Bayesian Retrieval of Thermodynamic Atmospheric Profiles from Ground-based Microwave Radiometer Data

Master's Thesis in Atmospheric Sciences

Submitted to the Faculty of Geo- and Atmospheric Sciences
in Partial Fulfillment of the Requirements for the Degree of Master of Science

by **Christopher Polster**

University of Innsbruck, 2016-08-12

Advisor: Prof. Dr. Mathias Rotach

Abstract

Ground-based microwave radiometers are increasingly used in the atmospheric sciences to retrieve vertical temperature, humidity and cloud information. Such information is valuable for boundary layer research and weather forecasting and efforts are undertaken to assimilate microwave radiometer observations into numerical weather prediction models. Multiple methods exist to perform the retrieval, differing in their data requirements, ease of use and flexibility to include measurements from sensors other than the radiometer.

A linear regression and an optimal estimation technique have been implemented as part of this thesis. They are derived from a Bayesian standpoint and important properties of these methods are discussed. Finally, their accuracy is evaluated with data from radiosoundings and radiometer measurements in Innsbruck. Standard deviations of temperature retrievals from an optimal estimation scheme incorporating forecasts from a numerical weather prediction model are found be be less than 1.2 K throughout the troposphere. The least accurate region is located between 1.5 and 3 km above ground level. At these heights the numerical forecasts are not as accurate as in the upper troposphere and the information content of the radiometer has already decreased substantially compared to the lower atmosphere therefore the retrieval scheme struggels to perform well.

Two case studies reveal that the optimal estimation scheme is promising for the retrieval of temperature inversions which are an often studied problem of microwave radiometer retrieval. An experiment shows that the quality of a-priori information, particularly its capability of providing a description of the features that an atmospheric state exhibits, has much influence on the accuracy of retrieved vertical profiles. The a-priori information are therefore a good place to start when trying to improve the retrieval performance.

Also presented in this thesis is a prototype of a numerical radiative transfer model for the microwave region. It is a minimalistic implementation in a high-level programming language and able to calculate linearizations of itself by utilizing automatic differentiation. The model is found to be sufficiently accurate for use in retrieval applications.

Table of Contents

Acknowledgements

I want to thank Mathias for giving me the freedom to take this research wherever I wanted it to go. Even though I reinvented the wheel a couple of times along the way and many of my ideas did not work out as I hoped, this freedom allowed me to learn a lot.

I am grateful to Reto, Szabo and Christian for providing the data I needed for my research.

Thank you to the developers of Python, numpy, the ridiculously awesome pandas and everyone answering questions on Stack Overflow. Programming and data analysis are much less frustrating because of you.

Thank you to everyone involved in the making of 30 Rock. You were the cause of many delays but kept me well entertained during the writing of this thesis.

Anna and Alex, thank you for enduring my hour-long rants about programming languages and for our fun and interesting discussions about science, life and TV.

Words can barely express my appreciation for everything my parents have done for me since I was born. Mom and Dad, you have always supported me and helped me become the person I am today. There is nothing greater I could have wished for.

Preface

This thesis was originally about the implementation of a neural network technique for the retrieval of temperature and humidity profiles but in the end it turned into an implementation of the optimal estimation method. This change of mind occurred during the initial literature review phase where I began to understand the limitations of regression models and started gaining more appreciation for the Bayesian approach that optimal estimation is based on. I have therefore chosen to present the theoretical work of this thesis, including the derivation of linear regression, from a Bayesian standpoint.

The implementation of a radiative transfer model was not part of the original plan for this work but the desire for an easy to use model arose after some frustrations with the rather antiquated user interface of an existing model. The simple but powerful idea of automatic differentiation had fascinated me since I first read about it in the context of neural network training, so I decided to use it in my model. Judging from the short presentation of radiative transfer in most publications, I naively assumed that the implementation of a numerical model would be a rather simple task and in terms of code it really is not very complicated. But I soon found that the devil is in the details. Getting the model right took a substantial amount of time, much more than I would have needed to set up an existing solution, but I learned a lot from the experience and that justifies the effort for me personally.

This actually seems to be the theme of this work: reinventing the wheel in order to get to know all aspects of the retrieval problem in depth. While this is not best scientific practice, I think a Master's thesis provides a good opportunity to take such an approach without wasting much of anyone's time (except possibly mine).

There is much more literature on retrievals and radiative transfer than referenced here. The plentiful citations in the text should provide a decent point of origin for anyone trying to find more information. A guiding influence that helped me orient myself in the field of microwave radiometer retrievals was the PhD thesis of Tim Hewison (2006) in which he describes the setup of an optimal estimation scheme for the retrieval of temperature and humidity from ground-based microwave radiometer observations with great detail.

Introduction

Information about the vertical state of the atmosphere is of high value in many areas of meteorology. Vertical profiles of temperature and humidity are used to determine atmospheric stability and the state of the boundary layer, have a history of use in day-to-day weather forecasting and are routinely assimilated into numerical weather prediction models. In-situ balloon soundings (radiosoundings), aircraft measurements and remote sensing data from radar or space-borne instruments are the main sources of information about the vertical state of the atmosphere today (e.g. Brousseau et al. 2014).

The microwave radiometer, a passive remote sensing instrument, has been used for decades in research to derive vertical thermodynamic information of the troposphere and is increasingly tested in operational scenarios where such information is useful. Ground-based microwave radiometers are available commercially and projects specialized on these instruments like the Atmospheric Radiation Measurement Program (Cadeddu et al. 2013) and MWRnet (http://cetemps.aquila.infn.it/mwrnet/) have emerged. The improvement of methods for the retrieval of vertical atmospheric profiles is a topic of current research. This thesis contributes to such efforts at the University of Innsbruck.

1 Properties and Applications of Microwave Radiometers

Measurements from a ground-based microwave radiometer (MWR) can be used to determine vertical profiles of temperature, humidity and cloud liquid water content (Westwater et al. 2004). In practice it has been found that temperature retrievals generally achieve higher accuracy than humidity retrievals, partially due to the properties of microwave radiative transfer in the atmosphere which is governed by absorption and emission of oxygen, water vapor and liquid water droplets (Löhnert et al. 2004; Xu et al. 2014). The absorption of microwave radiation by cloud droplets is sufficiently weak to allow penetration of the cloud and it is therefore possible to perform radiometer measurements even during cloudy skies, giving MWRs an advantage over infrared radiometers which are used for similar purposes but cannot see through clouds (Westwater et al. 2004; Kadygrov et al. 2015). The effect of ice particles on microwave radiation is negligible, but scattering caused by precipitation is strong and causes retrievals to lose accuracy.

A radiometer is a passive remote sensing instrument. It can therefore safely be used in settlements (Kadygrov et al. 2015) and generally requires less maintenance and than active remote sensors (Güldner and Spänkuch 2001). MWRs are much cheaper to operate than regular launches of balloon-borne in-situ instruments and provide a much higher temporal resolution. A disadvantage of being a passive instrument is the need for regular calibration (Westwater et al. 2004; Kadygrov et al. 2015). Common calibration methods for microwave radiometers are the tipping curve method and calibration with an

internal black-body reference target. Both methods only require a rotatable antenna, can be automatically carried out by the instrument and are typically complemented by less frequent manual calibrations with a liquid nitrogen target. The calibration procedure of MWRs overall is well understood (Cimini et al. 2004) and does not affect suitability for deployment in a long-term unattended mode (Westwater et al. 2004; Cimini et al. 2006).

It is instructive to compare the capabilities of various instruments commonly used to obtain the same atmospheric variables retrievable from a ground-based MWR. Radiosondes are usually treated as a reference measurement of the thermodynamic atmospheric state due to their high sensor accuracy and fine vertical resolution. However, balloons experience horizontal drift during their ascent and need a substantial amount of time to reach the tropopause which negatively affects their representativity for a given location. The relatively high costs of sensors and balloons causes radiosonde launch sites to be distributed only sparsely and launches typically occur only one to four times a day. This temporal resolution is insufficient to appropriately capture the daily boundary layer evolution. Aircraft soundings are available with much better temporal resolution but only where airports exist and they have similar representativity issues as radiosondes due to the horizontal displacement of the aircraft path during ascent and descent. Remote sensing instruments provide volume averaged profiles which have a better representativity of the atmospheric state at the cost of vertical resolution (Westwater et al. 2004). Remote sensors can achieve high temporal resolution depending on the need for averaging to reduce instrument noise. Instruments on a satellite in a polar orbit have global coverage but pass a fixed location twice only after long intervals and temperature and humidity sensors on satellites provide only coarse vertical resolution in the boundary layer (Peckham and Grippa 2000; Sánchez et al. 2013). The resolution of a ground-based microwave radiometer is highest in the boundary layer and decreases with altitude (Cadeddu et al. 2002; Löhnert et al. 2004). Low-level resolution can additionally be increased by scanning in off-zenith directions if the atmosphere is weakly inhomogeneous in the horizontal (Cimini et al. 2006; Crewell and Löhnert 2007; Güldner 2013). Together with their high temporal resolution this makes ground-based MWRs well suited for boundary layer research (Cimini et al. 2006). The spatial coverage of a network of ground-based microwave radiometers obviously cannot compete with the coverage of a satellite in a polar orbit but the instruments are mobile, inexpensive and relatively compact, allowing a wide range of applications (Kadygrov et al. 2015).

There are attempts to counter the lack of vertical information in the boundary layer with numerical weather prediction (NWP) models but so far mesoscale models alone have not been able to fill this data gap (Sánchez et al. 2013). In the contrary, the understanding of boundary layer processes is essential for the development of accurate convective-scale models (Martinet et al. 2015). Current research is concerned with the assimilation of observations from ground-based microwave radiometers into NWP models which is not being done operationally yet (Löhnert and Maier 2012). An experiment with assimilation

of temperature and humidity profiles retrieved from 13 MWRs by Cimini et al. (2014) resulted in a neutral-to-positive impact on forecast skills. Recently a numerical radiative transfer model has been developed by De Angelis et al. (2016) with the goal of direct assimilation of radiometer observations.

Not only numerical weather prediction benefits from additional information about the vertical state of the atmosphere, but also other forms of forecasts. Most prominent are short-term severe weather forecasts which depend on accurate vertical profiles of temperature, humidity and wind (Löhnert and Maier 2012). Madhulatha et al. (2013) and Cimini et al. (2015) used radiometer observations to derive indices of severe convective activity and fog. Chan and Lee (2013) analyzed the fluctuations of brightness temperatures measured by a MWR to assess low-level wind shear, crucial for the safe operation of an airport. Such information is also valuable for the safety of nuclear power plants and the assessment of local pollution propagation (Westwater 1972; Kadygrov et al. 2015).

A few notable publications shall be mentioned to conclude this section. A review of remote sensing by ground-based sensors in the microwave and millimeter-wave region was given by Westwater et al. (2004). Aside from the physical basics and aspects of instrument construction, the authors also discuss derivable thermodynamic variables and give an overview of retrieval techniques. An extensive and general review of remote sensing of the lower troposphere including a section about radiometers has recently been given by Wulfmeyer et al. (2015). Retrieval techniques are the focus of a yet unpublished article by Turner et al. (2013)[1]. The authors show example applications and discuss recent developments.

2 Prior Work in Innsbruck

The Institute of Atmospheric and Cryospheric Sciences (ACINN) at the University of Innsbruck operates a passive microwave radiometer as part of the Innsbruck Box project, whose objective is to study boundary layer dynamics in complex terrain. Consequently, there have been efforts to retrieve temperature and humidity profiles from brightness temperature measurements of the radiometer. Research in Innsbruck has been carried out by Massaro (2013) in his master's thesis and a subsequent publication (Massaro et al. 2015), followed by a contribution earlier this year by Meyer (2016).

Retrievals for the Innsbruck site have been done so far with linear regression techniques and a focus on retrieval performance during surface-based or elevated temperature inversions. The inclusion of regressors other than brightness temperatures in the retrieval process has been extensively studied. The desire to include additional information stems from Innsbruck's location in an Alpine valley, surrounded by mountains of over 2000 m

[1] ...which can be obtained from http://www.nssl.noaa.gov/users/dturner/public_html/metr5970/retrieval_uncertainty_paper.v4.pdf

height. This complex terrain provides a challenging environment for radiometer retrievals as the boundary layer evolution is strongly affected by the topography. But the location in the mountains also provides data not available in flat regions: weather stations exist at different heights in the mountains, providing in-situ profiles of the inner-Alpine atmosphere. Even though these measurements are affected by surface layer effects they contain valuable information about the state of the lower troposphere. Regression techniques allow easy integration of such information into a retrieval.

Massaro et al. (2015) found that the linear regression retrieval performance in Innsbruck is comparable to that of radiometers in flat regions. They found significant improvement in the accuracy of the temperature profile at heights where in-situ information is included and for specialized retrievals, categorizing the atmospheric state for example by stability and applying specifically trained regression models. The accurate retrieval of elevated temperature inversions remained a challenge. Meyer (2016) expanded on these results, taking into account a greater number of surface observations from various heights and distances from the radiometer. The inclusion of data from these stations combined with specialized retrievals allowed a better retrieval of elevated temperature inversions.

3 Objectives and Outline

In this thesis the retrieval problem is approached from a Bayesian perspective. One of the major benefits of the Bayesian approach is the assessment of uncertainty of a retrieved profile and the possibility to explicitly integrate a radiative transfer model into the retrieval.

Considering the good results of prior work in Innsbruck with linear regression models, there seems to be little incentive to fundamentally change the retrieval method. But there are aspects of regression techniques that are disadvantageous. They are a purely statistical method and any knowledge of the underlying physics of the retrieval problem is only considered implicitly. Deriving error estimates for retrieved profiles from a regression model is hard. The physically reasonable behavior of retrieved profiles is a consequence of a representative climatology but there is always the possibility of a situation appearing far away from the climatological range, leading to poor regression performance. Innsbruck for example has the problem that radiosondes are only launched at night, which likely leads to worse retrieval performance during the day (without a reasonably sized test dataset of daytime radiosondes this is not easy to verify). Other locations might not even have a training set of radiosonde profiles available for training.

The optimal estimation technique improves on such shortcomings of regression. Uncertainty assessment is natural to all Bayesian methods since distributions are propagated through the retrieval instead of just values. Physical knowledge from a radiative transfer model is explicitly incorporated, allowing accurate retrievals in situations not covered by a climatology. Observations from other sensors can be integrated into the method (see e.g.

Löhnert et al. 2004 who combined information from 5 sources with an optimal estimation method). However, computational costs of such retrievals are higher and any data used in the retrieval must have well characterized errors and known systematic biases. This makes the setup more complicated than that of a regression method.

The goal of this thesis is to provide an implementation of optimal estimation for retrievals of vertical temperature and humidity profiles, paving the way to operational retrievals useful for boundary layer research and weather forecasting in Innsbruck. The theoretical background of the method and radiative transfer is shown here in order to build a deep understanding of the method and its assumptions. Results of the implementation are shown and compared to retrievals by a regression model.

Retrieval Techniques

The retrieval problem of atmospheric remote sensing can be stated as: given a measurement of a sounding device, derive some part of the atmospheric state, for example a vertical profile of temperature. The inherent difficulty of this procedure is that the required inversion of radiative transfer is an ill-posed problem due to the complex nature of electromagnetic wave propagation in a medium like air. When a passive remote sensing instrument observes atmospheric radiation, its measurement is the result of emission and extinction by many layers of the atmosphere. In general, multiple atmospheric states exist that result in very similar observations by the instrument and additional information is necessary to constrain the state space such that a unique solution, ideally the true atmospheric state, exists. These constraints can be a climatology, a numerical forecast or measurements from other instruments.

Depending on the available information and resources, different retrieval techniques for the inversion are applicable. A variety of statistical regression models exists which try to directly describe the inversion problem. These models approximate the atmospheric state based on an observation by a remote sensing instrument after being trained on a data set representative of the relationship which the model tries to imitate. Regression methods are generally easy to set up if such a data set is available, but they are prone to large errors in situations not adequately represented in the training data. Other retrieval methods use a numerical model of the forward problem to iteratively determine a solution consistent with given constraints on the atmospheric state and physical processes as simulated by the forward model. These methods are called optimal or physical methods.

In this chapter, an optimal retrieval technique and the linear regression method are derived from a purely Bayesian approach to statistics. Although all following results have been given before, their derivations are repeated here in order to achieve consistent notation and discuss the benefits of operating in a purely Bayesian framework. The mathematics are based on the work of Rodgers (2000), who treats many aspects of the retrieval problem of atmospheric sounding, partially in the Bayesian framework. His results are used for the derivation of the optimal estimation technique. Fundamental results of Bayesian statistics, properties of the Gaussian distribution and the Bayesian approach to linear regression are taken from Bishop (2006). The introduction to Bayesian statistics is additionally influenced by Downey (2013).

4 Bayesian Statistics

A probability always refers to some entity, which may be a hypothesis, an event or the value of some physical variable. By convention, probability is scaled to the interval $[0, 1]$ representing a continuum of certainty about the state of the associated entity. For the

example of a hypothesis, a probability of 0 stands for the absolute certainty that the hypothesis is false, 1 stands for the absolute certainty that it is true and values between 0 and 1 express certainties in between these extremes.

Every probability has an associated probability density function, also called probability distribution, probability mass function or just PDF. Let $x \in U$ be a continuous variable. Its probability distribution is a function $U \to \mathbb{R}$ which is non-negative almost everywhere. By definition, integration of the PDF over a subset of U yields the probability that x is an element of that subset. Therefore all probability distributions must fulfil

$$\int_U p(x)\,\mathrm{d}x = 1\ , \tag{1}$$

i.e. the probability that x has any value of all possible values must be 1.

A fundamental concept in Bayesian statistics is that of joint and conditional probability. A joint probability is the probability that two variables are each elements of specified subsets of their respective domains at the same time. A conditional probability is the probability of one variable being an element of a given set when it is known that the second variable is an element of another given set. Let x, y be two variables of a statistical scenario. The PDFs associated with joint and conditional probabilities of x and y are written as $p(x, y)$ (read x and y) and $p(x\,|\,y)$ (read x given y) or $p(y|x)$, respectively. It is possible to extract the probability distribution of a variable from a joint distribution by integrating over all possible values of the other variables. For example:

$$p(y) = \int_U p(x, y)\,\mathrm{d}x$$

if $x \in U$. This process is called marginalization.

The fundamental theorem of Bayesian statistics (aptly named 'Bayes' theorem'), can be derived from the product rule of probability (Bishop 2006) which reads

$$p(x, y) = p(x\,|\,y)\,p(y)\ ,$$

i.e. the joint probability distribution of x and y is the product of the conditional probability distribution x given y and the PDF of y independent of x. Because x and y are interchangeable, symmetry dictates that

$$p(x, y) = p(y|x)\,p(x)$$

is valid too. Equating both expressions of the joint PDF and rearranging yields Bayes' theorem:

$$p(x\,|\,y) = \frac{p(y|x)\,p(x)}{p(y)} \sim p(y|x)\,p(x)\ . \tag{2}$$

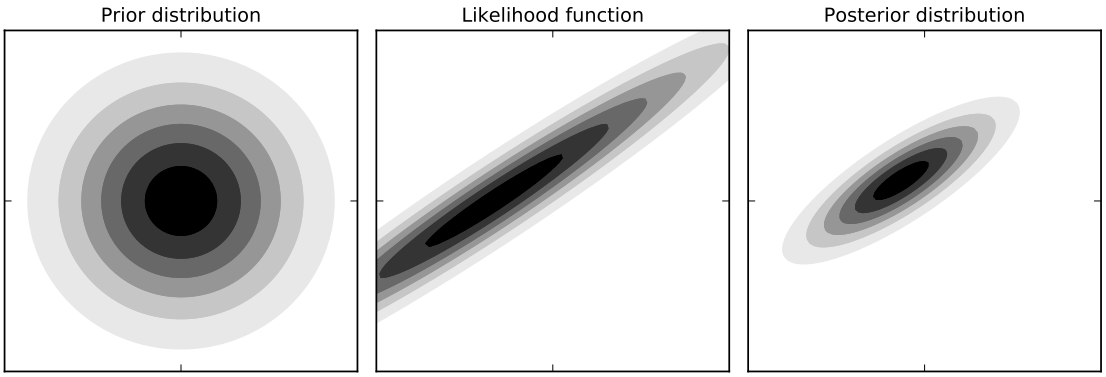

Figure 1: Visualization of a Bayesian update. A circular prior PDF is combined with an elliptical likelihood function according to Bayes' theorem, resulting in a posterior distribution combining the knowledge contained in both. Combining both information on the state of the considered variable has significantly constrained the high-probability state space (shaded) of the posterior relative to the prior.

Because all probability distributions fulfill (1), the denominator acts merely as a normalization and often does not need to be determined explicitly. Bayes' theorem allows to change the conditioning of a probability distribution from $p(y|x)$ to $p(x|y)$ which otherwise may be hard to compute directly. For retrieval applications it is useful to consider the diachronic interpretation of Bayes' theorem in which (2) is viewed as an update of the distribution of x in light of some data y (Downey 2013). At first only $p(x)$ is known, containing a-priori knowledge of x. This PDF is called a prior. Then, some new data y is observed and $p(y|x)$ expresses the likelihood of this observation under consideration of the prior knowledge of x. Multiplying prior and likelihood and normalizing the result with $p(y)$ yields the so called posterior distribution $p(x|y)$, which is the now updated distribution of x after having observed y. By applying Bayes' theorem, the prior knowledge of x is combined with new information contained in the distribution of y. This process is visualized in Figure 1 for two two-dimensionally distributed variables.

The Bayesian framework is valuable because it deals with entire probability distributions instead of plain values. Quantifying the uncertainty of a value or a fact by such distributions enables a user to make better decisions and obtain an accurate description of the state of some entity. A probability distribution also allows sampling, which is useful for the generation of synthetic data based on the distribution and for visualization of error margins of higher-dimensional variables. However, having to deal with probability distributions complicates calculations and analytical solutions to specific problems do not always exist. Such problems can be treated with discrete representations of the involved PDFs and with the use of Monte Carlo methods. Downey (2013) shows that these methods can work well, but they suffer from the high dimensionality of many problems which causes them to be too computationally expensive to be practical. Another way to deal with complicated problems in the Bayesian framework is by settling for a solution that

is just values but still based on the underlying distribution of the result. For example, determination of the maximum of a probability distribution is often possible with analytic methods even for complicated PDFs. Depending on which distribution the maximum solution is derived for, it is called the maximum likelihood (ML) or maximum a-posteriori (MAP) solution.

Another issue with the need for probability distributions in Bayesian statistics arises when no data are at hand with which to decide how the distribution of some variable should look like. At that point, decisions become subjective and distributions are chosen based on mathematical convenience or personal experience. Allowing subjectivity in a scientific method initially sounds like a bad idea, but here it is argued that the way this issue is treated in the Bayesian framework is favorable over the alternatives. The reason is not that subjectivity itself is good, but that it is unavoidable. Most alternative methods starting from a non-Bayesian approach make the same subjective assumptions as a substitute Bayesian technique, only that the assumptions are often made implicitly. This can be seen in the case of linear regression, which will be derived from a Bayesian approach in section 7. The assumption of Gaussian-distributed target variables and regression parameters leads to the same solution in the Bayesian framework as the classical approach of minimizing the sum of squared residuals with an L^2 regularization term. The equivalence of solutions implies that both methods rely on the same assumptions. In the Bayesian framework they are explicit and exposed and the extension to different assumptions is obvious. The classical approach hides these assumptions behind a sum of squares whose connection to the Gaussian distribution is not directly obvious. If a user is explicitly aware of the assumptions on which a model is built, he/she can include this information into the interpretation of results and assessments of applicability of the model to various scenarios, which clearly is an advantage.

5 The Multivariate Gaussian Distribution

For theoretical derivations in the Bayesian framework it is desirable to work with probability distributions for which analytic solutions to common problems exist. The most prominent example of such a distribution is the Gaussian distribution

$$\mathrm{N}(\mathbf{x}\,|\,\boldsymbol{\mu},\boldsymbol{\Sigma}) \;=\; \frac{1}{(2\pi)^{d/2}\det(\boldsymbol{\Sigma})^{1/2}}\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\top}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right) \tag{3}$$

with $\mathbf{x}, \boldsymbol{\mu} \in \mathbb{R}^d$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$. $\boldsymbol{\mu}$ is the mean of this distribution and it is also the location of the only extremum. This is particularly useful for the determination of $\boldsymbol{\mu}$ from a given Gaussian since a simple extremum search directly yields the unique maximum. The symmetric and positive definite covariance matrix $\boldsymbol{\Sigma}$ governs the shape of the distribution. If all off-diagonal elements of $\boldsymbol{\Sigma}$ are zero, all components of the multidimensional distribution are independent. If non-zero off-diagonal elements exist,

some or all components are correlated. The two-dimensional example distributions shown in Figure 1 are all Gaussian distributions. The prior has a diagonal covariance matrix and is therefore circular. The likelihood function has non-zero off-diagonal elements resulting in tilted, ellipsoidal isolines.

A Gaussian distribution is characterized by its mean and covariance. In particular, the normalization $((2\pi)^{d/2}\det(\mathbf{\Sigma})^{1/2})^{-1}$ can be determined from $\mathbf{\Sigma}$ alone. This simplifies derivations where it is known that the result is a Gaussian distribution, because the normalization can be inferred directly once $\mathbf{\Sigma}$ is known. Bayesian operations benefit from this simplification since the marginalization of a Gaussian yields a Gaussian and a conditional distribution of two Gaussian variables is again Gaussian if the conditioning is linear (Bishop 2006).

With the normalization out of the way it is furthermore practical to operate with Gaussian distributions in logarithmic space. There, (3) is given by the quadratic form

$$-2\ln(\mathrm{N}(\mathbf{x}\,|\,\boldsymbol{\mu},\mathbf{\Sigma})) \;=\; (\mathbf{x}-\boldsymbol{\mu})^{\top}\mathbf{\Sigma}(\mathbf{x}-\boldsymbol{\mu}) + c$$

where $c \in \mathbb{R}$ is a constant that contains the normalization. This form still contains all information from (3) and is used extensively in subsequent derivations.

Throughout this thesis it will be assumed again and again that the values of a continuous variable can be described by a Gaussian distribution. Its convenient analytic properties are a major reason for this choice but there are other arguments supporting this assumption. From theoretical considerations it is known that the Gaussian distribution is the best distribution that can be chosen in terms of information content when only the mean and variance of a variable are known because it comes with the least assumptions about the unknown properties of the variable (Rodgers 2000). Furthermore, due to the central limit theorem, the sum of a set of random variables has a distribution approaching a Gaussian when the number of terms in the sum increases (Bishop 2006). Physical measurements are often time averages of observations of higher frequency. The central limit theorem causes such averages to have error distributions that are to good approximation Gaussian.

5.1 Parameter Estimation

To approximate a data set with the a Gaussian distribution, the mean and covariance parameters have to be estimated from the data. For a given set of data $\{\mathbf{x}_1, ..., \mathbf{x}_N\}$, whose elements are assumed to be independent and identically distributed, appropriate values are found in the maximum likelihood estimators of $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$. The respective expressions are repeated here from the book of Bishop (2006) without derivations.

