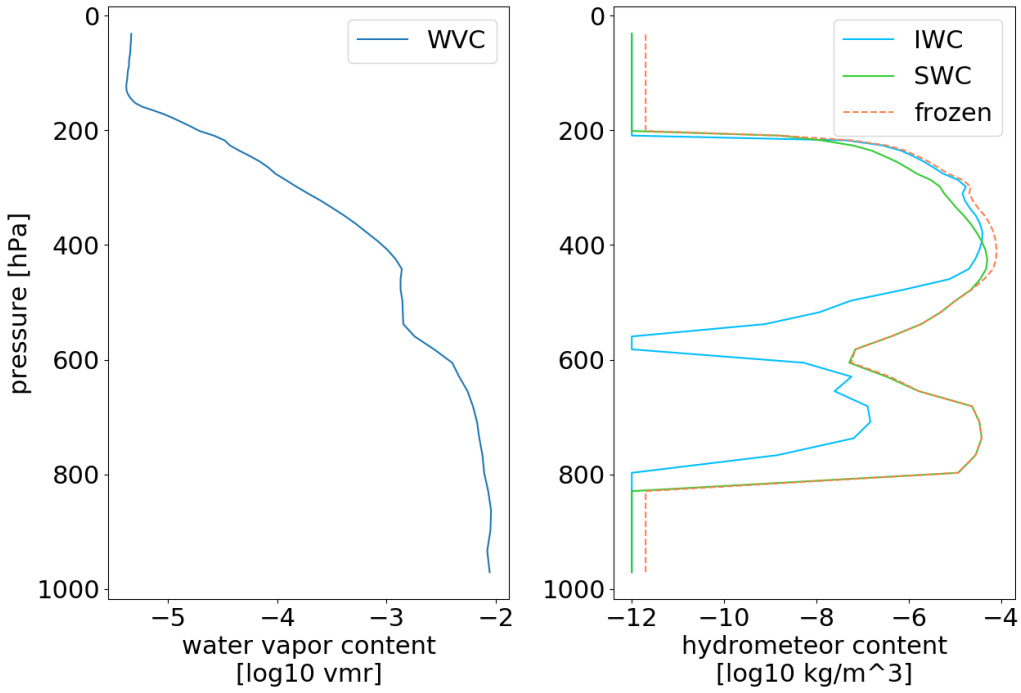

Figure 13: The WVC (left) and hydrometeor content (right) is shown for the chosen profile.

The specific profile was chosen, because it contains only frozen and no liquid hydrometeors.

This allows for a simple but thorough overview of the retrieval results for the chosen profile. Figure 13 shows the true quantities, which will be retrieved in the following step. On the left, the WVC is displayed, while the right side displays the hydrometeor content of SW and IW. One can see distinct ice clouds at heights of around 400 hPa and 750 hPa. The hydrometeor content value of $10^{-12}\,\mathrm{kg/m^3}$ is artificially introduced (see section 6.4.1) and resembles clear sky conditions. The additional orange line in the right graph is the frozen hydrometeor content, which is the sum of SWC and IWC. Later on in this chapter also liquid hydrometeor content will be used, referring to the sum of LWC and RWC. Even though the retrieval quantity is $\widetilde{\mathrm{x}}$ (see equation 2.1), the sum of the actual SWC and IWC x is used to calculate the frozen hydrometeor content. Thus, the deviation between the frozen hydrometeor content and SWC, which for the most part is several orders of magnitude higher than the IWC, is small.

Using the AKM calculated before (see section 4.2), it is possible to estimate the part of the hydrometeor content, which the retrieval can extract out of y. Because the AKM is calculated using the entire testing database, it can be used to give a mean estimate on the retrievable part of any given x. Therefore, it can be applied to a dataset x to get a first idea of what the retrieval result might look like. The actual retrieval should always perform better, because it is a more sophisticated procedure applied to y, while applying the AKM to the data resembles a linear interpolation. Also, the AKM $\mathrm{K_{EOF}}$ has to be applied to the atmospheric state $\widetilde{\mathrm{x}}_{\mathrm{EOF}}$ and can not be used to obtain x for a given y. So the AKM will be used in the following to gain a rough estimate on the null space error and the smoothing error. On one hand the AKM holds information about the retrievable content of a specific setup and can therefore be used to capture one part of the null space error. On the other hand the AKM is also influenced by the smoothing error introduced by the retrieval procedure itself. One can understand the result of applying the AKM to a given x as an estimate for what kind of information the retrieval might see within the corresponding y.

However, to further analyse the retrieval, an error measure is needed. Following the choice made within Holl et al. (2014) and Brath et al. (2018), the median fractional error (MFE) is used. Equation 6.1 describes how the fractional error (FE) is calculated. The median can be applied afterwards to reduce the complexity of the data by a chosen dimension.

$$\mathrm{FE} = \exp_{10}\left(\left|\widetilde{\mathrm{x}}_{\mathrm{retrieved}} - \widetilde{\mathrm{x}}_{\mathrm{truth}}\right|\right) - 1 \tag{6.1}$$

Utilizing the retrieval setup given within chapter 3 and the settings described within chapter 4 and chapter 5 the results displayed in figure 14 are achieved. Displayed is the SWC of the chosen profile together with the resulting FE. The graph on the left shows the retrieved SWC ($\widetilde{\mathrm{x}}_{\mathrm{retrieved}}$) in red, the circular markers representing the actually retrieved heights. Both the black ($\widetilde{\mathrm{x}}$) and the blue ($\widetilde{\mathrm{x}}_{\mathrm{recon}}$) line describe the actual SWC, while the latter is reconstructed after an EOF transformation with the previously specified cutoff. The green line shows the result for applying the AKM to the true SWC profile ($\widetilde{\mathrm{x}}_{\mathrm{AKM}}$). One can see that the retrieved SWC is closer to the truth than the smoothed one, as it was expected. For the scope of this section the only thing to note about the FE is the magnitude relative to the performance of the retrieval

results. The FE is later on (section 6.4.1) used to derive a precision until which the retrieval results are deemed reasonable. Furthermore, the orange shade indicates the spread of the NN ensemble. With increasing altitude the spread of the NN ensemble decreases, which is a result of the lower variability of SWC within the upper atmosphere.

The best accordance between retrieval and truth are in the upper parts of the profile starting at 450 hPa and above. Below 450 hPa the retrieval performs predominantly worse in comparison, which can be seen at the FE. Most likely this is caused by a masking effect of the dense ice cloud at 400 hPa. The simulated measurement is taken at the top of the atmosphere. Thus, the radiation from the lower levels of the atmosphere has to penetrate the dense ice cloud, which radiates such a distinct signal itself, partially masking the radiation coming from below. The radiation of the also dense cloud at 750 hPa is masked in such a way that the retrieval, given only the 18 channels, can not resolve the existing vertical structure. This is described as the smoothing error of the retrieval. Further discussion on the fundamental retrieval errors will follow within the next section.

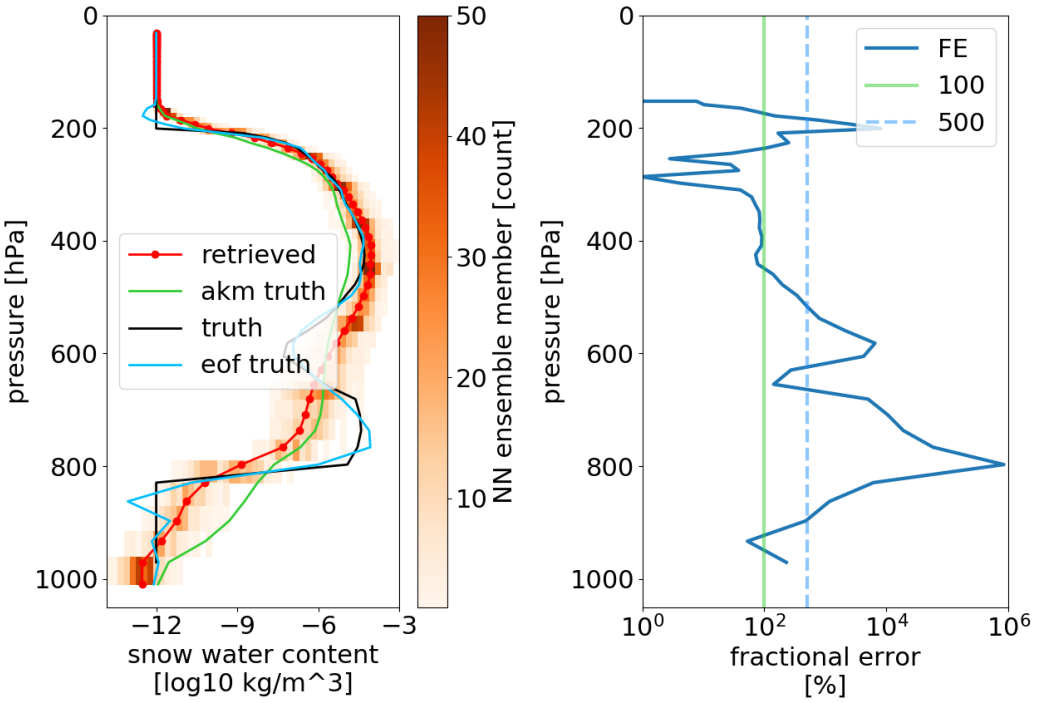

Figure 14: The left graph shows the SWC of the selected profile (black), as well as the reconstructed SWC based on the EOF cutoffs (blue), the retrieved SWC (red) and the SWC estimated with the AKM (green). The orange shade indicates the NN ensemble spread. Red dots indicate the specific height levels of the retrieved data. The right graph shows the corresponding FE and two selected boundary values.

6.2 Analysing the retrieved data

This section contains the achieved retrieval results using the testing dataset. The results will be discussed in respect to the fundamental retrieval errors. To simplify the analysis, the retrieved quantities are always considered as frozen and liquid hydrometeor contents, while the retrieval is processed on the single hydrometeors. Note that this chapter will only discuss the retrieved hydrometeor content. The WV retrieval is only used as an example for complex features.

