

We revised the manuscript in line with the recommendations from the referees. In the following we will list the most important modification with respect to the AMTD version of the manuscript. A pdf manuscript compiled with LatexDiff is attached to this “Reply to the Editor”.

We have modified Section 5. The intention is to better combine the part of the manuscript that shows the error estimation with the part that presents the comparison. We now much better discuss to what extent the agreement between MUSICA IASI and GRUAN is in line with the expected errors. We provide bias, scatter and  $R^2$  values for comparison of individual ensembles and for all 100 radiosondes. We empirically prove that EUMETSAT L2 temperature is indeed the leading error source and discuss possible reasons for observed biases.

We have adjusted the notation in the mathematical formulae to the recommendation of TUNER.

We very briefly discussed the limits of a comparison study using only 100 independent observations.

We better describe the collocation criteria and briefly mention the inconsistencies that can occur when using EUMETSAT L2 data belonging to different PPF software versions.

We add a reference to a publication that presents AIRS water vapour profiles and to another publication that discusses the evaluation of AIRS water vapour profiles by means of radiosonde data.

# Evaluation of MUSICA ~~MetOp~~/IASI tropospheric water vapour profiles by theoretical error assessments and comparisons to GRUAN Vaisala RS92 measurements

Christian Borger<sup>1,a</sup>, Matthias Schneider<sup>1</sup>, Benjamin Ertl<sup>1,2</sup>, Frank Hase<sup>1</sup>, Omaira E. García<sup>3</sup>, Michael Sommer<sup>4</sup>, Michael Höpfner<sup>1</sup>, Stephen A. Tjemkes<sup>5</sup>, and Xavier Calbet<sup>6</sup>

<sup>1</sup>Institute of Meteorology and Climate Research (IMK-ASF), Karlsruhe Institute of Technology, Karlsruhe, Germany

<sup>2</sup>Steinbuch Centre for Computing (SCC), Karlsruhe Institute of Technology, Karlsruhe, Germany

<sup>3</sup>Izaña Atmospheric Research Center, Agencia Estatal de Meteorología (AEMET), Santa Cruz de Tenerife, Spain

<sup>4</sup>Deutscher Wetterdienst, Meteorologisches Observatorium Lindenberg, Richard-Aßmann-Observatorium, Am Observatorium 12, 15848, Lindenberg/Tauche, Germany

<sup>5</sup>EUMETSAT, Eumetsat Allee 1, 64295 Darmstadt, Germany

<sup>6</sup>AEMET, C/Leonardo Prieto Castro 8, Ciudad Universitaria, 28071 Madrid, Spain

<sup>a</sup>now at: Satellite Remote Sensing Group, Max Planck Institute for Chemistry, Mainz, Germany

*Correspondence to:* C. Borger and M. Schneider

(Christian.Borger@mpic.de and matthias.schneider@kit.edu)

**Abstract.** Volume mixing ratio water vapour profiles have been retrieved from IASI (Infrared Atmospheric Sounding Interferometer) spectra by using the MUSICA (MULTi-platform remote Sensing of Isotopologues for investigating the Cycle of Atmospheric water) processor. The retrievals are made for IASI observations that coincide with Vaisala RS92 radiosonde measurements performed in the framework of the GCOS (Global Climate Observing System) Reference Upper-Air Network (GRUAN) in three different climate zones: the tropics (Manus Island, 2°S), mid-latitudes (Lindenberg, 52°N) and polar regions (Sodankylä, 67°N).

The retrievals show good sensitivity with respect to the vertical H<sub>2</sub>O distribution between ~~1-2 km~~ 1 km above ground and the upper troposphere. Typical DOFS (~~degree~~ degrees of freedom for signal) values are about ~~5.5-5.6~~ for the tropics, ~~5.0-5.1~~ for summertime mid-latitudes, ~~4.0-3.8~~ for wintertime mid-latitudes, and ~~4.5-4.4~~ for summertime polar regions. The errors of the MUSICA IASI water vapour profiles have been theoretically estimated considering the contribution of ~~a large number of many different~~ uncertainty sources. For all three climate regions unrecognized cirrus clouds and uncertainties in atmospheric temperature have been identified as the most important error sources ~~; whereby the total errors are estimated to be typically and they can reach about~~ 25%.

The MUSICA IASI water vapour profiles have been compared to 100 individual ~~coinciding-coincident~~ GRUAN water vapour profiles. The systematic difference between the ~~IASI and GRUAN~~ data is within ~~12% at all altitudes~~ 11% below 12 km altitude; however, at higher altitudes the MUSICA IASI data show a dry bias with respect to the GRUAN data of up to 21%. The scatter is largest close to the surface ~~and close to the tropopause, but does never exceed 30%. The (30%), but never exceeds 21% above 1 km altitude.~~ The comparison study documents that the MUSICA ~~MetOp~~/IASI retrieval processor provides H<sub>2</sub>O profiles ~~with good accuracy and captures the that capture well the large~~ variations in H<sub>2</sub>O volume mixing ratio profiles from

~~-2 km-1 km~~ above ground up to altitudes close to the tropopause ~~with a precision that is in accordance to the theoretical error assessment.~~

. Above 5 km the observed scatter with respect to GRUAN data is in reasonable agreement with the combined MUSICA IASI and GRUAN random errors. The increased scatter at lower altitudes might be explained by surface emissivity uncertainties at the summertime continental sites of Lindenberg and Sodankylä and the upper tropospheric dry bias might suggest deficits in correctly modeling the spectroscopic line shapes of water vapour.

## 1 Introduction

Atmospheric water plays a key role for the atmospheric energy balance and temperature distribution via radiative effects (clouds and vapour) and latent heat transport. Hence the distribution and transport of atmospheric moisture is closely linked to atmospheric dynamics on all scales and understanding its spatial and temporal variations is essential for weather and climate modeling. Also, understanding the coupling between moisture transport, clouds and atmospheric dynamics is seen as a major challenge for improving atmospheric models (Stevens and Bony, 2013). In this context the global monitoring of the water vapour distribution is important, whereby the large inhomogeneity in time and space (horizontally and vertically) is particularly challenging.

In the meantime, several in situ and remote sensing measurement techniques for the observation of water vapour have been established using platforms such as surface stations, balloons, aircraft ~~or~~ and satellites. The radiative properties of water vapour enable satellite remote sensing measurements in a large range of wavelength regimes from the visible (e.g. GOME; Grossi et al., 2015) GOME (Grossi et al., 2015), near-infrared (e.g. MODIS; Gao and Kaufman, 2003), infrared (MODIS (Gao and Kaufman), thermal infrared, e.g. TES; Worden et al., 2012 and IASI; Herbin et al., 2009; Schneider and Hase, 2011) AIRS (Susskind et al., 2003), TES (Worden et al., 2012) and IASI (Herbin et al., 2009; Schneider and Hase, 2011), to the microwave (e.g. AMSU; Rosenkranz, 2000). The IASI instrument (Infrared Atmospheric Sounding Interferometer; Clerbaux et al., 2009) AMSU (Rosenkranz, 2000). The instrument IASI (Infrared Atmospheric Sounding Interferometer Clerbaux et al., 2009) aboard EUMETSAT's MetOp satellites is particularly promising: it provides global observations with high resolution and accuracy twice a day on a long-term mission for more than 14 years. Furthermore, IASI follow-up missions have already been approved guaranteeing ~~respective~~ observations until the 2030s, which will offer great opportunities for studying the atmospheric composition over long time periods.

When using satellite data in research, it is important to understand their characteristics (sensitivity/representativeness and errors). Theoretical error assessments can be used to reveal the leading error sources. Ideally these error assessments should be accompanied by empirical data validation studies, in which the remote sensing data are compared to independent high quality reference data. ~~Temperature and humidity profile reference data are~~ Radiosonde measurements are a good candidate for providing references for validating the remote sensing profiles; however, great care is needed for constraining the uncertainties in the radiosonde data (McMillin et al., 2007). In particular promising in this context are the temperature and humidity profiles produced from Vaisala RS92 radiosonde measurements in the framework of the GCOS Reference Upper-Air Network (GRUAN, www.gruan.org), a subnetwork of the Global Climate Observing System (GCOS, [://www.wmo.int/pages/prog/gcos/index.php](http://www.wmo.int/pages/prog/gcos/index.php)).