The likelihood of a single data point $\mathbf{x}_i$ given the mean $\boldsymbol{\mu}$ and covariance $\mathbf{\Sigma}$ of the distribution it is assumed to be drawn from is $\mathrm{N}(\mathbf{x}_i\,|\,\boldsymbol{\mu},\mathbf{\Sigma})$. Because the data are assumed to be

independent and identically distributed, the likelihood of the entire data set $\{\mathbf{x}_1, ..., \mathbf{x}_N\}$ given $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is

$$p(\{\mathbf{x}_1, ..., \mathbf{x}_N\} \,|\, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \;=\; \prod_{i=1}^{N} \mathrm{N}(\mathbf{x}_i \,|\, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \;,$$

i.e. the product of all individual likelihood functions. This product of Gaussians is again Gaussian and its mean can be found at its maximum or equivalently at the minimum of its negative natural logarithm. Determination of the covariance matrix is more involved as it has to be symmetric and positive definite. The resulting ML estimates are

$$\boldsymbol{\mu}_{ML} \;=\; \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i$$

and

$$\boldsymbol{\Sigma}_{ML} \;=\; \frac{1}{N} \sum_{i=1}^{N} (\mathbf{x}_i - \boldsymbol{\mu}_{ML})^\top (\mathbf{x}_i - \boldsymbol{\mu}_{ML}) \;.$$

Notably, the mean is calculated independently from the covariance. The ML estimator of the covariance can be proven to systematically underestimate the true covariance. The correct estimator is

$$\tilde{\boldsymbol{\Sigma}}_{ML} \;=\; \frac{1}{N-1} \sum_{i=1}^{N} (\mathbf{x}_i - \boldsymbol{\mu}_{ML})^\top (\mathbf{x}_i - \boldsymbol{\mu}_{ML}) \;.$$

5.2 Bayesian Operations for Gaussians

Two important results for Gaussian distributions are derived in this section which are subsequently used in the optimal estimation and linear regression methods. The derivations are shortened by taking the facts that a product of two Gaussians and the marginalization of a Gaussian are again Gaussian for granted. Rigorous derivations of these results are given by Bishop (2006).

The goal of this section is to find the missing probability distributions $p(\mathbf{x}|\mathbf{y})$ and $p(\mathbf{y})$ in Bayes' theorem when the prior $p(\mathbf{x})$ and likelihood function $p(\mathbf{y}|\mathbf{x})$ are given by

$$p(\mathbf{x}) \;=\; \mathrm{N}(\mathbf{x} \,|\, \mathbf{a}, \mathbf{P}^{-1}) \;, \tag{4a}$$

$$p(\mathbf{y} \,|\, \mathbf{x}) \;=\; \mathrm{N}(\mathbf{y} \,|\, \mathbf{B}\mathbf{x} + \mathbf{b}, \mathbf{Q}^{-1}) \tag{4b}$$

respectively. The mean of the conditional likelihood (4b) is a linear function of $\mathbf{x}$ but the covariance is independent of $\mathbf{x}$. To reduce clutter in the logarithmic forms of these distributions, their covariances have been specified as the inverse of matrices $\mathbf{P}$ and $\mathbf{Q}$:

$$-2\ln(p(\mathbf{x})) = (\mathbf{x}-\mathbf{a})^\top\mathbf{P}(\mathbf{x}-\mathbf{a}) + \text{const} \,, \tag{5a}$$

$$-2\ln(p(\mathbf{y}|\mathbf{x})) = (\mathbf{y}-\mathbf{B}\mathbf{x}-\mathbf{b})^\top\mathbf{Q}(\mathbf{y}-\mathbf{B}\mathbf{x}-\mathbf{b}) + \text{const} \,. \tag{5b}$$

Terms denoted 'const' are containers for any constant expressions independent of $\mathbf{x}$ and will be repeatedly used as such in following calculations. Consider Bayes' theorem in logarithmic space

$$\begin{aligned}
-2\ln(p(\mathbf{x}|\mathbf{y})) &= -2\ln\left(\frac{p(\mathbf{x},\mathbf{y})}{p(\mathbf{y})}\right) \\
&= -2\ln\left(\frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}\right) \\
&= -2\ln(p(\mathbf{y}|\mathbf{x})) - 2\ln(p(\mathbf{x})) + \text{const} \\
&= (\mathbf{y}-\mathbf{B}\mathbf{x}-\mathbf{b})^\top\mathbf{Q}(\mathbf{y}-\mathbf{B}\mathbf{x}-\mathbf{b}) + (\mathbf{x}-\mathbf{a})^\top\mathbf{P}(\mathbf{x}-\mathbf{a}) + \text{const} \\
&= \mathbf{y}^\top\mathbf{Q}\mathbf{y} - \mathbf{y}^\top\mathbf{Q}\mathbf{B}\mathbf{x} - \mathbf{y}^\top\mathbf{Q}\mathbf{b} - \mathbf{x}^\top\mathbf{B}^\top\mathbf{Q}\mathbf{y} \\
&\quad + \mathbf{x}^\top\mathbf{B}^\top\mathbf{Q}\mathbf{B}\mathbf{x} + \mathbf{x}^\top\mathbf{B}^\top\mathbf{Q}\mathbf{b} - \mathbf{b}^\top\mathbf{Q}\mathbf{y} + \mathbf{b}^\top\mathbf{Q}\mathbf{B}\mathbf{x} \\
&\quad + \mathbf{x}^\top\mathbf{P}\mathbf{x} - \mathbf{x}^\top\mathbf{P}\mathbf{a} - \mathbf{a}^\top\mathbf{P}\mathbf{x} + \text{const}
\end{aligned} \tag{6}$$

where the last step expands the vector-matrix-vector products. The posterior is known to be a Gaussian distribution, therefore an equivalent representation of the form

$$\begin{aligned}
-2\ln(p(\mathbf{x}|\mathbf{y})) &= (\mathbf{x}-\mathbf{c})^\top\mathbf{S}(\mathbf{x}-\mathbf{c}) + \text{const} \\
&= \mathbf{x}^\top\mathbf{S}\mathbf{x} - \mathbf{x}^\top\mathbf{S}\mathbf{c} - \mathbf{c}^\top\mathbf{S}\mathbf{x} + \text{const}
\end{aligned} \tag{7}$$

must exist. Because the equivalence must hold for all $\mathbf{x}$, an expression for $\mathbf{\Sigma}$ is found by matching quadratic terms of $\mathbf{x}$ between both forms (6) and (7):

$$\begin{aligned}
\mathbf{x}^\top\mathbf{S}\mathbf{x} &= \mathbf{x}^\top\mathbf{B}^\top\mathbf{Q}\mathbf{B}\mathbf{x} + \mathbf{x}^\top\mathbf{P}\mathbf{x} \\
&= \mathbf{x}^\top(\mathbf{B}^\top\mathbf{Q}\mathbf{B} + \mathbf{P})\mathbf{x} \\
\Rightarrow \qquad \mathbf{S} &= \mathbf{B}^\top\mathbf{Q}\mathbf{B} + \mathbf{P} \,.
\end{aligned}$$

Matching the linear terms involving $\mathbf{x}^\top$ between both forms (6) and (7) results in an expression for the posterior mean $\mathbf{c}$:

$$\begin{aligned}
-\mathbf{x}^\top\mathbf{S}\mathbf{c} &= -\mathbf{x}^\top\mathbf{B}^\top\mathbf{Q}\mathbf{y} + \mathbf{x}^\top\mathbf{B}^\top\mathbf{Q}\mathbf{b} - \mathbf{x}^\top\mathbf{P}\mathbf{a} \\
&= -\mathbf{x}^\top(\mathbf{B}^\top\mathbf{Q}(\mathbf{y}-\mathbf{b}) + \mathbf{P}\mathbf{a}) \\
\Rightarrow \qquad \mathbf{c} &= \mathbf{S}^{-1}(\mathbf{B}^\top\mathbf{Q}(\mathbf{y}-\mathbf{b}) + \mathbf{P}\mathbf{a}) \,.
\end{aligned}$$

Equating terms linear in $\mathbf{x}$ yields the same result. With $\mathbf{c}$ (mean) and $\mathbf{S}$ (inverse covariance) determined, the posterior

$$p(\mathbf{x}|\mathbf{y}) \;=\; \mathrm{N}(\mathbf{x}|\mathbf{S}^{-1}(\mathbf{B}^\top\mathbf{Q}(\mathbf{y}-\mathbf{b})+\mathbf{P}\mathbf{a}),\mathbf{S}^{-1}) \tag{8}$$

is known. $p(\mathbf{y})$ could now be obtained by

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

a subtraction in linear space, assuming that $p(\mathbf{y})$ is a Gaussian distribution

$$-2\ln(p(\mathbf{y})) \;=\; (\mathbf{y}-\mathbf{d})^\top\mathbf{T}(\mathbf{y}-\mathbf{d})$$

and matching terms between both expressions like is was done for the posterior. Due to the rather complex mean and covariance of $p(\mathbf{x}|\mathbf{y})$ this is tedious. Alternatively, it is possible to integrate the joint distribution with respect to $\mathbf{x}$

$$p(\mathbf{y}) \;=\; \int p(\mathbf{x},\mathbf{y})\,\mathrm{d}\mathbf{x} \;=\; \int p(\mathbf{y}|\mathbf{x})\,p(\mathbf{x})\,\mathrm{d}\mathbf{x}\;,$$

again a quite tedious procedure. Instead, starting from the joint distribution and rewriting it in a blocked form, an intuitive argument is made which can be reinforced by rigorous derivation (given e.g. by Bishop 2006). Let $\mathbf{z}$ be the vector obtained by stacking $\mathbf{x}$ and $\mathbf{y}$

$$\mathbf{z} \;=\; \begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix}\;.$$

The joint distribution $p(\mathbf{z})$ is Gaussian

$$\begin{aligned} p(\mathbf{z}) \;&=\; \mathrm{N}(\mathbf{z}|\boldsymbol{\mu},\mathbf{R}^{-1})\;,\\ -2\ln(p(\mathbf{z})) \;&=\; (\mathbf{z}-\boldsymbol{\mu})^\top\mathbf{R}(\mathbf{z}-\boldsymbol{\mu})+\mathrm{const}\\ &=\; \mathbf{z}^\top\mathbf{R}\mathbf{z}-\mathbf{z}^\top\mathbf{R}\boldsymbol{\mu}-\boldsymbol{\mu}^\top\mathbf{R}\mathbf{z}+\mathrm{const} \end{aligned} \tag{9}$$

and by matching terms with $p(\mathbf{x},\mathbf{y}) \;=\; p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$, which was evaluated previously in (6), $\boldsymbol{\mu}$ and $\mathbf{R}$ can be determined. Again, the quadratic terms are matched to obtain the inverse of the covariance

$$\begin{aligned} &\mathbf{y}^\top\mathbf{Q}\mathbf{y}-\mathbf{y}^\top\mathbf{Q}\mathbf{B}\mathbf{x}-\mathbf{x}^\top\mathbf{B}^\top\mathbf{Q}\mathbf{y}+\mathbf{x}^\top\mathbf{B}^\top\mathbf{Q}\mathbf{B}\mathbf{x}+\mathbf{x}^\top\mathbf{P}\mathbf{x}\\ =\;&\mathbf{z}^\top\begin{bmatrix}\mathbf{P}+\mathbf{B}^\top\mathbf{Q}\mathbf{B} & -\mathbf{B}^\top\mathbf{Q}\\ -\mathbf{Q}\mathbf{B} & \mathbf{Q}\end{bmatrix}\mathbf{z}\\ =\;&\mathbf{z}^\top\mathbf{R}\mathbf{z} \end{aligned} \tag{10}$$

and linear terms are matched to determine the mean (here shown for the transposed terms):

$$
\begin{aligned}
& -\mathbf{y}^\top\mathbf{Q}\mathbf{b} + \mathbf{x}^\top\mathbf{B}^\top\mathbf{Q}\mathbf{b} - \mathbf{x}^\top\mathbf{P}\mathbf{a} \\
= \; & -\mathbf{z}^\top \begin{bmatrix} \mathbf{P}\mathbf{a} - \mathbf{B}^\top\mathbf{Q}\mathbf{b} \\ \mathbf{Q}\mathbf{b} \end{bmatrix} \\
= \; & -\mathbf{z}^\top\mathbf{R}\boldsymbol{\mu} \; .
\end{aligned}
$$

Inversion of the block matrix $\mathbf{R}$ from (10) (e.g. Petersen and Pedersen 2012) yields the covariance matrix

$$
\mathbf{R}^{-1} \; = \; \begin{bmatrix} \mathbf{P}^{-1} & \mathbf{P}^{-1}\mathbf{B}^\top \\ \mathbf{B}\mathbf{P}^{-1} & \mathbf{Q}^{-1} + \mathbf{B}\mathbf{P}^{-1}\mathbf{B}^\top \end{bmatrix} \; . \tag{11}
$$

Knowing (11), the mean is found to be

$$
\boldsymbol{\mu} \; = \; \mathbf{R}^{-1} \begin{bmatrix} \mathbf{P}\mathbf{a} - \mathbf{B}^\top\mathbf{Q}\mathbf{b} \\ \mathbf{Q}\mathbf{b} \end{bmatrix} \; = \; \begin{bmatrix} \mathbf{a} \\ \mathbf{B}\mathbf{a} + \mathbf{b} \end{bmatrix} \; . \tag{12}
$$

Comparison with the probability distribution $p(\mathbf{x}) \; = \; \mathrm{N}(\mathbf{x}\,|\,\mathbf{a}, \mathbf{P}^{-1})$ from (4) reveals that the upper left block of the covariance matrix (11) and the upper block of the mean vector (12) of the joint distribution $p(\mathbf{z})$ correspond directly to the covariance and mean of $p(\mathbf{x})$. This makes intuitive sense: When the distribution is evaluated, these are the blocks that interact with $\mathbf{x}$ exclusively while the off-diagonal blocks of $\mathbf{R}^{-1}$ result in interaction of $\mathbf{x}$ and $\mathbf{y}$. It seems reasonable that the blocks interacting exclusively with $\mathbf{x}$ are the only ones necessary to describe the (marginalized) distribution with respect to $\mathbf{x}$ alone. The same idea leads to the expectation that the lower left block of $\mathbf{R}^{-1}$ and the lower block of $\boldsymbol{\mu}$ are the only ones necessary to describe the (marginalized) distribution with respect to $\mathbf{y}$ only. For symmetry, $p(\mathbf{y})$ should therefore be given by

$$
p(\mathbf{y}) \; = \; \mathrm{N}(\mathbf{y}\,|\,\mathbf{B}\mathbf{a} + \mathbf{b}, \mathbf{Q}^{-1} + \mathbf{B}\mathbf{P}^{-1}\mathbf{B}^\top) \tag{13}
$$

and this is indeed the result found in a rigorous derivation.

6 Optimal Estimation

This retrieval technique incorporates radiometer observations into a prior estimate of the atmospheric state with a Bayesian update. Because this update explicitly makes use of the physics behind radiative transfer as simulated by a numerical model, optimal estimation is often called a physical retrieval technique (e.g. Güldner and Spänkuch 2001; Turner et al. 2007). With knowledge of the uncertainties of the forward model and the prior information which are essential to solve the non-uniqueness problem of retrieval, atmospheric states consistent with these error characteristics are derived by the method (Löhnert et al. 2004). Because uncertainties are propagated in a Bayesian approach, an estimate of the error of the retrieved state is derived as well.

The optimal estimation method is more complex to set up than regression techniques and more expensive in terms of computational time due to the requirements of error estimates and a forward model. However, the computational cost has been offset by increasing hardware capabilities and retrievals with a frequency suitable for real-time operational use are possible with any modern desktop computer. Setup can be simplified by software such as Qpack (Eriksson et al. 2005) which provides not only a forward model (ARTS) but also the routines necessary for the optimal estimation procedure.

Recent research involving optimal estimation methods is often connected to the desire of assimilating retrieved profiles into numerical weather prediction models (Cimini et al. 2014 and references therein). Martinet et al. (2015) employed a two step assimilation procedure: vertical atmospheric profiles are retrieved from radiometer observations first and then assimilated into a weather prediction model in a second step. De Angelis et al. (2016) recently presented a new radiative transfer model called RTTOV-gb, providing the capability to assimilate radiometer observations directly into NWP model fields.

The standard reference for the theory behind optimal estimation schemes is a book by Rodgers (2000). The presentation here also follows this book but is restricted to a purely Bayesian approach. All probability distributions involved in the optimal estimation procedure are assumed to be Gaussian allowing analytical derivations. The result is a retrieval scheme efficient even for atmospheric state vectors of high dimensionality for which discrete approaches are very costly.

All knowledge about the state vector $\mathbf{x}$ independent of a radiometer observation is contained in the prior distribution $p(\mathbf{x})$. This knowledge might originate from climatological records, output from a NWP model, other instruments than the radiometer or any combination of these sources. The construction of an optimal prior is not an easy task and will be discussed in sections 9 and 10. Wherever from, due to the assumption that the prior PDF is Gaussian it is of the form

$$p(\mathbf{x}) \;=\; \mathrm{N}(\mathbf{x}\,|\,\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a)\;, \tag{14}$$

where the subscript 'a' stands for 'a-priori'. The likelihood function $p(\mathbf{y}\,|\,\mathbf{x})$ expresses the probability that an observation $\mathbf{y}$ is the result of some state $\mathbf{x}$. It is quantified with the help of a forward model of the radiative transfer governing the radiometer observation

$$\mathbf{y} \;=\; \mathbf{F}(\mathbf{x}) + \boldsymbol{\epsilon}\;, \tag{15}$$

where $\mathbf{F}$ is the forward model operator and $\boldsymbol{\epsilon}$ is an error term, arising from measurement noise, inaccuracies of the forward model, discretization errors and possibly other terms (see section 18). The distribution of the error term is Gaussian, centered around a systematic bias $\boldsymbol{\mu}_{\boldsymbol{\epsilon}}$ with a covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$

$$p(\boldsymbol{\epsilon}) \;=\; \mathrm{N}(\boldsymbol{\epsilon}\,|\,\boldsymbol{\mu}_{\boldsymbol{\epsilon}}, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}})\;. \tag{16}$$

Substituting (15) into (16) and changing the independent variable of the PDF to $\mathbf{y}$ yields in the likelihood function

$$\begin{aligned} p(\mathbf{y}-\mathbf{F}(\mathbf{x})) &= \mathrm{N}(\mathbf{y}-\mathbf{F}(\mathbf{x})\,|\,\boldsymbol{\mu}_{\boldsymbol{\epsilon}},\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}) \\ &= \mathrm{N}(\mathbf{y}\,|\,\mathbf{F}(\mathbf{x})+\boldsymbol{\mu}_{\boldsymbol{\epsilon}},\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}) \;=\; p(\mathbf{y}\,|\,\mathbf{x})\;. \end{aligned} \tag{17}$$

Knowing (14) and (17), the posterior PDF (which is the result of the retrieval) is

$$p(\mathbf{x}\,|\,\mathbf{y}) \;=\; \frac{p(\mathbf{y}\,|\,\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})} \;=\; \frac{\mathrm{N}(\mathbf{y}\,|\,\mathbf{F}(\mathbf{x})+\boldsymbol{\mu}_{\boldsymbol{\epsilon}},\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}})\;\mathrm{N}(\mathbf{x}\,|\,\boldsymbol{\mu}_a,\boldsymbol{\Sigma}_a)}{p(\mathbf{y})}\;. \tag{18}$$

It expresses the probability of an atmospheric state $\mathbf{x}$ given a measurement $\mathbf{y}$ and incorporates the forward model with its error characteristics and a-priori knowledge of the atmospheric state independent of the observation. All involved distributions are Gaussians, therefore $p(\mathbf{x}\,|\,\mathbf{y})$ is also a Gaussian distribution and the normalization $p(\mathbf{y})$ does not have to be determined explicitly but can be inferred from the product in the numerator.

The results from section 5.2 cannot be used yet with (18), because the forward model operator $\mathbf{F}$ is non-linear which is why the product of $p(\mathbf{y}\,|\,\mathbf{x})$ and $p(\mathbf{x})$ cannot be carried out analytically (in general). The model operator $\mathbf{F}$ has to be linearized first, resulting in an iterative approach.

6.1 Iterative Solutions

To linearize $\mathbf{F}$, a reference state to linearize at must be chosen. Let this state be $\mathbf{x}_i$. Applying a Taylor series expansion to the forward model and chopping off all non-linear terms yields the linear approximation

$$\mathbf{F}(\mathbf{x}) \;\approx\; \mathbf{F}(\mathbf{x}_i)+\mathbf{K}_i(\mathbf{x}-\mathbf{x}_i)\;.$$

$\mathbf{K}_i$ is the Jacobian of $\mathbf{F}$ at $\mathbf{x}_i$ containing the derivatives of the forward model output with respect to each element of the state vector $\mathbf{x}$. After replacing $\mathbf{F}$ in (18) by this linearization, the likelihood function

$$p(\mathbf{y}\,|\,\mathbf{x}) \;=\; \mathrm{N}(\mathbf{y}\,|\,\mathbf{K}_i\mathbf{x}+\mathbf{F}(\mathbf{x}_i)-\mathbf{K}_i\mathbf{x}_i+\boldsymbol{\mu}_{\boldsymbol{\epsilon}},\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}})$$

and the prior (14) fulfill the assumptions of (4) and the result (8) can be used by substituting

$$\{\mathbf{a}\;\rightarrow\;\boldsymbol{\mu}_a,\;\mathbf{P}^{-1}\;\rightarrow\;\boldsymbol{\Sigma}_a,\;\mathbf{B}\;\rightarrow\;\mathbf{K}_i,\;\mathbf{b}\;\rightarrow\;\mathbf{F}(\mathbf{x}_i)-\mathbf{K}_i\mathbf{x}_i+\boldsymbol{\mu}_{\boldsymbol{\epsilon}},\;\mathbf{Q}^{-1}\;\rightarrow\;\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}\}$$

to obtain an expression for the posterior:

$$
\begin{aligned}
p(\mathbf{x}|\mathbf{y}) &= \mathrm{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \ , \\
\boldsymbol{\mu} &= \boldsymbol{\Sigma}(\mathbf{K}_i^\top \boldsymbol{\Sigma}_\boldsymbol{\epsilon}^{-1}(\mathbf{y} - \mathbf{F}(\mathbf{x}_i) + \mathbf{K}_i \mathbf{x}_i - \boldsymbol{\mu}_\boldsymbol{\epsilon}) + \boldsymbol{\Sigma}_a^{-1}\boldsymbol{\mu}_a) \ , \\
\boldsymbol{\Sigma} &= (\mathbf{K}_i \boldsymbol{\Sigma}_\boldsymbol{\epsilon}^{-1} \mathbf{K}_i^\top + \boldsymbol{\Sigma}_a^{-1})^{-1} \ .
\end{aligned}
$$

Due to the linear approximation of $\mathbf{F}$, the solution for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ will not be optimal. Instead, starting from a first guess $\boldsymbol{\mu}_0$ the optimal solution is found by fixed-point iteration, using the mean of the previous iteration as the linearization point in each step

$$
\boldsymbol{\mu}_{i+1} = \boldsymbol{\Sigma}_i(\mathbf{K}_i^\top \boldsymbol{\Sigma}_\boldsymbol{\epsilon}^{-1}(\mathbf{y} - \mathbf{F}(\boldsymbol{\mu}_i) + \mathbf{K}_i \boldsymbol{\mu}_i - \boldsymbol{\mu}_\boldsymbol{\epsilon}) + \boldsymbol{\Sigma}_a^{-1}\boldsymbol{\mu}_a) \ , \tag{19a}
$$

$$
\boldsymbol{\Sigma}_i = (\mathbf{K}_i^\top \boldsymbol{\Sigma}_\boldsymbol{\epsilon}^{-1} \mathbf{K}_i + \boldsymbol{\Sigma}_a^{-1})^{-1} \ . \tag{19b}
$$

In practice more robust iteration schemes are required to obtain good convergence properties. The easiest approach to find such schemes is to replace (19a) by an algorithm searching for the position of the global maximum of (18) which is the same as the mean due to the posterior being Gaussian. This search is best carried out in logarithmic space where the location of the extremum is found at the the root of the gradient of the quadratic form of (18) with respect to $\mathbf{x}$

$$
\begin{aligned}
0 &= \nabla_\mathbf{x}(-2\ln(p(\mathbf{x}|\mathbf{y}))) \\
&= (\mathbf{y} - \mathbf{F}(\mathbf{x}) - \boldsymbol{\mu}_\boldsymbol{\epsilon})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{F}(\mathbf{x}) - \boldsymbol{\mu}_\boldsymbol{\epsilon}) + (\mathbf{x} - \boldsymbol{\mu}_a)^\top \boldsymbol{\Sigma}_a (\mathbf{x} - \boldsymbol{\mu}_\mathbf{x})^\top \ .
\end{aligned} \tag{20}
$$

Usually one settles for a least squares solution to this minimization problem[2] as methods for these problems are simpler to implement (e.g. Newton's method for finding the root of a function requires the Hessian, which is often expensive to calculate). Rodgers (2000) shows resulting iterative schemes for the Gauss-Newton method and a more robust variant, the Levenberg-Marquardt iteration. With the identity

$$
\boldsymbol{\mu}_a - \boldsymbol{\Sigma}_i \mathbf{K}_i^\top \boldsymbol{\Sigma}_\boldsymbol{\epsilon}^{-1} \mathbf{K}_i \boldsymbol{\mu}_a = \boldsymbol{\Sigma}_i(\boldsymbol{\Sigma}_i^{-1} - \mathbf{K}_i^\top \boldsymbol{\Sigma}_\boldsymbol{\epsilon}^{-1} \mathbf{K}_i)\boldsymbol{\mu}_a = \boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_a^{-1} \boldsymbol{\mu}_a \ ,
$$

it is easily seen that (19a), the solution obtained from fixed point iteration with the linearized forward model, is equivalent to equation 5.9 from Rodgers (2000) which he obtained by applying the Gauss-Newton method to (20). This underlines the equivalence of the two approaches to determine the posterior mean.

The flowchart in Figure 2 visualizes the resulting iterative retrieval scheme. Given prior knowledge, a measurement and an initial guess of the state vector, minimization steps with linearizations of the forward model are taken until the state vector has converged. The mean and covariance matrix of the last iteration are the result of the retrieval.

[2] The maximum search is translated to a minimum search after application of the logarithm and multiplication by -2.

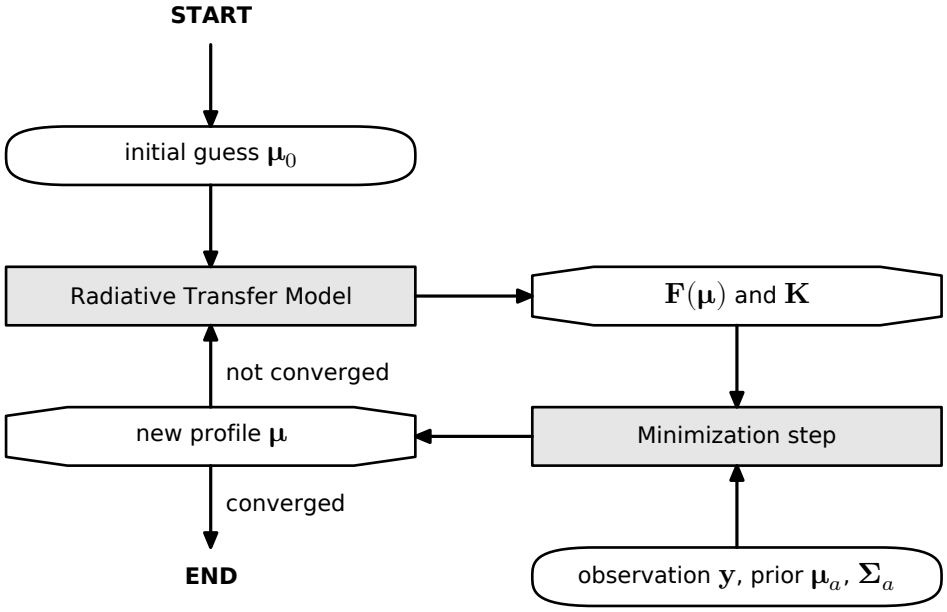

Figure 2: Visualization of the iterative retrieval algorithm. Missing is the forward model error distribution, which is used in the Levenberg-Marquardt step. Square boxes represent numerical calculations, rounded boxes data input and octagonal boxes data being produced during the retrieval.