The retrieval is processed using the settings listed at the end of chapter 4 and mentioned in chapter 5. After training the retrieval with the training dataset the testing dataset is processed, resulting in the data shown in figure 15. On the left side the results for the frozen hydrometeor content are shown, while the liquid content is on the right. All profiles are sorted to range from lowest hydrometer paths to highest. The path describes the column integrated quantity.

The null space error can not be found within figure 15, since the training and testing dataset are based on the exact same dataset. By randomly splitting said dataset into two parts the statistical basis of both resulting parts stays approximately the same (see section 5.2). As a result the training database will cover most atmospheric conditions contained within the testing dataset. So the part of the null space error originating from the statistical disagreement between the training and testing database can not be assessed. This aspect of the null space error is based on information that is not given within the training dataset.

However, there is a second aspect of the null space error, which describes the inability of the retrieval to reproduce a specific feature of the true atmospheric state. This error is different from the smoothing error (described in the following paragraph), because it addresses only atmospheric features that can not be picked up by the retrieval and not such features that can not be fully resolved. Such a feature is the part of the atmospheric profile that is cut off by the EOF transformation. The difference between the black and the blue line on the left-hand side of figure 14 resembles this kind of null space error for the SWC retrieval.

The smoothing error contains all features that can not be fully resolved by the retrieval. For example within figure 14 the actual SWC profile shows two distinct peaks, which the retrieval result (red) does not capture. To get an estimate on the effect of the smoothing error one can apply the corresponding AKM to a profile, as described in detail in section 4.2. The result of the smoothing error is also displayed in the second row of figure 15. One can find the effect of the smoothing error by comparing the retrieved hydrometeor content and the AKM transformed truth to the original truth. The retrieved and smoothed content never change as abruptly as the true content does. A prominent example is seen on the lower and upper edges of the clouds, looking at the fourth row of figure 15. It results out of the masking effect described before and the shape of the Jacobians, describing the limitations of the measurement. Both effects are also visible in figure 14.

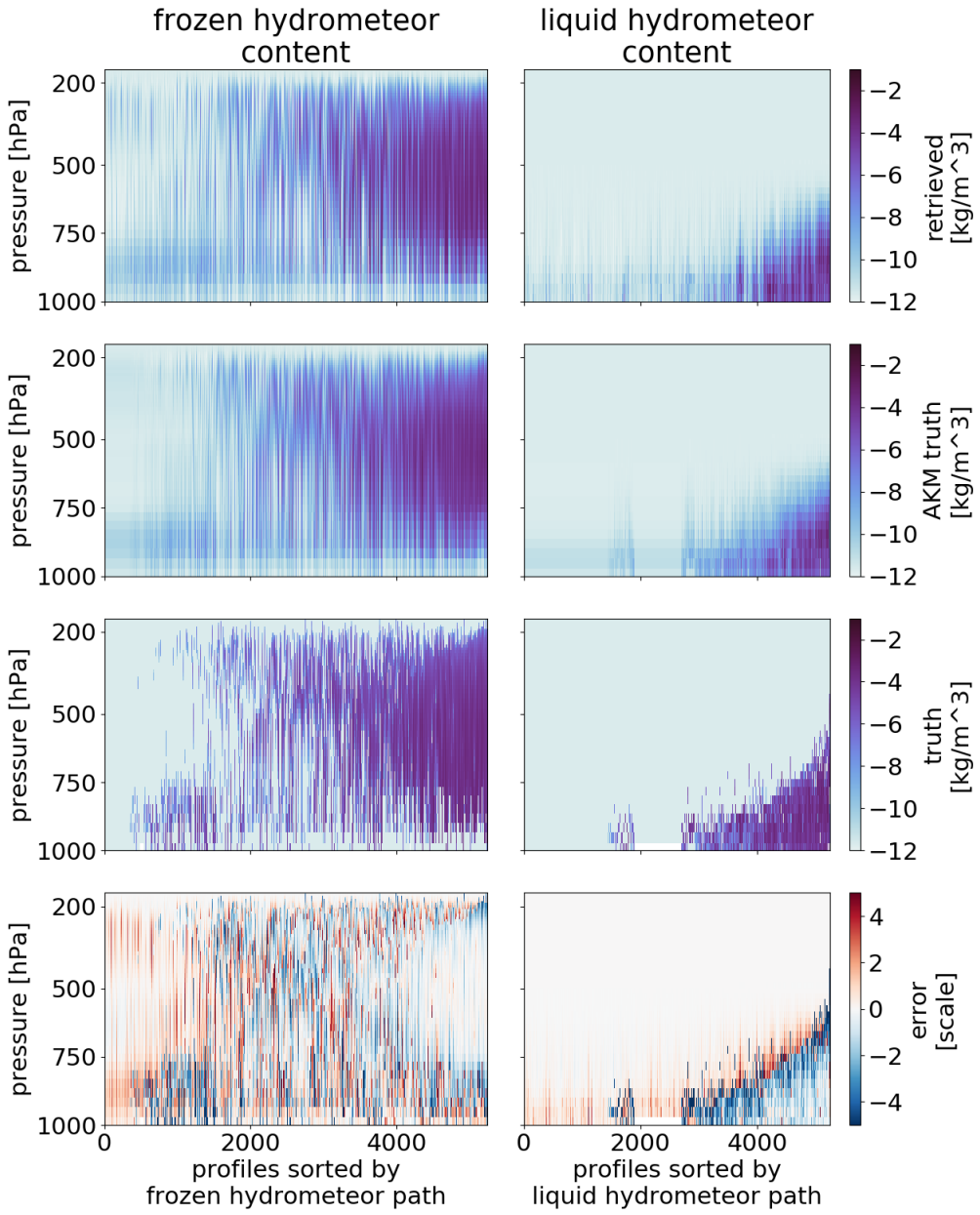

Figure 15: Retrieval results for frozen (left) and liquid (right) hydrometeor content are shown. From top to bottom are shown: the retrieval result, the AKM transformed truth, the truth and the residual of the retrieval. The residual is calculated based on the decadic values.

Furthermore, the smoothing error is captured well within the covariance matrix of the WV retrieval error. The error covariance matrix is shown in figure 16. The most prominent feature to note about the covariance matrix are the high values along its diagonal. They relate to the fact that consecutive height level are dependent on one another. As the retrieval process is executed in eof-space, the EOFs are connecting neighbouring height level to each other. Thus, an error at a certain height does always imply an error around it. The smoothing error relates to the second feature within the error covariance matrix. The distinct negative pattern around the diagonal. At a height of 500 hPa the pattern is most dominant. It is caused by the retrieval errors occurring

at cloud edges. If a cloud edge is to sharp, meaning the hydrometeor content drops or increases several orders of magnitude, the retrieval can not resolve such an edge. The result can be seen in figure 14 and figure 15. The clear sky part of the edge is overestimated, while the density of the cloud is underestimated at the edge. Thus, resulting in a negative coupling of neighbouring height levels, which results in the pattern seen in the error covariance matrix. This effect is also described as the smoothing error.

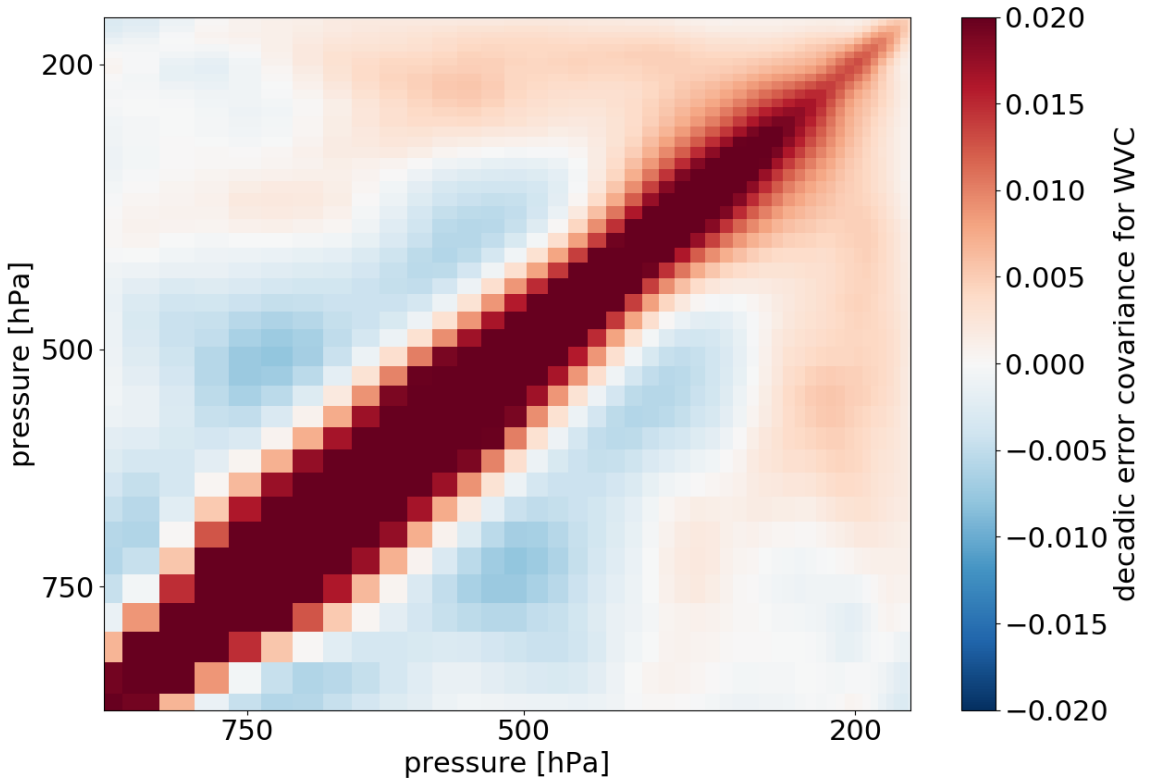

Figure 16: The error covariance for WVC is shown, which is calculated based on the decadic error. While the maximum of the decadic error covariance is 0.058, the upper limit of the colorbar is only set to 0.02. Therefore, the strong red region appears on the diagonal.