Currently GRUAN consists of about 30 reference sites and provides humidity and temperature profiles of a high and well documented quality (Dirksen et al., 2014).

In this paper we perform a detailed theoretical error assessment and an empirical validation of the water vapour profiles as generated by the MUSICA ~~MetOp/~~(MUlti-platform remote Sensing of Isotopologues for investigating the Cycle of Atmospheric water Sch  
5 retrieval processor. The retrievals are made for three different climate regions (tropics, ~~midlatitudes~~mid-latitudes, polar regions) and for coincidences with GRUAN in situ radiosonde measurements, which we use as the reference for the empirical validation study. Our investigations will give ~~a comprehensive overview of~~ an overview on the retrieval’s capability of profiling atmospheric water vapour. The paper is ~~organised~~ organized as follows: Section 2 will give a brief overview of the MUSICA ~~MetOp/~~IASI processor by describing general retrieval and error estimation principles, by presenting the particularities of the  
10 MUSICA retrieval setup and by discussing the MUSICA retrieval output. Section 3 presents the sites and time periods for which the data evaluation is performed. Section 4 shows the theoretical IASI data characterisation and Sect. 5 ~~gives presents~~ and discusses the results of the comparison between the remote sensing data and the GRUAN in situ reference data. In Sect. 6 we summarize the ~~outeome~~ outcomes of the study.

## 2 MUSICA ~~MetOp/~~IASI data

### 15 2.1 Atmospheric remote sensing retrieval principles

In this subsection we give a very brief introduction ~~into to~~ the principles of the optimal estimation retrieval method. It is a standard retrieval method in atmospheric remote sensing. For more details please refer to Rodgers (2000) and for a general introduction on vector and matrix algebra dedicated textbooks are recommended.

Atmospheric remote sensing means that the atmospheric state is retrieved from the radiation measured after having interacted  
20 with the atmosphere. This interaction of radiation with the atmosphere is modeled by a radiative transfer model (also called the forward model,  $F$ ), which enables relating the measurement vector and the atmospheric state vector by:

$$\mathbf{y} = F(\mathbf{x}, \mathbf{b}). \quad (1)$$

We measure  $\mathbf{y}$  (the measurement vector, e.g. a thermal nadir spectrum in the case of IASI) and are interested in  $\mathbf{x}$  (the atmospheric state vector). Vector  ~~$\mathbf{p}$~~   $\mathbf{b}$  represents auxiliary parameters (like surface emissivity) or instrumental characteristics (like  
25 the instrumental line shape), which are not part of the retrieval state vector. However, a direct inversion of Eq. (1) is generally not possible, because there are many atmospheric states  $\mathbf{x}$  that can explain one and the same measurement  $\mathbf{y}$ .

For solving this ill-posed problem a cost function  $J$  is set up, that combines the information provided by the measurement with a priori known characteristics of the atmospheric state:

$$\underline{J} = [\mathbf{y} - F(\mathbf{x}, \mathbf{b})]^T \underline{S_\epsilon}^{-1} [\mathbf{y} - F(\mathbf{x}, \mathbf{p})] \underline{S_{\mathbf{y}, \text{noise}}}^{-1} [\mathbf{y} - F(\mathbf{x}, \mathbf{b})] + [\mathbf{x} - \mathbf{x}_a]^T \mathbf{S}_a^{-1} [\mathbf{x} - \mathbf{x}_a]. \quad (2)$$

30 Here, the first term is a measure of the difference between the measured spectrum (represented by  $\mathbf{y}$ ) and the spectrum simulated for a given atmospheric state (represented by  $\mathbf{x}$ ), while taking into account the actual measurement noise ~~level~~ ( $S_\epsilon$  ( $S_{\mathbf{y}, \text{noise}}$

is the measurement noise covariance matrix). The second term of the cost function (Eq. 2) constrains the atmospheric solution state ( $x$ ) towards an a priori most likely state ( $x_a$ ), whereby the kind and strength of the constraint are defined by the a priori covariance matrix  $S_a$ . The constrained solution is reached at the minimum of the cost function (Eq. 2). Due to the nonlinear behavior of  $F(x, p)F(x, b)$ , the minimisation is generally achieved iteratively. For the  $(i + 1)$ th iteration it is:

$$5 \quad x_{i+1} = x_a + \underline{G_i}[\underline{y} - F(x_i, p) + \underline{K_i}(x_i - x_a)]\underline{G_i}[\underline{y} - F(x_i, b) + \underline{K_i}(x_i - x_a)]. \quad (3)$$

$K$  is the Jacobian matrix (derivatives that capture how the measurement vector will change for changes in the atmospheric state  $x$ ).  $G$  is the gain matrix (derivatives that capture how the retrieved state vector will change for changes in the measurement vector  $y$ ).  $G$  can be calculated from  $K$ ,  $S_e S_{y, noise}$  and  $S_a$  as:

$$G = (K^T \underline{S_e} \underline{S_{y, noise}}^{-1} K + S_a^{-1})^{-1} K^T \underline{S_e} \underline{S_{y, noise}}^{-1}. \quad (4)$$

10 The averaging kernel is an important component of a remote sensing retrieval and it is calculated as:

$$A = GK. \quad (5)$$

The averaging kernel  $A$  reveals how a small change of the real atmospheric state vector  $x$  affects the retrieved atmospheric state vector  $\hat{x}$ :

$$\hat{x} - x_a = A(x - x_a). \quad (6)$$

15 The propagation of errors due to parameter uncertainties  $\epsilon_p \Delta b$  can be estimated analytically with the help of the parameter Jacobian matrix  $K_p K_b$  (derivatives that capture how the measurement vector will change for changes in the parameter  $p, b$ ). According to Eq. (3), using the parameter  $p + \epsilon_p b + \Delta b$  (instead of the correct parameter  $p, b$ ) for the forward model calculations will result in an error in the atmospheric state vector of:

$$\Delta \hat{x} = -G K_p \epsilon_p K_b \Delta b. \quad (7)$$

20 The respective error covariance matrix  $S_e$  is:  $S_{\hat{x}, b}$  is:

$$\underline{S_e} \underline{S_{\hat{x}, b}} = \underline{G} \underline{K_p} \underline{S_p} \underline{K_p} \underline{K_b} \underline{S_b} \underline{K_b}^T \underline{G}^T, \quad (8)$$

where  $S_p S_b$  is the covariance matrix of the uncertainties  $\epsilon_p \Delta b$ .

Noise on the measured radiances also affects the retrievals. The error covariance matrix for noise can be analytically calculated as:

$$25 \quad \underline{S_e} \underline{S_{\hat{x}, noise}} = \underline{G} \underline{S_y} \underline{S_{y, noise}} \underline{G}^T, \quad (9)$$

where  $S_y S_{y, noise}$  is the covariance matrix for noise on the measured radiances  $y$ .

## 2.2 The MUSICA retrieval setup

The MUSICA ~~MetOp~~/IASI retrieval is based on a nadir version of the retrieval code PROFFIT (~~PROFile FIT; Hase et al., 2004~~) (PROFile FIT Hase et al., 2004) and on the corresponding radiative transfer model PRFFWD (~~PRoFitForWarD model; Hase et al., 2004~~) (PRoFitForWarD model; Hase et al., 2004) (P

The nadir code has been developed in support of the project MUSICA (MUlti-platform remote Sensing of Isotopologues for  
5 investigating the Cycle of Atmospheric water, <http://www.imk-asf.kit.edu/english/musica.php>). The PRFFWD-nadir code has been recently updated by including water continuum calculations according to the model “MT\_CKD” v2.5.2 (~~MLawer et al., 2012; Delamer~~

For the MUSICA ~~MetOp~~/IASI retrieval calculations a single broad spectral window ~~ranging~~ from  $1190\text{ cm}^{-1}$  to  $1400\text{ cm}^{-1}$  is used. The spectral signatures of  $\text{H}_2^{16}\text{O}$ ,  $\text{H}_2^{18}\text{O}$  and  $\text{H}_2^{17}\text{O}$  are fitted together as a single species ~~and HDO~~ (from now on called  $\text{H}_2\text{O}$ ) and  $^1\text{H}^2\text{H}^{16}\text{O}$  (from now on called HDO) as a separate species. Furthermore, the retrieval’s spectral window  
10 contains spectroscopic features of  $\text{CH}_4$  and  $\text{N}_2\text{O}$  as well as weak spectroscopic features of  $\text{HNO}_3$  and very weak spectroscopic features of  $\text{CO}_2$  ~~and  $\text{O}_3$~~ . All these trace gases (~~except  $\text{O}_3$~~ ) are simultaneously fitted during the retrieval process whereby the spectroscopic parameters are taken from the HITRAN 2016 database (Gordon et al., 2017) with small modifications for HDO parameters (similar to Schneider et al., 2016, the line intensity parameters of HDO have been increased by 10%).