6.2 Convergence Testing

Considering the magnitude of errors expected from a retrieval, it is unnecessary to iterate until the atmospheric state vector has been determined with machine precision. Therefore, a criterion which indicates if the state vector has been determined with sufficient accuracy is required. Rodgers (2000) presents multiple possibilities which are repeated here.

One possibility is to assess convergence directly with the value of the posterior distribution (18) evaluated in logarithmic space (the cost function):

$$(\mathbf{y} - \mathbf{F}(\boldsymbol{\mu}_i) - \boldsymbol{\mu}_{\boldsymbol{\epsilon}})^\top \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}^{-1} (\mathbf{y} - \mathbf{F}(\boldsymbol{\mu}_i) - \boldsymbol{\mu}_{\boldsymbol{\epsilon}}) + (\boldsymbol{\mu}_i - \boldsymbol{\mu}_a)^\top \boldsymbol{\Sigma}_a^{-1} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_a) + \text{const} \ . \tag{21}$$

The constant term arising from the normalization can safely be ignored for the purposes of convergence detection. The value of the gradient (20) can be incorporated in this criterion too (Hewison 2006). Also used are criteria based on the change between two iteration steps measured in observation or state space. It is important to scale such measures to account for the naturally different magnitudes of the elements of the observation or state vector. The error covariance matrices of measured vs. simulated brightness temperatures,

$$\boldsymbol{\Sigma}_{\delta y}^{-1} \ = \ \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} (\mathbf{K}_i \boldsymbol{\Sigma}_a \mathbf{K}_i^\top)^{-1} \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} \ ,$$

(Rodgers 2000) and of the state vector, $\mathbf{\Sigma}_{i+1}$ given by (19b), provide appropriate scalings. Because covariance matrices are positive definite the resulting measures are always positive. They are

$$(\mathbf{F}(\mathbf{x}_i) - \mathbf{F}(\mathbf{x}_{i+1}))^\top \mathbf{\Sigma}_{\delta y}^{-1} (\mathbf{F}(\mathbf{x}_i) - \mathbf{F}(\mathbf{x}_{i+1})) \tag{22}$$

in observation space and

$$(\mathbf{x}_i - \mathbf{x}_{i+1})^\top \mathbf{\Sigma}_{i+1}^{-1} (\mathbf{x}_i - \mathbf{x}_{i+1}) \tag{23}$$

in state space. When these terms are sufficiently small, iteration can be stopped and the retrieved atmospheric state returned.

In practice, the issue of convergence is more complicated because not only the choice of criterion affects convergence. The minimization technique, method of forward model linearization and included radiometer channels are all relevant components affecting convergence. The first guess of the iteration procedure also influences the convergence properties. Different strategies can be applied to increase the rate and likelihood of convergence. Cimini et al. (2011) found that retrieving temperature first and then humidity in a second step improved convergence efficiency. Hewison (2006) relaxed the chosen criterion after the first 10 iterations to encourage convergence. Turner and Löhnert (2014) additionally coupled the criterion to a fixed sequence of parameters that controlled their minimization algorithm.

6.3 1D-VAR, Cost Functions and Regularization

There is an alternative way to derive the optimal estimation retrieval scheme. Instead of directly working with probability distributions, the retrieval problem can be set up as the minimization of the expected value of the retrieval error variance. This leads to the same solution as the Bayesian approach when assuming that all variables are Gaussian-distributed. Because of the minimization of a variance in the alternative approach, optimal estimation is also called a 'variational' method, often abbreviated as 1D-VAR, due to its application to a vertical profile representing one dimension of the atmospheric state. A derivation starting from the variational perspective is found in chapter 4.2 of the book by Rodgers (2000).

The quadratic form $-2\ln(p(\mathbf{x}|\mathbf{y}))$ minimized during the iteration is often referred to as a cost function, terminology that is also common in regression problems. So-called regularization or penalty terms can be added to this cost function in order to make the retrieval favor physically realistic solutions. In atmospheric profile retrieval such a penalty might be a term involving the lapse rate (Peckham and Grippa 2000). Hewison (2006) discussed penalizing superadiabatic layers and also considered a regularization of excessively high liquid water content in clouds.

There is a problem with including information in the retrieval in the form of regularization terms: while the information is included in the retrieved mean profile, it is absent from its uncertainty estimate given by $\mathbf{\Sigma}_i$. This issue can be resolved with the Bayesian perspective. Viewing the cost function as the combination of the likelihood function with prior distribution, it appears that each additional penalty term corresponds to another prior distribution involved in the Bayesian update. Quadratic penalty terms are associated with Gaussian distributions. For these, an analytic form of the uncertainty can be derived. This is further discussed in section 10. Other types of distributions may not permit an analytic form of the posterior.

Regularization is a powerful tool for the inclusion of additional knowledge but penalties more complex than quadratic forms generally violate the Gaussian assumptions leading to inaccurate uncertainty estimates. For many end users a more accurately retrieved mean profile is more valuable than a theoretically consistent uncertainty assessment and such errors are tolerable. Other ways to include additional knowledge of the atmospheric state are discussed in section 10.

7 Linear Regression

Linear regression is a simple and computationally cheap retrieval method. It uses an approximate model of the inversion process of radiative transfer to predict the atmospheric state based on a radiometer observation. This model is fitted to a representative set of training data, for example historic data from radiosonde ascents paired with collocated radiometer measurements. Predictions of a regression model do not require the evaluation of a forward model during retrieval and are therefore quickly evaluated. The computational effort is concentrated in the training phase where the model parameters are determined. When generating a linear regression model, three aspects affect its subsequent performance: the choice of basis functions, the quality of the training data set and the measures taken to avoid overfitting.

The linear regression model consists of a set of basis functions dependent on the input variables (predictors) which are combined linearly to make a prediction of the target variables. Applied to the retrieval problem, the observed brightness temperatures are the model's input and an atmospheric profile of temperature, humidity or liquid water is the output. A linear regression model can approximate a non-linear relationship if non-linear basis functions are chosen. The 'linear' in its name refers to the linearity in the model coefficients acting as weights with which the evaluated basis functions are combined. In retrieval applications it is common to use a basis directly corresponding to individual brightness temperatures. Crewell and Löhnert (2007) and Massaro et al. (2015) also included quadratic terms to better approximate the non-linear nature of the inversion problem. Because the coefficients affect the model only linearly they can be determined from a training data set in a single fitting step without the need for iteration.

In order to obtain a model that provides good predictions in many situations, a large training data set is necessary that represents the relationship which is to be approximated well. Regression models in general are only as good as the data they are trained with and extrapolations to states not included in the training data are prone to errors and have no guarantee of physical reasonability (Cimini et al. 2006; Turner et al. 2007; Chan 2010). This behavior is one of the factors limiting the transferability of a trained regression model between different radiometer sites.

The set of basis functions and the number of training data pairs both affect if a model is in danger of overfitting. A model is overfitted if it performs well for the training data but makes poor predictions for any other reasonable input data. It is common practice to have a test data set independent of the training data for the evaluation of model performance or to use cross validation techniques. Overfitting of a linear regression model can be controlled by regularization. This is preferable to the addition of noise to predictors during training like some authors have done (e.g. Churnside et al. 1994; Del Frate and Schiavon 1998; Löhnert et al. 2004) since it works also with small data sets and is mathematically more elegant and deterministic. In the following derivation of Bayesian linear regression regularization will appear naturally from the prior of the model parameters.

Due to their simplicity, linear regression models are commonly used. Statistical evaluations of linear regression models used for retrievals of temperature and humidity have shown that their accuracy generally decreases with increasing altitude (e.g. Solheim et al. 1998). This is mainly due to the information content of radiometer observations which is lower at greater heights (see section 15). Atmospheric structures predicted by linear regression models are often much smoother than the actual structures and models have been found to be able to resolve at most the lowest temperature inversion of many (Crewell and Löhnert 2007). Studies with other regression models, most notably neural networks, suggest that these aspects might be improved upon by models better able to approximate the non-linearity of the inversion problem (see also section 8 or Churnside et al. 1994).

The classical approach to linear regression is the least squares framework but the Bayesian framework provides an equivalent derivation. It is presented here based on the book by Bishop (2006). Let x be a component of the atmospheric state vector and $\mathbf{y} \in \mathbb{R}^d$ a vector containing a radiometer observation. Consider a model f, a linear combination of n fixed basis functions $\boldsymbol{\phi} = \begin{bmatrix} \phi_1 & ... & \phi_n \end{bmatrix}$, approximating a scalar function t of the input $\mathbf{y}$

$$t(\mathbf{y}) \approx f(\mathbf{y}, \mathbf{w}) = w_0 + \sum_{i=1}^{n} w_i \phi_i(\mathbf{y}) = \sum_{i=0}^{n} w_i \phi_i(\mathbf{y}) = \boldsymbol{\phi}(\mathbf{y})\mathbf{w} \,,$$

where $\mathbf{w} \in \mathbb{R}^n$ and an additional constant basis function $\phi_0(\mathbf{y}) = 1$ is introduced in the third step to account for constant term w_0 without the need to explicitly consider it in

the following derivations. A common choice for basis functions in retrievals are indentity functions for individual components, i.e. $n = d$ and

$$\phi_0(\mathbf{y}) = 1,\ \phi_1(\mathbf{y}) = y_1,\ ..., \phi_d(\mathbf{y}) = y_d \tag{24}$$

with $\mathbf{y}^\top = \begin{bmatrix} y_1 & ... & y_d \end{bmatrix}$, but in general any set of linear and non-linear functions can be chosen. The objective when fitting the regression model is to find a parameter vector $\mathbf{w}$ from a given set of $(t(\mathbf{y}), \mathbf{y})$ pairs that results in a good approximation of t by f. The Bayesian approach provides a distribution of these parameters.

In order to obtain an analytical result for the distribution of $\mathbf{w}$, deviations between the true function t and its approximation f are assumed to follow a Gaussian distribution with zero mean and covariance β^{-1}

$$\begin{aligned} p(t(\mathbf{y}) - f(\mathbf{y}, \mathbf{w}) \,|\, \beta) &= \mathrm{N}(t(\mathbf{y}) - f(\mathbf{y}, \mathbf{w}) \,|\, 0, \beta^{-1}) \\ \Rightarrow \qquad p(x \,|\, \mathbf{y}, \mathbf{w}, \beta) &= \mathrm{N}(x \,|\, \boldsymbol{\phi}(\mathbf{y})\mathbf{w}, \beta^{-1}) \ , \end{aligned} \tag{25}$$

where $x = t(\mathbf{y})$ is the so-called target variable and its dependency on $\mathbf{y}$ is implied by the conditioning of its probability distribution. If data from pairwise corresponding datasets $\mathbb{Y} = \{\mathbf{y}_1, ..., \mathbf{y}_m\}$ and $\mathbb{X} = \{x_1, ..., x_m\}$ are independent, the likelihood function of the target variables is given by

$$p(\mathbb{X} \,|\, \mathbb{Y}, \mathbf{w}, \beta) = \prod_{i=1}^{m} \mathrm{N}(x_i \,|\, \boldsymbol{\phi}(\mathbf{y}_i)\mathbf{w}, \beta^{-1})$$

using (25) or equivalently by

$$p(\mathbb{X} \,|\, \mathbb{Y}, \mathbf{w}, \beta) = \mathrm{N}(\mathbf{x} \,|\, \boldsymbol{\Phi}\mathbf{w}, \beta^{-1}\mathbf{I})$$

with the definitions

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \text{ and } \boldsymbol{\Phi} = \begin{bmatrix} \phi_0(\mathbf{y}_1) & ... & \phi_n(\mathbf{y}_1) \\ \vdots & \ddots & \vdots \\ \phi_0(\mathbf{y}_m) & ... & \phi_n(\mathbf{y}_m) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\phi}(\mathbf{y}_1) \\ \vdots \\ \boldsymbol{\phi}(\mathbf{y}_m) \end{bmatrix} .$$

In order to use Bayes' theorem to obtain the probability distribution of $\mathbf{w}$ given the training data pairs, a prior for the parameter vector is necessary. It is possible to assume that any value of $\mathbf{w}$ is equally likely since no information about $\mathbf{w}$ is available at this point. Experience shows however that it is benefical to make the regression favor values of $\mathbf{w}$ close to $\mathbf{0}$ as this results in a model less prone to overfitting and with better numerically stability. An isotropic Gaussian with mean $\mathbf{0}$ is therefore considered

$$p(\mathbf{w} \,|\, \alpha) = \mathrm{N}(\mathbf{w} \,|\, \mathbf{0}, \alpha^{-1}\mathbf{I}) \tag{26}$$

where $\alpha > 0$ governs the covariance of the prior distribution. With likelihood function and prior determined, Bayes' theorem can be written down:

$$p(\mathbf{w}\,|\,\mathbb{X},\mathbb{Y},\alpha,\beta) \;=\; \frac{p(\mathbb{X}\,|\,\mathbb{Y},\mathbf{w},\beta)\,p(\mathbf{w}\,|\,\alpha)}{p(\mathbb{X}\,|\,\mathbb{Y},\beta)}\;.$$

Application of the result (8) with the substitutions

$$\{\mathbf{x}\;\rightarrow\;\mathbf{w},\;\mathbf{a}\;\rightarrow\;\mathbf{0},\;\mathbf{P}\;\rightarrow\;\alpha\mathbf{I},\;\mathbf{y}\;\rightarrow\;\mathbf{x},\;\mathbf{B}\;\rightarrow\;\mathbf{\Phi},\;\mathbf{b}\;\rightarrow\;\mathbf{0},\;\mathbf{Q}\;\rightarrow\;\beta\mathbf{I}\}$$

identified from (4) yields the desired posterior probability distribution for the parameter vector $\mathbf{w}$:

$$p(\mathbf{w}\,|\,\mathbb{X},\mathbb{Y},\alpha,\beta) \;=\; \mathrm{N}(\mathbf{w}\,|\,\boldsymbol{\mu},\boldsymbol{\Sigma}) \tag{27a}$$

$$\boldsymbol{\Sigma} \;=\; (\beta\mathbf{\Phi}^{\top}\mathbf{\Phi}+\alpha\mathbf{I})^{-1} \tag{27b}$$

$$\boldsymbol{\mu} \;=\; \beta\boldsymbol{\Sigma}\mathbf{\Phi}^{\top}\mathbf{x}\;. \tag{27c}$$

This concludes the training of the regression model. Predictions of a target x based on a new input $\mathbf{y}$ are obtained from the predictive distribution

$$\begin{aligned} p(x\,|\,\mathbf{y},\mathbb{X},\mathbb{Y},\beta,\alpha) \;&=\; \int\limits_{\mathbb{R}^{n+1}} p(x\,|\,\mathbf{y},\mathbf{w},\beta)\,p(\mathbf{w}\,|\,\mathbb{X},\mathbb{Y},\alpha,\beta)\,\mathrm{d}\mathbf{w} \\ &=\; \int\limits_{\mathbb{R}^{n+1}} \mathrm{N}(x\,|\,\boldsymbol{\phi}(\mathbf{y})\mathbf{w},\beta^{-1})\,\mathrm{N}(\mathbf{w}\,|\,\boldsymbol{\mu},\boldsymbol{\Sigma})\,\mathrm{d}\mathbf{w} \end{aligned}$$

which is the likelihood of the measurement given the model (25), marginalized with respect to the posterior parameter vector distribution (27a). Result (13) is applicable[3] with the substitution

$$\{\mathbf{x}\;\rightarrow\;\mathbf{w},\;\mathbf{a}\;\rightarrow\;\boldsymbol{\mu},\;\mathbf{P}\;\rightarrow\;\boldsymbol{\Sigma}^{-1},\;\mathbf{y}\;\rightarrow\;x,\;\mathbf{B}\;\rightarrow\;\boldsymbol{\phi}(\mathbf{y}),\;\mathbf{b}\;\rightarrow\;0,\;\mathbf{Q}\;\rightarrow\;\beta\}$$

yielding

$$p(x\,|\,\mathbf{y},\mathbb{X},\mathbb{Y},\beta,\alpha) \;=\; \mathrm{N}(x\,|\,\boldsymbol{\phi}(\mathbf{y})\boldsymbol{\mu},\beta^{-1}+\boldsymbol{\phi}(\mathbf{y})\boldsymbol{\Sigma}\boldsymbol{\phi}^{\top}(\mathbf{y}))$$

for the predictive distribution. The predicted mean is the sum of products between the basis functions evaluated for the predictors and the the parameter vector's MAP solution. This result is consistent with the expectation of a prediction from a non-Bayesian approach to linear regression. The covariance of the predicted value is a combination of the uncertainty inherent in the target variable and uncertainty in the model parameters.

At first sight, a full description of the predictive distribution seems useful for error estimates of predictions. But care has to be taken in its interpretation because the distribution

[3] The reduction in dimension in (4b) does not invalidate the result. The distribution's mean is still linear.

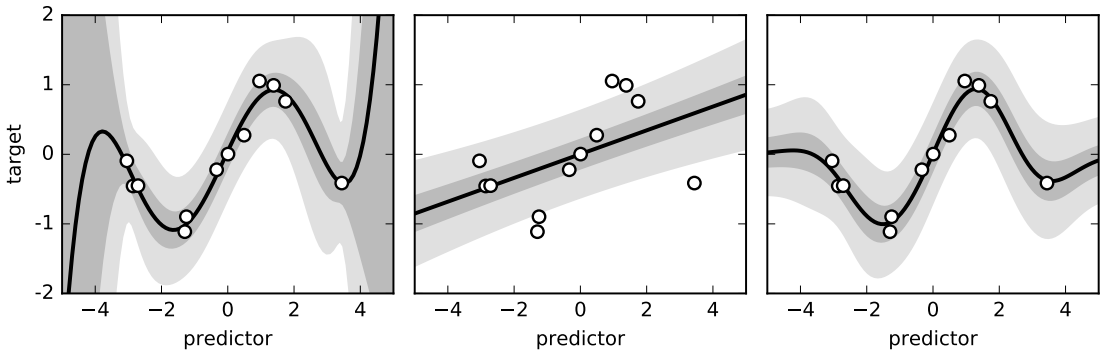

Figure 3: Predictive distributions of Bayesian regression models trained on 12 data points obtained from a sine signal with added Gaussian noise of standard deviation 0.2 (circles). The black line is the mean of the predictive distribution, the gray and light gray shaded areas mark ± one and three standard deviations around this mean. The left and center models are fifth and first order polynomials respectively, the right model uses seven Gaussian basis functions located between -3 and 3.

describes uncertainty of the model parameters but not of the choice of model itself, i.e. the choice of basis functions. Uncertainty obtained from the predictive distribution is just one term in the overall error estimate of predictions and on its own only useful if the chosen model can accurately reproduce the variability of the target data. Figure 3 illustrates how the choice of basis functions affects the predictive distribution with a constructed example in one dimension. The model shown in the left frame is a polynomial of order five and able to describe the variability of the target data well. It gives a good fit inside the data domain and an appropriate uncertainty range outside the domain can be derived from the predictive distribution. The model in the middle is a line with only slope and offset as parameters. Not only is it a bad fit for the data inside the domain but the predictive distribution massively overstates the model's accuracy. For a line, the fit is good and the parameters can be determined with high accuracy, but considering all possible models it is a bad fit because a line cannot express variability of the target data sufficiently. On the right, seven Gaussian basis functions are used in a linear regression model, resulting in a good fit inside the data domain with appropriate uncertainty estimates. However, the error estimate derived from the predictive distribution reduces to the inherent uncertainty of the target data outside the training domain. No contribution to the estimate from the model exists anymore due to the localized bases functions (the covariance is dominated by β^{-1} since $\boldsymbol{\phi}(\mathbf{y})$ vanishes). Like for the line model, this obviously does not represent the true uncertainty of the model predictions.

The treatment of linear regression so far considered multiple predictors but was restricted to a scalar target. For the retrieval of an atmospheric profile from radiometer measurements, each atmospheric layer is a target variable resulting in a one dimensional target vector. The common approach in this situation is to use the same basis functions for each target component and derive a parameter matrix instead of a parameter vector $\mathbf{w}$. During the derivation of this matrix it becomes apparent that the problem decouples into

independent, scalar regression problems for each target component (Bishop 2006). The results from above are therefore directly applicable to multiple targets.

8 Other Retrieval Techniques

The optimal estimation method has been the starting point for integrated approaches, that combine observations from multiple instruments. For example, the integrated profiling technique by Löhnert et al. (2004) applies optimal estimation to combine data from a microwave radiometer, a cloud radar, a lidar-ceilometer, radiosonde ascents and results from a microphysical cloud model. Lopez et al. (2006) used observations from a ground-based microwave radiometer in a two-dimensional variational scheme together with a cloud radar to retrieve cloud liquid water content. Recent efforts by De Angelis et al. (2016) aim at the direct use of brightness temperature observations in a 4D-VAR assimilation scheme of a numerical weather prediction model.

Besides linear regression models, neural networks are used frequently to retrieve vertical profiles of thermodynamic variables from radiometer observations (e.g. Churnside et al. 1994; Solheim et al. 1998; Xu et al. 2015). Neural networks are non-linear models. Their training phase is more complicated than that of linear regression models but evaluation is easy once the model parameters have been found. Most neural networks for retrieval applications are rather simple, typically with a single hidden layer and one type of activation function. Their overall performance in terms of the root mean square error has not been found to be consistently better than that of linear regression models. There are indications that retrieval performance is better in terms of the representation of atmospheric structures such as temperature inversions despite worse root mean square errors overall (Churnside et al. 1994). However, non-linear regression models have been found to react stronger to a bias (Martinet et al. 2015) and noise (Cimini et al. 2006) in the input data. Many other regression models exist that have not been used for retrievals yet. All regression models are generally easy to use but measures have to be taken to avoid overfitting. They often lack methods to estimate uncertainty and, as shown for the case of linear regression (Figure 3), even if an uncertainty estimate can be derived, its information content has to be studied carefully. In contrast to optimal estimation retrievals which explicitly use a forward model, atmospheric states determined by regression methods are in general not consistent with the brightness temperatures they are derived from.

Individual components of the retrieval can be tweaked next to the method itself. Del Frate and Schiavon (1998) and Tan et al. (2011) have used principal component analysis/empirical orthogonal functions to reduce the dimensionality of the retrieval problem. Such dimensionality reducing techniques can be applied to the radiometer measurements or the atmospheric state vector. It can also be beneficial for the retrieval accuracy to apply variable transformations, e.g. to enforce positivity. The choice of variables that quantify the atmospheric state is of importance too. While regression models are able to retrieve

temperature, humidity or liquid water isolated from one another, optimal estimation techniques are more restricted. This is discussed further in section 14. Regression methods are also able to retrieve only partial profiles, while optimal estimation is dependent on a forward model which often requires a comprehensive description of the atmospheric state in order to be accurate.

9 Priors and Training Data Sets

The performance of optimal estimation and regression methods strongly depends on the data that drive them. Error estimates and the prior distribution used in optimal estimation are important for the uncertainty assessment of the retrieved atmospheric state and constrain the ill-posed inversion problem of retrieval. A first guess for the iteration procedure can be derived from the prior distribution. The information on the atmospheric state contained in an observation of a ground-based microwave radiometer originates mainly from the lower troposphere (Cadeddu et al. 2013; Cimini et al. 2006; Löhnert and Maier 2012). The prior therefore governs the retrieval performance at higher altitudes. This makes numerical weather forecasts a good source of prior information as these are generally more accurate at higher levels, where larger scales of high short-term predictability dominate the atmospheric state, than in the boundary layer. Thus, a NWP model prior provides complementary information to the radiometer observation and the retrieval additionally benefits from all data sources assimilated into the NWP model (Cimini et al. 2015). The result is good accuracy throughout the entire troposphere (see e.g. results from Cimini et al. 2011). The covariance of a NWP model prior can be extracted from the model's assimilation procedure or be obtained from a comparison of radiosonde ascents with previous forecasts. In complex terrain, where the model topography often significantly differs from the actual topography, additional processing is required before model output can be used in an optimal estimation framework.

Priors may also be determined directly from a radiosonde climatology. This generally leads to distributions with a large uncertainty because the atmospheric state can vary greatly over the course of a year at a fixed location. The atmospheric state space accessible to the optimal estimation procedure is therefore only weakly constrained and a good first guess should be chosen to ensure convergence on an appropriate state. A possible choice are profiles retrieved by a regression model in a first step. It is worthwhile to consider using regression retrievals also for the construction of the prior distribution. This likely yields a smaller uncertainty compared to the climatology. A disadvantage of this approach is that the independent estimation of the regression model parameters and the prior covariance requires a large climatological data set. The covariance of a data-based prior can further be reduced by categorization of the data and subsequent calculation of specialized priors. It is possible to specialize on seasons (e.g. Peckham and Grippa 2000; Chan 2010; Tan et al. 2011; Massaro et al. 2015), months (e.g. Turner et al. 2007), time of day (e.g.

Massaro et al. 2015) or on meteorological situations such as a distinction between clear, cloudy and rainy cases or atmospheric stability (Meyer 2016). Categorization requires that the category can be determined prior to the retrieval.

Regression models relate brightness temperature measurements to atmospheric states based on the data pairs that were used for the training of the model. The representativity of actual atmospheric states of a training data set is of high importance because the extrapolation behavior of a regression model is not guaranteed to be physically reasonable. A model can become biased toward a certain meteorological situation if the situation is overrepresented in the training data, which has to be considered when collecting these (Cadeddu et al. 2002). A climatology of radiosonde ascents and brightness temperature measurements simulated by a radiative transfer model from these ascents are commonly used for training. Del Frate and Schiavon (1998) used an entirely synthetic training data set from an atmospheric profile generator. Few sites have radiometer measurements over a long enough time period to allow measurement-based regression, i.e. training with collocated and simultaneous measurements by balloon and radiometer. Güldner and Spänkuch (2001) and Cimini et al. (2006) found that measurement-based regression models have significantly less bias due to the absent bias of the radiative transfer model. Where radiosonde data are not available, numerical weather prediction model output can be used alternatively (Güldner 2013). Regression models based on specialized training data sets may exhibit 'jumps' in the the time evolution of retrieved profiles when a procedure managing the transition between categories is missing (Massaro et al. 2015). In section 7 it was mentioned that regression models are not arbitrarily transferable due to the danger of extrapolation. Models for measurement sites at different altitudes or in different climates must therefore be trained with different data sets. Most authors only use data directly from the radiometer site to train their models. An exception is a study by Sánchez et al. (2013), who combined radiosonde ascents from Denver (USA), Madrid and Coruña (both Spain) which are stations located at similar latitude and altitude.

10 Inclusion of Additional Information

Many sites that operate a microwave radiometer also operate other instruments such as infrared spectrometers, ceilometers, cloud radars or surface in-situ sensors. Radiosonde climatologies and NWP model forecasts can provide external information for retrievals. How best to include such sources of additional information into retrieval methods is a current and important topic of research.