Another error to assess originates from the measurement error. An effective way of understanding this error is to compare the instruments accuracy to the covariance matrix of the original $\mathbf{y}$, calculated by the radiative transfer model. However, within this study the measurement error is not explicitly analysed. A detailed explanation for the reasoning can be found in section 5.2. In order to be able to assess the effect of the measurement error, another error measure is examined within section 6.4.2.

To compare the achieved retrieval results with the results presented in Brath et al. (2018), the hydrometeor paths are calculated. Figure 17 shows the results. The error is calculated by subtracting the true hydrometeor path from the retrieved one. Both hydrometeor paths are therefore kept as decadic values. Thus, the resulting error describes the scale of the residual. From top to bottom the results for SWP, IWP, LWP, and RWP are displayed. Each of the graphs contains a distinct group of atmospheric states in the upper left corner. Those states are clear sky cases. The length of the resulting line gives a hint about the retrieval's performance on low

hydrometeor paths. Focusing on the LW panel, the retrieval shows a clear overestimation of the LWP for clear sky cases, while at the same time, cases with a low LWP (below $10^{-4}\,\mathrm{kg/m^2}$) are underestimated. Thus, the retrieval does not provide reliable results below a LWP of $10^{-4}\,\mathrm{kg/m^2}$. If the retrieval yields a LWP of $10^{-4}\,\mathrm{kg/m^2}$ a clear sky scenario might already be prevalent.

The other prominent feature is the underestimation of the hydrometeor path. All retrieval quantities, except for RW, show this behaviour. It could be related to the underestimation of dense clouds due to a masking effect. However, this feature continues throughout the entire range of retrieved hydrometeor paths. Thus, it is most likely not caused by a masking effect of dense clouds, but rather a underestimation of the hydrometeor content in general. This proves to be true, considering the results in the following section.

The RWP seems to be such a difficult retrieval quantity, that below $10^{-3}\,\mathrm{kg/m^2}$ the retrieval result is uncertain. Up to that point most of the atmospheric cases are only underestimated, while after that point the resulting error is in the order of 4 to -3. The results for RWP shown in Brath et al. (2018) are in line with the ones found within this study.

Considering the other three hydrometeors, some characteristic patterns can be found. For a retrieved hydrometeor path below $10^{-5}\,\mathrm{kg/m^2}$ the atmospheric state relates most likely to clear sky. Between a retrieved value of $10^{-5}\,\mathrm{kg/m^2}$ and $10^{-4}\,\mathrm{kg/m^2}$ the atmospheric state can not be determined with certainty. However, if the retrieved hydrometeor path is above $10^{-4}\,\mathrm{kg/m^2}$ one can be certain, that a cloud is present and most likely underestimated in its density.

These results match the ones in Brath et al. (2018) for the LWP, but vary greatly for the frozen hydrometeors. In Brath et al. (2018) the SWP is retrieved more consistently in a way, that it is neither under- nor overestimated. Also the scale of the error is smaller by one order of magnitude. The retrieval in Brath et al. (2018) directly retrieved the hydrometeor paths. Thus, resulting in a more uniformly distribution around the statistically most likely hydrometeor path. As the retrieval examined within this study is based on the hydrometeor content and rarely overestimates it (see section 6.3) the resulting hydrometeor paths can also not be overestimated.

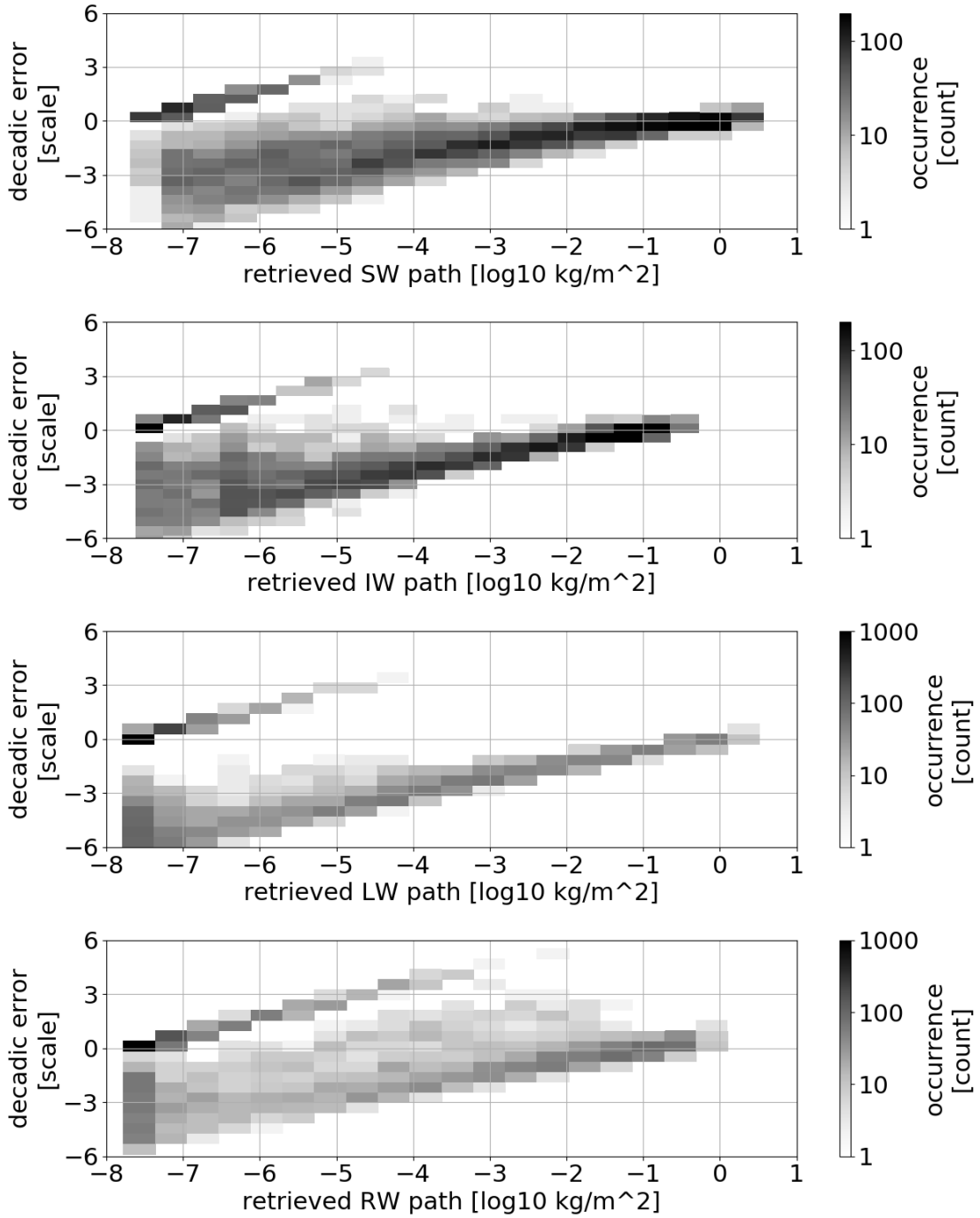

Figure 17: The occurrence of decadic errors for the testing dataset is shown. The decadic errors are compared to the retrieved hydrometeor path for all hydrometeors.

6.3 Applying error measures

This section will further discuss the precision of the retrieval results. Therefore two measures are used and their limitations discussed. One is the MFE and the other one is the standard deviation resulting out of the NN ensemble.

The first fundamental approach within this section is the binning of the retrieved and true dataset. By introducing bins for the hydrometeor content and the pressure scale the retrieval results of the entire testing dataset can be analysed at once. The upper graphs in figure 18 show the number of profiles that fall into a specific bin. Results for the frozen hydrometeor content are shown on the left, while for the liquid they are shown on the right side.

$$e_\mathrm{b} = \mathrm{mean}\left(\widetilde{x}_{\mathrm{retrieved,b}} - \widetilde{x}_{\mathrm{truth,b}}\right) \tag{6.2}$$

The second row contains the mean retrieval error within a specific bin. It is the mean error described by equation 6.2. By using x as normed decadic values within every bin b the error measure e describes the scale of the error. e also allows for an assessment on whether the retrieval is over- or underestimating the true content at a specific height and for a specific hydrometeor content. The third row within figure 18 contains the MFE (see euqation 6.1) for the frozen and liquid hydrometeor content. In the same way as for the mean error e, the median for the FE is calculated over every bin individually. A more detailed description on the MFE for all retrieved quantities will follow in section 6.4.1. The last row of graphs shows the standard deviation of the NN ensemble. The MFE and the std of the NN ensemble will be further discussed in detail in the following section.

The most important detail to keep in mind, when analysing the results in figure 18 is the distribution of the data. At a hydrometeor content of around $10^{-10}\,\mathrm{kg/m^3}$ the dataset becomes sparse. Within every height the number of occurrences drops below 50. Thus, training on this extreme events is not only difficult because of the low hydrometeor content, but also due to the fact that those cases are rare. The same holds true for high hydrometeor contents. Therefore, the retrieval can not be expected to work in such cases.

Focusing on the liquid hydrometeor content in figure 18, the mean error reveals the effect of the null space. The retrieval consequently underestimates the liquid hydrometeor content above 700 hPa to 600 hPa. However, the standard deviation (std) of the NN ensemble is constantly low above 650 hPa. The deciding factor is the sparseness of the data, which is shown in the upper right panel. The existence of liquid hydrometeor content above a certain height is not at all captured by the retrieval and thus lies within the null space.