~~The water isotopologues are retrieved~~ For the water isotopologues,  $\text{CH}_4$ ,  $\text{N}_2\text{O}$  and  $\text{HNO}_3$  profile retrievals are performed on  
15 a logarithmic scale. For  $\text{CO}_2$  and  $\text{HNO}_3$  the a priori profiles are scaled and for  $\text{CH}_4$  and  $\text{N}_2\text{O}$  an ad hoc regularisation is applied allowing profile retrievals on a logarithmic scale (García et al., 2017). A single a priori profile is used for all the retrievals for each of the different trace gases (~~Schneider et al., 2016; García et al., 2017~~), i.e. the used a priori are the same for all locations and time periods (Schneider et al., 2016; García et al., 2017). For  $\text{CH}_4$ ,  $\text{N}_2\text{O}$ ,  $\text{HNO}_3$  and  $\text{CO}_2$  the a priori profiles are averaged low latitude profiles from WACCM (Whole Atmosphere Community Climate Model-version 5, <http://waccm.acd.ucar.edu>) and  
20 are provided by NCAR (National Center for Atmospheric Research, J. Hannigan, private communication). The water vapour isotopologue a priori data are averages obtained from the isotopologue incorporated global general circulation model LMDZ (Risi et al., 2012).

The retrieval also fits the surface temperature and the atmospheric temperature profile, whereby the a priori temperatures are taken from the EUMETSAT IASI level 2 (L2) products. There is no constraint on the surface temperature. The allowed atmo-  
25 spheric temperature variations are 1 K at ground, 0.5 K in the free troposphere, and 0.75 K above the tropopause. This altitude dependency follows roughly the altitude dependency of uncertainties in the EUMETSAT IASI L2 atmospheric temperature profiles (August et al., 2012).

The MUSICA ~~MetOp~~/IASI water vapour retrieval only works for pixels that are not contaminated by clouds, whereby we rely on the IASI L2 cloud flag (we require zero for the flag “cldfrm”). Ground elevations are from GTOPO30 devel-  
30 oped by the US Geological Survey and provided by the Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC). GTOPO30 is a global digital elevation model with a horizontal grid spacing of 30-arc seconds (approximately 1 km). The land surface emissivities are from the “global database of infrared land surface emissivity” (IREMIS; <http://cimss.ssec.wisc.edu/iremish>/<http://cimss.ssec.wisc.edu/iremish/>; Seemann et al., 2008) and the sea surface emissivities are calculated according to the model of Masuda et al. (1988).

Figure 1 depicts an example of a typical radiance spectrum in the retrieval’s spectral range as measured by IASI (upper graph) and the corresponding differences compared to the simulated spectra (the residuals, lower graph). The residuals are mostly within the order of the instrument’s  $1\sigma$  measurement noise (Pequignot et al., 2008). However, there are also distinctive spectral signatures that are not well understood, specifically at  $1250\text{ cm}^{-1}$  and at  $1280\text{ cm}^{-1}$ .

- 5 For further information on the retrieval setup ~~a more detailed description is and its evolution~~ more detailed descriptions are available in Schneider and Hase (2011) ~~and Wiegele et al. (2014)~~, Wiegele et al. (2014), Schneider et al. (2016) and (García et al., 2017).

### 2.3 The MUSICA retrieval output

The output of the retrieval refers to the  $\{\ln[\text{H}_2\text{O}], \ln[\text{HDO}]\}$  basis system. In this basis system the state vector  $\mathbf{x}$  consists of the vector for the  $\text{H}_2\text{O}$  profile extended by the vector for the HDO profile:

$$10 \quad \mathbf{x} = \begin{pmatrix} \mathbf{x}_{\text{H}_2\text{O}} \\ \mathbf{x}_{\text{HDO}} \end{pmatrix}. \quad (10)$$

Correspondingly, the averaging kernel matrix  $\mathbf{A}$  has  $2 \times 2$  blocks

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}. \quad (11)$$

- $\mathbf{A}_{11}$  and  $\mathbf{A}_{22}$  describe how the retrieved  $\text{H}_2\text{O}$  and HDO states depend on the actual atmospheric  $\text{H}_2\text{O}$  and HDO variations, respectively, and  $\mathbf{A}_{12}$  and  $\mathbf{A}_{21}$  reveal the cross-dependencies of the retrieved  $\text{H}_2\text{O}$  on the actual atmospheric HDO and of  
15 the retrieved HDO on the actual atmospheric  $\text{H}_2\text{O}$ , respectively. Since  $\text{H}_2\text{O}$  and HDO vary largely in parallel, we use in the following the  $\mathbf{A}_{11} + \mathbf{A}_{12}$  as the kernel for  $\text{H}_2\text{O}$  (see also Sect. 4.3 in Barthlott et al., 2017).

Similarly, retrieval error covariance matrices consist of  $2 \times 2$  blocks where the blocks in the diagonal represent the  $\text{H}_2\text{O}$  and HDO covariances. For this study only the  $\text{H}_2\text{O}$  covariance block is of interest (i.e. we are only interested in the  $\text{H}_2\text{O}$  error covariances). The outer diagonal blocks represent the error covariances between  $\text{H}_2\text{O}$  and HDO.

## 20 3 Reference data and sites

The theoretical and empirical assessment studies are made for cloud-free IASI measurements that coincide with GRUAN processed Vaisala RS92 radiosonde measurements. Useful coincidences are defined in accordance to Pougatchev et al. (2009) and Calbet et al. (2017).

- We identified three different sites with coincidence between IASI and GRUAN measurements: Manus Island (Papua New  
25 Guinea;  $2^\circ 5'S$ ,  $146^\circ 58'E$ ) for the tropics, Lindenberg (Germany;  $52^\circ 12'N$ ,  $14^\circ 7'E$ ) for the mid-latitudes and Sodankylä (Finland;  $67^\circ 25'N$ ,  $26^\circ 35'E$ ) for the polar region.

Figure 2 depicts all the GRUAN  $\text{H}_2\text{O}$  profiles that coincide with IASI ~~observation~~ observations made for cloud-free conditions. There are 25 individual GRUAN profiles for Manus Island (during different months in 2011-2013), 58 for Lindenberg (26 during summer 2007 and 32 during different months in 2008), and 17 for Sodankylä (during summer 2007), i.e. in total

there are 100 individual GRUAN radiosonde measurements that coincide with IASI cloud-free measurements. ~~This ensemble~~  
~~These four ensembles~~ of GRUAN profiles ~~is-are~~ well representative for the highly varying tropospheric H<sub>2</sub>O distributions. In  
the free middle/upper troposphere the data show variations of up to two orders of magnitude. At the tropical site of Manus  
Island we observe up to 10 000 ppmv (at 5 km a.s.l.) and up to 1000 ppmv (at 10 km a.s.l.), whereas at the ~~mid-latitudinal~~  
5 ~~mid-latitude~~ and polar sites of Lindenberg and Sodankylä the H<sub>2</sub>O concentrations can be as small as 100 ppmv and 10 ppmv,  
respectively. In this context using ~~this ensemble-the four ensembles~~ of GRUAN data enables us to make an evaluation of the  
retrieval performance that has a good global validity.

The coincidences at the three sites are for different time periods ~~meaning that and~~ there is not a strictly uniform data set for  
creating the retrieval input files(~~: EUMETSAT L2 data are not available for all the time periods or are generated by different~~  
10 ~~EUMETSAT L2 product processing facility (PPF) software version (for~~ more details see Sects. 3.2-3.4 and the summary of  
Table 1).