Regression methods make it easy to include other data as additional regressors. Most regression models however do not consider the uncertainties of their predictors and additional information about the atmospheric state may not be used optimally. A high number of regressors also increases the possibility of overfitting.

While additional information should contribute to the prior of optimal estimation retrievals, it is often not immediately obvious how exactly the prior has to be set up to accommodate various kinds of such information. A NWP forecast provides complete profiles of atmospheric variables which can be used directly in a prior. A cloud base height or surface measurement does not lead to an entire profile from which a prior can be determined. Such isolated measurements may be included as fixed points of the prior by modifying the mean and covariance matrix (Bleisch and Kämpfer 2012). In-situ measurements of the atmospheric state can be used to explicitly set a value of the prior whereas a cloud base height may be translated to a saturated layer and cloud radar observations can be used to prescribe the vertical distribution of liquid water (Löhnert et al. 2004; Turner et al. 2007). When modifying the prior, care has to be taken not to make the covariance matrix singular. If no easy way of including an additional information into the prior distribution can be found, it might be sufficient to use such information for the first guess of the optimal estimation iterative procedure in order to encourage convergence to a profile that is consistent with the additional knowledge. Using regularization terms as discussed in section 6.3, has a similar effect.

Instruments from which entire atmospheric profiles can be retrieved and for which numerical forward models exists can be included in the optimal estimation framework by multiple Bayesian updates. These updates can either be performed sequentially one model after the other or simultaneously. The latter case is discussed in the book by Rodgers (2000) in chapter 4.1.1. He presents a general approach for the determination of the posterior PDF from multiple measurements with associated forward models. If n observations $\mathbf{y}_i$ are available that can be related to the state vector by forward models $\mathbf{F}_i$ and the forward model errors are pairwise independent, then Bayes' theorem takes the form

$$\begin{aligned} p(\mathbf{x}|\mathbf{y}_1,...,\mathbf{y}_n) &= \frac{p(\mathbf{y}_1,...,\mathbf{y}_n|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y}_1,...,\mathbf{y}_n)} \\ &= \frac{p(\mathbf{x})}{p(\mathbf{y}_1,...,\mathbf{y}_n)} \prod_{i=1}^{n} p(\mathbf{y}_i|x) \\ &= \frac{p(\mathbf{x})}{p(\mathbf{y}_1,...,\mathbf{y}_n)} \prod_{i=1}^{n} \mathrm{N}(\mathbf{y}_i|\mathbf{F}_i(\mathbf{x}) + \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \ , \end{aligned}$$

where the independence of forward model errors is used for the second identity.

There is some controversy regarding the usefulness of surface observations as additional information. Some authors claim these are too strongly influenced by surface layer processes and not representative of the air volume responsible for the radiometer observations (e.g. Crewell and Löhnert 2007). Research in Innsbruck has been focused on the inclusion of in-situ measurements from surface stations at different altitudes as they are available in mountainous locations. Massaro (2013) and Meyer (2016) found that such observations have a positive effect on the retrieval accuracy if included as additional regressors in a linear regression model despite not directly probing the volume that affects the radiometer.

Radiative Transfer and State Representation

The only common retrieval technique that does not depend on a numerical radiative transfer model is measurement based regression. But time series of simultaneous radiometer observations and radiosonde ascents to train the model with are rare. The ability to simulate radiometer measurements provides a way to generate data pairs from radiosonde climatologies or NWP forecasts. The optimal estimation approach directly incorporates a numerical radiative transfer model (RTM) into the retrieval procedure and additionally makes use of its linearization. The development of microwave radiative transfer models is therefore closely linked to the development of retrieval techniques.

This chapter reviews the fundamental principles of microwave radiative transfer in the atmosphere, discusses issues related to radiative transfer simulations and the representation of atmospheric state for retrieval purposes and ends with the presentation of a numerical model implementation.

11 The Radiative Transfer Equation

The differential form of the radiative transfer equation in a non-scattering medium is given by

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}s} \;=\; -I_\nu \alpha + F \tag{28}$$

where I_ν is radiative power per frequency, solid angle and area in a specific direction (spectral radiance), α is the absorption coefficient of the medium, F contains all sources of radiation and s is distance in beam direction (e.g. Buehler et al. 2005). The only source term considered here is the emission of the medium itself in thermal equilibrium according to Plank's law

$$F \;=\; \epsilon\, B_\nu(T)$$

where ϵ is the emission coefficient of the gray body and $B_\nu(T)$ is the Planck function describing the emitted spectral radiance of a black body at a given temperature T. The emission coefficient quantifies the difference of the medium to a perfect black body as described by the Planck function and dependent on the considered frequency. Because thermodynamic equilibrium is assumed, Kirchhoff's law allows to equal the absorption and emission coefficients of the medium

$$0 \;\leq\; \alpha \;=\; \epsilon \;\leq\; 1\ .$$

Substituting the results of Plank's law and Kirchhoff's law into the radiative transfer equation (28) yields

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}s} \;=\; -\alpha(I_\nu - B_\nu(T)) \;.$$

For temperatures occurring in the atmosphere and the spectral region of microwave radiation, the Planck function $B_\nu(T)$ is approximated well by the Rayleigh-Jeans approximation which relates the spectral radiance at a given frequency linearly to temperature T

$$B_\nu(T) \;\approx\; \frac{2\nu^2 kT}{c^2} \;,$$

where c is the speed of light and k is the Boltzmann constant. Microwave radiometers measure brightness temperatures instead of spectral radiances. Brightness temperature T_B is the temperature of a black body that emits the same spectral radiance as an observed gray body. In the Rayleigh-Jeans limit the relationship between temperature and brightness temperature is particularly simple: $T_B \;=\; \epsilon\, T$. By expressing observed spectral radiance in terms of Planck emission, again applying the Rayleigh-Jeans approximation

$$I_\nu \;:=\; \epsilon\, B_\nu(T) \;\approx\; \epsilon\, \frac{2\nu^2 kT}{c^2} \;=\; \frac{2\nu^2 k}{c^2}\, T_B \;,$$

and noting that the term $\frac{2\nu^2 k}{c^2}$ is constant for a fixed frequency, the radiative transfer equation (28) for a given frequency can be expressed exclusively in terms of temperature and brightness temperature

$$\frac{\mathrm{d}T_B}{\mathrm{d}s} \;=\; -\alpha(T_B - T) \;. \tag{29}$$

Because $T \;=\; T(s)$, analytical solutions of equation (29) can only be found in a few special cases and numerical solutions have to be sought in general.

12 A Solution for Ground-based Radiometer Applications

For a radiometer looking up from the surface or at an angle to the side, the absorbing and emitting medium is the atmosphere. To simulate a radiometer measurement by solving the radiative transfer equation, the atmospheric state has to be projected onto the light path observed by the radiometer. Since atmospheric state is usually known as a function of altitude, the first step towards a solution of (29) is a coordinate transformation into height coordinates

$$\frac{\mathrm{d}s}{\mathrm{d}z} \;=\; -\frac{1}{\cos(\theta)} \;.$$

z is the altitude coordinate and θ the zenith angle describing the deviation of the radiometers viewing direction from zenith. Because z increases with height but s is the distance

along the light path which points downward (the observed radiation is downwelling), $\frac{\mathrm{d}s}{\mathrm{d}z}$ must be negative. It is convenient for the solution of (29) to introduce optical depth

$$\begin{aligned} \tau(s) &:= \int_{s_0}^{s} \alpha(u)\mathrm{d}u \\ \Rightarrow \quad \tau(z) &= -\frac{1}{\cos(\theta)} \int_{z_0}^{z} \alpha(u)\mathrm{d}u \\ \Rightarrow \quad \frac{\mathrm{d}\tau}{\mathrm{d}z} &= -\frac{\alpha}{\cos(\theta)} \ . \end{aligned}$$

Applying the coordinate transformation to (29), multiplying by e^{τ} and applying the product rule of differentiation yields

$$\frac{\mathrm{d}T_B}{\mathrm{d}z}e^{\tau} + \frac{\alpha T_B e^{\tau}}{\cos(\theta)} = \frac{\mathrm{d}(T_B e^{\tau})}{\mathrm{d}z} = -\frac{\alpha T e^{\tau}}{\cos(\theta)} \ .$$

Integrating this differential equation from the surface ($z = 0$) to space ($z = \infty$) gives an expression for the brightness temperature observed by the (virtual) radiometer

$$T_B(0) = T_B(\infty)e^{\tau(\infty)} + \frac{1}{\cos(\theta)} \int_0^{\infty} \alpha(z)T(z)e^{\tau(z)}\mathrm{d}z \ . \tag{30}$$

$T_B(\infty)$ is the brightness temperature of the cosmic background radiation which has a value of approximately 2.75 K for frequencies in the microwave spectrum (Westwater et al. 2004). It is important to note that (30) is only valid at off-zenith angles if the vertical profiles of T and α are horizontally homogeneous and if refraction of light and Earth's curvature are neglected. These restrictions are further discussed in section 17.

13 Atmospheric Extinction and Absorption Models

The calculation of brightness temperatures using (30) requires the determination of atmospheric absorption/emission coefficients α. The main properties of atmospheric absorption in the microwave region from 20 to 200 GHz are given here based on a review by Westwater et al. (2004) and some additions from other authors.

The three main constituents in the atmosphere contributing to extinction and emission of microwave radiation are oxygen, water vapor and liquid water. A typical absorption spectrum of an atmospheric layer is shown in Figure 4. Water vapor has a relatively weak absorption band around 22.235 GHz (K band) and oxygen a stronger band around 60 GHz (V band). Outside of the shown frequency range exist an oxygen absorption line at 118.75 GHz and a stronger water vapor line at 183.31 GHz. The wings of these lines and so called continuum absorption, associated with pressure broadening processes and

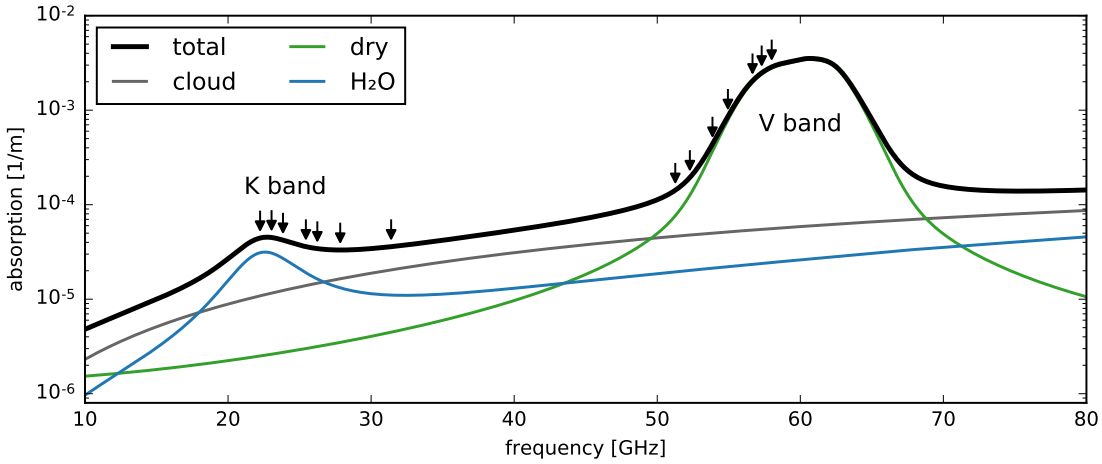

Figure 4: Absorption coefficients (black) and contributions of individual terms (other) according to the models of Liebe et al. (1993) (gaseous) and Turner et al. (2016) (cloud) for an atmospheric layer of 850 hPa pressure, 0°C temperature and saturation, containing 0.1 g/kg of liquid water. The arrows mark frequencies of the channels of the HATPRO radiometer which is used for subsequent retrievals.

other atmospheric species, also contribute to the overall absorption (Liebe et al. 1993; Hewison 2006). Finally, in the presence of clouds there is absorption by water droplets. Due to the size-wavelength-ratio of droplets in non-precipitating clouds and microwave radiation, scattering can be neglected and the Rayleigh-approximation is appropriate for the description of absorption of radiation by liquid water. Absorption and scattering by snow and ice in the atmosphere has negligible effect on radiation in the K and V bands (Sánchez et al. 2013). During rain events however, scattering of microwave radiation by precipitation cannot be neglected and rainwater accumulation on the radiometer's radome must be accounted for as well. Since these effects are much more difficult to quantify than absorption, they are usually not included in radiative transfer models and the resulting errors are accepted or retrievals are not performed during precipitation at all.

The absorption spectrum in Figure 4 was calculated by a model from Liebe et al. (1993) for gaseous absorption and a model from Turner et al. (2016) for absorption by cloud liquid water. There are other such models in use, common are a model by Rosenkranz (1998), MonoRTM (Clough et al. 2005) and ARTS (Buehler et al. 2005) (the latter two are not just absorption models but entire radiative transfer simulators). These models usually account for the most important absorption lines explicitly, correct for missing ones and parameterize the continuum term by an empirical function. The Liebe et al. (1993) model for example describes oxygen line absorption in the microwave region by 44 lines and water vapor by 34 lines. Cloud absorption is most often parameterized by a double-Debey model (e.g. Turner et al. 2016). Instead of exclusively describing atmospheric absorption, absorption models are formulated in terms of the more general refractivity N which is the complex refractive index reduced by one. With knowledge of N, absorption can be calculated from

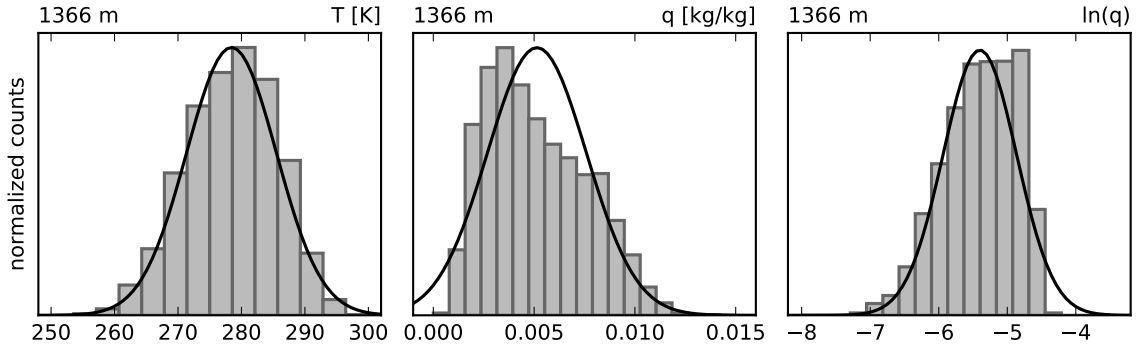

Figure 5: Histograms of temperature (left), specific humidity (center) and the natural logarithm of specific humidity (right) for the 1366 m level of the radiosonde training data set (section 20). The black curves are Gaussian distributions fitted to respective data.

$$\alpha = \frac{4\pi\nu}{c} \operatorname{Im}(N) ,$$

where c is the speed of light and ν is its frequency.

14 State Vector Variables

The emission and absorption properties of the atmosphere depend on its state which has to be represented in some form for the retrieval. Directly using the quantities that are to be retrieved (temperature, humidity, liquid water content) as state vector variables usually works well for regression retrievals. Any method that makes the assumption of Gaussian distributed state vector variables however, might require variable transformations to achieve value distributions that approximately follow a Gaussian curve. For optimal estimation retrievals it is common to combine water vapor and liquid water into a single quantity and then retrieve the logarithm of this quantity (e.g. Löhnert et al. 2004; Hewison 2006; Cimini et al. 2011). Figure 5 shows the benefits of using the logarithm of humidity with an example from a single height level of a climatology based on radiosonde measurements from Innsbruck airport. The distribution of temperature is very close to Gaussian, but humidity[4] is a non-negative variable with a skewed distribution. Applying the logarithm reduces the skewness of the distribution, brings it closer to a Gaussian shape and has the additional positive effect of enforcing the positivity of humidity values during retrieval.

Reducing specific humidity and liquid water content to a combined specific water content[5] variable is practical in multiple ways. In terms of the distribution functions the asymmetry found for liquid water content is even worse than for water vapor since many

[4] In this case specific humidity but the same applies to water vapor density (absolute humidity) and relative humidity.

[5] Specific water content is meant to be specific to a unit mass of air similar to specific humidity.

atmospheric profiles are cloud-free resulting in a second mode at value zero. Because the logarithm is not defined at zero this particular variable transformation to reduce skew is not applicable. But since there always is some water vapor in the atmosphere this mathematical problem can be avoided by adding both components of water content and then applying the logarithm transformation. Another advantage of the combined variable is the reduction of state vector dimensionality resulting in computationally less expensive optimal estimation retrievals.

The combination of all atmospheric water into a single variable is only sensible if there is a way to separate the components again for the calculation of the absorption coefficients. The effect of ice on microwave radiative transfer is negligible (see section 13) therefore it is sufficient to only include water vapor and liquid water in the specific water content. The governing quantity of the partitioning scheme for these components is the saturation vapor pressure of air over water as it determines when the capabilities of the gaseous water content are exceeded and cloud droplets start to form[6]. Because saturation vapor pressure is a function of temperature and temperature is included in the retrieved state vector, the water separation threshold can be calculated during the retrieval. The chosen partition function should be differentiable to stabilize the retrieval. Hewison (2006) approached this problem by introducing a transition region to a function of Deblonde and English (2003), which gradually introduced liquid water content starting at a given relative humidity threshold. The chosen form of partition function here is

$$\frac{\mathrm{d}q_{liq}}{\mathrm{d}s} = q_{sat} \begin{cases} 0 & s < 0.95 \\ \cos\left(\frac{s-1.05}{0.1}\frac{\pi}{2}\right)^2 & 0.95 \leq s \leq 1.05 \\ 1 & 1.05 < s \end{cases},$$

where

$$s = \frac{q_{tot}}{q_{sat}} = \frac{q_{vap} + q_{liq}}{q_{sat}}$$

is the fraction of specific water content to specific water content at saturation, a generalization of relative humidity that includes liquid water. Integration yields the desired expressions for liquid water content and specific humidity

$$q_{liq}(s) = \begin{cases} 0 & s < 0.95 \\ \frac{q_{sat}}{2}\left(s - 0.95 - \frac{0.1}{\pi}\cos\left(\frac{\pi s}{0.1}\right)\right) & 0.95 \leq s \leq 1.05 \\ q_{liq}(1.05) + q_{sat}(s - 1.05) & 1.05 < s \end{cases} \tag{31}$$

and

[6] The possibility of supersaturated air complicates this relation as other factors such as cloud condensation nuclei become important. This is neglected here.

$$q_{vap}(s) \;=\; q_{tot}(s) - q_{liq}(s)\;. \tag{32}$$

95 % relative humidity is the threshold for the introduction of cloud liquid water and saturation is reached at 105 % relative humidity. Any remaining water contributes to the liquid component.

It is important to note that because temperature is critical for the determination of the saturation pressure of water vapor the introduction of such a partition of water content results in a strong dependency of the humidity retrieval on the temperature retrieval (Bleisch and Kämpfer 2012). It is therefore impossible to retrieve humidity independently of temperature and the direct connection between relative humidity and cloud liquid water restricts the possible states of the atmosphere. The Karstens et al. (1994) cloud model used in this thesis to extract information on liquid water content from profiles obtained by a radiosonde ascent assumes cloud liquid water exists wherever relative humidity exceeds 95 % but then calculates the liquid water content independently of the exact value of relative humidity. The output of the numerical weather prediction model COSMO-7 whose forecasts are used later to provide prior distributions for optimal estimation retrievals fixes relative humidity at 100 % in all clouds. In both cases given combinations of humidity and cloud water may not be representable by the partition (31) and (32). In practice the two components are nevertheless added and any changes after separation are accepted. The uncertainty in the determination of cloud liquid water from models like the one by Karstens et al. (1994) is hard to assess (Ebell et al. 2010) and potentially large, so the convenience of the partition function outweighs the additional inaccuracies resulting from the restriction of humidity state space by the variable combination.

For optimal estimation retrievals there is additionally the need to update pressure information after each iteration since pressure is important for the calculation of absorption coefficients and depends on temperature and humidity but is not part of the explicitly retrieved atmospheric state variables. Pressure is calculated after each iteration by numerical integration of the hydrostatic equation

$$p(z) \;=\; p(z_0) + \exp\left(-g \int_{z_0}^{z} \frac{1}{R(h)\,T(h)}\,\mathrm{d}h\right)$$

where z_0 is the height of the radiometer and $p(z_0)$ is provided by a pressure sensor at that height (which is built into the radiometer) and R is the specific gas constant of air which is a function of height due to its dependence on water vapor content. The errors of this approximation have been assessed by a comparison of hydrostatic approximations with the pressure information from radiosonde ascents and were found to be negligible.

15 Retrieval Information Content and Weighting Functions

Two important properties associated with atmospheric absorption govern the frequencies

of channels which a microwave radiometer observes in order to retrieve vertical profiles of thermodynamic variables. Temperature information can be obtained from emission and absorption of oxygen which are directly proportional to local temperature and independent of mixing ratio because oxygen is a well mixed gas throughout the atmosphere. Water vapor amounts on the other hand vary strongly with height but if temperature is known, the emission of water vapor is proportional to its partial density (Westwater et al. 2004). To obtain vertical profiles, channels along the sides of absorption bands like the K and V band are observed. Each channel allows to look differently far into the atmosphere due the different opacity of the atmosphere at each frequency (Churnside et al. 1994). Unfortunately, measurements at different frequencies along the same absorption band are not independent of one another therefore the information content of observations is not proportional to the number of radiometer channels (Hewison 2006). To increase the vertical resolution without adding new channels, the radiometer antenna can be pointed away from zenith, resulting in longer attenuation paths which provide more information on the lower layers of the atmosphere. This procedure is called elevation scanning and is further discussed in section 17.

A common visualization of the information content of a radiometer observation are the so-called weighting functions each corresponding to one brightness temperature observation. Weighting functions are obtained from radiative transfer models by taking the derivative of the measured brightness temperature with respect to the atmospheric state vector. In the case of a vertically discretized atmosphere the weighting functions are the rows of the Jacobian of the forward model, containing the derivatives with respect to each state vector element (Rodgers 2000). The weighting functions therefore express the linearized response of the radiative transfer to a change in the atmospheric state. Because weighting functions have the same dimension as the atmospheric state vector, they can conveniently be visualized on the retrieval grid. Their magnitude can be interpreted as the sensitivity of the measurement to the considered state vector component at the given height. Hence, weighting functions can be used to assess from which heights in the atmosphere a radiometer channel obtains its information (Martinet et al. 2015).

The optimal estimation framework directly incorporates the forward model Jacobian into the inversion process. The individual weighting functions act similarly to basis functions which determine where and at which scales the profile can be modified during the retrieval. The functions' smoothness is therefore a limiting factor on the vertical resolution of features of the atmospheric state which can be recovered by the retrieval from a radiometer observation. This highlights the importance of the first guess in the optimal estimation framework. Its small scale features may be fixed during the retrieval if the weighting functions are smooth.

Derivation of the Jacobian from any numerical radiative transfer model is possible by finite differencing. This brute force approach is very flexible but computationally expensive and can result in inaccurate linearizations if perturbations are selected without care.

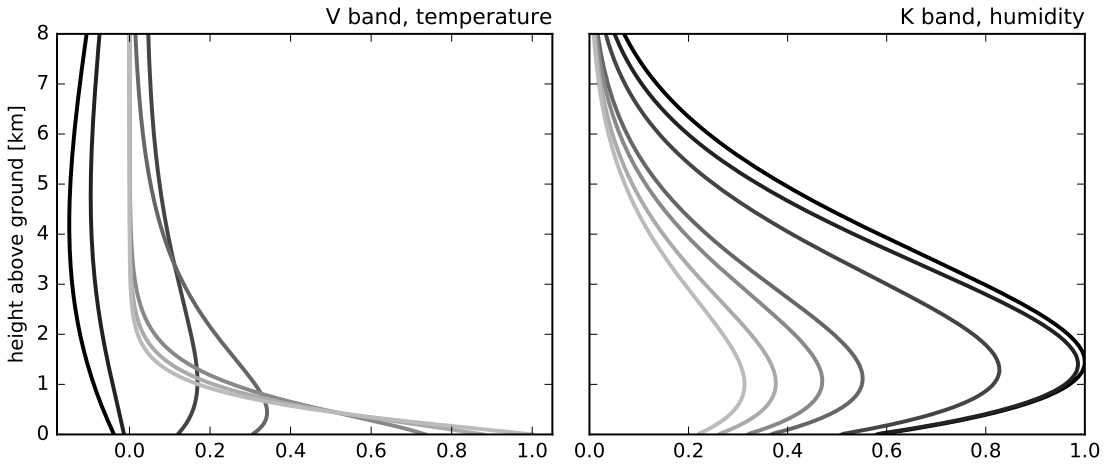

Figure 6: Normalized weighting functions for the V band (left) and K band (right) channels of the HATPRO instrument (see section 22) simulated by the radiative transfer model introduced in section 19 for an upwards looking view. The ground is at an altitude of 612 m. V band weighting functions are given with respect to temperature, K band weighting functions with respect to the natural logarithm of humidity. Lighter colors correspond to channels of higher frequency. The assumed atmospheric profile is the U.S. standard atmosphere with relative humidity decreasing from 70 % at the surface to 10 % at 11 km above which it is constant.

Model developers have therefore started to implement routines dedicated to the accurate calculation of Jacobians based on other differentiation techniques. ARTS (Buehler et al. 2005) and the recently developed RTTOV-gb (De Angelis et al. 2016) provide such procedures and the numerical model implemented at the end of this chapter is also able to calculate exact Jacobians without finite differencing.

Figure 6 shows weighting functions for HATPRO channels in the V band with respect to temperature and in the K band with respect to humidity. Temperature weighting functions are more diverse in their shape than those of humidity. This indicates that the vertical resolution of temperature retrievals is generally better than that of humidity retrievals (Cimini et al. 2011). Information on humidity mainly comes from the lower part of the troposphere. This is largely due to the much higher concentrations of water vapor at these heights as the saturation vapor pressure decreases exponentially with decreasing temperature. The V band channels of lower frequency show sensitivity to mid-tropospheric temperatures while information from channels of higher frequency is restricted to the lower troposphere.

16 Retrieval Grid

Section 14 discussed issues regarding the choice of variables used to represent the atmospheric state and used the concept of a state vector. This vector is the result of a vertical discretization of the atmospheric profile. This discretization is based on the retrieval grid, a set of altitudes at which the state variables are determined.

At first it seems sensible to choose a retrieval grid with a resolution as high as possible in order to obtain the most accurate retrievals. Most studies however use only around 40 to 60 vertical levels for the discretization of the entire tropopause (e.g. Solheim et al. 1998; Hewison 2006; Martinet et al. 2015). Considering the shape of the weighting functions in Figure 6, the retrieval grid should have a higher resolution in the lower troposphere. At a certain point increases in resolution will not be reflected by the retrieval performance because the ability of a retrieval to resolve small scale features of the atmospheric state is limited by the smoothness of the weighting functions. This was also found by Meyer (2016) who observed no improvement of accuracy in an experiment with a retrieval grid of enhanced resolution. Practically, the vertical discretization is most often chosen so that it is not much finer than the actual vertical resolution of the radiometer since less levels result in higher computational performance of the retrieval technique. The cost of calculating the Jacobians in particular scales with the number of levels of the retrieval grid and is usually the most expensive process in an optimal estimation retrieval.