Another prominent feature is the systematic underestimation of existing hydrometeor content at any given height, while for clear sky cases the hydrometeor content is overestimated. The overestimation is small compared to the relevant range of retrievable hydrometeor content, indicated by the MFE panels. If the indicated two orders of magnitude are added to the $10^{-12}\,\mathrm{kg/m^3}$ one still ends up well below the retrievable $10^{-4}\,\mathrm{kg/m^3}$.

Looking at the std of the NN ensemble, only a distinct area within the frozen hydrometeor

content provides a sufficient certainty. That area aligns with the small MFEs, the small mean errors, and the statistically most likely atmospheric states. Thus, one can assume the retrieval for the frozen hydrometeor content to work reasonably well in most scenarios. Especially, because the most likely case is clear sky, which also possess a MFE around 100 %. Considering the small hydrometeor content of the clear sky cases, a MFE of 100 % is negligible.

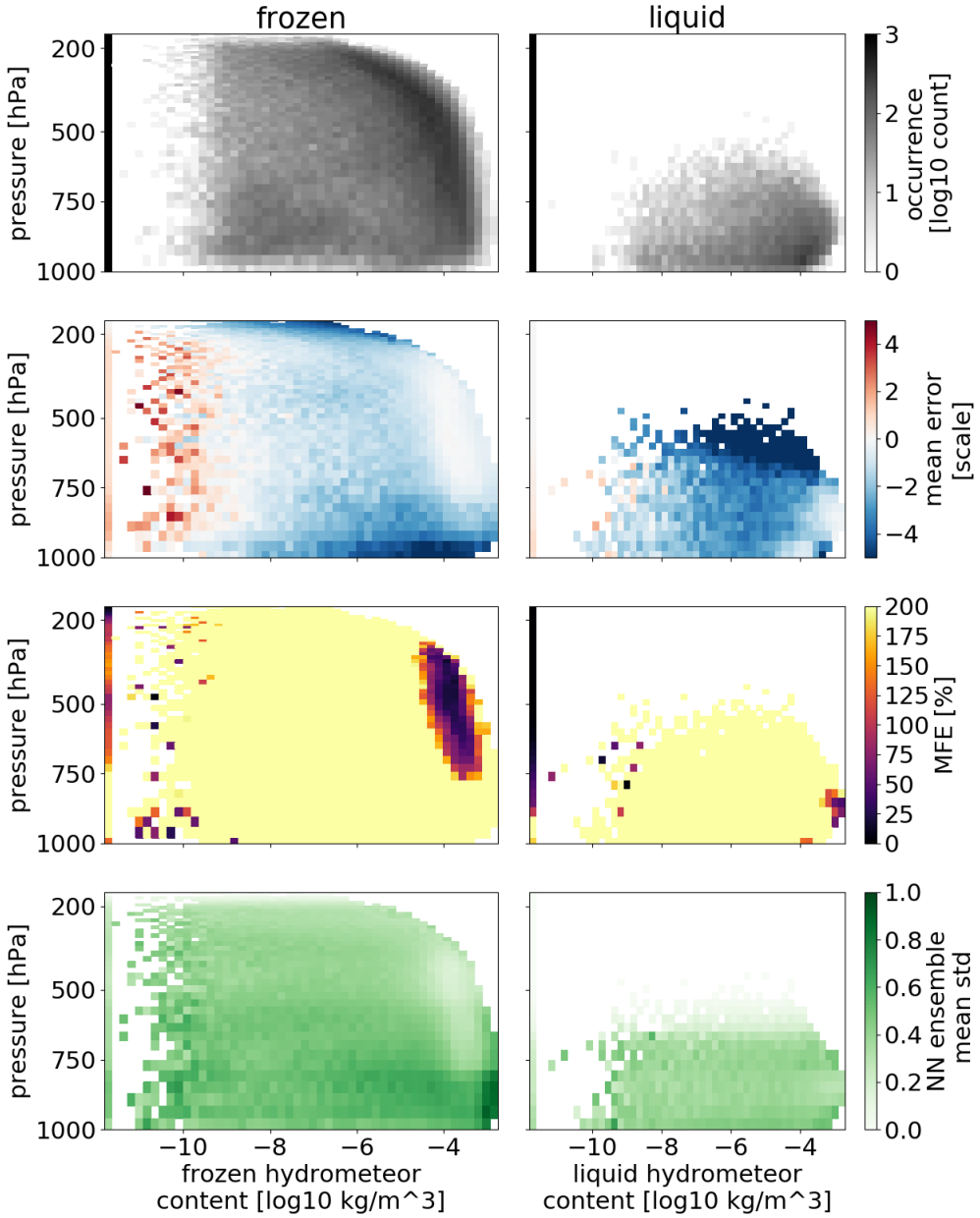

Figure 18: A statistical analysis is shown for the retrieval of the testing dataset. On the left the results for the frozen hydrometeor content are shown, while the results for liquid hydrometeor content are on the right. From top to bottom the graphs show the occurrence of a certain hydrometeor content within the testing dataset, the mean error for the decadic results ('bias'), the MFE, and the mean standard deviation of the NN ensemble members. The y-axis shows the true hydrometeor content.

6.4 Estimating the retrieval error

This section is split into two parts. Both are describing a measure that can be used to estimate the retrieval's performance, while the actual true atmospheric state is unknown. This proofs to be essential for analysing the flight data later on in section 7.1.

6.4.1 Minimum content boundary

This subsection describes the definition of a minimum content for every quantity, that needs to be present at a specific height to achieve a decent retrieval result. The specifics of this process are also explained. At first, the entire process will be explained using SWC as an example. Thereafter the WVC will be used to discuss the use of the minimum content boundary. Finally, the results for all hydrometeors and occurring issues are discussed.

The goal is to define a hydrometeor content as a threshold for the retrieval to perform adequately. A more detailed definition will be given later on, but in essence it is thus possible to judge whether a retrieval result is more precise than a specified error. Important is the fundamental assumption that the retrieval performance within any given height is predominantly depended on its respective hydrometeor content. Later on the truthfulness of this assumption is shown.

The first step in order to find the minimum content boundary, is to assess the retrieval's performance. This is done using the FE introduced by equation 6.1. After calculating the FE for the entire testing dataset one receives the results shown within figure 19. To understand the process of acquiring the minimum content boundaries (orange lines), the following paragraphs will explain the necessary steps. The explanations are all based on the corresponding figure 19.

The profiles are sorted by their hydrometeor content within each height separately. Thus, the profiles are not left as a whole but rather picked apart within each height. The example is based on the SWC retrieval. However, the sorting is not only done for SWC (right panel), but also for the FE (left panel). Both are sorted in the same way, so that each corresponding FE within a certain height matches the respective true hydrometeor content. The sorting is reflected by the constantly increasing SWC within the right panel and the distinct contours of the FE within the left panel. These contours within the FE agree with the fundamental assumption, which states that, if the hydrometeor content increases within a given height the FE decreases.

Now, the first important thing to address is the dashed red line. This line is used as a threshold to determine whether the profile contains any retrievable hydrometeor content or if it is a clear sky case. The threshold was set to -12 within the decadic system. This threshold is also in line with respect to equation 2.1. The retrieval performance was found to consistently differ between cases with a higher SWC than $10^{-12}\,\mathrm{kg/m^3}$, compared to cases with that exact SWC. The FE was found to remain nearly unchanged for most of the retrieved values at the threshold. Due to this fact all cases to the left of the dashed red line are suspected to be clear sky cases. As such, the FE for all those cases is calculated by taking the median within every height over all profiles placed at the left side of the dashed red line. This leads to the uniform FE to the left of the dashed red line (left panel).

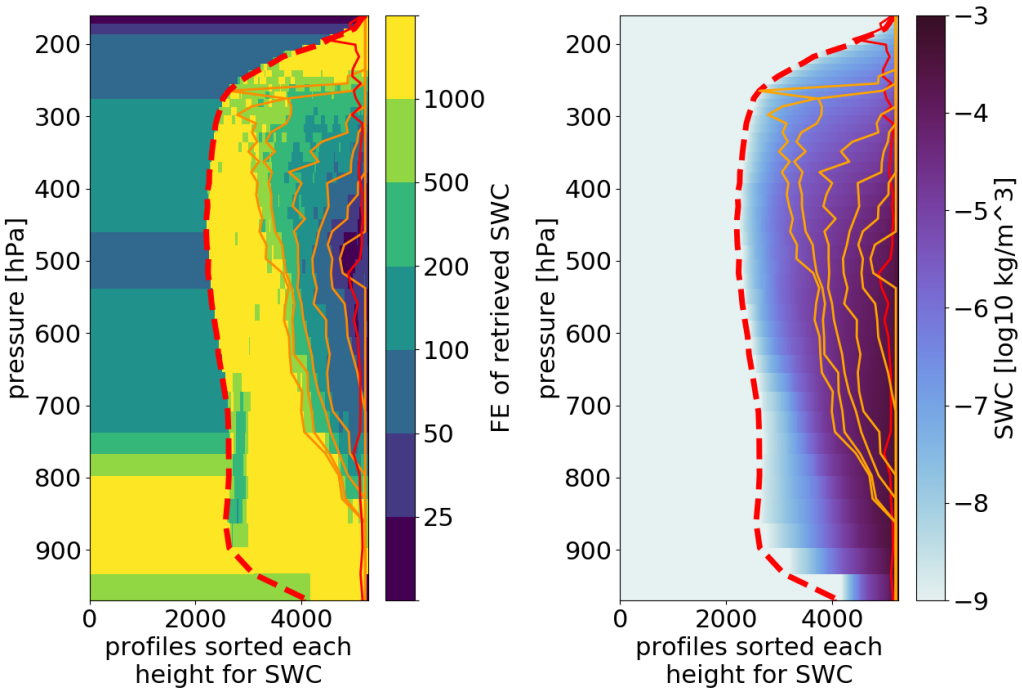

Figure 19: The figure contains the data which is necessary to define the minimum content boundary for SW. The right graph shows the true SWC, while each height is sorted for the SWC independently. The FE (on the left) is sorted accordingly. The dashed red line separates the clear sky cases (to the left of the line, equal to 10^{-12} kg/m^3) from the others. The FE for the clear sky cases is shown as the median over all the clear sky cases. The thin red line marks the FE minimum with respect to a maximal amount of SWC. The orange lines show the identified minimum content boundary for different maximal allowed FEs.