### 3.1 GRUAN processed Vaisala RS92 in-situ profiles

The Vaisala RS92 radiosonde is equipped with a wire-like capacitive temperature sensor (“Thermocap”), two polymer capaci-  
tive moisture sensors (“Humicap”), a silicon-based pressure sensor and a GPS receiver to measure position, altitude and winds.  
15 Each second the RS92 transmits sensor data, which are received, processed and stored by the ground station equipment.  
The Humicap consists of a hydro-active polymer thin film as dielectric between two electrodes applied on a glass substrate. The  
humidity sensors are not covered by protective caps, but they are alternately heated to prevent icing. To prevent overheating,  
the heating of the humidity sensors is switched off below −60 °C, or above 100 hPa, whichever is reached first. Humicaps show  
good performance over a wide range of temperatures but suffer from systematic errors such as dry bias due to solar radiative  
20 heating and a response lag below −40 °C. Known main error sources affecting the humidity profile are daytime solar heat-  
ing of the Humicaps introducing a dry bias, sensor time-lag at temperatures below about −40 °C and temperature-dependent  
calibration correction.

We work with Vaisala RS92 data that have been processed by the GRUAN lead centre (<http://www.gruan.org>). The GRUAN  
data processing assures that the obtained humidity, pressure and temperature profiles are well-calibrated and highly accurate  
25 (Dirksen et al., 2014; Sommer et al., 2016).

### 3.2 Manus Island (MI)

At Manus Island we have coincidences in [2011](#), 2012 and 2013 ~~to-with~~ 25 individual GRUAN radiosonde profiles. The collo-  
cation of IASI and GRUAN measurements has been performed by EUMETSAT in the framework of a planned IASI retrieval  
comparison study (Calbet et al., 2017) [allowing a spatial and temporal window of 25 km and 30 min respectively](#).

30 For our retrieval we use the a priori temperatures (atmosphere and surface skin) as well as surface emissivities from the  
EUMETSAT IASI L2 product ~~-generated with the IASI L2 PPF software version 5~~. Since most of the ground scenes are over  
the ocean surface, the emissivity values are mainly according to the model of Masuda et al. (1988). The satellite pixels have



been carefully examined for clouds by EUMETSAT according to the cloud flags as provided in the IASI L2 data and in addition by visual inspection (Calbet et al., 2017).

### 3.3 Lindenberg 2008 (LI08)

For Lindenberg there are coincidences ~~to with~~ 32 individual GRUAN profiles in 2008 (representative ~~for of~~ all seasons). We performed the collocation and required that the satellite pixel has to be within a distance of 25 km with respect to the starting position of the radiosonde and that the satellite's pixel sensing time has to be within the sensing time period of the radiosonde.

~~Like As~~ for Manus Island we rely on the IASI L2 data for our retrieval input data (surface and atmospheric temperatures, surface emissivity, cloud filter, etc.). However, while for the ~~2012/13~~ 2011-13 time period (Manus Island) the IASI L2 data are generated with the IASI L2 PPF (~~Product Processing Facility~~) software version 5, for the 2008 retrievals we work with L2 data generated by the IASI L2 PPF software version 4. As shown in García et al. (2017) there can be inconsistencies between the MUSCA IASI products that are generated using different IASI L2 PPF software versions.

### 3.4 Lindenberg 2007 (LI07) and Sodankylä 2007 (SK07)

In 2007 we have 26 individual GRAUN profiles for Lindenberg and 17 individual GRUAN profiles for Sodankylä (details on the Sodankylä campaign are available in Calbet et al., 2011) that coincide with IASI observations. This dataset is limited to the summer observations. We performed the collocation using the same criteria as for the Lindenberg 2008 coincidences, i.e. we required that the satellite pixel has to be within a distance of 25 km with respect to the starting position of the radiosonde and that the satellite's pixel sensing time has to be within the sensing time period of the radiosonde.

For summer 2007 IASI L2 data have not been available for our retrieval input. Thus data from the radiosonde measurements have been used as the a priori temperatures ~~and this might cause some minor inconsistencies to a retrieval that uses~~. Above the top altitude of the radiosonde we use zonally and monthly averaged temperature climatologies (COSPAR International Reference Atmosphere). Using the GRUAN temperatures (and climatologies at higher altitudes) instead of the IASI L2 temperatures as a priori temperature (the a priori for the atmospheric temperatures might cause some inconsistency between the LI07 and SK07, on the one hand, and the MI and LI08 retrievals);, on the other hand.

Surface emissivities are taken from the “global database of infrared land surface emissivity” (IREMIS; [~~http://cimss.ssec.wisc.edu/iremisl/~~](http://cimss.ssec.wisc.edu/iremisl/) <http://cimss.ssec.wisc.edu/iremisl/>; Seemann et al., 2008), i.e. in agreement to the retrievals for 2008 (IASI L2 emissivities are based on IREMIS for land surfaces and use the Masuda model for sea surfaces). Because there are no IASI L2 cloud products for summer 2007, we use the radiosonde measurements and the cloud detection algorithm according to the model of Zhang et al. (2010) for identifying cloud-free situations.

## 4 Theoretical MUSICA MetOp/IASI data characterisation

### 4.1 Averaging Kernels

Figure 3 illustrates examples of H<sub>2</sub>O row kernels ( $A_{11} + A_{12}$ , logarithmic state vector entries according to Sect. 2.3) for the three reference sites. Grey lines show all row kernels and the thick coloured lines highlight the kernels for some selected altitudes. The peaks of the row kernels for the heights of 1.8 km, 4.9 km, 8.0 km and 10.9 km altitudes of 1.8 km, 3.6 km, 6.4 km and 9.8 km close to their nominal altitudes, meaning that the values retrieved for this these altitudes well represent the situation at the nominal altitude. In contrast the row kernel for the ~~lowermost troposphere (kernel with 0.4 km nominal altitude) peaks at around 1 km at every reference site and is not well separated from the row kernels with nominal altitudes just above the boundary layer (kernel with 1.2 km nominal altitude).~~ This shows that the retrieval has a poor sensitivity for the atmospheric layers near the ground. ground altitude peaks typically 1 km above ground altitude, meaning that data retrieved at the ground level mainly represent the altitudes about 1 km above the ground level. The row kernel for the 13.6 km altitude peaks close to its nominal altitude only for the tropical site of Manus Island. At Lindenberg and Sodankylä the respective kernels peak at 9-10 km altitude, i.e. at these sites variations in the 13.6 km retrieval data are mainly driven by the actual atmospheric variations at 9-10 km.

~~For higher altitudes the sensitivity is best in the tropics and poorest in the polar region. In agreement with the location of the tropopause we observe good sensitivity in the tropics up to 13 km, in mid-latitudes up to 11-12 km in summer and 8 km in winter and in polar regions in up to 11 km in summer. This becomes evident by looking on the altitudes where the example kernels of Fig. 3 have their peak levels. Whereas for Manus Island the 13.6 km kernel is~~ The thick black dashed line represents the sum along the row of the uppermost kernel that peaks close to its nominal altitude, for averaging kernel matrix and is a measure for the remote sensing systems sensitivity. This value is typically between 0.9 and 1.1 from 1 km up to 11 km at Lindenberg and Sodankylä ~~the respective kernels are the 10.9 km kernels, and from 1 km up to 13 km at Manus Island.~~

The seasonal dependency of the ~~sensitivity averaging kernels~~ is indicated in Fig. 64, which depicts the seasonal variations in the degree of freedom for signal (DOFS) values. The DOFS values are calculated as the trace of the averaging kernel matrix and the higher the DOFS values the more profile information is in the retrieved atmospheric state. In the tropics we observe no seasonal dependency. In the mid-latitudes the DOFS values are distinctively higher in summer than in winter ~~(for polar regions we analyse only summertime observations).~~ The seasonal variations in the mid-latitudes are in line with the seasonal . More details on this seasonal dependency are provided in Fig. 5, which depicts typical wintertime and typical summertime row kernels for Lindenberg. It seems that the seasonal variation is mostly a variation of the sensitivity at higher altitudes. In summer the remote sensing system can detect the actual atmospheric variations up to 11-12 km, whereas in winter the sensitivity is limited to altitudes below 8-9 km. This is connected to the variation of the tropopause altitude. As shown in Schneider et al. (2017) the averaging kernels depend strongly on the atmospheric temperature and humidity profiles (as well as on the surface temperature and emissivity).