Special consideration must be given to the altitude of the highest level of the retrieval grid. Radiometer channels measuring in the more transparent regions of the microwave spectrum can have significant contributions by emission of the tropopause and lower stratosphere. If such parts of the profile are not represented by the atmospheric state vector they should be corrected for in radiative transfer calculations otherwise simulated brightness temperatures can be significantly biased. This issue will be discussed further in sections 19.5 and 19.6.

17 Elevation Scanning

Many radiometers are able to rotate their antenna in order to measure brightness temperatures at off-zenith angles, a procedure called elevation scanning. Changing the angle of measurement changes the associated weighting functions and therefore the heights from which the channels get their information on the atmospheric state. This can increase the overall information content of an observation and result in a higher vertical resolution of the retrieval in the boundary layer (Westwater et al. 2004) while the impact on retrieval performance at higher levels is only small (Cimini et al. 2006). Figure 7 shows weighting functions for two frequencies in the V band at different angles. When the zenith angle is larger, i.e. when the radiometer view is more horizontal, the sensitivity of the channels increases near the surface while decreasing higher up.

Xu et al. (2014) remarked that due to the U-shape of the radomes used by many microwave radiometers, measurements from elevation scans are less affected by rainwater accumulation on the instrument than those at zenith. This effect was previously observed by Cimini et al. (2011) who noticed that off-zenith signals during precipitation were less noisy than zenith observations.

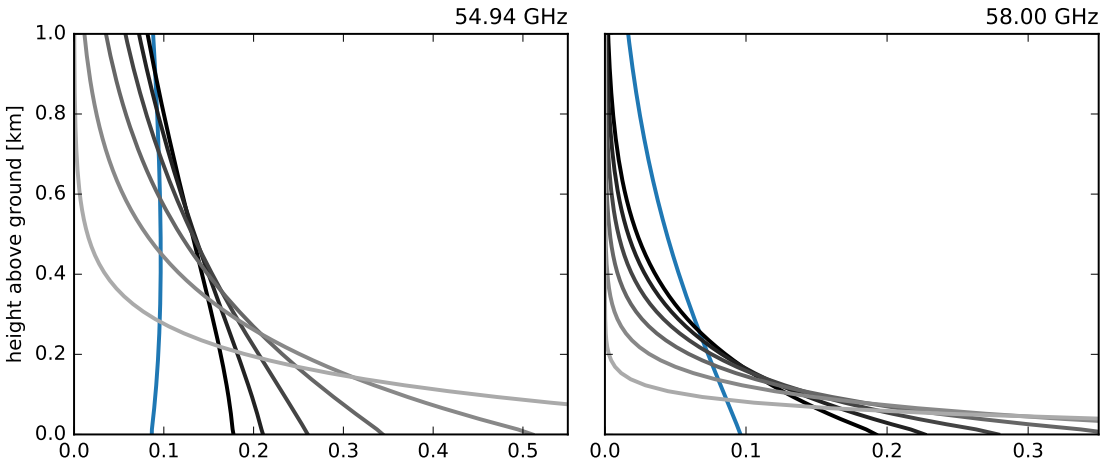

Figure 7: Normalized weighting functions for 54.94 GHz (left) and 58.00 GHz (right) for different zenith angles simulated by the radiative transfer model introduced in section 19. The 0° weighting function (upward looking) is the blue line, angles 60°, 65°, 70°, 75°, 80°, 85° are the gray lines, with lighter colors referring to larger zenith angles. The atmospheric profile and radiometer altitude are the same as in Figure 6.

Elevation scanning puts additional demands on the instrument, retrieval and radiative transfer model. In order to accurately observe brightness temperatures at large zenith angles, radiometers have to be very sensitive. High sensitivity is usually achieved by wide bandwidths of the channels used at off-zenith angles (Cadeddu et al. 2002; Crewell and Löhnert 2007). Because the execution of an elevation scan requires additional time, a trade-off between temporal resolution and averaging time of a measurement is necessary (Cadeddu et al. 2002). Measurements in transparent channels are very sensitive to the misalignment of angles so the antenna adjustment must be precise (Hewison 2006). In complex terrain the radiometer view might intersect with a mountain and surface emission by the slope must be considered. Violations of the assumption of horizontal homogeneity are a major problem of elevation scanning in general, because the likelihood that the radiometer view includes different air masses rises the more horizontal it is (Cimini et al. 2006). This is particularly an issue in situations of inhomogeneous cloud cover (Güldner 2013). Effects of inhomogeneity can be reduced by scanning in multiple directions and averaging the measurements from same zenith angles. This increases the time to make an observation further. Cimini et al. (2006) noticed a strong impact of horizontal inhomogeneity on neural network retrievals of temperature and humidity, likely due to the non-linearity of the method amplifying fluctuations in the brightness temperatures. Chan (2010) used information from off-zenith angles only for the lowest levels of the atmosphere when retrieving humidity profiles by linear regression.

Many of the mentioned complications can be avoided by excluding transparent channels in retrievals using elevation scan data. Crewell and Löhnert (2007) and Massaro et al. (2015) used only the most opaque channels in the V band at off-zenith angles for this reason.

Aside from these considerations, subsequent retrievals in this work use only the four most opaque channels in the V band due effects of refraction being missing in the radiative transfer model. Off-zenith light paths in the atmosphere are affected by refraction caused by density changes of air with height and Earth's curvature must be taken into account as well (Hewison 2006). The radiative transfer model introduced in section 19 currently does not adjust light paths for refraction which causes large errors in simulated brightness temperatures of transparent channels at off-zenith angles. Paths lengths of light in the opaque region of the V band are very short however (300 - 2000 m, Crewell and Löhnert 2007) and refraction of light is negligible.

18 Error Characterization

The likelihood function (17) requires an error estimate which is a combination of instrumental error and forward model uncertainties. Hewison (2006) provides a thorough assessment of terms contributing to this estimate. Only a short review is given here.

The instrumental error is mainly due to random fluctuations in the measured brightness temperatures. The effect of these can be reduced by longer averaging times of the radiometer as long as the atmospheric situation can be considered stationary. Hewison (2006) determined radiometric noise from observations of a black body reference target but many authors simply assume uncorrelated errors between 0.2 and 0.5 K for the radiometer channels (e.g. Churnside et al. 1994; Del Frate and Schiavon 1998; Löhnert et al. 2004; Martinet et al. 2015). Additional instrumental errors can occur from misaligned angles during elevation scans and radio frequency interference caused by sources of microwave radiation which are not the atmosphere.

The second component of the likelihood covariance matrix are forward model errors. Cimini et al. (2004) claimed that uncertainties in the radiative properties of the atmosphere are the main limiting factor for the retrieval of water vapor. Such uncertainties are related to inaccurate representations of the atmospheric state and errors in the description of physics, particularly absorption, in a numerical radiative transfer model. Because retrieval grids are usually much coarser than the vertical resolution of a radiosounding, a discretization error occurs when the atmospheric state is represented on the retrieval grid. As mentioned in section 16, the assumptions made for extensions of the vertical profile which are not explicitly included in the state vector can contribute to errors in simulated brightness temperatures too. Absorption model errors can be separated into a contribution from uncertainties in the description of absorption processes and a contribution from mathematical simplifications which are sometimes made to speed up radiative transfer calculations.

Uncertainties related to implementation choices like the number of levels of the retrieval grid, the restriction of radiative transfer to monochromatic frequencies or the use of a fast absorption predictor can be easily quantified from a comparison of model simulation

with and without the evaluated component. Uncertainties related to the description of absorption are harder to assess because a reference absorption model does not exist and modeled brightness temperatures are not easily comparable to ones measured by a radiometer. Hewison (2006) compared simulations of two radiative transfer models with different descriptions of absorption in order to quantify this error term.

Aside from the covariance there is also a bias term in (17) compensating for systematic differences in brightness temperatures obtained from radiometer and numerical model. Turner et al. (2007), Löhnert and Maier (2012) and Martinet et al. (2015) observed significant changes in retrieval performance if this offset was not corrected for. Cimini et al. (2011) noticed only little sensitivity of the optimal estimation method to a bias correction but remarked that this was likely related to the choice of radiometer channels used for their retrievals. For long time operation of a radiometer it should be regularly tested if the offset is stationary over time. Changes are especially likely after a recalibration of the radiometer (Löhnert and Maier 2012).

19 Implementation of a Radiative Transfer Model Prototype

A numerical microwave radiative transfer model has been implemented as part of this thesis. It is based on the absorption model by Liebe et al. (1993) for absorption due to oxygen and water vapor and a model by Turner et al. (2016) for cloud liquid water absorption. Given an atmospheric profile of pressure, temperature and humidity, it first calculates absorption coefficients and then evaluates equation (30) numerically. The prototype's distinguishing features are its accessible implementation in a high-level, dynamic programming language (Python 3) and the ability to determine temperature and humidity Jacobians by forward-mode automatic differentiation. The model will subsequently be referred to as MWRTM (MicroWave Radiative Transfer Model).

19.1 Motivation

With established radiative transfer software such as MonoRTM (Clough et al. 2005) or ARTS (Buehler et al. 2005) freely available, another implementation of a numerical radiative transfer model should be well motivated. The development of a model prototype for this thesis was driven by these interests:

Current RTMs either resort to expensive finite difference approaches when calculating Jacobians or use hand-crafted routines mixing analytically derived expressions and finite difference approximations where such expressions are unavailable. MonoRTM currently has no implementation of derivatives, so finite differencing is the only option, while ARTS has specially written modules for calculating Jacobians. Automatic differentiation, a method with which any numerical program can be differentiated to machine precision without the need to explicitly write code for the derivatives has proven its usefulness for

the efficient training of machine learning models. It is applied in the RTM prototype for the fast and exact calculation of weighting functions without the necessity of writing hand-crafted routines.

Radiative transfer models are typically written in statically typed, compiled languages such as Fortran or C++ in order to achieve the best computational performance. Data analysis however is usually done in a high-level language such as Python. A question of interest is if it is computationally feasible to implement a RTM in a dynamic language. This would make the model code more accessible to a user who already uses this language for data analysis and simplifies the setup procedure. Finally there is a personal motivation of the author associated with the decision to implement a radiative transfer model. It was hoped that the implementation exercise would expand the author's understanding of radiative transfer simulations and result in a better grasp on the retrieval problem.

The expectation is not to write a 'perfect' model in the first try. Is rather about building a flexible, easy to understand but reasonably fast and accurate prototype adequate for optimal estimation retrievals. The next sections discuss important aspects of development and evaluate the overall model performance.

19.2 Numerical Considerations

In order to obtain numerically accurate results the radiative transfer model has to operate internally with a much finer discretization than that of typical retrievals grids. The input profiles of pressure, temperature and specific water content must therefore be interpolated to the internal model grid. The model provides a linear interpolation scheme for this task. Absorption coefficients are then calculated on the finer model grid where the numerical integration also takes place. A first implementation calculated absorption coefficients on the retrieval grid and then interpolated these for the integration. However, this resulted in large errors in the model output. The quadrature scheme used for integrations is the second order trapezoidal rule.

It was found that the internal model and the retrieval grid can be coarser at greater altitudes without a loss of accuracy due to the generally greater vertical changes of absorption coefficients and temperature in the lower troposphere.

19.3 Automatic Differentiation

The interest in efficient and accurate computation of gradients has risen immensely in recent years because the training of many machine learning algorithms requires such information. Automatic differentiation is established as the standard approach to this problem. The necessity of efficiently calculating Jacobians of numerical models (often called adjoint modeling) is also a problem often encountered in atmospheric sciences. Data assimilation for numerical weather prediction models was one of the first applications of

automatic differentiation (Griewank 2012). However, the technique was mentioned only in one short paragraph in a review of adjoint models by Errico (1997) and even a modern radiative transfer model such as ARTS still has hand-written code for the calculation of Jacobians.

The concept of automatic differentiation has been known at least since the 1960s and was rediscovered multiple times since then (Griewank 2012). At its core, it is nothing more than the application of the chain rule of differentiation to the computational graph of a program. It is not the same as symbolic differentiation though, which often requires significant effort from a human and has the problem of combinatorial blowup of expressions. Implementations of automatic differentiation are able to compute derivatives simultaneously with the forward code or in a second pass, by deriving each edge of the computational graph individually and combining the results directly at each node. A decent implementation has to visit each edge of the computational graph only once resulting in similar complexities of forward and Jacobian calculations. A good (although not peer-reviewed) explanation of the idea behind automatic differentiation is found at http://colah.github.io/posts/2015-08-Backprop.

There are two ways of performing automatic differentiation, characterized by the direction in which the computational graph is traversed. The forward-mode moves from the input nodes towards the output, computing derivatives with respect to each input variable along the way. Implementations of forward-mode automatic differentiation are easy in programming languages supporting dispatch on types. Only minor or no changes to the program are necessary aside from replacing numeric variables with instances of custom types that carry the derivatives and overloading the used operators and functions. The custom types and operators have to be implemented once and can then be used in any program. Reverse-mode automatic differentiation moves from the output leaves toward the input nodes, computing the derivative of each output with respect to each node of the computational graph. Because the reverse-mode requires a second pass through the graph, implementations need to keep track of intermediate results calculated in the first pass, which requires additional infrastructure compared to the forward mode. Both techniques yield the same results but can be very different in terms of computational cost for a given problem. Generally speaking, reverse-mode is suitable for problems with many inputs and only few outputs while the forward mode is more efficient for problems with more outputs than inputs.

A radiative transfer model with many atmospheric layers as input and only a few brightness temperatures as output is better suited for reverse-mode differentiation. The prototype here nevertheless uses the easier to implement forward-mode. Transformation of the forward model code required two custom types, the overloading of a few operators and functions and minimal changes to the existing code. The only performance optimization is a special consideration of derivatives associated with the calculation of the absorption coefficients on the internal model grid.

19.4 Fast Absorption Prediction

The radiative transfer model spends the majority of its time calculating absorption coefficients. Although the models of Liebe et al. (1993) and Turner et al. (2016) are fairly simple, it is convenient to replace them by a fast absorption predictor (FAP) in order to reduce the overall model computation time. FAPs are regression models approximating the absorption coefficients as a function of the state vector. A FAP is trained specifically for each simulated frequency based on data representing typical atmospheric conditions. They have previously been used by Löhnert et al. (2004), Hewison (2006) and others.

Following the choice of state vector variables from section 14, the predictors of the regression models are chosen to be pressure, temperature and the natural logarithm of specific water content. A first implementation modeled both gaseous and cloud absorption together with a polynomial of the state vector variables. But even for polynomials of degree five including all interaction terms, large errors were found in the approximated absorption coefficients. Investigation of this issue revealed that the polynomial was unable to adequately represent the strong non-linearity of absorption near the cloud threshold. Polynomials of higher order suffer from combinatorial blowup of interaction terms and numerical stability issues, therefore a more refined approach was taken.

Instead of fitting a polynomial to the combined gaseous and cloud absorption, the components are approximated separately after splitting the specific water content of the state vector into vapor and liquid components according to (31) and (32). Specific cloud absorption following Turner et al. (2016) is only a function of temperature and modeled by the FAP as a 5th order polynomial. Gaseous absorption depends on p, T and q_{vap}. Löhnert et al. (2004) and Hewison (2006) used third order polynomials without interaction terms as FAPs for gaseous absorption and claimed good accuracy. Here the necessity was found to approximate absorption coefficients by third order polynomials including all interaction terms for V band frequencies and third order polynomials including interactions terms up to second order for K band frequencies or else the FAPs were inaccurate at low pressures particularly in the more opaque channels of the K band. Modeling the logarithm of the absorption coefficient to enforce its positivity was tested but found to negatively affect the accuracy, likely due to the stronger non-linearity introduced by the transformation.

In general, low-order polynomial FAPs perform rather badly at low pressures. More accurate, higher-order FAPs are however much more expensive to evaluate, defeating their original purpose. The retrieval grid used in this thesis end at 12 km above ground and the accuracy of the FAPs described in the previous section is a major reason for this choice.

19.5 Correcting for Missing Parts of the Profile

As discussed in section 16, it is important to consider contributions from the tropopause and lower stratosphere to brightness temperatures measured by a radiometer at the ground

in transparent channels. With the retrieval grid ending at 12 km, the radiative transfer model must account for the emission and absorption in the upper atmosphere in order to make accurate predictions particularly for the more transparent V band oxygen channels. The transparent K band channels are not significantly affected by the upper atmosphere because water vapor concentrations are low and emission is weak.

MWRTM provides the possibility to correct for missing parts of the profile by adjusting the cosmic background temperature to account for the emission and opacity properties of the upper atmosphere. This is a static correction based on a reference profile of pressure, temperature and humidity. Here, values from the US standard atmosphere with no water vapor are used to determine corrections for emission from heights between 12 and 40 km.

19.6 Model Intercomparison and Verification

The implemented model should be compared to other established radiative transfer models to ensure its correctness and to quantify spectroscopy errors which are an important part of the forward model error covariance. Brightness temperature simulations from two other models are available for the evaluation: MonoRTM version 5.2 (Clough et al. 2005) and a model using gas absorption according to Rosenkranz (1998), previously used by Massaro et al. (2015). The latter is not available as an executable computer program but calculated values of brightness temperatures are.

The test data set for the intercomparison is made up of 1568 radiosonde profiles from Innsbruck airport. Because the parameterization that calculates cloud liquid water content from radiosonde ascents for the data of Massaro et al. (2015) is different from the one described in sections 14 and 20 and the aim of this intercomparison is to quantify differences due to the radiative transfer model formulation, all profiles of the test data set are clear sky cases.

Figure 8 shows the results of the model comparison in terms of means and standard deviations of model differences for each channel of a HATPRO radiometer (Rose et al. 2005). The radiative transfer models agree best in opaque spectral regions, which was also found in studies by Westwater et al. (2004) and Cimini et al. (2004). Biases change sign between V and K bands and are small in the opaque channels of the V band. MonoRTM and the Rosenkranz (1998) model agree better in the K band than in the transparent V band channels. As found before by Cimini et al. (2004), the Liebe et al. (1993) absorption model used by MWRTM gives brightness temperatures systematically warmer than the other models in the K band. MWRTM is closer to MonoRTM for all water vapor channels but closer to the Rosenkranz (1998) model in the oxygen channels.

Magnitudes of the model biases overall are similar to those from a comparison by Cimini et al. (2004) who evaluated four absorption models including all of the ones used here, at similar frequencies but with an older version of MonoRTM. Looking at individual channels, there are however some differences that should be addressed. The magnitude

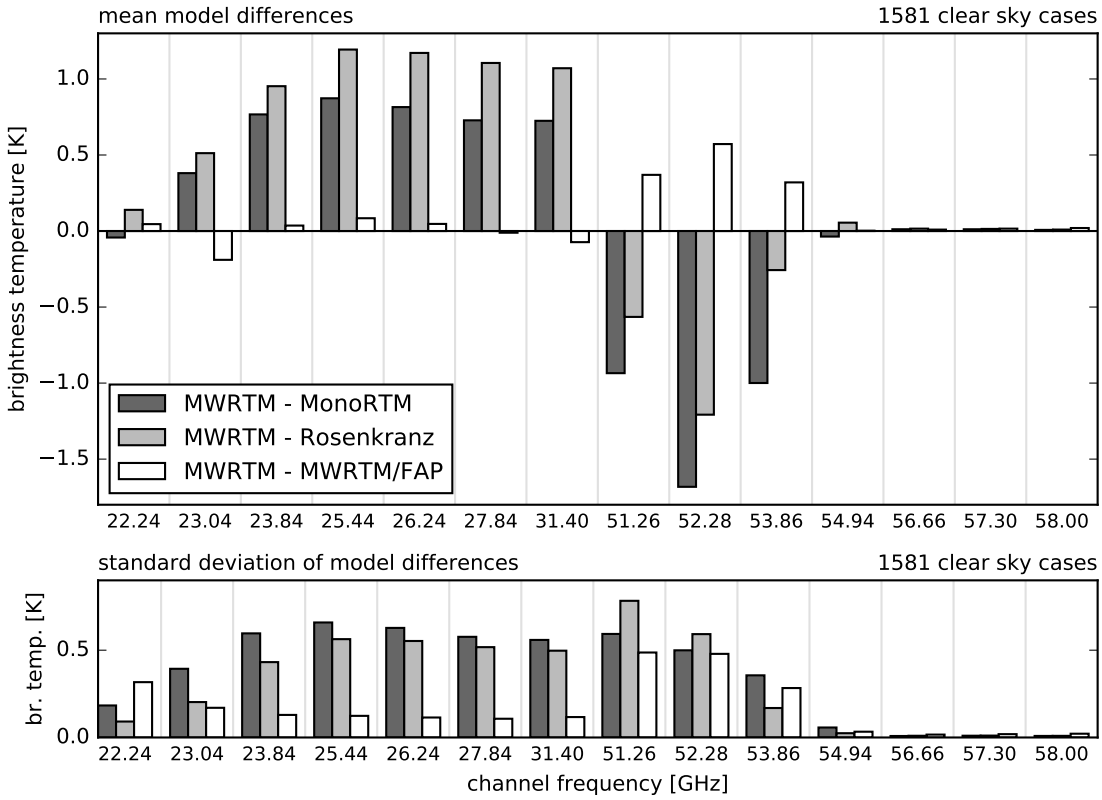

Figure 8: Mean (top) and standard deviation (bottom) of brightness temperature differences from three different radiative transfer models. High resolution simulations by MWRTM provide the reference state to which the differences are calculated. Version 5.2 of MonoRTM (dark gray) is used and Rosenkranz (light gray) is a model used by Massaro et al. (2015) based on gaseous absorption according to Rosenkranz (1998). MWRTM/FAP (white) is a variant of MWRTM with a coarser vertical discretization than the reference and uses a fast absorption predictor. The test data set are 1581 clear sky profiles derived from radiosondes launched at Innsbruck airport. All models simulate a zenith view and the standard deviations are obtained by sampling the full covariance matrices of each model pair.

of the mean differences between MWRTM and Rosenkranz in channels 52.28 and 53.85 GHz are much smaller in the comparison by Cimini et al. (2004) than here. Further investigation of this issue revealed that the differences in the more transparent V band channels are strongly dependent on the chosen profile extension discussed in section 19.5. It is likely that the model used by Massaro et al. (2015) makes a different assumption about the state of the upper atmosphere than MWRTM, resulting in larger differences here. In a comparison carried out between absorption models according to Rosenkranz (1998) and Liebe et al. (1993) by Löhnert et al. (2004), mean brightness temperature differences at frequencies 51.8 and 52.8 GHz were around 0.6 - 0.8 K. These values are closer to ones from the transparent channels of the V band here (this comparison assumes that the bias is a continuous function of frequency). A second discrepancy in comparisons between these two models is found in the 22.24 GHz channel which shows good agreement

between MWRTM and Rosenkranz here but a bias value of about 2.5 K in the study by Cimini et al. (2004) and 1.7 K in the comparison by Löhnert et al. (2004). It was not possible to determine where this discrepancy originates from.

For evaluations of retrieval accuracy with simulated brightness temperatures model biases are negligible because they affect simulated observations, training data for a regression model and the radiative transfer calculations during optimal estimation equally. When working with actual radiometer data, model biases are explicitly accounted for and therefore of little importance.

Standard deviations are larger for the more transparent channels but generally not significantly higher than values expected for instrument uncertainties.

The low resolution variant of MWRTM with fast absorption predictors shows large mean values and standard deviations of brightness temperature differences only in the transparent V band channels. These are again strongly dependent on the assumed state of the upper atmosphere. MWRTM uses the entire radiosonde profiles while MWRTM/FAP cuts profiles at 12 km, which is the highest altitude of the retrieval grid and then assumes a US standard atmosphere profile above. With this in mind, the FAP accuracy is seen as adequate for retrieval use.

As claimed in section 17, errors due to missing refraction in MWRTM for simulations at off-zenith angles were found to be small in the opaque V band channels based on a comparison between MWRTM and the Rosenkranz model (not shown). Weighting functions calculated from the automatic differentiation scheme were validated by finite difference approximations and not a model comparison. Overall, the performance of MWRTM is found to be good enough for retrieval applications.

Data and Methodology

The optimal estimation and linear regression methods are used for retrievals of temperature and humidity profiles based on data collected in Innsbruck. A radiosonde climatology provides reference profiles for the evaluation of the performance of the retrieval techniques based on simulated brightness temperatures and actual radiometer data. Additional prior information is provided by forecasts and analyses of a regional numerical weather prediction model. Figure 9 gives an overview of the Innsbruck region and the locations of these data sources.

20 Radiosonde Climatology

Radiosondes are launched regularly once per day from Innsbruck airport at the location of marker 1 in Figure 9. Launch time is usually between 01:00 and 04:00 UTC, which corresponds to 02:00 to 05:00 local time in winter and 03:00 to 06:00 local time during daylight saving time. Here, the climatological dataset of temperature, humidity and pressure spanning the years 1999 to 2005 and 2009 to 2012, previously used by Massaro (2013) and Meyer (2016), is extended with profiles obtained between 2013 and 2016. In total, data from 3561 radiosonde are available which are separated into a training data set of 3296 profiles and a test data set of 265 profiles which contains all data between February 2015 and January 2016. This partitioning results in the largest overlap of test data with other data sources.

Because liquid water content is not measured by any instrument during a balloon sounding, cloud water content is calculated by the empirical cloud model of Karstens et al. (1994) based on humidity and temperature information. This model is frequently used for microwave radiometer retrieval studies, e.g. by Löhnert and Maier (2012) and Martinet et al. (2015). After application of the model, 52 % of all training profiles and 53 % of all test profiles have at least one layer with cloud water content larger than zero and are therefore categorized as cloudy.

The profiles of pressure, temperature, humidity and liquid water content are linearly interpolated to the 50 levels of the retrieval grid. Figure 10 shows the climatology's mean profile of temperature and total water content together with intervals of ± 1 and ± 2 standard deviation. Temperature variability is almost constant throughout the troposphere, while water variability is greater near the surface and vanishes at the mean height of the tropopause. The latter is due to the small water vapor content in the upper atmosphere. Total water content uncertainty intervals are asymmetric because the underlying distribution is a Gaussian in logarithmic space.