To get the minimum content boundaries (orange lines) a starting point to search for the boundary value has to be defined. This starting point is resembled by the red line for each height. These starting points are calculated, by searching for the first minimum within the FE (left panel) starting from the highest hydrometeor content within each height. To exclude fluctuations within the FE a running mean of 50 profiles is applied. Thus, the starting points resemble a conservative estimate of the smallest FE with regard to the fundamental assumption.

Searching for the minimum content boundary is thereafter a matter of searching for the FE that matches the one, that is supposed to be the maximum FE allowed. To exclude fluctuations once again a running mean over 50 profiles is applied, thus outliers within the FE will not stop the search. The search continues to lower SWC values (to the left within the left panel) until the maximum FE is reached. At this point one receives one of the orange lines, which are calculated for a maximum FE of 25 %, 50 %, 100 %, 200 %, 500 % and 1000 % respectively. If the FE at the starting point already exceeds the maximum FE, the search is conducted to towards higher hydrometeor contents.

As a last step, the minimum hydrometeor content can be extracted. This is done by following the orange line within the right panel of figure 19. The resulting hydrometeor content has to be present at a specific height for the retrieval to perform better than the chosen maximum FE. This hydrometeor content profile is called the minimum content boundary. Within the rest

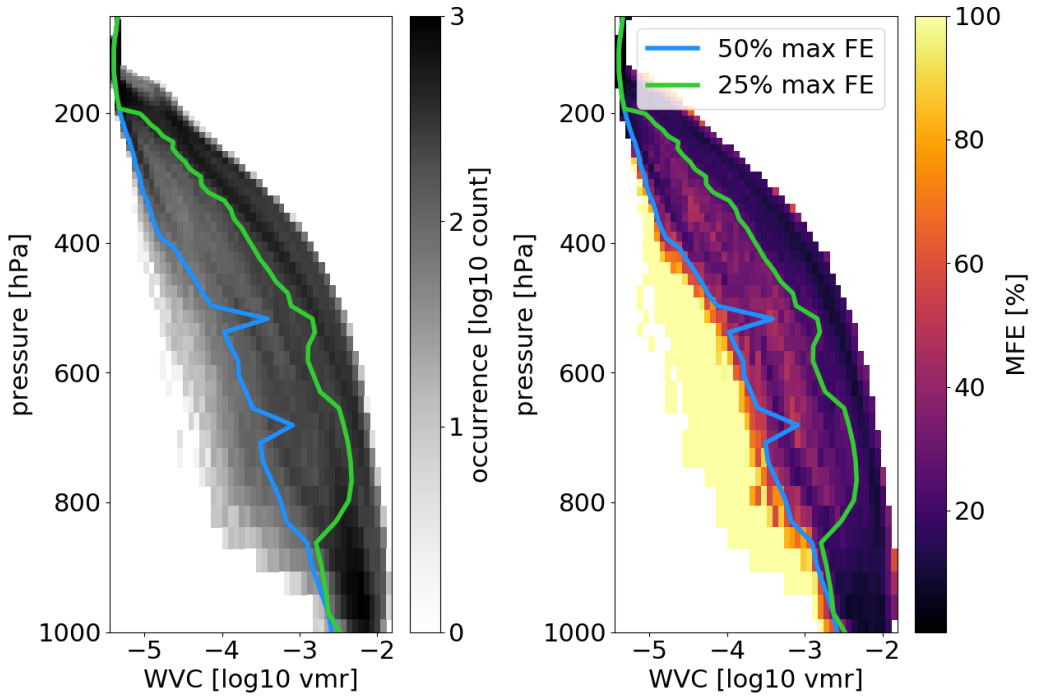

Figure 20: The occurrence of a certain WVC at a specific height within the testing dataset is shown on the left. On the right the MFE for the retrieval result of WVC is shown, as well as the minimum content boundary for 25 % (green line) and 50 % (blue line) maximum FE. The y-axis shows the true WVC.

of this subsection the use of this minimum content boundary will be discussed.

Using WVC as an example, the acquired minimum content boundary for 25 % and 50 % maximum FE is shown in figure 20. The left graph shows the number of profiles which fall into one bin and the MFE for each bin is shown on the right. Thus, one can see where the minimum retrieval boundaries (blue and green lines) end up with respect to the MFE. Their position can be roughly estimated by going from left to right, until the specified maximum FE is reached. Therefore, they are a conservative estimate on the WVC that needs to be present at a certain height for the retrieval to achieve a MFE lower than the predefined maximum FE.

Up to about 900 hPa the retrieved WVC is mostly below a FE of 25 %. Most of the cases are already captured to the right of the 25 % minimum content boundary (left panel, green line). Ultimately, just a few hundred extremely dry atmospheric states end up to the right of the 50 % minimum content boundary. Only the extreme cases that are not well represented in the database show a higher FE. The high FE error can be explained by the small statistical basis of the training database for those cases. The generally small FE can be explained by the small variability of WVC within the atmosphere (see figure 6). For the upper end of the atmosphere the variability within the given training dataset nearly disappears (see figure 6). Thus, almost all profiles end up in the same bin in the left panel of figure 20. This leads to the fact that the mean profile, which is used to reconstruct the retrieval results from eof-space back to normal-space, already fits the true WVC. Therefore, the retrieval itself does not need to be as precise, while adjusting the mean profile. After looking at the selected EOF modes for WV, it is possible to

proof, that the retrieval can not even change the WVC within the upper part of the atmosphere. So one could state that this part of the retrieval is only based on the a priori information, which is the mean WVC within the testing dataset.

Figure 21 shows the same data as figure 20, but for the different hydrometeors. The first thing to note is the new scale for the MFE which goes from 0 % to 500 % instead of up to 100 %. Also, the maximum FEs chosen to calculate the minimum content boundaries are changed to 100 % (green line) and 500 % (blue line). For a bin containing a low number of cases (left graph) the MFE (right graph) sometimes ends up being better than it would be for a bigger sample size. This effect is visible at around $10^{-9}\,\mathrm{kg/m^3}$ for all heights and hydrometeors.

For the frozen hydrometeors, one can see a distinct region, in which the retrieval performs admirably (lower MFE than 100 %). This region follows the most common cloud structures (respective left graphs). Raising the question why the retrieval is able to pick up the radiation signal of less dense ice clouds in the upper atmosphere. High clouds emit a colder radiation signal compared to lower clouds, resulting in a bigger difference to the background radiation coming from the lower atmosphere and the ground. Thus, higher ice clouds are more easily detected. In a hight of 300 hPa ice clouds can be detected down to a hydrometeor content of $10^{-5}\,\mathrm{kg/m^3}$

Another feature of the retrieval is, that frozen hydrometeor content close to the ground can not be retrieved. The reason for this can also be found within figure 14. While the dense ice clouds in the upper atmosphere are well retrieved, they mask any radiation signal coming from the lower atmosphere, such that the sensor can't detect it. Therefore, the retrieval can not resolve the lower edge of the clouds. Those clouds stretch into the lower atmosphere, resulting in high MFEs. This effect can be contributed to the smoothing error (see section 6.1) and also accounts for the high MFE for clear sky cases. The clear sky cases in this graphs are also not left as a whole. As the binning is based on hydrometeor content and height, the profiles are picked apart. Thus, a dense cloud might be above any given clear sky value within the figure. Resulting in the fact that the frozen hydrometeor clear sky cases come with such high MFEs.

The smoothing of the lower cloud edges can also be found within the results for the liquid hydrometeors. At the point, where the good region of the retrieval ends even the clear sky retrieval gets worse than it normally is. While the good region for the liquid hydrometeor retrieval is way smaller than for the frozen one, the number of non clear sky cases is smaller as well. The RWC can be retrieved with an MFE lower than 500 % even up to the ground, while the LWC can only be retrieved around a height of 750 hPa. Also the retrieval of LWC is only possible if the LWC is above $10^{-4}\,\mathrm{kg/m^3}$. This is reasonable for the used sensor, given that most channels are sensitive to ice clouds and not to liquid clouds.

By running additional radiative transfer simulations with specific changes in the IWC one can find, that the radiation signal of this change is nearly undetectable for the sensor. This analysis was conducted by Brath et al. (2018). The fact that the retrieval is able to capture IWC with a MFE smaller than 100 % at all is thus questionable. Reason for these good retrieval results is the correlation between IWC and SWC within the training dataset. While an increase in SWC is measurable, it also goes most likely along with an increase in IWC. This relation allows for the retrieval to detect the IWC, while not detecting the radiation of the IWC, but radiation