In summary, at all three different sites the MUSICA MetOp/IASI retrieval provides H<sub>2</sub>O profile information from ~~-2 km about~~ 1 km above ground up to about the tropopause altitude.

## 4.2 Calculation of error Jacobians

The error Jacobians ( $\mathbf{K}_{\mathbf{p}} \mathbf{K}_{\mathbf{b}}$  from Eqs. 7 and 8) are calculated by the forward model PRFFWD (Hase et al., 2004) as follows: PRFFWD is executed running on a vertical grid of 28 levels from surface altitude to approximately 55 km above mean sea level. For every site reference forward calculations are performed for all cloud-free situations. The input (i.e. temperature, trace gas concentrations, etc.) for the reference forward model runs is the same as the input used in the forward calculation of the last iteration step of the MUSICA MetOp/IASI retrievals, i.e. the reference radiances are given by  $\mathbf{F}(\hat{\mathbf{x}}, \mathbf{p}) \mathbf{F}(\hat{\mathbf{x}}, \mathbf{b})$ . Then for each reference scenario we make additional forward calculations with slightly modified parameters, i.e. we calculate  $\mathbf{F}(\hat{\mathbf{x}}, \mathbf{p} + \epsilon_{\mathbf{p}}) \mathbf{F}(\hat{\mathbf{x}}, \mathbf{b} + \Delta \mathbf{b})$ . For a measurement vector  $\mathbf{y}$  having  $m$  elements and a parameter vector  $\mathbf{p}, \mathbf{b}$  having  $n$  elements the Jacobian matrix  $\mathbf{K}_{\mathbf{p}} \mathbf{K}_{\mathbf{b}}$  will have the dimension  $m \times n$ . The individual matrix elements are calculated as:

$$\mathbf{K}_{\mathbf{p}_{k,l}} \mathbf{K}_{\mathbf{b}_{k,l}} = \frac{\frac{F_k(\hat{\mathbf{x}}, \mathbf{p} + \epsilon_{\mathbf{p}_l}) - F_k(\hat{\mathbf{x}}, \mathbf{p})}{\epsilon_{\mathbf{p}_l}} \frac{F_k(\hat{\mathbf{x}}, \mathbf{b} + \Delta \mathbf{b}_l) - F_k(\hat{\mathbf{x}}, \mathbf{b})}{\Delta \mathbf{b}_l}}{\quad}, \quad (12)$$

where  $k$  is the index for the  $k$ th element of the measurement state vector  $\mathbf{y}$  (simulated by vector function  $F$ ) and  $l$  is the index for the  $l$ th element of the parameter vector  $\mathbf{p}, \mathbf{b}$ , respectively.

Table 2 gives an overview of the uncertainty assumptions  $\epsilon_{\mathbf{p}} \Delta \mathbf{b}$  used for calculating the Jacobians and for performing the error estimation. The calculations of the error Jacobians for water vapour continuum and clouds require a specific treatment, which is detailed in the following two subsections.

### 4.2.1 Water vapour continuum

We hypothetically assume that calculations based on the model “MT\_CKD” v2.5.2 (Mlawer et al., 2012; Delamere et al., 2010; Payne et al., 2011) only partly capture the full water vapour continuum effect. For the respective Jacobian calculation we perform forward calculations without considering the water vapour continuum ( $\mathbf{F}_{\text{noWVC}}(\hat{\mathbf{x}}, \mathbf{p}) \mathbf{F}_{\text{noWVC}}(\hat{\mathbf{x}}, \mathbf{b})$ ). Then we calculate the Jacobian matrix as  $\mathbf{K}_{\text{noWVC}} = \mathbf{F}_{\text{noWVC}}(\hat{\mathbf{x}}, \mathbf{p}) - \mathbf{F}(\hat{\mathbf{x}}, \mathbf{p}) \mathbf{K}_{\text{noWVC}} = \mathbf{F}_{\text{noWVC}}(\hat{\mathbf{x}}, \mathbf{b}) - \mathbf{F}(\hat{\mathbf{x}}, \mathbf{b})$ . The spectral response for an underestimation of 10% of the water vapour continuum effect is then  $\mathbf{K}_{\text{noWVC}} \epsilon_{\text{noWVC}} \text{ with } \epsilon_{\text{noWVC}} = 0.1 \mathbf{K}_{\text{noWVC}} \Delta \mathbf{b}_{\text{noWVC}} \text{ with } \Delta \mathbf{b}_{\text{noWVC}} = 0.1$ .

### 4.2.2 Opaque clouds (cumulus)

We estimate the influence of fractional coverage with by opaque liquid cumulus clouds with different cloud top altitudes (1.3 km, 3.0 km and 4.9 km). The radiance at top of the cloudy atmosphere  $\mathbf{F}_{\text{cum}}(\hat{\mathbf{x}}, \mathbf{p}) \mathbf{F}_{\text{cum}}(\hat{\mathbf{x}}, \mathbf{b})$  is calculated by starting PRFFWD at the cloud’s top height, assuring that no radiation from below the cloud contributes to  $\mathbf{F}_{\text{cum}}(\hat{\mathbf{x}}, \mathbf{p}) \mathbf{F}_{\text{cum}}(\hat{\mathbf{x}}, \mathbf{b})$ . Additionally it is assumed that the surface emissivity of the cloud is 1.0 and that the skin temperature of the cloud’s upward looking surface is in thermal equilibrium with the surrounding air temperature. The Jacobian matrix for opaque cumulus clouds is then  $\mathbf{K}_{\text{cum}} = \mathbf{F}_{\text{cum}}(\hat{\mathbf{x}}, \mathbf{p}) - \mathbf{F}(\hat{\mathbf{x}}, \mathbf{p}) \mathbf{K}_{\text{cum}} = \mathbf{F}_{\text{cum}}(\hat{\mathbf{x}}, \mathbf{b}) - \mathbf{F}(\hat{\mathbf{x}}, \mathbf{b})$  and the spectral response of a 10% fractional cloud cover is  $\mathbf{K}_{\text{cum}} \epsilon_{\text{cum}} \text{ with } \epsilon_{\text{cum}} = 0.1 \mathbf{K}_{\text{cum}} \Delta \mathbf{b}_{\text{cum}} \text{ with } \Delta \mathbf{b}_{\text{cum}} = 0.1$ .

### 4.2.3 Transmitting clouds (mineral dust and cirrus)

Some clouds are not opaque and we have to consider partial attenuation by the cloud particles. This is the case for cirrus clouds and mineral dust clouds. We consider these clouds by introducing them as an additional species in the forward model calculations. The extinction of these clouds is the sum of absorption and scattering. Since PRFFWD does not include the simulation of scattering clouds we calculate the attenuated radiances using forward model calculations from KOPRA (Karlsruhe optimized and precise radiative transfer algorithm; Stiller, 2000) and consider single scattering.

The frequency dependency of the extinction cross sections, the single scattering albedo, and the scattering phase functions of the clouds are calculated from OPAC v4.0b (Optical Properties of Aerosol and Clouds; Hess et al., 1998; Koepke et al., 2015). For cirrus clouds we assume the particle composition as given by OPAC's "Cirrus 3" ice cloud example (see Table 1b in Hess et al., 1998) and for mineral dust clouds a particle composition according to OPAC's "Desert" aerosol composition example (see table Table 4 in Hess et al., 1998).

We make cirrus cloud forward calculations  $\mathbf{F}_{\text{cir}}$  considering cirrus clouds with a vertical cloud layer thickness of 1 km and cloud top at different altitudes ranging from 6 km to 14 km. The Jacobians are calculated as  $\mathbf{K}_{\text{cir}} = \mathbf{F}_{\text{cir}}(\hat{\mathbf{x}}, \mathbf{p}) - \mathbf{F}(\hat{\mathbf{x}}, \mathbf{p})$   $\mathbf{K}_{\text{cir}} = \mathbf{F}_{\text{cir}}(\hat{\mathbf{x}}, \mathbf{b}) - \mathbf{F}(\hat{\mathbf{x}}, \mathbf{b})$  and for a cloud coverage of 50% the spectral response is  $\mathbf{K}_{\text{cir}} \epsilon_{\text{cir}}$  with  $\epsilon_{\text{cir}} = 0.5$   $\mathbf{K}_{\text{cir}} \Delta b_{\text{cir}}$  with  $\Delta b_{\text{cir}} = 0.5$ .