Whenever the accuracy of a retrieval method is evaluated in subsequent sections, these radiosonde data are the reference that retrieved profiles are compared to. It must be noted

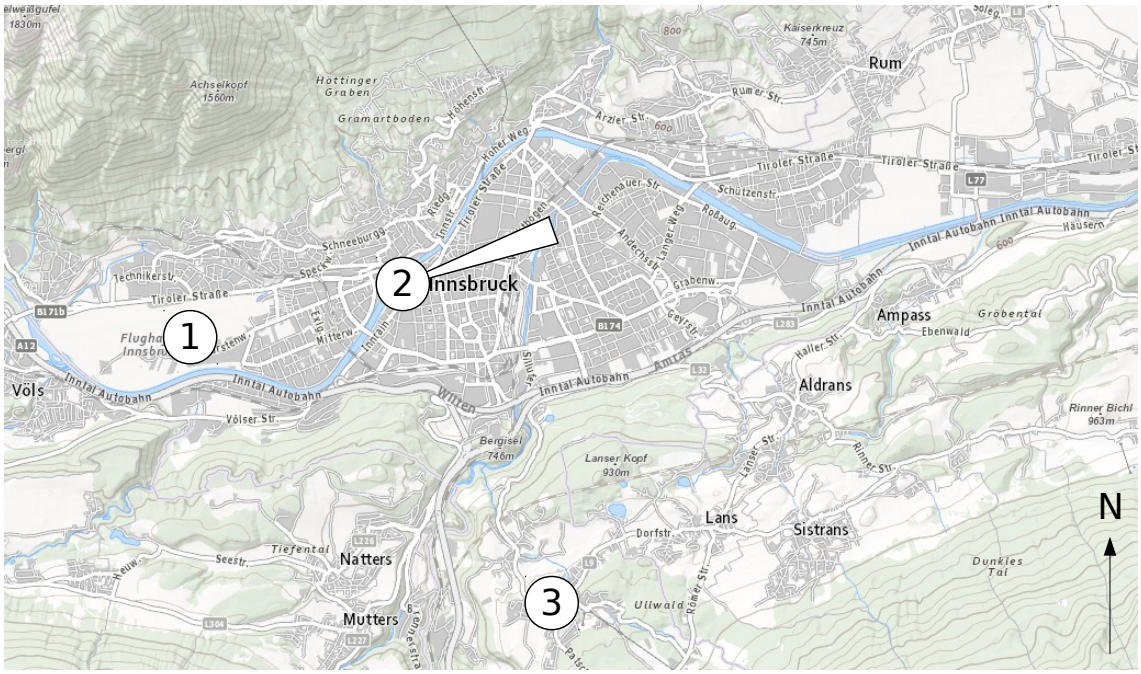

Figure 9: Data locations in the Inn-valley around Innsbruck. 1: Innsbruck Airport (577 m), radiosonde launch location. 2: HATPRO location with approximate elevation scanning direction (612 m). 3: COSMO-7 model grid point. Map source: http://www.basemap.at

that radiosondes provide point measurements, are affected by horizontal drift and take a substantial amount of time to reach their final height. These properties of balloon-borne sensors raise concerns regarding the representativity of the measured vertical profiles for a fixed location. Only 4 % of all radiosonde profiles were obtained during day time. However, because the test data set only consist of nighttime soundings this bias of climatology should not be noticeable in the subsequent performance assessments of retrieval methods based on the training set of radiosonde profiles.

21 COSMO-7 Simulated Soundings

COSMO-7 is a regional weather prediction model for western and central Europe operated by MeteoSwiss. Its terrain-following grid has a horizontal mesh size of 6.6 km and 60 vertical levels up to 21453 m altitude. Vertical profiles of temperature, humidity and cloud water content for a grid point near Innsbruck (marker 3 in Figure 9) have been obtained for the timespan 2015-02-06 to 2016-01-20 from forecasts initiated at 00 UTC. Liquid water content q_{liq} is extracted from the total cloud water content q_{cloud} based on the temperature T of the cloud according to the formula

$$q_{liq} = q_{cloud} \begin{cases} 0 & 233.15\,\mathrm{K} \leq T \\ \frac{T - 233.15\,\mathrm{K}}{40\,\mathrm{K}} & 233.15\,\mathrm{K} < T < 273.15\,\mathrm{K} \\ 1 & 273.15\,\mathrm{K} \leq T \end{cases} .$$

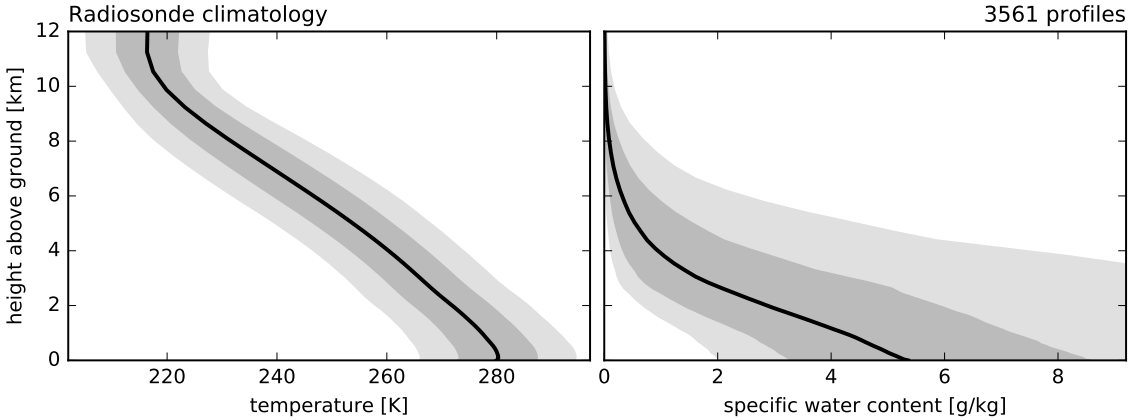

Figure 10: The radiosonde climatology of temperature (left) and total water content (vapor and liquid water) derived from soundings launched at Innsbruck airport. The black lines indicate the climatological mean values, gray shadings mark the one (dark) and two (light) standard deviation intervals obtained from sampling the distribution.

Forecasts are available in 6 h steps and up to +30 h. Profiles are linearly interpolated to match any given time of retrieval between two lead times.

Because the surface at the grid point near Innsbruck is located in the model topography more than 400 m too high at an altitude of 1093 m, the COSMO-7 profiles have to be extrapolated to the true surface height. At first it was considered to simply connect the lowest level of the COSMO-7 profile to actual surface measurements of temperature and humidity. This however disregards information from the boundary layer processes resolved by the model which would then be elevated features. Instead, profiles are stretched to the ground by replacing the original height grid of COSMO-7 by one that ends at the same altitude but starts at the true surface.

The new grid is chosen such that lower levels are stretched more than upper levels but the overall relative spacing of the model levels is approximately preserved. Because this procedure does not take temperature gradients into account, the resulting profiles likely have a significant bias relative to the true atmospheric state. To counter this, the mean bias of the model profiles with respect to the radiosondes from Innsbruck airport is removed from the dataset. Because there are only 253 times where a radiosounding exists parallel to a COSMO-7 forecast this bias is determined from the entire data set. An important implication of this choice is that COSMO-7 profiles used for the retrieval are not independent of the radiosonde reference data set. This inconvenience is accepted in favor of a more robust bias determination based on a larger data set.

In order to use the COSMO-7 forecasts as a prior for optimal estimation retrievals a covariance matrix quantifying the model's uncertainty must be known. Like for the bias, this matrix is calculated from a comparison with the reference radiosonde profiles. Figure 11 shows the uncertainty range of a model profile attached to a forecast interpolated to 2015-09-11 03:48 UTC in the lower half of the troposphere. Temperature uncertainty is

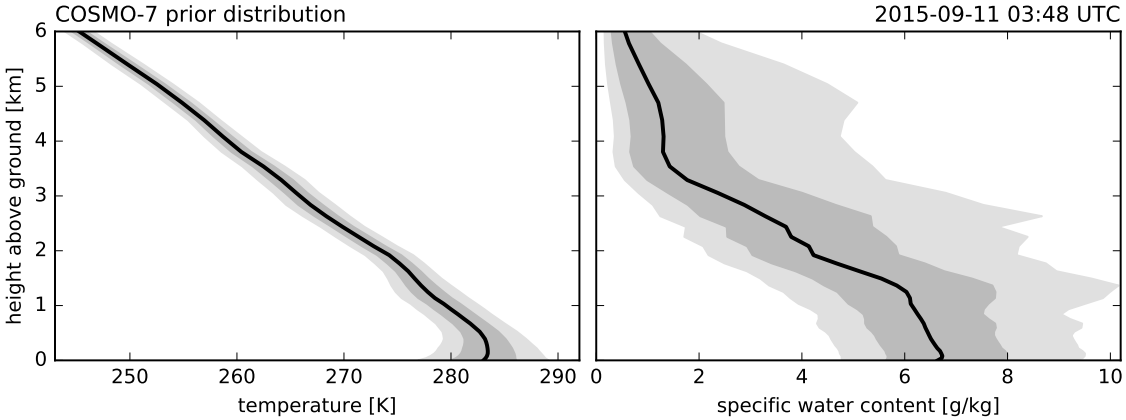

Figure 11: A COSMO-7 prior distribution of temperature (left) and total water content humidity (right) valid at 2015-02-06 03:02 UTC. The means (black lines) are interpolated from lead times +00 and +06 of the 2015-02-06 00:00 UTC run. The gray shadings show the one (dark) and two (light) standard deviation intervals obtained by sampling. Shown is the lower half of the troposphere.

highest in the lower troposphere, but smaller outside the boundary layer. Uncertainty increases above 6 km only in the tropopause region. The water content uncertainty is approximately constant up to 5 km above which it is decreasing toward zero. The example profile indicates that COSMO-7 is able to simulate temperature inversions in the lower atmosphere.

22 HATPRO Observations

The microwave radiometer operated by the the Institute of Atmospheric and Cryospheric Sciences at the University of Innsbruck is a Humidity and Temperature Profiler (HATPRO), built by the Radiometer Physics GmbH (Rose et al. 2005). It records brightness temperatures in 14 channels with a temporal resolution of 1 s. Channels in the K band correspond to frequencies of 22.24, 23.04, 23.84, 25.44, 26.24, 27.84 and 31.40 GHz. Channels in the V band correspond to frequencies of 51.26, 52.28, 53.86, 54.94, 56.66, 57.30 and 58.00 GHz. These are marked with arrows in Figure 4. The receiver antenna can be rotated to perform elevation scans. Environmental pressure, temperature and humidity are additionally measured by the instrument and it has a precipitation sensor which is translated to a rain/no rain flag in the output data.

The radiometer is set up on the roof of a university building at an altitude of 612 m above sea level. Figure 9 shows the location of the radiometer site in the Inn Valley (marker 2). The distance to the radiosonde launch site at Innsbruck airport is approximately 2.5 km. The height difference between these sites is 35 m with the HATPRO located higher. The elevation scan direction (indicated by the triangle) is down-valley, in the opposite direction of the airport.

Data from the instrument are available in the timespan from August to October 2015. In this period, HATPRO was configured to perform boundary layer scans every 10 minutes

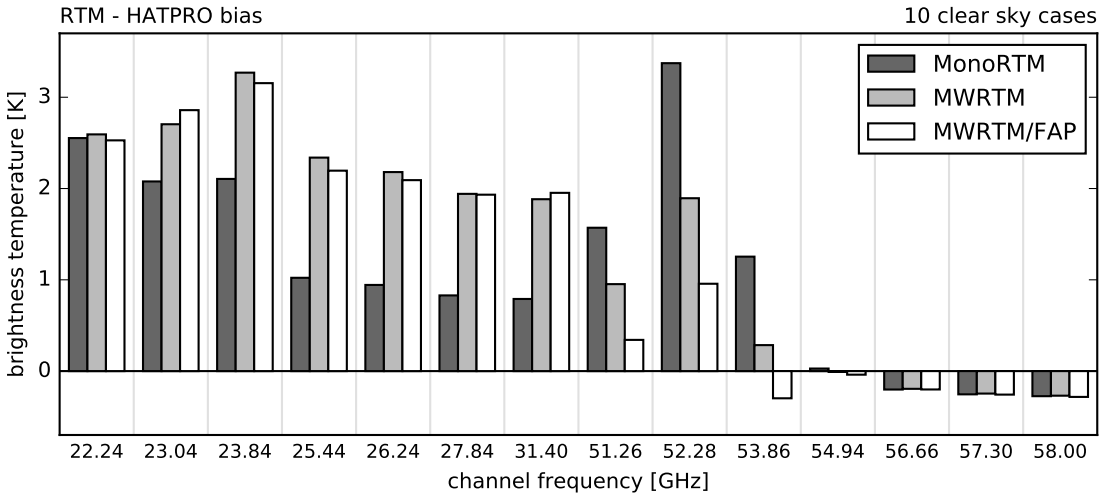

Figure 12: Radiative transfer model biases relative to HATPRO observations determined from 10 clear sky cases of matching radiometer and radiosonde measurements. Compared are simulations of MonoRTM (Clough et al. 2005) and MWRTM (section 19) with (MWRTM/FAP) and without a fast absorption predictor.

with averaging times of approximately 30 s each at zenith angles of 0°, 60°, 70.8°, 75.6°, 78.6°, 81.6°, 83.4°. HATPRO observations were matched to 26 radiosonde launches by taking the closest boundary layer scan with at most 30 minutes time difference to the launch of the radiosonde.

Mean and covariance of the error distribution (16) are governed by the bias and uncertainty of forward model error and instrumental noise. Because no observation-based estimate of the instrumental noise was available for the ACINN HATPRO, a radiometric accuracy of 0.5 K standard deviation is assumed for all channels. This value is likely too high for the instrumental error alone considering that Rose et al. (2005) claim at most 0.4 K error for 1 s integration times but fluctuations of temperature and humidity in the atmosphere cause measurement noise as well. In any case, a conservative estimate of the instrumental noise is certainly favorable to an underestimation. The bias of radiative transfer models with respect to the HATPRO observations is determined from clear sky observations matched to radiosonde ascents. Unfortunately, only 11 profiles were classified as clear-sky from the radiosonde data and web cam pictures of the Inn valley revealed that only 10 of these profiles were actually cloud free. Bias values calculated from this rather poor set of data are shown in Figure 12.

MWRTM has the largest bias values in the K band channels and simulates brightness temperatures systematically warmer than measurements. This behavior was also found by Cimini et al. (2004) for a model based on the Liebe et al. (1993) absorption description. In the V band, MWRTM with and without the fast absorption predictor is closer to the HATPRO observations than MonoRTM for the transparent channels. Biases of both models are very similar and small in the opaque V band channels. The depicted bias

values are used for the mean of the error distribution (16) in optimal estimation retrievals based on HATPRO observations.

23 Retrieval and Radiative Transfer Setup

Synthetic brightness temperature data used as input for retrievals and the training of regression models are simulated with MWRTM using the full Liebe et al. (1993) and Turner et al. (2016) absorption models based on high-resolution radiosonde data. Zero-mean random noise with a standard deviation of 0.5 K is added to all brightness temperature simulations used as input for retrievals when assessing the synthetic performance of optimal estimation and regression methods but not for regression training data (regularization is used instead). This is effective for the detection of overfitted regression models and is supposed to recreate the uncertainty of real measurements from a radiometer in order to increase the significance and representative value of synthetic retrieval performance assessments. As dicussed in section 17, off-zenith data are only used from the four most opaque V band channels (54.94, 56.66, 57.30, 58.00 GHz).

The optimal estimation scheme uses MWRTM with fast absorption predictors for the forward and Jacobian calculations. The observation error covariance $\mathbf{\Sigma}_{\boldsymbol{\epsilon}}$ is determined from a comparison of MWRTM simulations with values from the Rosenkranz (1998) model from section 19.6. Because the Rosenkranz data are based on high resolution radiosonde data, use no fast absorption predictor and are based on slightly different assumptions about the state of the upper atmosphere and liquid water content, the determined covariance matrix should include contributions from most forward model uncertainties described in section 18 and a detailed assessment of each individual term as done by Hewison (2006) is passed on. An uncorrelated instrument error of 0.5 K standard deviation is added to the thus determined observation error covariance, consistent with the noise added to brightness temperature input data from MWRTM and the instrumental error assumed for HATPRO observations in section 22.

The optimal estimation scheme is implemented with a Levenberg-Marquardt minimization scheme adapted from a form given by Rodgers (2000) (equation 5.3.6)

$$((1+\gamma)\mathbf{\Sigma}_a^{-1} + \mathbf{K}_i^\top \mathbf{\Sigma}_{\boldsymbol{\epsilon}}^{-1}\mathbf{K}_i)\,(\boldsymbol{\mu}_{i+1} - \boldsymbol{\mu}_i) \;=\; \mathbf{K}_i^\top \mathbf{\Sigma}_{\boldsymbol{\epsilon}}^{-1}(\mathbf{y} - \mathbf{F}(\boldsymbol{\mu}_i) - \boldsymbol{\mu}_{\boldsymbol{\epsilon}}) - \mathbf{\Sigma}_a^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_a)\ .$$

This equation is solved with a linear algebra routine for the difference $(\boldsymbol{\mu}_{i+1} - \boldsymbol{\mu}_i)$ which is then used to determine $\boldsymbol{\mu}_{i+1}$, the new mean of the atmospheric state vector distribution.

$\gamma \geq 0$ is a parameter controlling the step size taken with each iteration. The initial value of γ is set to 3000 and is then adapted after each iteration based on the change of the cost function (21). If the cost function increases, γ is increased by a factor of 5, if the cost function decreases, γ is halved. This strategy is not optimal but has been found to work well for retrievals here and was also used by Schneebeli (2009).

Convergence is determined from the change of the cost function (21). If the cost function does not decrease by more than 2 % over 3 iterations, the iteration state with the smallest cost function value is returned as the retrieval result. If no convergence is reached after 20 iterations, the iteration with the smallest cost function value is still returned but a quality flag is set indicating that the corresponding profile did not converge properly. The threshold of 2 % was chosen subjectively based on experiences during experimentation.

The basis functions chosen for all regression models are identity functions for the regressors and a constant function as shown in (24). The regressors for temperature retrievals are brightness temperatures from V band channels, surface pressure and surface temperature. Specific water content retrievals use brightness temperatures from K band channels, surface pressure and surface specific humidity as regressors. Regularization was found to be important during training in order to avoid overfitting. The chosen regularization maximizes the performance of the regression models on the test data set.

Retrieval Results and Discussion

In this chapter, the accuracy of linear regression and optimal estimation methods is assessed and the methods' behavior in selected case studies investigated. Performance is determined based on simulated brightness temperatures based on radiosonde data but also based on observations from the HATPRO radiometer. The measures used to quantify accuracy are the mean and standard deviation of retrieved profiles with respect to radiosonde data. This measure gives an estimation of the average overall performance of a retrieval method. Two case studies are also shown which give a more detailed look into the behavior of the optimal estimation scheme and an experiment with continuously propagated temperature retrievals is carried out for a day of radiometer observations.

The focus of research into microwave radiometers in Innsbruck in the past has been on the retrieval of vertical temperature profiles (Massaro 2013; Meyer 2016). Here, temperature profiles are also treated with additional care while humidity and cloud liquid water content are treated together as the specific water content of the atmosphere.

24 Predictions without Radiometer Data

Before the accuracy of retrieval methods is determined, the accuracy of the data these methods are based on should be assessed. Figure 13 shows the standard deviation with respect to the radiosonde test data set of forecasts from the COSMO-7 numerical weather prediction model, of the climatological mean temperature and water content and of a persistence forecast that predicts the atmospheric state at a given time as state the from the previous day. The COSMO-7 forecasts are available for two different lead times which are interpolations based on the analysis field and a 6 hour forecast (+00+06) or a 24 and a 30 hour forecast (+24+30).

The climatological standard deviation indicates that the variability of temperature in the troposphere above Innsbruck is almost constant with height and declines at altitudes where the tropopause is found. The variability of water content is highest near the surface and decreases linearly up to 5 km above which it appears to asymptotically approach zero. This behavior is mostly governed by the generally with height decreasing temperature in the tropopause which causes the saturation water vapor pressure to decrease exponentially so that less and less water can be contained by the air at greater heights. The persistence forecast strategy has a fairly constant temperature accuracy of a little more than 3 K throughout the troposphere. It has better water content accuracy than the climatology forecast in the lowest 4 km of the atmosphere and performs similar aloft.

The best accuracy exhibit COSMO-7 forecasts which are as accurate as a persistence forecast at the surface but have a standard deviation smaller than 1.5 K for all levels above 1 km height above the ground. The shorter +00+06 forecasts reach less than 1 K

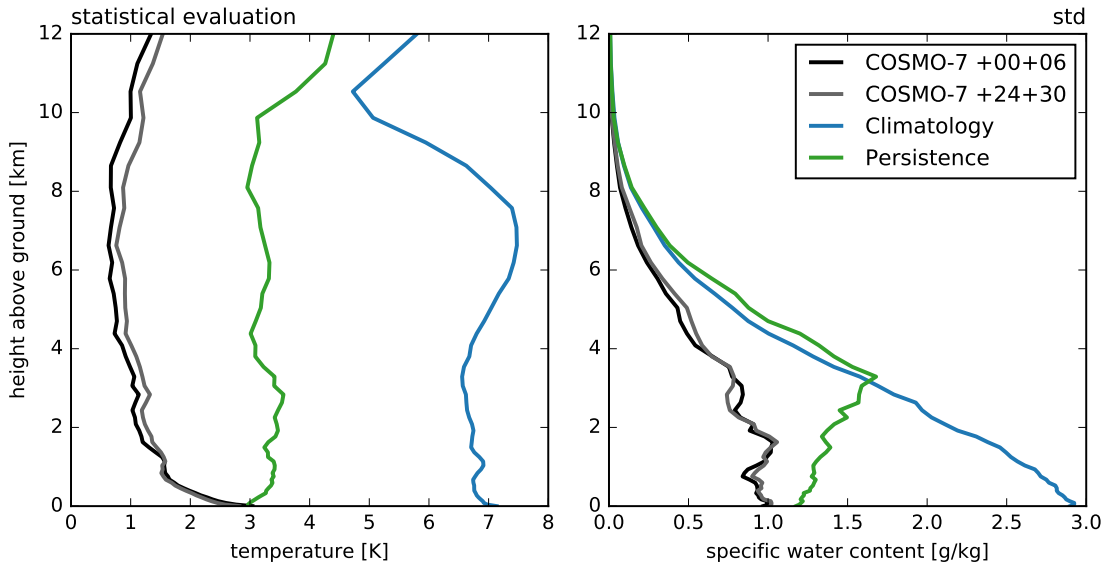

Figure 13: Standard deviations of temperature (left) and total water content (right) profiles from COSMO-7 forecasts (black and grey), the climatological mean (blue) and persistence forecasts (green) with respect to the radiosonde test data set. The +00+06 COSMO-7 profiles are interpolated to the reference radiosonde launch times based on forecasts with lead times of 0 (the analysis fields) and 6 hours while the +24+36 profiles are interpolated based on profiles with lead times of 24 and 36 hours. The evaluation of the persistence forecasts is performed on a reduced data set of 212 profiles due to missing days in the test data set.

standard deviation error in the middle and upper troposphere while the +24+30 forecasts are not as accurate. Water content forecasts from COSMO-7 do not vary in accuracy between the the two lead times evaluated here. They have a standard deviation of 1 g/kg at the surface which decreases approximately linearly with height and is overall substantially smaller than the variability of the climatology.

As expected, the COSMO-7 NWP model provides temperature forecasts of good accuracy at higher altitudes but with a comparatively poor performance in the lowest layers of the atmosphere. Because performance does not differ not substantially between the considered lead times, the +00+06 forecasts will be used as priors for all subsequent optimal estimation retrievals.

25 Statistical Retrieval Performance

The statistical performance of different retrieval schemes is first assessed based on simulations of brightness temperatures by MWRTM and then investigated with data from an actual radiometer.

25.1 Linear Regression Retrievals

Figure 14 shows the statistical evaluation of linear regression models used to retrieve

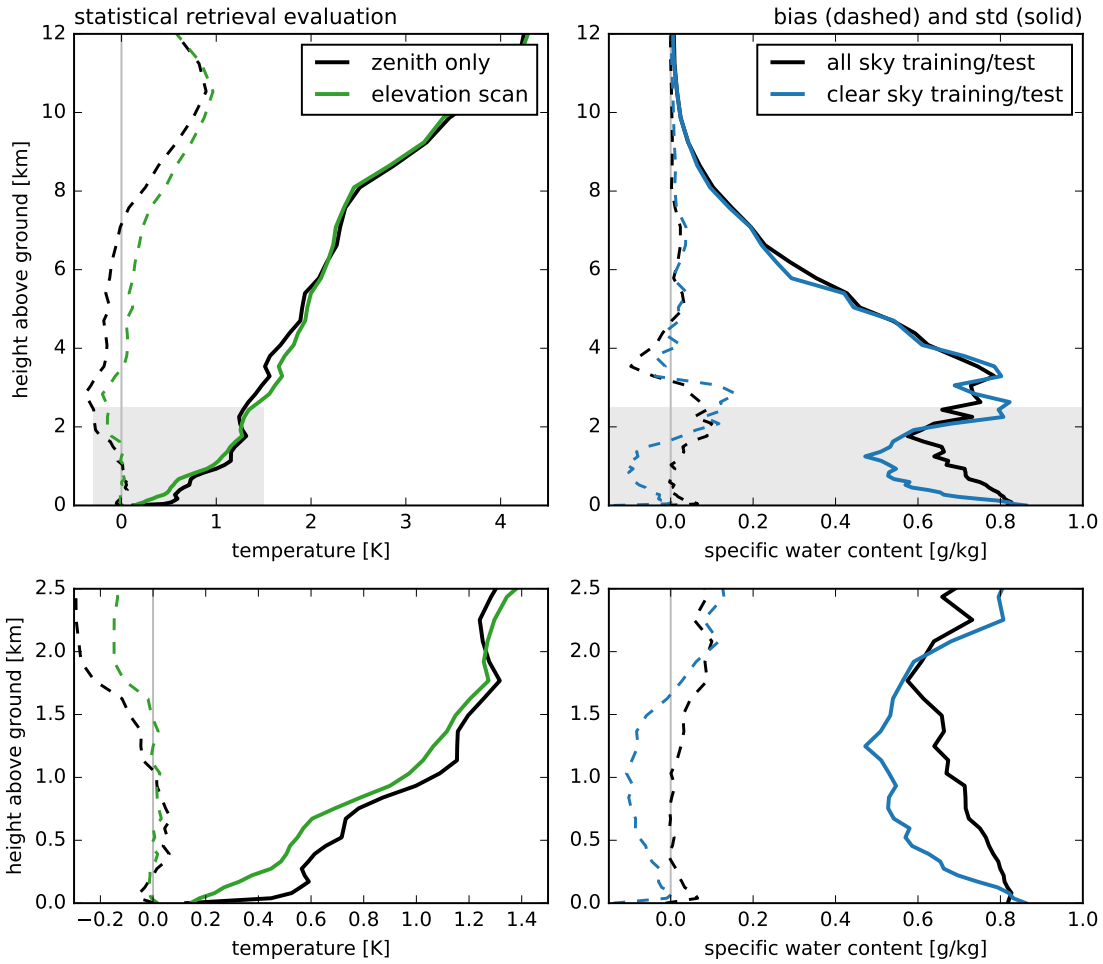

Figure 14: Performance evaluation of linear regression retrievals of temperature (left) and specific water content (right) in terms of bias (dashed) and standard deviation (solid). Temperature retrievals are based on MWRTM-simulated brightness temperatures in the V band at zenith (black) or zenith with additional elevation scan information from 4 channels (green). Water content retrievals are based on MWRTM-simulated brightness temperatures in the K band at zenith. The black model uses all training and test data while the blue model is only trained with and tested on clear sky profiles. Surface values of pressure and temperature/humidity are used as additional regressors. The lower panels are enlarged views of the shaded regions in the upper panels.

temperature and specific water content. These models use surface pressure and a surface value of the retrieved quantity as additional regressors which are taken from the lowest level of the retrieval grid for the purposes of this synthetic performance evaluation.

For temperature retrievals accuracy is generally decreasing with height starting from an accuracy of less than 0.4 K at the surface and reaching 3 K at approximately 9 km height above ground. The use of elevation scan information improves the retrieval accuracy in the lowest 2 km and reduces the bias in the middle troposphere. Accuracy of the model using elevation scan data is worse between 2 and 5 km. This issue could be resolved by gradually removing elevation scan brightness temperatures as regressors with height as

the information content of the boundary layer scans is restricted to the lower atmosphere.