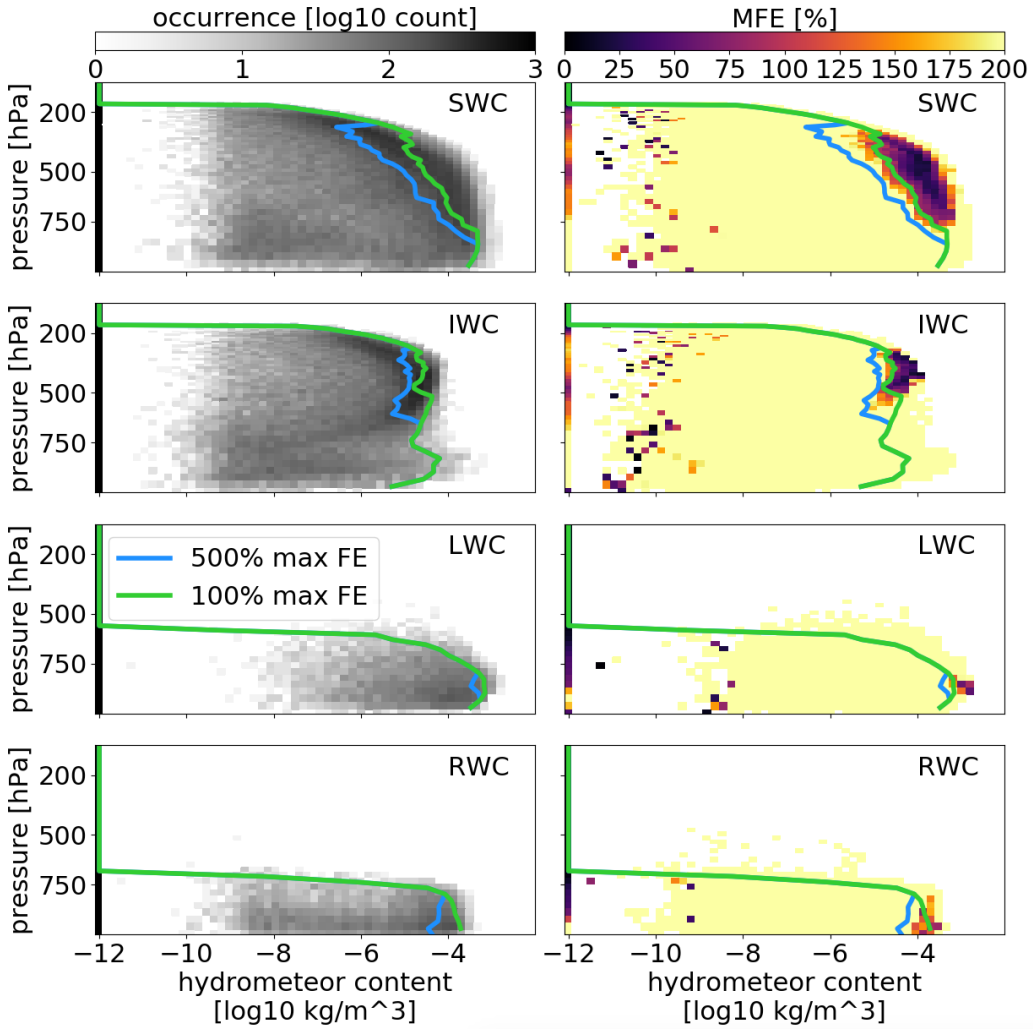

Figure 21: The occurrence of a certain hydrometeor content at a specific height within the testing dataset is shown on the left. On the right the MFE of the retrieval result is shown for all retrieved hydrometeors, as well as the minimum content boundary for 100 % (green line) and 500 % (blue line) maximum FE. The y-axis shows the true hydrometeor content.

based on the SWC.

The presented definition of the minimum content boundary is further used to determine whether the retrieval achieved good results or not. To do so, one has to compare the true hydrometeor content to the minimum content boundary. Thus, it is only possible for cases where the truth is already known. In the following analysis of the flight measurement (see section 7.1) the error measure shall be applied, while only knowing the measurement. Therefore, the assessment of the retrieval performance is made, based on the retrieved data. The only values marked as an error by this approach are those, that fall below the minimum content boundary. If the retrieved value is higher than the minimum content boundary, even though it is a mistake, the value wont be detected as an error. However, the chosen minimum content boundary is conservative, declaring only the best parts of the retrieval as good.

The minimum content boundary has one strong advantage. It can be applied after the retrieval results are processed and does not need any additional computation, after the boundaries

are set. However, the drawback of this is the inflexibility of the set minimum content boundary. Once the boundary is set it never changes. Thus, is does not correspond to the retrieval of an individual profile. It does only represent a fixed cutoff based on the retrieval setup and the utilized testing dataset.

6.4.2 Noise ensemble standard deviation

This chapter will examine another error measure which can be used when operating NNs. To test the stability of the output of the NN ensemble one can slightly vary the input data by adding some noise to it. As this is a common practise in testing the robustness of NNs, which is also used in Uliny et al. (2016).

The idea is to introduce noise on the dataset y which is feed into the NN ensemble. Thereby, estimating the full retrieval error. Instead of only adding one fixed set of randomly generated noise based on the measurement accuracy of the respective channel, multiple sets of noise are generated to assess the reliability of the NN ensemble based on different realisations of the measurement. The assumption is, that this 'noise ensemble' yields additional information on the retrieval error without the need for the true atmospheric state.

If one were to use only the standard deviation of the NN ensemble to assess the reliability, a part of the error would be missing. The standard deviation of the NN only holds information about the error introduced by the retrieval itself. Meaning, the smoothing error and the null space error are covered by it, while the measurement error is not. Smoothing and null space error will contribute an increase within the standard deviation of the NN ensemble. This happens due to fact that the neural nets are only trained on a fraction of the entire dataset. Therefore, the different NNs only converge on one solution if it is well covered by the training dataset. Otherwise the NNs can be expected to act randomly, increasing the overall std of the NN ensemble. By adding the introduced noise ensemble, the measurement error will be fed through the NN ensemble. Thus, the std of the NN ensemble and noise ensemble together is higher, containing all fundamental retrieval errors.

The effect of the noise ensemble is shown in figure 22. The topmost panel displays the MFE for SWC as seen before and is used for comparison. Continuing from top to bottom, the panels contain the std of the NN ensemble, the std of NN and noise ensemble, and a resulting 'full std mask'.

Additionally, all panels show the minimum content boundary with a maximum FE of 100 %. By comparing the second with the third panel, one can see the increase in the ensemble spread introduced by the noise ensemble. The region in which the retrieval results possess the least variability does not change, when introducing the noise ensemble. It does neither change its location, nor does it change the resulting std. Therefore, the method proves to be working as intended.

If a feature is well captured by the majority of NNs within the ensemble, even noise on the input data wont change the result. The std only increases in regions in which the results are not based on learned features, but are rather randomly estimated.

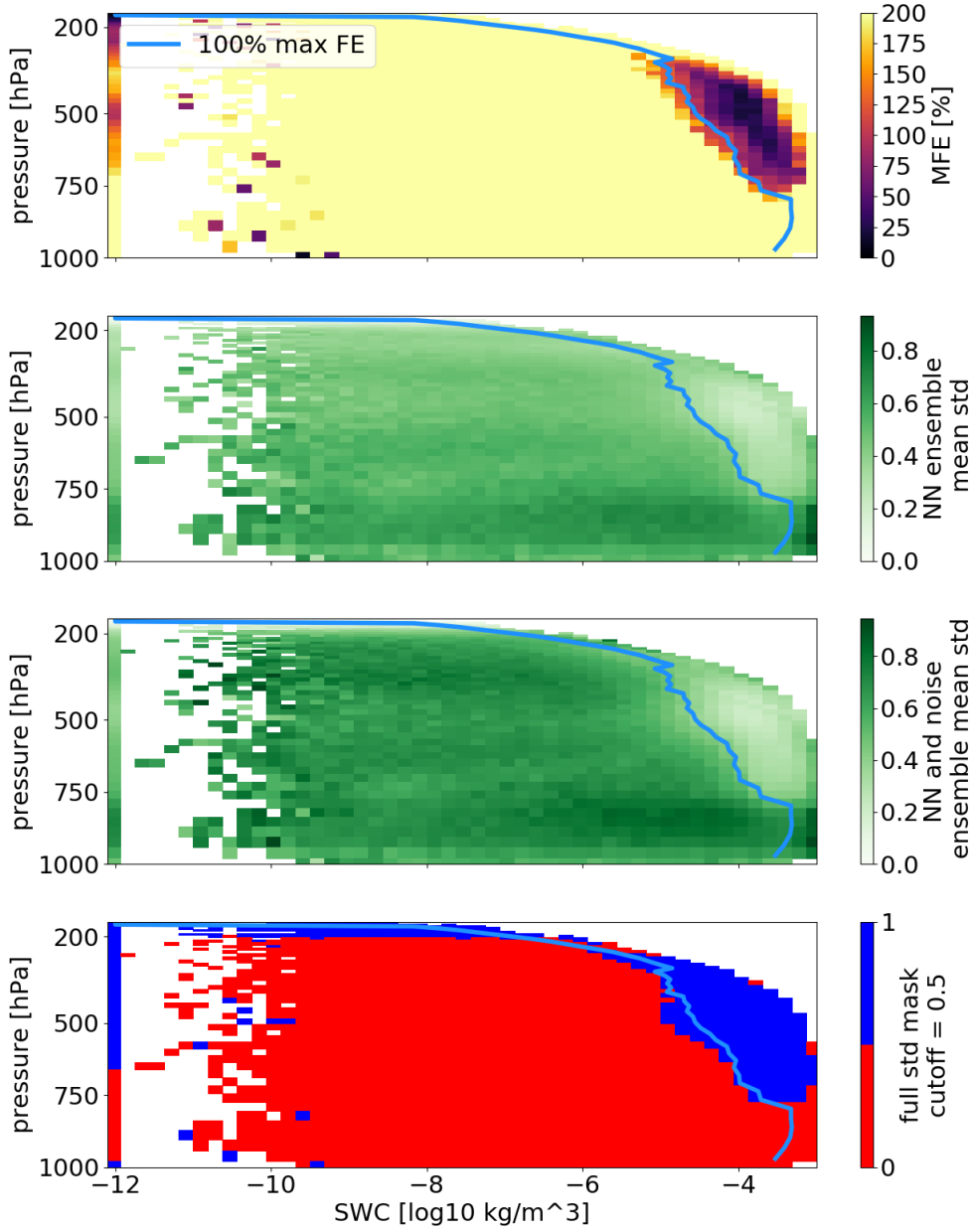

Figure 22: The figure contains from top to bottom the MFE, the std of the NN ensemble, the std of the NN and noise ensemble, and a mask based on the std of NN and noise ensemble. The blue line shows the minimum content boundary for a maximum FE of 100 %. The y-axis shows the true hydrometeor content.