For the dust clouds we make forward calculations  $\mathbf{F}_{\text{dust}}$  for homogeneous 2 km thick layers between the ground and 6 km altitude. The Jacobians are then given as  $\mathbf{K}_{\text{dust}} = \mathbf{F}_{\text{dust}}(\hat{\mathbf{x}}, \mathbf{p}) - \mathbf{F}(\hat{\mathbf{x}}, \mathbf{p})$   $\mathbf{K}_{\text{dust}} = \mathbf{F}_{\text{dust}}(\hat{\mathbf{x}}, \mathbf{b}) - \mathbf{F}(\hat{\mathbf{x}}, \mathbf{b})$ .

### 4.3 Spectral response on uncertainty

Figure 6 depicts the spectral responses (i.e.  $\mathbf{K}_{\text{p}} \epsilon_{\text{p}} \mathbf{K}_{\text{b}} \Delta b$ ) for an example of different uncertainty sources for a typical situation at the tropical reference site. The left panel shows that ~~surface-skin~~ lower tropospheric temperature uncertainties mainly affect the spectra between  $1190 \text{ cm}^{-1}$  to  $1250 \text{ cm}^{-1}$  (which is also the spectral region of an "atmospheric window"), but is negligible for higher wavenumbers. This is ~~similar for lower tropospheric temperature uncertainties and~~ in contrast to upper tropospheric temperature uncertainties, which have highest spectral responses for wavenumbers larger than  $1250 \text{ cm}^{-1}$ .

The right panel of Fig. 6 illustrates that uncertainties in dust layers and uncertainties due to cirrus clouds have the highest impact at the lower end of wavenumbers and that a cirrus cloud has a ~~weaker~~ different dependency on wavenumber than a dust layer. Furthermore unrecognized clouds have the opposite effect on the spectrum than increasing the ~~surface-skin-and~~ atmospheric temperatures although affecting the spectrum in the same order of magnitude.

### 4.4 Estimated errors

Figure 4 shows a certain seasonal variability in the DOFS values (in particular at the ~~mid-latitudinal~~ mid-latitude site), indicating varying sensitivities of the remote sensing system. This variation is also present in the sensitivity with respect to uncertainty sources. For this reason we present the estimated errors for all the Manus Island and Sodankylä retrievals (MI and SK07) and

for all the Lindenberg 2008 retrievals (LI08). The Lindenberg 2008 error estimations are representative for all seasons, hence they well cover the full sensitivity variation with respect to uncertainty sources.

Table 2 gives an overview on the different uncertainty sources we consider for our error estimation. We distinguish between random uncertainty sources (the uncertainty affecting an observation is uncorrelated with the uncertainty affecting another observation), systematic uncertainty sources (the uncertainty is the same for all observations) and uncertainty sources that are always positive but with a random amplitude (clouds: the sky is either cloud free or covered by a random amount of clouds).

#### 4.4.1 Measurement noise Errors caused by random uncertainty sources

Figure ?? depicts 7 depicts from the top to the bottom the H<sub>2</sub>O error profiles for noise on the measured radiances due to the random uncertainties instrumental noise, emissivity and atmospheric temperatures (from the left to the right for Manus Island, Lindenberg, and Sodankylä). The shown error profiles are the square root of the diagonal elements of the error covariance matrix  $\mathbf{S}_e$  calculated  $\mathbf{S}_{x,noise}$  calculated for the instrumental noise according to Eq. (9) and of the error covariance matrix  $\mathbf{S}_{x,b}$  calculated for emissivity and atmospheric temperature according to Eq. (8).

For the calculations of  $\mathbf{S}_{x,noise}$  we assume a noise covariance  $\mathbf{S}_e$   $\mathbf{S}_{y,noise}$  of the IASI radiances according to Pequignot et al. (2008).

The measurement noise errors vary around 2-10% near the ground, but decrease to approximately 2-3% above the boundary layer and remain there throughout the free troposphere. Above and close to the tropopause errors increase again to values of around 10%. For Manus Island we observe similar errors for all the different observations. For Sodankylä and in particular for Lindenberg the errors vary. For instance, in the lower troposphere at Lindenberg the error is 10% for some days, but only 1-3% for other days. The varying sensitivity with respect to the uncertainty sources is due to the varying atmospheric conditions and in agreement with the varying DOFS values as documented by Fig. 4 (the Lindenberg data cover all mid-latitude seasons).

#### 4.4.2 Temperature and surface emissivity

The H<sub>2</sub>O error pattern profiles due to positive errors in surface emissivity, surface skin temperature and atmospheric temperatures are calculated as  $x_e$  according to Eq. (8). The patterns are plotted in For the calculating the error covariances  $\mathbf{S}_{x,b}$  due to surface emissivity uncertainties, we assume a 1% emissivity uncertainty and a spectral correlation length of this uncertainty of 100 cm<sup>-1</sup>. The resulting errors are highest close to the ground and for the continental sites of Lindenberg and Sodankylä, where they can reach 30%. Above 5 km altitude these errors are generally below 2%.

For the calculating the error covariances  $\mathbf{S}_{x,b}$  due to atmospheric temperatures uncertainties, we assume 2 K uncertainty from ground-2 km and 1 K uncertainty above 2 km altitude with correlation lengths increasing from 2 km at ground to 10 km in the stratosphere. The errors are typically 10-15%, but can occasionally reach 25%.

#### 4.4.2 Errors caused by systematic uncertainty sources

Fig. ?? (top panels for B shows from the top to the bottom the H<sub>2</sub>O error profiles due to systematic uncertainties in surface emissivity and surface skin temperature and bottom panels for uncertainties in atmospheric temperatures), atmospheric temperature and spectroscopic parameters. The error profiles are calculated as  $\Delta\hat{x}$  according to Eq. (7).

- 5 **Uncertainties in the surface emissivity** We assume two patterns of surface emissivity uncertainty. The first pattern means a  $-1\%$  uncertainty at the spectral grid points  $1185\text{ cm}^{-1}$  and  $1240\text{ cm}^{-1}$  and  $0\%$  uncertainty at the grid points  $1295\text{ cm}^{-1}$ ,  $1350\text{ cm}^{-1}$  and  $1405\text{ cm}^{-1}$  (this means that between  $1240\text{ cm}^{-1}$  and  $1295\text{ cm}^{-1}$  the uncertainty is linearly changing from  $-1\%$  to  $0\%$ ). The second pattern means a  $-1\%$  uncertainty at the spectral grid points  $1405\text{ cm}^{-1}$  and  $1350\text{ cm}^{-1}$  and  $0\%$  for the rest (with a linear change  $-1\%$  to  $0\%$  between  $1350\text{ cm}^{-1}$  and  $1295\text{ cm}^{-1}$ ). The top row of panels in Fig. B shows that surface
- 10 emissivity uncertainties are mainly important for the wavenumber region below  $1300\text{ cm}^{-1}$  (the first uncertainty pattern). An emissivity uncertainty of  $+2\%$  have a strong impact near the ground (errors of approx.  $-20\text{--}40\%$ ). For Manus Island, the error is always negative in the lower troposphere and similar for all observations, whereas for Lindenberg and Sodankylä a  $+2\%$  uncertainty in the surface emissivity causes sometimes positive and sometimes negative errors in the lower troposphere, i.e. the sensitivity with respect to this uncertainty source is strongly varying. Surface skin temperature is fitted during the retrieval
- 15 process, i.e. uncertainties in the first guess surface skin temperatures are partly corrected for during the retrieval process. The remaining error due to an  $+2\text{ K}$  surface skin temperature uncertainty is largest in the lower troposphere, where it can reach up to  $20\%$ . It  $-1\%$  has a rather uniform response at Manus Island: positive error of up to  $5\%$  close to the ground and a weak negative error around  $3\text{ km}$  altitude. At the continental sites of Lindenberg and Sodankylä the response on a systematic  $-1\%$  emissivity uncertainty can be positive or negative. Above  $4\text{ km}$  altitude the surface emissivity and surface skin temperature uncertainties
- 20 are smaller than  $5\%$  close to the ground (it varies between about  $-25\%$  and  $+20\%$ ). Around  $3\text{ km}$  the error response is generally negative and between  $-2\%$  and  $-20\%$ .