The regression retrievals of temperature have no significant bias in the lowest 1 km of the atmosphere where their standard deviations are less than 1 K. Accuracy decreases to 2 K at 4 km height and increases further with height throughout the rest of the troposphere. The accuracies found by Massaro et al. (2015) and Meyer (2016) for linear regression models used with Innsbruck data were of similar magnitude although the decrease of accuracy of their regression models was not as fast as the one observed here. Possible reasons for this discrepancy are the use of regression models with quadratic terms by these authors and different assumptions about noise added to the test data.

The regression retrievals have higher accuracy for water content in the lowest 2 km than COSMO-7 forecasts but do not perform better than the NWP model at higher levels. If a regression model is trained with clear sky cases only, it shows better accuracy in the lowest 2 km of the atmosphere than an all-sky trained model when evaluated on clear sky cases only but the specialized model has an increased bias. Specializing on cloudiness for temperature models caused no change of accuracy or bias of the retrieved profiles.

25.2 Optimal Estimation Retrievals with COSMO-7 Prior

Figure 15 shows the statistical evaluation of retrievals by an optimal estimation scheme using COSMO-7 forecasts as a prior for the atmospheric state variables. Retrievals with all brightness temperature information were found to converge 96 % of the time with an average of 11.9 iterations. Retrieval calculations took between 8 and 15 seconds on a modern multi core processor showing that MWRTM simulations are fast enough for operational retrieval applications.

The accuracy of temperature profiles obtained from this method is less than 1 K throughout the troposphere with the exception of a few layers at 3 km height above ground. Comparison with the COSMO-7 performance shows that the optimal estimation scheme relies on the temperature forecast of the NWP model exclusively for all levels above 2.5 km. The radiometer information however improve accuracy compared to the model output substantially in the lower levels. The use of zenith observations alone reduces the temperature accuracy by approximately 0.2 K in the lowest 2 km which was also found for the linear regression model. Performance of the retrieval scheme is worst at heights between 1.5 and 3 km where the radiometer information content decreases and the NWP model forecasts are not as accurate as in the upper troposphere. The temperature bias of the optimal estimation retrievals is very low. This can partially be attributed to the bias-corrected COSMO-7 priors.

The temperature retrieval accuracy can be compared to studies by Cimini et al. (2011) and Martinet et al. (2015) who used optimal estimation techniques together with analyses and forecasts from a regional NWP model respectively. An accuracy of 1 K and better was found in both studies for the troposphere. The highest standard deviation is located

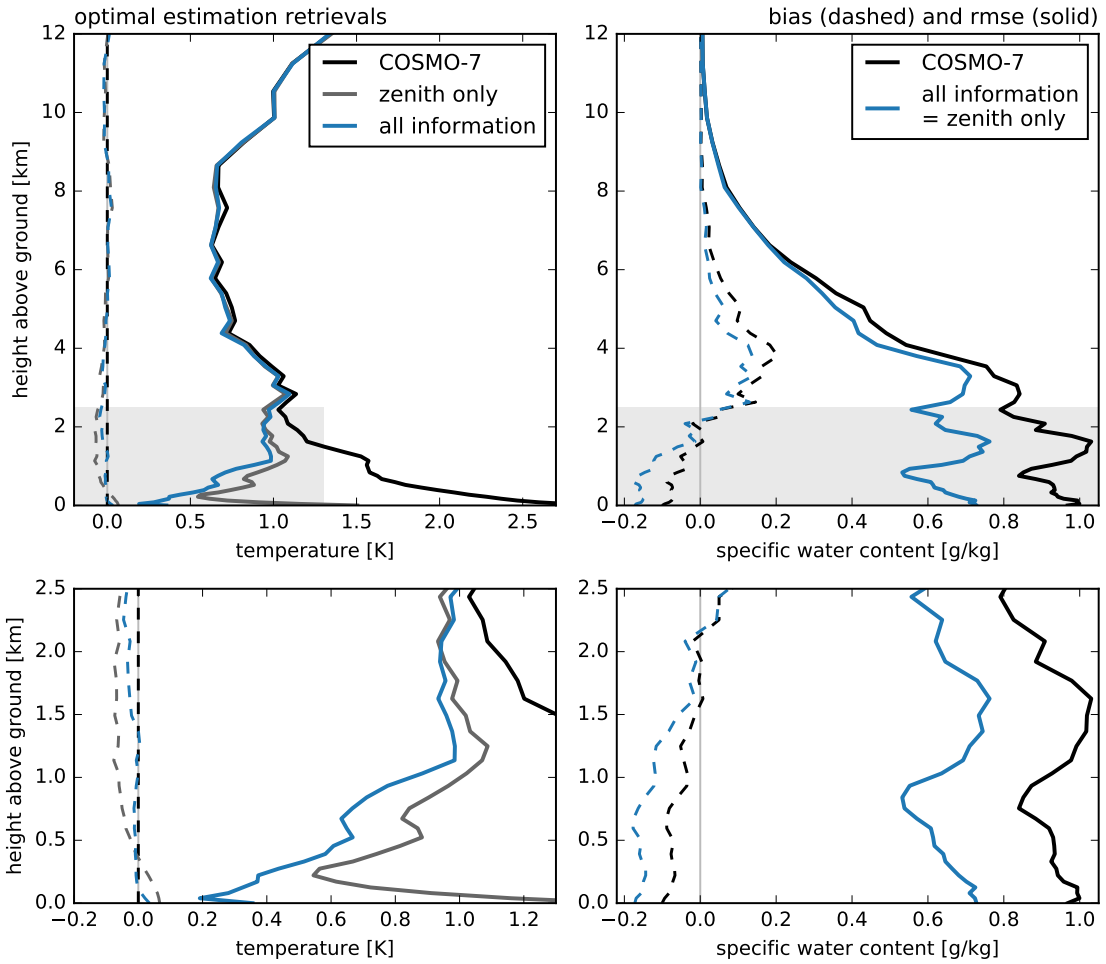

Figure 15: Performance evaluation of optimal estimation retrievals of temperature (left) and specific water content (right) in terms of bias (dashed) and standard deviation (solid). An evaluation of COSMO-7 profiles interpolated from analysis fields and 6 h forecasts is shown for comparison in black. The optimal estimation retrievals are based on all available brightness temperature information (simulated by MWRTM) (blue) or only zenith data (gray, only relevant for temperature retrievals). The lower panels are enlarged views of the shaded regions in the upper panels.

around 2 km height in the evaluation of Cimini et al. (2011) and around 1.7 km height in the study of Martinet et al. (2015) similar to the results shown here.

The radiometer observations improve the water content accuracy of COSMO-7 forecasts by more than 0.2 g/kg in the lowest 3 km of the atmosphere when incorporated into the prior information by the optimal estimation technique and have a positive effect up to 6 km height above ground. Bias values of water content are however higher below 2 km compared to the bias of the COSMO-7 forecasts. Note that the COSMO-7 bias of water content was corrected in logarithmic space during the processing of the model output and is therefore not zero in a non-logarithmic atmospheric state space.

25.3 Combined Approaches

It is of interest to see if a combination of regression and optimal estimation methods exists that improves upon the retrieval result of optimal estimation alone. Two schemes were set up to investigate this, one using the atmospheric state retrieved from a linear regression model as the first guess of the iteration procedure and one using a prior distribution constructed from the error characteristics of a linear regression model using COSMO-7 forecasts as first guesses of the iteration.

As seen in Figure 16 from the black and gray curves, temperature accuracy is not improved in the lowest 1.5 km by either combined approach. The scheme using a prior distribution based on regression retrievals has no better accuracy than regression alone at levels higher than 1.5 K. Water content retrievals with regression-derived first guesses show no significant improvements of statistical accuracy too. The retrieval scheme with the regression prior inherits the smaller bias of water content in the lower troposphere.

Also shown in Figure 16 is the performance evaluation of a linear regression method using COSMO-7 forecast data from the middle and upper troposphere for temperature and water content data in the lower atmosphere. The temperature forecasts make a big difference in terms of accuracy and bias for the temperature retrievals at height. The regression model is however not able to use the full potential of the COSMO-7 data at height and still has less statistical accuracy than the optimal estimation scheme or even the COSMO-7 forecasts by themselves. Reducing the number of brightness temperature regressors at higher levels would likely amend this issue.

Because the COSMO-7 forecasts are not more accurate in the lower atmosphere than linear regression retrievals of water content alone, no improvement is seen when incorporating such forecasts into a linear regression model.

25.4 Retrievals with Real Radiometer Data

Retrieval results from elevation scan brightness temperature measurements by the HATPRO radiometer in Innsbruck are evaluated in Figure 17. Temperature profiles obtained from optimal estimation have similar accuracy to the ones determined from synthetic retrievals, with highest standard deviations at heights around 2 km and overall accuracy better than 1 K almost everywhere. A notable difference from the synthetic retrievals is the standard deviation maximum in the 300 m directly above the ground. A similar spike in standard deviation is found for the water content at heights up to 600 m. Considering the differing locations and altitudes of the radiometer site and the radiosonde launch site, it seems likely that this is a representativity issue. The balloon probably experiences a different nocturnal boundary layer profiles near the airport than the radiometer which is located higher and in an urban area.

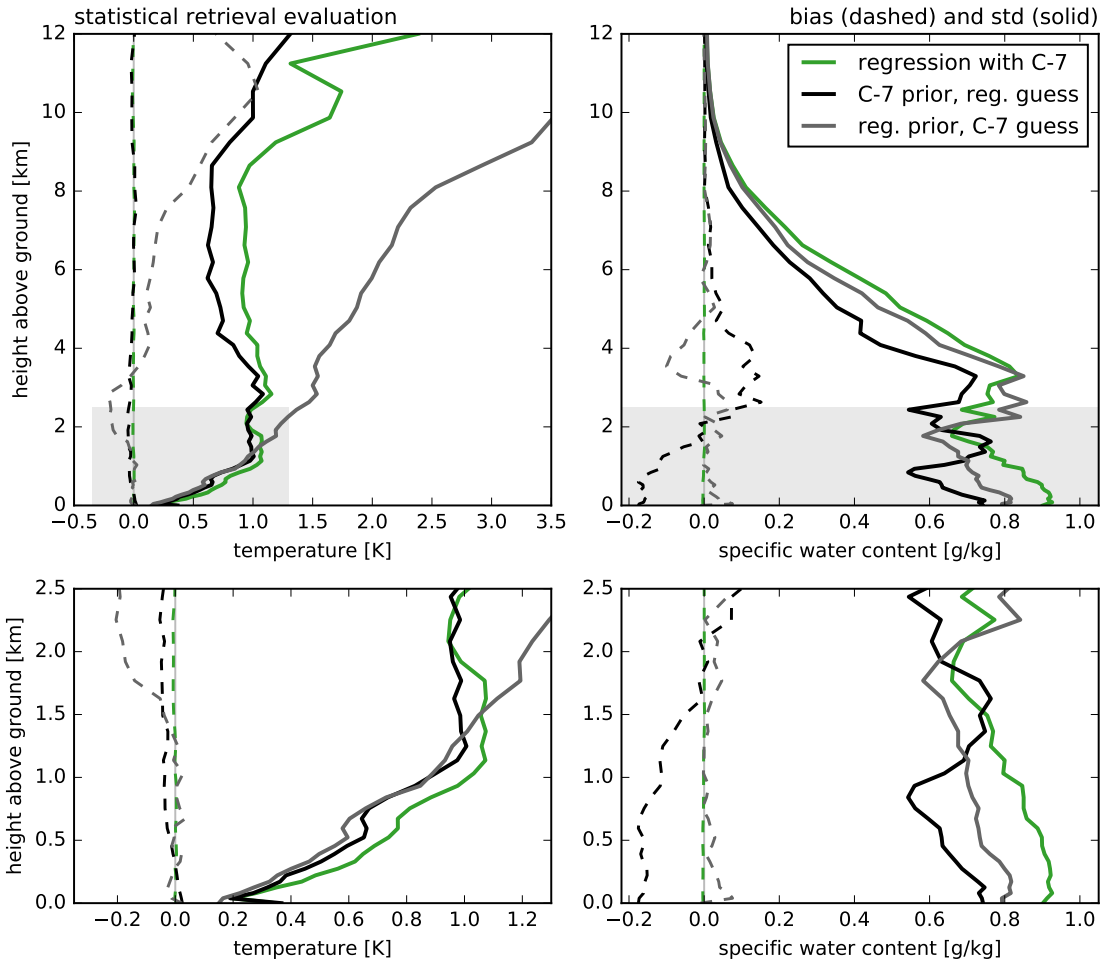

Figure 16: Performance evaluation of combined linear regression/optimal estimation retrievals of temperature (left) and specific water content (right) using either prior distributions from COSMO-7 forecasts and a first guess determined from a regression model (black) or a prior constructed from a set of linear regression model retrievals and a COSMO-7 forecast as the first guess (gray). The green model is a linear regression model using COSMO-7 forecasted temperatures from the 2104, 2694, 3448, 4415, 5651, 7235, 9262 and 11857 m levels or specific water content from the 612, 783, 1003, 1284, 1644 levels as additional regressors. Because the regression model can only be trained on the data set, 4-fold cross validation was used for the performance assessment. The lower panels are enlarged views of the shaded regions in the upper panels.

Bias values of the statistical temperature and water content comparison are high, particularly those of temperature in the lowest 2 km. The gray curve shows that the bias correction determined in section 22 affects temperature retrievals strongest in the lowest 500 m of the atmosphere and reduces the bias of water content retrievals by up to 0.5 g/kg in the lower troposphere. The same behavior is found for the performance of a linear regression model (not shown): if the brightness temperatures from HATPRO are not treated with a bias correction, water content retrievals are strongly biased towards higher amounts.

Definitive conclusions from these results cannot be drawn due to the small sample size of

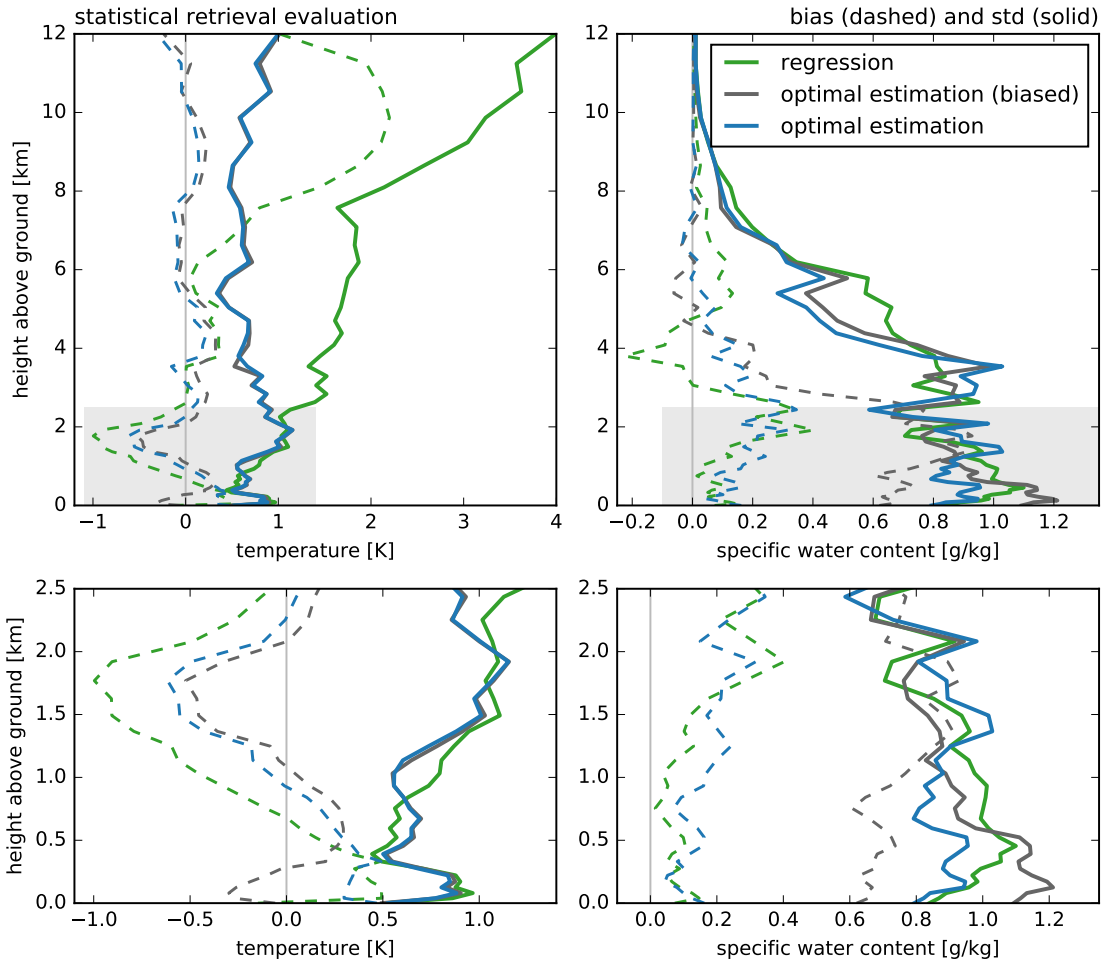

Figure 17: Performance evaluation of linear regression (green) and optimal estimation (blue, gray) retrievals of temperature and specific water content based on brightness temperature measurements from a radiometer (HATPRO). The test data set are 26 radiosonde ascents with simultaneous radiometer observations. COSMO-7 forecasts are used as the prior and first guess of the optimal estimation retrievals. The regression models use brightness temperatures and surface pressure and temperature/humidity observations as regressors. The bias between MWRTM and HATPRO determined in section 22 has been corrected for the green and blue models but not for the gray. The lower panels are enlarged views of the shaded regions in the upper panels.

data. The results indicate however that the retrieval methods, and particularly the radiative transfer model MWRTM, are useful when applied to actual radiometer measurements. For temperature retrievals comparable accuracy can be achieved to the synthetic statistical performance. The accuracy of water content retrievals is not as good with standard deviation values of 1 g/kg in the lower troposphere.

26 Selected Case Studies

Now that the general reasonability of retrieval results from the optimal estimation method has been shown, its behavior is examined for two selected cases: a ground-based and an

elevated temperature inversion. These test scenarios are commonly investigated (e.g. by Hewison et al. 2004; Crewell and Löhnert 2007) and have been at the focus of research in Innsbruck before (Massaro 2013; Meyer 2016). The presentation of results is concluded with a visualization of a boundary layer evolution over 24 h as retrieved by a linear regression and optimal estimation method from HATPRO data.

26.1 Ground-based Temperature Inversion

A ground-based temperature inversion of 5 K strength was measured during a radiosonde ascent started on the 2015-10-28 at 02:15:05 UTC. Aside from some cirrus no liquid clouds existed that night at Innsbruck[7].

Figure 18 shows temperature and water content retrievals from a linear regression model and an optimal estimation scheme based on a COSMO-7 prior. Both methods are able to retrieve a temperature inversion of appropriate strength although the regression model places the maximum about 100 m too high. The optimal estimation scheme is better able to retrieve the temperature curve above the maximum and starts to deviate from the radiosonde profile at 1.5 km height above ground while the regression model is consistently about 1 K warmer than the radiosounding up to the weak second inversion at 2 km which both methods fail to capture. It is however included in the one standard deviation uncertainty range of the optimal estimation retrieval while no uncertainty estimate is available for the regression model.

Figure 19 shows temperature profiles from intermediate steps of the optimal estimation iterative procedure. The mean of the COSMO-7 prior from which the iteration starts has a ground-based temperature inversion which is rather weak and too close to the ground compared to the radiosonde profile. In the first 6 iterations, the optimal estimation retrieval strengthens this inversion by 2 K and lifts its maximum to the appropriate height. At this stage temperatures are significantly lower than measured by the radiosonde. The magnitude of the inversion is adjusted with the next 7 iterations that add another 2 K to the maximum and adjust the temperature above the inversion to values similar to the radiosonde profile. Convergence is then reached at iteration 13.

The water content profile has a moist layer from 2 to 4 km that is not captured by any of the methods. Both models resort to a curve that has too high values of humidity at lower levels and too low values at higher levels. The moist layer is covered by the uncertainty range of the optimal estimation scheme but the dryer layer below falls outside the one standard deviation range.

[7] These information are based on web cam pictures found at http://www.foto-webcam.eu/webcam/innsbruck/

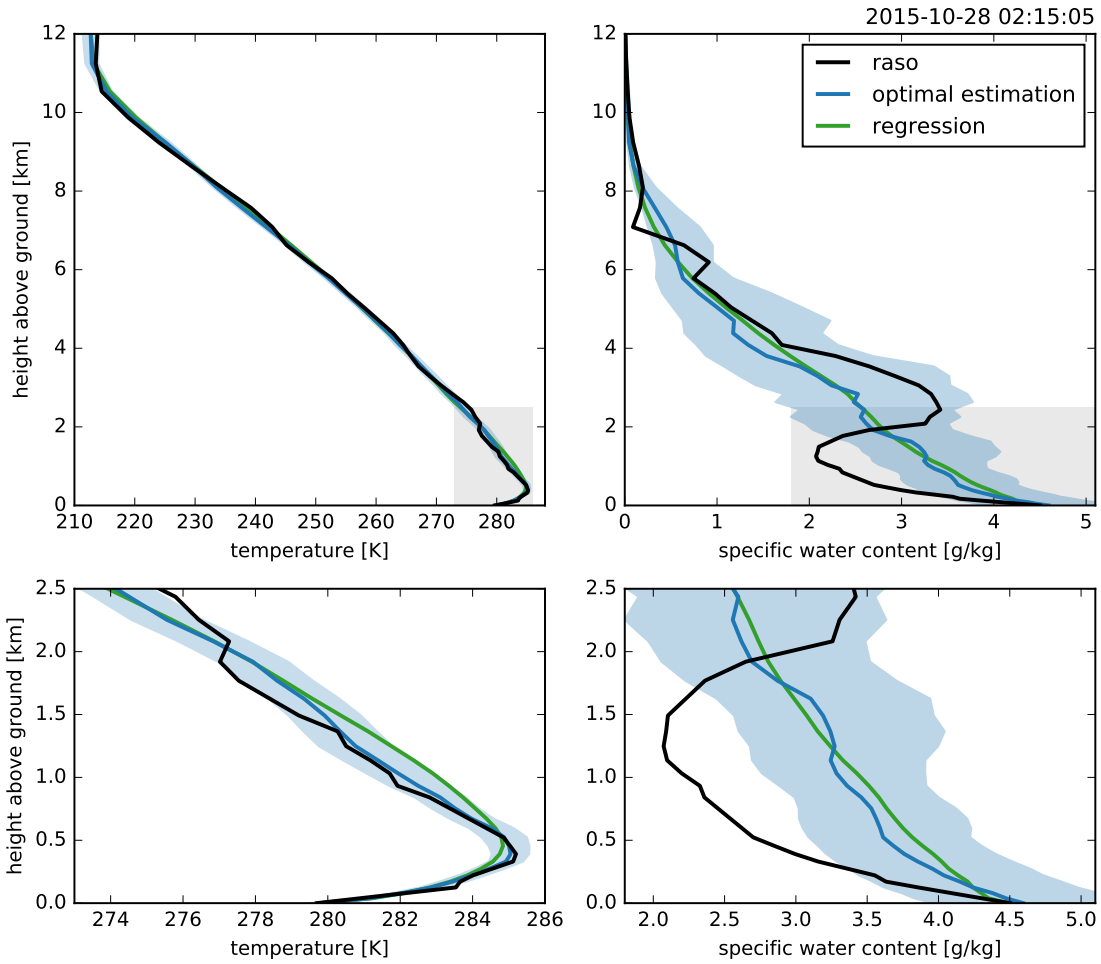

Figure 18: Retrievals of temperature (left) and specific water content (right) by a linear regression model (green) and the optimal estimation technique (blue) compared to the radiosounding (black) from which brightness temperatures were simulated by MWRTM. The radiosonde was launched on 2015-10-28 at 02:15:05 UTC from Innsbruck airport. The blue shaded region shows the ± one standard deviation uncertainty bounds obtained from sampling the posterior distribution of the optimal estimation retrieval. The lower panels are enlarged views of the shaded regions in the upper panels.

26.2 Elevated Temperature Inversion

The second case of interest is an elevated temperature inversion of 2 K magnitude measured by a radiosonde launched on 2015-09-11 at 03:48:00 UTC at Innsbruck airport. Web cam pictures of the night show patches of low-level clouds at the height of this inversion but a full stratus deck is never built up. The radiosonde profile was classified as cloudy based on its measured humidity values at the inversion height.

The left panel of Figure 20 shows that the COSMO-7 prior mean does not have an elevated inversion and much warmer temperatures in the lowest 500 m above the ground. The optimal estimation retrieval is able to derive an elevated inversion at the appropriate height from this first guess but fails to adequately capture its magnitude. A regression retrieval

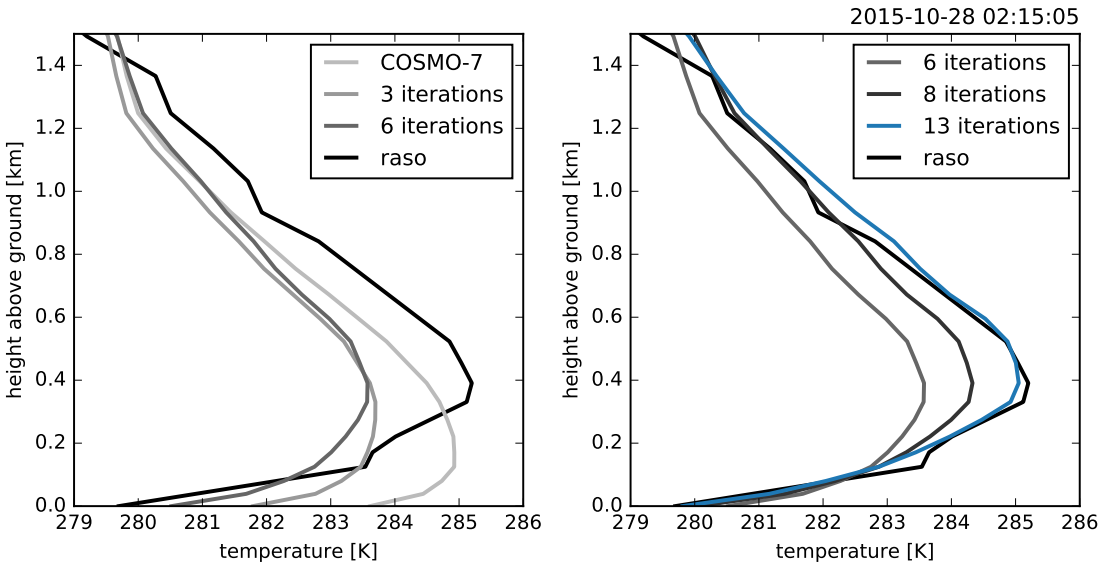

Figure 19: Visualization of intermediate profiles of temperature from the iteration procedure of the optimal estimation retrieval from Figure 18. The line labeled COSMO-7 shows the first guess and the blue line the profile at convergence. The profile at iteration 6 is shown in both plots for comparison purposes.

also indicates an elevated inversion with its smoothed out S-shape but the magnitude is even weaker than that of the optimal estimation profile.