As it is to be expected, the region with the smallest std also lines up perfectly with the smallest MFEs. A threshold is needed, to specify whether one can trust the retrieval results or not, based on the given std. The threshold is set to be 0.5. This choice is made, to match the masking implied by the minimum content boundary. The resulting 'full std mask' is better suited to judge the retrieval's performance of individual profile retrievals. The minimum content boundary does not depend on the individual measurement, while the full std mask does. The std threshold, which yields the full std mask, can be applied to an individual profile retrieval, to

assess the retrieval's performance on that individual set of brightness temperatures.

A strong drawback of this error measure is the computational cost. The computational cost increases with the size of the noise ensemble. The dataset, which one wants to retrieve, has to be processed through every NN. All computations have to be executed again for every introduced realisation of the dataset with a different noise. Therefore, the computational cost does drastically increase. The results shown in this study are achieved by calculating 50 realisations of the dataset.

Chapter 7

Applying the Retrieval

This chapter is concerned with the concept and analysis of applying the examined retrieval to the flight dataset. Within the first section the retrieval will be applied to ICON model data and analysed using the introduced error measures. The model data is chosen to match flight B897, since the measurements used within the second section of this chapter are taken on flight B897. While the second section presents a comparison between the retrieval results and ICON model data, it serves as an outlook of what can be achieved with the given retrieval configuration.

7.1 Retrieving a simulated flight

After outlining the problems which come with airborne datasets, the retrieval will be applied to ICON model data matching flight B897. This section aims to clarify the benefits of the beforehand introduced error measures. Therefore, these error measures are dealt with, pretending not to be in possession of further information about the true atmospheric state.

As explained in section 5.1 a new retrieval is necessary if the training dataset changes. This is also the case for a changing flight altitude. Thus, a new retrieval has to be trained for every altitude in which measurements are taken. Additionally, the flight dataset has to be preprocessed as described in chapter 2 to ensure the correct viewing angle of the sensors. Therefore, the B897 flight dataset used in the second section of this chapter is not continuous. The corresponding model data, which is discussed within this section, is chosen to match the flight dataset, as such shares the characteristic of not being continuous in space and time. For the sake of simplicity the simulated sensors are kept at a constant height, although the aircraft was operated at different altitudes. The model data used in the second section of this chapter captures the changes in altitude.

The benefits of the different error measures are showcased utilizing only the retrieval results for SWC. Figure 23 displays the relevant information. The upper most panel shows the true SWC followed by the retrieved SWC, a mask based on the minimum content boundary, the decadic retrieval error of the SWC and a mask based on the std of NN and noise ensemble. Focusing on the upper two panels, the retrieval result seems reasonable. All strong features are well covered by the results. Especially, the lower cloud edges line up perfectly.

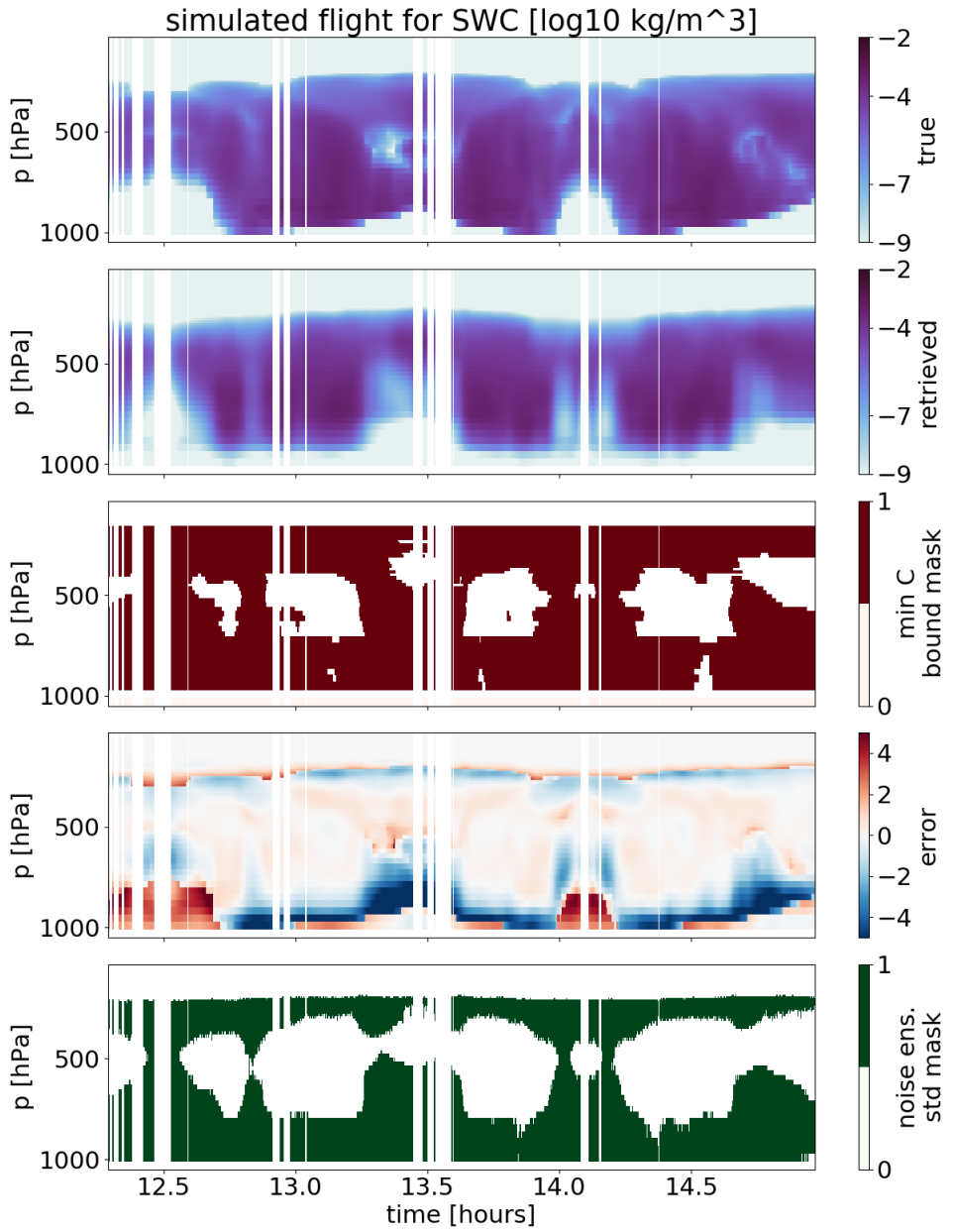

Figure 23: The figure contains the retrieval results for the simulated B897 flight. From top to bottom the true SWC, the retrieved SWC, a masking based on the std of the NN, the error based on the decadic results, and a mask based on the minimum content boundary are displayed. The SWC is given in decadic kg/m^3. The x-axis shows the pressure in hPa, while the y-axis shows the time of the day in hours.

However, the retrieval error, shown in the fourth panel, does reveal the underestimation of the density of lower cloud edges. Even though the cloud edges seem to line up, the masking effect does not allow for the radiation of the lower clouds to be received. Additionally on either cloud edge (top and bottom) the smoothing error is visible as a small over- and underestimation around the cloud.

Another shortcoming of the retrieval is visible. High dense clouds, as they are present at 12.5 h, mask the radiation signal from below. Thus, the retrieval assumes clouds to be below

those high clouds, even thought there are none. This effect occurs again at 14.2 h. Furthermore, the retrieval can not handle the presence of frozen and liquid hydrometeors. This is noticeable at 13.4 h and 14.7 h and is further discussed in the following section.

However, the provided masks, do cover up all of the mentioned shortcomings. While the mask based on the minimum content boundary is very conservative, the mask based on the std of NN and noise ensemble is not. The content boundary mask, displayed in the third panel, only considers the inner most core of dense clouds to be valid. Additionally some features are present, which are especially prominent at 14.5 h. Those artefacts are most likely caused by the retrieved SWC exceeding the minimum content boundary, even though the true SWC is underestimated. The more sophisticated mask based on the std of the NN and noise ensemble is shown in the lowermost panel. The mask is calculated on the different realisations of $\mathrm{y}_{\mathrm{noise}}$, individually calculated for every profile. Therefore, only the outer cloud edges, as well as areas affected by the smoothing error and masking effect are considered invalid. Those areas are all masked, because the retrieval results are not certain enough.

Ultimately, the mask based on the std of the NN and noise ensemble has to be considered more reliable. If one can not spare the needed computational power, the minimum content boundary could also be used. The properties of the std mask make it an excellent choice to assess the retrieval's performance for any individual profile. In contrast, the boundary mask yields a better understanding of the underlying physical requirements on the atmospheric profiles.

7.2 Applying the retrieval to measurements

This section serves as an outlook on potential applications and benefits, originating out of the examined retrieval setup. The retrieval results are evaluated utilizing the residual of the B897 flight retrieval results and ICON model data. As the flight data originates out of multiple altitudes, multiple NN ensembles are necessary for the retrieval. Therefore, the introduced error measures con not be simply applied to the B897 flight retrieval results.

Figure 24 contains the comparison between ICON model data and retrieved measurements taken on flight B897. The panels show from top to bottom the model data, the retrieved data, and the residual. To compute the residual the model data is subtracted from the retrieved data. While the panels on the left relate to the frozen hydrometeor content, the panel one the right show the liquid hydrometeor content.

The most dominant retrieval issues are also found within this figure. Even thought the model can not be considered the truth in this comparison, the masking effect on the lower cloud edge is visible for the frozen hydrometeor content. The residual of the liquid hydrometeor content displays another known feature of the retrieval: liquid hydrometeor content can not be detected above a height of 700 hPa. Thus, the modelled liquid hydrometeor content above 700 hPa can not be confirmed by the measurement results. However, the hydrometeor content does agree reasonably for the centre of the clouds.