Positive atmospheric temperature uncertainties cause large positive errors in the retrieved tropospheric H<sub>2</sub>O profiles (we assume a systematic uncertainty of  $+2\text{ K}$  up to  $2\text{ km}$  altitude and  $+1\text{ K}$  at higher altitudes). The errors can reach  $+30\%$ , whereby these errors are largest for the heights (and atmospheric layers respectively) atmospheric layers where the atmospheric temperature uncertainty is assumed. For instance, uncertainties in lower tropospheric temperature ( $0\text{--}2\text{ km}$ , upper row ground  $2\text{ km}$ , black lines) cause maximal errors from ground up to  $3\text{ km}$  and decrease rapidly with altitude onwards, whereas uncertainties in upper tropospheric temperature ( $5\text{--}10\text{ km}$ , bottom row green lines) are negligible from ground up to  $6\text{ km}$ , but then increase to values of around  $+20\%$  at  $8\text{ km}$ .

#### 4.4.3 Spectroscopic parameters

- 30 Figure ?? illustrates the error pattern profiles ( $x_e$  according to Eq. 7) for uncertainties in the spectroscopic parameters. We consider Concerning spectroscopic parameters we consider systematic uncertainties in the H<sub>2</sub>O line intensity and pressure broadening parameters and an uncertainty in the applied water continuum model.

The uncertainty in the water vapour continuum model causes error profiles having small oscillations. For a water continuum model that underestimates the water continuum effect by 10% (see Sect. 4.2.1), the error is positive near ground (about +2%), negative at around 3 km altitude (about -4%) and negligible for altitudes above 5 km.

A positive uncertainty of  $+5\%$  in the linestrength-water vapour ( $\text{H}_2\text{O}$  and  $\text{HDO}$ ) line strength parameter causes a negative error of about  $-5\%$  in the retrieved  $\text{H}_2\text{O}$  values, ~~with slight oscillations (the slight oscillation is an effect of the constraint between  $\text{H}_2\text{O}$  and  $\text{HDO}$  applied for the MUSICA retrievals).~~ The impact of  $+5\%$  uncertainties in the pressure broadening parameter depends on the reference site: At Manus Island the resulting errors are negligible above 3 km, but at Lindenberg and Sodankylä the error profiles contain strong oscillations with ~~an amplitude of approximately maximal error of about +10% and maximal errors close above 10 km altitude.~~

This behavior of the errors due to uncertainties in the line shape modeling might be explained as follows: most of thermal nadir spectra's information about the vertical  $\text{H}_2\text{O}$  distribution is a consequence of the vertical atmospheric gradients of temperature and humidity. Without this gradients the spectral emission from a lower atmospheric layer is widely canceled out by the absorption at a higher layer. The gradients are generally strong up to the tropopause, i.e. up to the tropopause the remote sensing system's sensitivity is widely determined by these gradients. At Manus Island the tropopause is generally above 15 km, whereas at Lindenberg and Sodankylä is can be at much lower altitudes. This can be observed in Fig.2, which indicates a decrease of  $\text{H}_2\text{O}$  concentration up to 16 km above Manus Island, but only up to about 13 km and 12 km above Lindenberg and Sodankylä, respectively. Due to the weaker gradients above Lindenberg and Sodankylä and the relatively good spectral resolution of the IASI spectra, the line shapes do also provide information on the vertical distribution of  $\text{H}_2\text{O}$ . This is due to the pressure broadening effect, i.e. the broadness of the line decreases with decreasing pressure. As a consequence the  $\text{H}_2\text{O}$  profiles retrieved at Lindenberg and Sodankylä are much more affected by uncertainties in the line shape modeling than the profiles retrieved at Manus Island.

~~Uncertainties in the spectroscopic parameters of  $\text{HDO}$  and due to varying  $\text{CH}_4$  and  $\text{N}_2\text{O}$  concentrations account for absolute error values of less than 1% and are not depicted in the Figure.~~

#### 4.4.3 ~~Clouds~~Errors due to unrecognized clouds

Figure 9 shows the influence of different cloud types on the ~~retrieval~~errors. Uncertainties due to unrecognized cirrus clouds (top row in Fig. 9) lead to errors of -20% from 3-6 km at all sites and then decrease with altitude. However their impact on the ~~humidity-WVMR~~ profiles in the boundary layer shows large variation, especially at Lindenberg and Sodankylä, which is a result of the more variable atmospheric conditions at these sites (compared to the tropical site of Manus Island).

The influence of a 10% fractional cloud cover of opaque clouds depends on the height where the clouds are assumed (middle row in Fig. 9): Clouds at 1.3 km show only a small impact on the humidity profiles in the boundary layer with error magnitudes of 5-10%, but clouds at 3.0 km account for errors of more than 10% up to 5 km above mean sea level. Yet similarly to cirrus clouds their effect in the boundary layer shows large variation at Lindenberg and Sodankylä.

The error ~~pattern~~-profiles due to mineral dust layers (bottom row in Fig. 9) show that such layers have almost no impact if they are situated in the boundary layer, however if they are situated in the middle troposphere the errors ~~can account for~~



~~values-of~~ are more than 10%. The effect of dust clouds can be in particular large for the ~~mid-latitudinal~~ mid-latitude site of Lindenberg, where we also observe the largest variability in the calculated error ~~patterns~~ profiles.

## 5 ~~Empirical-validation-study~~ Comparison of GRUAN and IASI data

We use GRUAN processed Vaisala RS92 radiosonde measurements as reference for empirically validating the retrieved  
5 MUSICA IASI H<sub>2</sub>O profiles. The radiosonde ascents are collocated temporally and spatially with MetOp overpasses (for  
details see Sect. 3), which is essential for a meaningful comparison.

### 5.1 Regridding and smoothing of the high resolution GRUAN in-situ profiles

The in situ profiles have a high vertical resolution. This ~~is different to~~ differs from the remote sensing profiles, which can only  
detect the major characteristics of the vertical H<sub>2</sub>O distribution. Before comparing the data we have to account for ~~this~~ these  
10 different characteristics by regridding and smoothing the in situ profiles.

While the remote sensing retrieval provides atmospheric states and averaging kernels on a coarse atmospheric grid (between  
ground level and about 55 km a.s.l. 28 grid points are defined), the radiosonde reports data about every ~~10 m~~ 5 m. So we have  
to regrid the radiosonde data to the coarse vertical grid used by the remote sensing retrieval. In order to guarantee that the  
regridding does not significantly affect the H<sub>2</sub>O partial columns, the regridding is performed in two steps.

15 First, the radiosonde data points between the 28 MUSICA retrieval grid points are averaged by using a triangle inverse-  
distance weighted function resulting in a first estimate of the regridded radiosonde data. In the second step this first estimate  
is corrected by requiring that the partial columns between adjacent grid levels remain almost the same in the original high  
resolution data and in the regridded data. In the correction process a constraint is put ~~to~~ on the smoothness of the profile,  
thereby ~~avoiding that the correction produces~~ preventing the correction from producing strongly oscillating profiles. The results  
20 are regridded data consisting of reasonably smooth profiles having practically the same partial columns as the original high  
resolved radiosonde profiles. For the high altitudes that are not detected by the GRUAN radiosonde we use the retrievals a  
priori data ( $x_a$ ).

~~In order to get the in situ profile data that are comparable to the remote sensing data we have to smooth the regridded~~ The  
regridded GRUAN in situ profiles  ~~$\hat{x}_{\text{GRUAN}} g$~~  may be smoothed according to the averaging kernels of the remote sensing  
25 retrieval. The regridded and smoothed GRUAN in situ profile  ~~$\hat{x}_{\text{GRUAN}} \hat{g}$~~  is then comparable to the remote sensing profile,  
whereby:

$$\text{GRUAN} \hat{g} = (\mathbf{A}_{11} + \mathbf{A}_{12})(\text{GRUAN} g - x_a) + x_a. \quad (13)$$

Here  $\mathbf{A}_{11}$  is the H<sub>2</sub>O block of the averaging kernel matrix and  $\mathbf{A}_{12}$  the block that describes the response of the retrieved H<sub>2</sub>O  
on atmospheric HDO (see Sect. 2.3) and the vector  $x_a$  is the a priori state vector. An example illustrating the effects of the  
30 regridding and the smoothing is given in Fig. 10.