If it were known that an elevated temperature inversion exists, e.g. by seeing that a partial low stratus deck has formed on web cam pictures, the first guess for the optimal estimation could be modified to include such an inversion. This has been done in the center panel of Figure 20: the COSMO-7 prior mean was disturbed by + 0.5 K in a few layers above 400 m and by - 0.5 K in the layers below 400 m. This profile has been used as the first guess for the shown optimal estimation retrieval. No improvement in the depiction of the inversion results from this change. If however the prior mean is changed together with the first guess, the elevated inversion is captured much better by the optimal estimation method (right panel). This shows that the prior distribution is important even at low levels where the retrieved profile is governed mostly by the radiometer observations. It also shows that there is much potential for retrievals of elevated temperature inversions if an appropriate prior can be found. The absolute values of the prior do not seem to matter as much as the shape of its mean which should fit that of the actual atmospheric state for best results.

26.3 Boundary Layer Evolution

Starting from the ground based inversion case discussed in section 26.1, the temperature evolution in the boundary layer has been retrieved by an optimal estimation scheme based on HATPRO elevation scan observations of 10 minute resolution. Instead of using

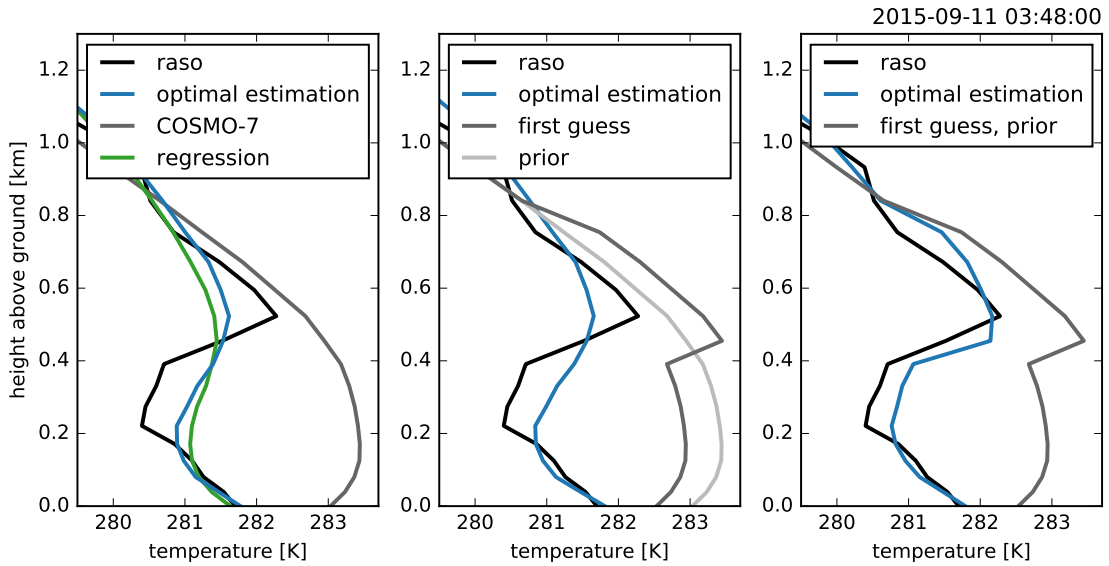

Figure 20: Optimal estimation retrievals of temperature for an elevated boundary layer inversion case from 2015-09-11 at 03:48:00 UTC (time of radiosonde launch). The right panel shows the temperature profile (blue) retrieved default the model setup with a prior and first guess from a COSMO-7 forecast (gray) as well as a retrieval from a linear regression model (green). The middle panel shows an optimal estimation retrieval (blue) with the same COSMO-7 prior (now light gray) but a first guess that has been modified to include an elevated temperature inversion of 1 K strength (dark gray). The right panel shows an optimal estimation retrieval (blue) with a prior that has its covariance from the COSMO-7 forecast but whose mean is the first guess from the middle panel (dark gray) which is also the first guess here.

a COSMO-7 prior and first guess, the scheme starts with the radiosonde profile from 2015-10-28 02:15:05 UTC and retrieves temperature from each HATPRO observation starting from the atmospheric state retrieved from the previous measurement. Because a covariance for the prior is needed and cannot be sampled from any available data, the COSMO-7 covariance is chosen as the best available alternative. Such a continuous retrieval scheme could be used operationally when no NWP forecasts are available. It should however only be used for profiling of the boundary layer because the information content of the radiometer is small at mid- and upper-tropospheric levels and no reasonable evolution of the profile will take place there.

The retrieved temperature field is visualized in the right panel of Figure 21. It can be compared to the temperature field retrieved by a linear regression model shown in the left panel. Overall, the fields of both retrieval schemes are similar. The nocturnal ground-based inversion is dissolved around 9:00 UTC in both fields and a new inversion forms after 19:00 UTC. The radiosonde profile from 2015-10-29 02:15 UTC, i.e. the end of the shown timespan, has a ground-based temperature inversion of 5 K strength with a maximum of 285 K at approximately 400 m height above ground (not shown). Both retrieval schemes do not show such properties but the representativity issue found in section 25.4 has to be kept in mind and this discrepancy might not be a failure of the retrieval methods.

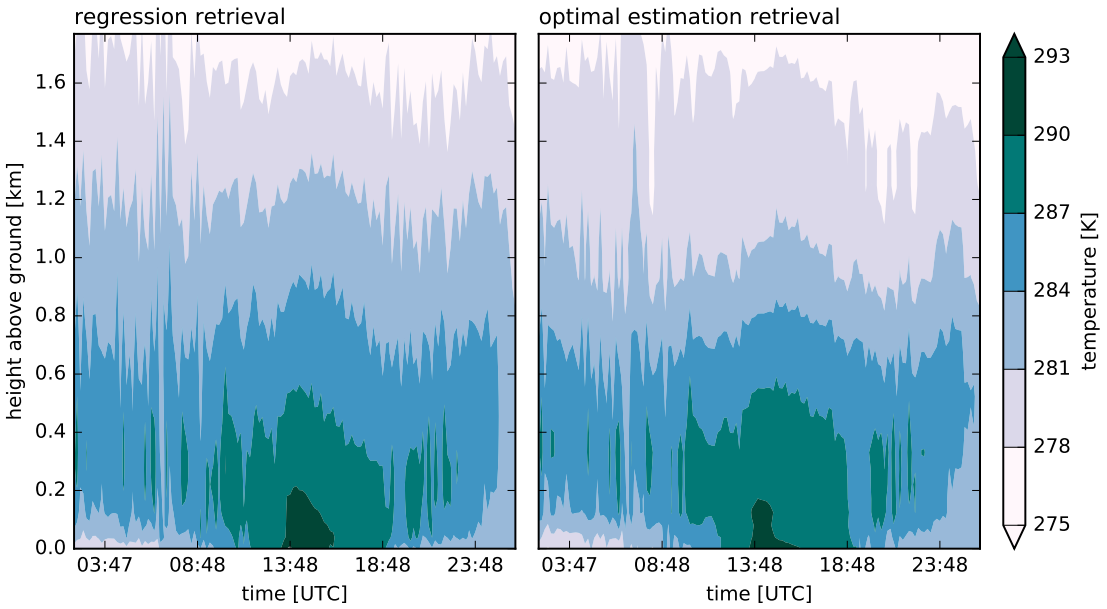

Figure 21: Visualization of the temperature evolution as retrieved from a linear regression model (left) and the optimal estimation technique (right) based on actual radiometer observations in the time span from 2015-10-28 02:15 UTC to 2015-10-29 02:15 UTC. Profiles are retrieved with a temporal resolution of 10 minutes. The prior covariance of the optimal estimation retrievals is that of COSMO-7 forecasts and the prior means are the retrieved atmospheric states from the previous timestep. The initial profile is a radiosonde profile from Innsbruck airport.

Because no temperature soundings exist for the day, the temperature evolution has to be judged qualitatively and cannot be verified quantitatively. A temperature maximum at the ground around 16:00 local time seems reasonable and both retrieval methods agree on the timing and value of this maximum. Based on the height of the 284 K isoline, the boundary layer grows by approximately 200 m in both fields although its top is more stable in the optimal estimation retrievals. This can be seen from the distance between the 284 and 281 K isolines.

Conclusions and Outlook

27 Conclusions

An optimal estimation method for the retrieval of vertical profiles of temperature and atmospheric water content from brightness temperature measurements of a microwave radiometer has been implemented. Part of the method is a radiative transfer model for the microwave region which was also developed based on existing models for absorption of microwave radiation by oxygen, water vapor and cloud liquid water. The radiative transfer model is written in a high-level programming language, uses automatic differentiation for the calculation of exact linearizations of the radiative transfer and has been found to be sufficiently accurate for use in the optimal estimation scheme.

Retrieval results from this scheme suggest an average accuracy of temperature retrievals better than 1 K throughout the troposphere. The retrieved temperature profiles benefit greatly from forecasts of a regional weather prediction model in the middle and upper troposphere while the information contained in the radiometer observations provide complementary information in the lower atmosphere where the numerical model has its largest uncertainties.

Two case studies of nocturnal temperature inversions showed that the optimal estimation scheme is able to accurately resolve such structures if the prior information is adequate. It was found that the atmospheric features exhibited by the prior are very important for the retrieval of an elevated temperature inversion likely because it is difficult for the optimal estimation scheme to introduce a fine scale feature into the first guess due to the relatively smooth weighting functions of the radiometer channels. A general deviation of the prior from the true atmospheric state in terms of absolute values is easier to correct for the technique than misrepresented features.

No major advancements in the retrieval of atmospheric water content were achieved with the considered optimal estimation scheme. This can partially be attributed to the smaller number of radiometer observations available in the humidity-sensitive region of the microwave spectrum and to larger uncertainties in the radiative transfer due to a strong dependency on the temperature retrieval and the effects of cloud liquid water absorption.

28 Outlook

Because a lot of time was spent on the development of the radiative transfer model and the setup of the optimal estimation procedure, results presented here were rather general and only basic variants of the methods were used. Particularly the application to observations from HATPRO suffers from the small number of cases included in the performance assessment.

The foundation for experiments with the optimal estimation scheme has been laid and the case studies have shown that the method has much promise for future applications. The inclusion of additional information into the retrieval has been the focus of previous studies in Innsbruck and should be continued also with the optimal estimation technique whose prior distribution is the ideal place for such information.

A region between 1.5 and 3 km height above ground level was found in the performance evaluation where temperature retrieval accuracy is worse than in other regions of the troposphere. In this region, COSMO-7 forecasts are not highly accurate yet and the information content from radiometer observations has already decreased substantially. These heights are however in the reach of sensors placed on mountains and the observations that Massaro (2013) and Meyer (2016) included in their regression retrievals could help to improve the accuracy of this region specifically. Forecasts from a NWP model with higher horizontal and vertical resolution can also help to improve the accuracy in this region and could also provide priors that better resolve fine-scale features of the atmospheric state.

Aside from the method itself there is also a demand for better evaluation tools with which the accuracy of retrieval schemes is measured. Standard deviations and biases have little significance for the performance in specific atmospheric situations. It would be nice to have feature-based measures to better assess the resolution of inversions by the retrieved profiles. Standard deviations are also not universally comparable due to differences in the natural variability of atmospheric state variables between locations (Tan et al. 2011).

If the implemented radiative transfer model should be a part of future research in Innsbruck is questionable. While the model works, future studies could benefit from better comparability to other research if an established model like ARTS (Buehler et al. 2005) or the hopefully soon available RTTOV-gb (De Angelis et al. 2016) is used. These models are also closer to the current state of research in atmospheric radiative transfer because experts in this field contribute to the development.

References

All online references were last accessed 2016-08-07.

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

Bleisch, R. and Kämpfer, N. (2012). Adding constraints by in situ informations to optimal estimation retrievals of tropospheric water vapour profiles from microwave radiometry. *J. Quant. Spectrosc. Radiat. Transf., 113*(16), 1993–2003. doi:10.1016/j.jqsrt.2012.07.002.

Blumberg, W. G., Turner, D. D., Löhnert, U. and Castleberry, S. (2015). Ground-based temperature and humidity profiling using spectral infrared and microwave observations. Part II: Actual retrieval performance in clear-sky and cloudy conditions. *Journal of Applied Meteorology and Climatology, 54*(11), 2305–2319. doi:10.1175/JAMC-D-15-0005.1.

Brousseau, P., Desroziers, G., Bouttier, F. and Chapnik, B. (2014). A posteriori diagnostics of the impact of observations on the AROME-France convective-scale data assimilation system. *Quarterly Journal of the Royal Meteorological Society, 140*(680), 982–994. doi:10.1002/qj.2179.

Buehler, S., Eriksson, P., Kuhn, T., von Engeln, A. and Verdes, C. (2005). ARTS, the atmospheric radiative transfer simulator. *Journal of Quantitative Spectroscopy and Radiative Transfer, 91*(1), 65–93. doi:10.1016/j.jqsrt.2004.05.051.

Cadeddu, M., Peckham, G. and Gaffard, C. (2002). The vertical resolution of ground-based microwave radiometers analyzed through a multiresolution wavelet technique. *IEEE Trans. Geosci. Remote Sens., 40*(3), 531–540. doi:10.1109/TGRS.2002.1000313.

Cadeddu, M. P., Liljegren, J. C. and Turner, D. D. (2013). The atmospheric radiation measurement (ARM) program network of microwave radiometers: Instrumentation, data, and retrievals. *Atmospheric Measurement Techniques, 6*(9), 2359–2372. doi:10.5194/amt-6-2359-2013.

Chan, P. W. (2010). The Use of Neural Network Retrieval for Thermodynamic Profiles of a Ground-based Microwave Radiometer. In *11th Spec. Meet. Microw. Radiom. Remote Sens. Environ.* . Washington, D. C..

Chan, P. W. and Lee, Y. F. (2013). Application of brightness temperature data from a ground-based microwave radiometer to issue low-level windshear alert over Hong Kong International Airport. *, 3*(July), 457–466.

Churnside, J. H., Stermitz, T. A. and Schroeder, J. A. (1994). Temperature Profiling with Neural Network Inversion of Microwave Radiometer Data. *J. Atmos. Ocean. Technol., 11*(1), 105–109. doi:10.1175/1520-0426(1994)011<0105:TPWNNI>2.0.CO;2.

Cimini, C., Marzano, F. S. and Ciotti, P. (2004). Atmospheric Microwave Radiative Models Study Based on Ground-Based Multichannel Radiometer Observations in the 20-60 GHz Band. In *Fourteenth ARM Sci. Team Meet. Proc.*, pages 10. Albuquerque, New Mexico.

Cimini, D., Campos, E., Ware, R., Albers, S., Giuliani, G., Oreamuno, J., Joe, P., Koch, S. E., Cober, S. and Westwater, E. (2011). Thermodynamic Atmospheric Profiling During the 2010 Winter Olympics Using Ground-Based Microwave Radiometry. *IEEE Transactions on Geoscience and Remote Sensing, 49*(12), 4959–4969. doi:10.1109/TGRS.2011.2154337.

Cimini, D., Caumont, O., Lohnert, U., Alados-Arboledast, L. and Huet, T. et al. (2014). A data assimilation experiment of temperature and humidity profiles from an international network of ground-based microwave radiometers. In *2014 13th Specialist Meeting on Microwave Radiometry and Remote Sensing of the Environment (MicroRad)*, pages 81–84.

Cimini, D., Hewison, T. J., Martin, L., Güldner, J., Gaffard, C. and Marzano, F. S. (2006). Temperature and humidity profile retrievals from ground-based microwave radiometers during TUC. *Meteorol. Zeitschrift, 15*(1), 45–56. doi:10.1127/0941-2948/2006/0099.

Cimini, D., Nelson, M., Güldner, J. and Ware, R. (2015). Forecast indices from a ground-based microwave radiometer for operational meteorology. *Atmospheric Measurement Techniques, 8*(1), 315–333. doi:10.5194/amt-8-315-2015.

Clough, S., Shephard, M., Mlawer, E., Delamere, J., Iacono, M., Cady-Pereira, K., Boukabara, S. and Brown, P. (2005). Atmospheric radiative transfer modeling: a summary of the AER codes. *J. Quant. Spectrosc. Radiat. Transf., 91*(2), 233–244. doi:10.1016/j.jqsrt.2004.05.058.

Crewell, S. and Löhnert, U. (2007). Accuracy of boundary layer temperature profiles retrieved with multifrequency multiangle microwave radiometry. *IEEE Trans. Geosci. Remote Sens., 45*(7), 2195–2201. doi:10.1109/TGRS.2006.888434.

De Angelis, F., Cimini, D., Hocking, J., Martinet, P. and Kneifel, S. (2016). RTTOV-gb - Adapting the fast radiative transfer model RTTOV for the assimilation of ground-based microwave radiometer observations. *Geoscientific Model Development Discussions*, pp. 1–29. doi:10.5194/gmd-2016-65.

Deblonde, G. and English, S. (2003). One-Dimensional Variational Retrievals from SSMIS-Simulated Observations. *Journal of Applied Meteorology, 42*(10), 1406–1420. doi:10.1175/1520-0450(2003)042<1406:OVRFSO>2.0.CO;2.

Del Frate, F. and Schiavon, G. (1998). A combined natural orthogonal functions/neural network technique for the radiometric estimation of atmospheric profiles. *Radio Sci., 33*(2), 405–410. doi:10.1029/97RS02219.

Downey, A. (2013). *Think Bayes*. O'Reilly Media, Inc..

Ebell, K., Löhnert, U., Crewell, S. and Turner, D. D. (2010). On characterizing the error in a remotely sensed liquid water content profile. *Atmospheric Research, 98*(1), 57–68. doi:10.1016/j.atmosres.2010.06.002.

Eriksson, P., Jimnez, C. and Buehler, S. A. (2005). Qpack, a general tool for instrument simulation and retrieval work. *Journal of Quantitative Spectroscopy and Radiative Transfer, 91*(1), 47–64. doi:10.1016/j.jqsrt.2004.05.050.

Errico, R. M. (1997). What Is an Adjoint Model?. *Bulletin of the American Meteorological Society, 78*(11), 2577–2591. doi:10.1175/1520-0477(1997)078<2577:WIAAM>2.0.CO;2.

Griewank, A. (2012). Who Invented the Reverse Mode of Differentiation?. *Documenta Methamatica, I*(Extra Volume: Optimization Stories), 389–400.

Güldner, J. (2013). A model-based approach to adjust microwave observations for operational applications: results of a campaign at Munich Airport in winter 2011/2012. *Atmospheric Measurement Techniques, 6*(10), 2879–2891. doi:10.5194/amt-6-2879-2013.

Güldner, J. and Spänkuch, D. (2001). Remote Sensing of the Thermodynamic State of the Atmospheric Boundary Layer by Ground-Based Microwave Radiometry. *Journal of Atmospheric and Oceanic Technology, 18*(6), 925–933. doi:10.1175/1520-0426(2001)018<0925:RSOTTS>2.0.CO;2.

Hewison, T., Gaffard, C., Nash, J., Ruffieux, D. and Nater, R. et al. (2004). Monitoring Inversions From Ground-Based Remote Sensing Instruments During Temperature, Humidity, and Cloud Profiling Campaign (TUC). In *8th Specialist Meeting on Microwave Radiometry and Remote Sensing Applications* .

Hewison, T. J. (2006). *Profiling Temperature and Humidity by Ground-based Microwave Radiometers*. PhD Thesis, University of Reading.

Kadygrov, E. N., Ganshin, E. V., Miller, E. A. and Tochilkina, T. A. (2015). Ground-based microwave temperature profilers: Potential and experimental data. *Atmos. Ocean. Opt., 28*(6), 598–605. doi:10.1134/S102485601506007X.

Karstens, U., Simmer, C. and Ruprecht, E. (1994). Remote sensing of cloud liquid water. *Meteorol. Atmos. Phys., 54*(1-4), 157–171. doi:10.1007/BF01030057.

Liebe, H. J., Hufford, G. A. and Cotton, M. G. (1993). Propagation modeling of moist air and suspended water/ice particles at frequencies below 1000 GHz. In *Atmospheric Propagation Effects through Natural and Man-Made Obscurants for Visible through MM-Wave Radiation*, pages 3–11.

Löhnert, U., Crewell, S. and Simmer, C. (2004). An Integrated Approach toward Retrieving Physically Consistent Profiles of Temperature, Humidity, and Cloud Liquid Water. *J. Appl. Meteorol., 43*(9), 1295–1307. doi:10.1175/1520-0450(2004)043<1295:AIATRP>2.0.CO;2.

Löhnert, U. and Maier, O. (2012). Operational profiling of temperature using ground-based microwave radiometry at Payerne: prospects and challenges. *Atmos. Meas. Tech., 5*(5), 1121–1134. doi:10.5194/amt-5-1121-2012.

Lopez, P., Benedetti, A., Bauer, P. and Janisková, M. (2006). Experimental 2D-Var assimilation of ARM cloud and precipitation observations. *Quarterly Journal of the Royal Meteorological Society, 132*(617), 1325–1347. doi:10.1256/qj.05.24.

Madhulatha, A., Rajeevan, M., Venkat Ratnam, M., Bhate, J. and Naidu, C. V. (2013). Nowcasting severe convective activity over southeast India using ground-based microwave radiometer observations. *Journal of Geophysical Research: Atmospheres, 118*(1), 1–13. doi:10.1029/2012JD018174.

Martinet, P., Dabas, A., Donier, J.-M., Douffet, T., Garrouste, O. and Guillot, R. (2015). 1D-Var temperature retrievals from microwave radiometer and convective scale model. *Tellus A, 67*, 1–14. doi:10.3402/tellusa.v67.27925.

Massaro, G. (2013). *Temperature and humidity profiles in complex terrain by ground-based microwave radiometry*. Master's Thesis, University of Padova.

Massaro, G., Stiperski, I., Pospichal, B. and Rotach, M. W. (2015). Accuracy of retrieving temperature and humidity profiles by ground-based microwave radiometry in truly complex terrain. *Atmos. Meas. Tech., 8*(8), 3355–3367. doi:10.5194/amt-8-3355-2015.

Meyer, D. (2016). *Microwave Sensing of Atmospheric Temperature - Investigating the retrieval performance by including additional profile information*. Master's thesis, ETH Zürich.

Peckham, G. E. and Grippa, M. (2000). Improved retrieval of tropospheric temperatures from remote measurements of thermal radiation using the adiabatic lapse rate as a constraint. *Q. J. R. Meteorol. Soc., 126*(563), 749–760. doi:10.1002/qj.49712656317.

Petersen, K. B. and Pedersen, M. S. (2012). *The Matrix Cookbook.*.

Rodgers, C. D. (2000). *Inverse Methods for Atmospheric Sounding - Theory and Practice*, volume 2 of *Series on Atmospheric Oceanic and Planetary Physics*. World Scientific Publishing Co. Pte. Ltd..

Rose, T., Crewell, S., Löhnert, U. and Simmer, C. (2005). A network suitable microwave radiometer for operational monitoring of the cloudy atmosphere. *Atmospheric Research, 75*(3), 183–200. doi:10.1016/j.atmosres.2004.12.005.

Rosenkranz, P. W. (1998). Water vapor microwave continuum absorption: A comparison of measurements and models. *Radio Science, 33*(4), 919–928. doi:10.1029/98RS01182.

Sánchez, J., Posada, R., García-Ortega, E., López, L. and Marcos, J. (2013). A method to improve the accuracy of continuous measuring of vertical profiles of temperature and water vapor density by means of a ground-based microwave radiometer. *Atmos. Res., 122*, 43–54. doi:10.1016/j.atmosres.2012.10.024.

Schneebeli, M. (2009). *Advancements in Ground-Based Microwave Remote Sensing of the Troposphere - Calibration , Data Retrieval and Applications*. PhD thesis, Universität Bern.

Solheim, F., Godwin, J. R., Westwater, E. R., Han, Y., Keihm, S. J., Marsh, K. and Ware, R. (1998). Radiometric profiling of temperature, water vapor and cloud liquid water using various inversion methods. *Radio Science, 33*(2), 393–404. doi:10.1029/97RS03656.

Tan, H., Mao, J., Chen, H., Chan, P. W., Wu, D., Li, F. and Deng, T. (2011). A Study of a Retrieval Method for Temperature and Humidity Profiles from Microwave Radiometer Observations Based on Principal Component Analysis and Stepwise Regression. *J. Atmos. Ocean. Technol., 28*(3), 378–389. doi:10.1175/2010JTECHA1479.1.

Turner, D. D., Clough, S. A., Liljegren, J. C., Clothiaux, E. E., Cady-Pereira, K. E. and Gaustad, K. L. (2007). Retrieving Liquid Water Path and Precipitable Water Vapor From the Atmospheric Radiation Measurement (ARM) Microwave Radiometers. *IEEE Trans. Geosci. Remote Sens., 45*(11), 3680–3690. doi:10.1109/TGRS.2007.903703.

Turner, D. D., Kneifel, S. and Cadeddu, M. P. (2016). An Improved Liquid Water Absorption Model at Microwave Frequencies for Supercooled Liquid Water Clouds. *J. Atmos. Ocean. Technol., 33*(1), 33–44. doi:10.1175/JTECH-D-15-0074.1.

Turner, D. D. and Löhnert, U. (2014). Information Content and Uncertainties in Thermodynamic Profiles and Liquid Cloud Properties Retrieved from the Ground-Based Atmospheric Emitted Radiance Interferometer (AERI). *Journal of Applied Meteorology and Climatology, 53*(3), 752–771. doi:10.1175/JAMC-D-13-0126.1.

Turner, D. D., Mace, G. G., Löhnert, U., Ebell, K. and Comstock, J. M. (2013). Retrievals and Their Uncertainties : What Every Atmospheric Scientist and Meteorologist Should Know. *Submitted to: Bulletin of the American Meteorological Society.*

Westwater, E. R. (1972). Ground-Based Determination of Low Altitude Temperature Profiles by Microwaves. *Monthly Weather Review, 100*(1), 15–28. doi:10.1175/1520-0493(1972)100<0015:GDOLAT>2.3.CO;2.

Westwater, E. R., Crewell, S. and Matzler, C. (2004). A review of surface-based microwave and millimeter-wave radiometric remote sensing of the troposphere. *Radio Sci. Bull., 310*, 59–80.

Westwater, E. R., Crewell, S. and Mätzler, C. (2005). Surface-based Microwave and Millimeter Wave Radiometric Remote Sensing of the Troposphere: a Tutorial. *IEEE Geoscience and Remote Sensing Society Newsletter, 134*, 16–33.

Wulfmeyer, V., Hardesty, R. M., Turner, D. D., Behrendt, A., Cadeddu, M. P., Di Girolamo, P., Schlüssel, P., Van Baelen, J. and Zus, F. (2015). A review of the remote sensing of lower tropospheric thermodynamic profiles and its indispensable role for the understanding and the simulation of water and energy cycles. *Reviews of Geophysics, 53*(3), 819–895.

Xu, G., Ware, R. S., Zhang, W., Feng, G., Liao, K. and Liu, Y. (2014). Effect of off-zenith observations on reducing the impact of precipitation on ground-based microwave radiometer measurement accuracy. *Atmos. Res., 140-141*(3), 85–94. doi:10.1016/j .atmosres.2014.01.021.

Xu, G., Xi, B., Zhang, W., Cui, C., Dong, X., Liu, Y. and Yan, G. (2015). Comparison of atmospheric profiles between microwave radiometer retrievals and radiosonde soundings. *Journal of Geophysical Research: Atmospheres, 120*(19), 10,313–10,323. doi:10.1002/2015JD023438.

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Alle Stellen, die wörtlich oder inhaltlich den angegebenen Quellen entnommen wurden, sind als solche kenntlich gemacht.

Die vorliegende Arbeit wurde bisher in gleicher oder ähnlicher Form noch nicht als Magister-/Master-/Diplomarbeit/Dissertation eingereicht.

Innsbruck, 2016-08-12

Christopher Polster