Another feature can be found within the figure, which is the effect of liquid hydrometeor content on the retrieval of frozen hydrometeor content. The effect is noticeable at 13.4 h and 14.7 h. While liquid clouds emit a warmer radiation signal, ice clouds emit a cooler radiation signal than their environment. At the two mentioned times a liquid cloud is beneath an ice cloud. The warm radiation signal of the liquid could partially masks the cooler radiation signal emitted by the ice cloud. Thus, the frozen hydrometeor content in the lower part of the ice cloud is heavily underestimated. Vice versa the same could be said for the liquid hydrometeor content, but the retrieval can not detect the upper part of the liquid cloud anyhow (see section 6.3).

Overall the agreement between the to datasets is high. However, the model data contains distinct features that the given retrieval can not be expected to resolve. The masking effect on the lower edge of dense clouds has also been examined before and therefore is to be expected.

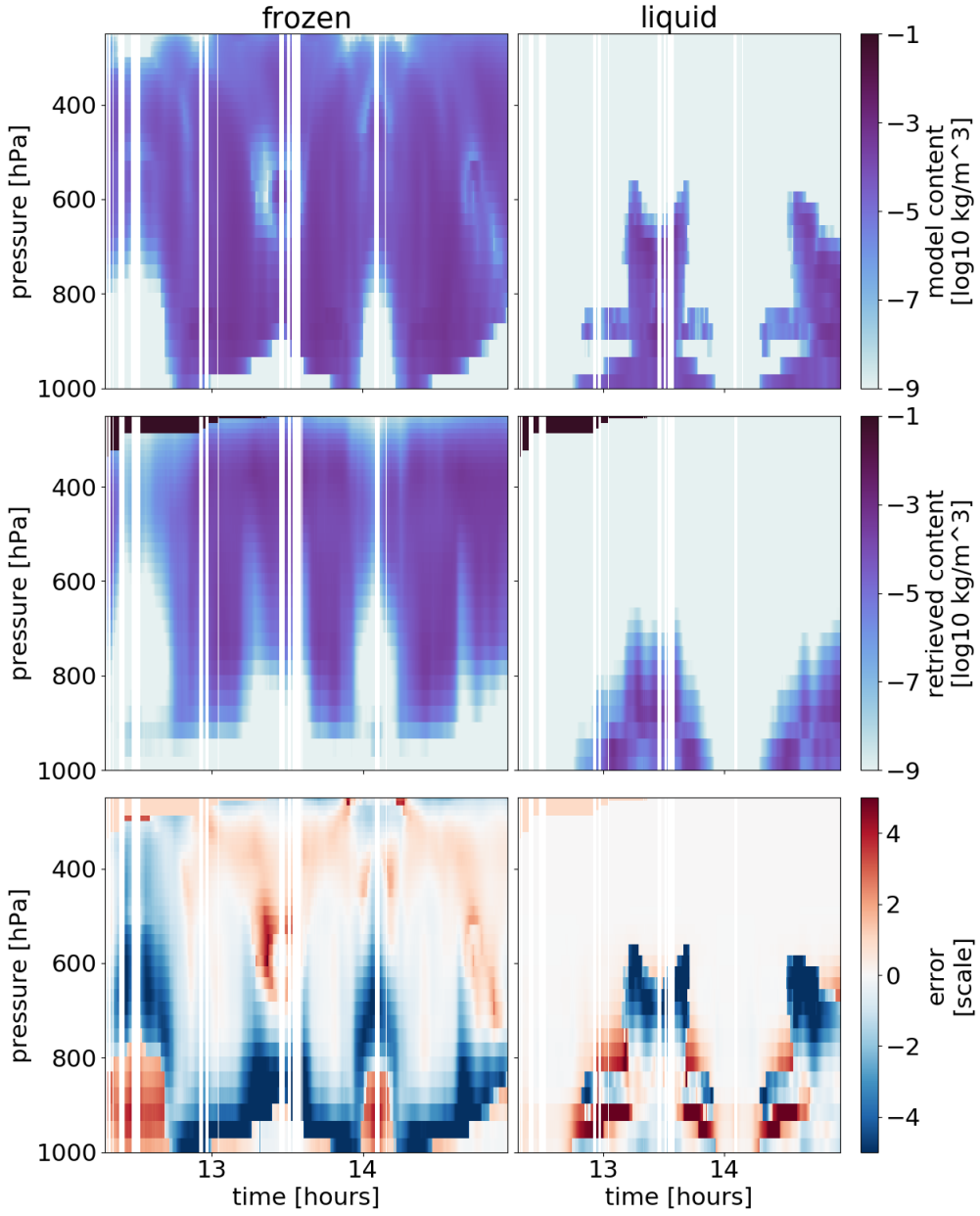

Figure 24: The figure contains the hydrometeor retrieval results for the B897 flight (top row) together with the corresponding ICON model data (middle row). The last row contains the residual of the decadic retrieval results and model data. On the left is the frozen hydrometeor content displayed. The liquid hydrometeor content is displayed on the right. The x-axis shows the pressure in hPa, while the y-axis shows the time of the day in hours.

Chapter 8

Conclusion

Above all, this study explores the possibilities of the retrieval of atmospheric profiles, while solely relying on a passive remote sensing instrument. This is done by applying an EOF transformation to the measurement and state space and the use of a NN ensemble. While the EOF transformation decrease the complexity of the retrieval problem, the NN ensemble bridges the gap between both spaces allowing for a simple connection, that can be further trained.

Although the settings for the EOF transformation are carefully chosen, some residuals remain. One way of handling this includes the introduction and optimisation of a cost function, determining the best number of EOF modes to use. While the simple NN structure chosen for the retrieval does provide the bridge between the two spaces, it also leaves room for improvement. With a bigger training database, better chosen neuron structure and an optimised training algorithm the performance of the NN element of the retrieval is expected to greatly improve.

In terms of accuracy, the introduced retrievals displays some deficits. However, a through assessment of the retrieval's performance is possible, using the introduced error measures. The performance reached for individual profiles can be assessed by utilizing the std of the NN and noise ensemble, adding some computational cost to the analysis. Additionally, information about the physical limitations of the retrieval can be easily assessed, using the minimum content boundary.

One of the retrieval's biggest assets can be found when comparing its runtime (after the set up) to the one of other retrievals. Neither a large database, nor any radiative transfer simulation needs to be consulted at runtime. Additionally, a few matrix and weighting operations allow the retrieval of multiple atmospheric profiles at once. However, this is only possible by defining a non-iterative function f^{-1} to process the measurements, after the system's set up. Furthermore, the functional connection proves to also eliminate the influence of measurement noise on the retrieval.

Considering the simplicity of the used NN, in addition to the already existing deep understanding of their current and future applications, this study yields promising results. The retrieval method examined within this thesis is most useful if considered as a fast way of retrieving multiple atmospheric states at the cost of a bigger uncertainty, compared to retrieval methods with higher computational costs.

References

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer. 23, 25

Brath, M., Fox, S., Eriksson, P., Harlow, R. C., Burgdorf, M., and Buehler, S. A. (2018). Retrieval of an ice water path over the ocean from ISMAR and MARSS millimeter and submillimeter brightness temperatures. *Atmospheric Measurement Techniques*, 11(1):611–632. 1, 3, 4, 5, 25, 30, 34, 35, 42

Buehler, S. A., Mendrok, J., Eriksson, P., Perrin, A., Larsson, R., and Lemke, O. (2017). ARTS, the atmospheric radiative transfer simulator – version 2.2, the planetary toolbox edition. *Geoscientific Model Development Discussions*, 2017:1–28. 4

Fox, S., Lee, C., Moyna, B., Philipp, M., Rule, I., Rogers, S., King, R., Oldfield, M., Rea, S., Henry, M., Wang, H., and Harlow, R. C. (2017). ISMAR: an airborne submillimetre radiometer. *Atmospheric Measurement Techniques*, 10(2):477–490. 3

Holl, G., Eliasson, S., Mendrok, J., and Buehler, S. A. (2014). SPARE-ICE: Synergistic ice water path from passive operational sensors. *Journal of Geophysical Research: Atmospheres*, 119(3):1504–1523. 30

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2014). *An Introduction to Statistical Learning: With Applications in R*. Springer Publishing Company, Incorporated. 25

Kangas, V., D'Addio, S., Betto, M., Barre, H.and Loiselet, M., and Mason, G. (2012). Metop Second Generation microwave sounding and microwave imaging missions. *Proceedings of the 2012 EUMETSAT Meteorological Satellite Conference, Sopot, Poland*, pages 3–7. 3

Kingma, D. and Ba, J. (2014). Adam: A Method for Stochastic Optimization. 25

McGrath, A. and Hewison, T. (2001). Measuring the Accuracy of MARSS—An Airborne Microwave Radiometer. *Journal of Atmospheric and Oceanic Technology*, 18(12):2003–2012. 3

Richards, J. and Jia, X. (2006). *Remote Sensing Digital Image Analysis: An Introduction*. Springer. 11

Rodgers, C. D. (2000). *Inverse Methods For Atmospheric Sounding: Theory And Practice*. Series On Atmospheric, Oceanic And Planetary Physics. World Scientific Publishing Company. 7, 8

Strang, G. (2014). *Linear Algebra and Its Applications*. Elsevier Science. 11

Uliny, M., Lundstrom, J., and Byttner, S. (2016). Robustness of Deep Convolutional Neural Networks for Image Recognition. 44

Versicherung an Eides statt

Hiermit versichere ich, Maximilian Schaper, an Eides statt, dass ich die vorliegende Arbeit im Studiengang M. Sc. Meteorologie selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel - insbesondere keine im Quellenverzeichnis nicht genannten Internet-Quellen - benutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen wurden, sind als solche kenntlich gemacht. Ich versichere weiterhin, dass ich die Arbeit vorher nicht in einem anderen Prüfungsverfahren eingereicht habe und die eingereichte schriftliche Fassung der auf dem elektronischen Speichermedium entspricht.
Einer Veröffentlichung dieser Arbeit stimme ich zu.

Ort, Datum	Unterschrift