We would like to note that by using Eq. (13) we assume that H<sub>2</sub>O and HDO variations are fully correlated. However, H<sub>2</sub>O  
and HDO do not vary fully in parallel, i.e. calculating  ~~$\hat{x}_{\text{GRUAN}} \hat{g}$~~  according to Eq. (13) implies an uncertainty that can be



estimated by the uncertainty covariance matrix  $\mathbf{S}_{\mathbf{e},\text{GRUAN}}\mathbf{S}_{\hat{\mathbf{g}}}$  according to (see also Sect. 4.3 of Barthlott et al., 2017):

$$\mathbf{S}_{\mathbf{e},\text{GRUAN}}\mathbf{S}_{\hat{\mathbf{g}}} = \mathbf{A}_{12}\mathbf{S}_{\mathbf{a},\delta\text{D}}\mathbf{A}_{12}^T. \quad (14)$$

Here  $\mathbf{S}_{\mathbf{a},\delta\text{D}}$  describes the actual atmospheric  $\delta\text{D}$  covariances. Because  $\mathbf{A}_{12}$  and  $\mathbf{S}_{\mathbf{a},\delta\text{D}}$  have small entries only, this uncertainty is below 1% and can be neglected for our comparison.

## 5.2 Comparison of GRUAN and IASI Metric for quantifying data agreement

In this section we present the comparison between For a better statistical quantification of the deviations of the remote sensing data from the GRUAN reference data, we introduce a skill score DL describing the difference of the logarithmic values of the respective water vapour concentrations. Because  $\Delta\ln(x) \approx \frac{\Delta x}{x}$ , we interpret the logarithmic scale difference between IASI and GRUAN as the relative difference (and use the GRUAN data in the denominator). DL then becomes:

$$\begin{aligned} \text{DL} &= \frac{\ln([\text{H}_2\text{O}]_{\text{retrieval}}) - \ln([\text{H}_2\text{O}]_{\text{GRUAN}})}{[\text{H}_2\text{O}]_{\text{GRUAN}}}, \\ &\approx \frac{[\text{H}_2\text{O}]_{\text{retrieval}} - [\text{H}_2\text{O}]_{\text{GRUAN}}}{[\text{H}_2\text{O}]_{\text{GRUAN}}}, \end{aligned} \quad (15)$$

where  $[\text{H}_2\text{O}]_{\text{GRUAN}}$  is the regridded and smoothed GRUAN  $\text{H}_2\text{O}$  profiles and the IASI  $\text{H}_2\text{O}$  profiles. Thereby we illustrate the remote sensing data quality at the three reference sites that belong to radiosonde  $\text{H}_2\text{O}$  data (i.e.  $\hat{\mathbf{g}}$  from Eq. 13) and  $[\text{H}_2\text{O}]_{\text{retrieval}}$  is the retrieved IASI  $\text{H}_2\text{O}$  data. The so defined skill score DL is a good measure for the relative difference between the GRUAN and IASI data.

As a good measure for the mean relative difference between GRUAN and IASI we can use the mean difference of logarithmic values (MDL):

$$\begin{aligned} \text{MDL} &= \frac{1}{N} \sum_{i=1}^N \text{DL}_i \approx \frac{1}{N} \sum_{i=1}^N [\ln([\text{H}_2\text{O}]_{\text{retrieval}}) - \ln([\text{H}_2\text{O}]_{\text{GRUAN}})]_i \\ &\approx \frac{1}{N} \sum_{i=1}^N \left( \frac{[\text{H}_2\text{O}]_{\text{retrieval}} - [\text{H}_2\text{O}]_{\text{GRUAN}}}{[\text{H}_2\text{O}]_{\text{GRUAN}}} \right)_i. \end{aligned} \quad (16)$$

Similarly, we can use the standard deviation of the logarithmic differences as a measure for the relative scatter between GRUAN and IASI and introduce  $\sigma_{\text{MDL}}$  as

$$\sigma_{\text{MDL}} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\text{DL}_i - \text{MDL})^2}. \quad (17)$$

For illustrating the variation of the atmospheric state we introduce  $\sigma_{\hat{\mathbf{g}}}$  as

$$\sigma_{\hat{\mathbf{g}}} = \sqrt{\frac{1}{N} \sum_{i=1}^N \left[ \ln([\text{H}_2\text{O}]_{\text{GRUAN}})_i - \overline{\ln([\text{H}_2\text{O}]_{\text{GRUAN}})} \right]^2}. \quad (18)$$

We want to document to what extent the differences between GRUAN and MUSICA IASI data can be explained by the estimated errors. In Sect. 4 we estimate in detail the error in the MUSICA IASI H<sub>2</sub>O profiles for three different climate zones. It should be noted that the radiosonde measurements are also affected by several uncertainties which have to be taken into account during the evaluation.

- 5 Uncertainties in the GRUAN H<sub>2</sub>O profiles have also to be considered. In general the uncertainty of the GRUAN data increases with altitude. For the regridded and smoothed GRUAN profiles  $\Delta\hat{g}$  is about 3-5% near the surface and 5-20% at around 10 km altitude. For further information details on the radiosonde uncertainty we refer to Appendix A. If we assume that the MUSICA IASI and the GRUAN errors are uncorrelated random errors we can calculate the  $1\sigma$  scatter of DL around MDL as expected from the MUSICA IASI and GRUAN errors by:

$$10 \quad \Delta_{\text{MDL}} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\Delta\hat{x}_i^2 + \Delta\hat{g}_i^2)}. \quad (19)$$

### 5.2.1 Correlation plots

- We use scatter plots for representing the results at selected heights where the retrieval has a good sensitivity at the respective reference sites. Figure 14 illustrates correlations between the retrieval results and the smoothed GRUAN data for selected retrieval altitude levels representative for the lower, middle and upper troposphere, respectively. Here the index  $i$  stands for an individual observation and  $N$  is the number of all observations.

## 5.3 Data agreement for individual ensembles

- In this section we present the comparison between the regridded and smoothed GRUAN H<sub>2</sub>O profiles and the IASI H<sub>2</sub>O profiles using the metric as described in the previous subsection. The statistical quantifications are made individually for the four different ensembles as given in Table 1. The aim is to illustrate the remote sensing data quality for the three different climate zones.

### 5.3.1 MUSICA IASI standard retrieval

- Figure 11 depicts the vertical distribution of the data agreement for the MI and LI08 ensembles. These ensembles correspond to IASI measurements with available EUMETSAT L2 data and we can execute the standard MUSICA IASI retrieval, which uses the EUMETSAT L2 temperature data as the a priori atmospheric temperatures. For the lower and middle troposphere we use the same altitude levels (1.8 km and 4.9 km, respectively) at all reference sites. MI ensemble the MDL value (thick black line) oscillates between  $-23\%$  and  $-7\%$  below 10 km altitude and is close to zero at higher altitudes. The scatter  $\sigma_{\text{MDL}}$  is indicated by the black error bars and it is generally within 20% except for the altitudes around 12 km where it is slightly higher. For the upper troposphere we illustrate the values for the highest altitude level at which the row kernels in Fig. 3 still indicate good sensitivity. This is about 13.5 km for Manus Island and about 11 km for Lindenberg and Sodankylä. LI08 ensemble the MDL values oscillates between  $-20\%$  and  $+16\%$  and the scatter  $\sigma_{\text{MDL}}$  is up to 42% below 5 km and about 15% at higher altitudes.

For both ensembles (MI and LI08) the  $\sigma_{\text{MDL}}$  values are significantly smaller than the  $1\sigma$  variation in the smoothed radiosonde data ( $\sigma_{\hat{q}}$ ).

~~Near the ground and in the lower troposphere (top row) The red shaded area around the MDL value represents the  $\Delta_{\text{MDL}}$  values, i.e. the scatter in the MDL value we expect due to the errors in the MUSICA IASI and GRUAN  $\text{H}_2\text{O}$  data. The  $\Delta_{\text{MDL}}$  values are calculated according to Eq. (19) by considering MUSICA IASI random errors due to measurement noise, emissivity and atmospheric temperature (actually we work with the error estimations as depicted in Fig. 14) the retrieval has a slightly higher variability than the radiosonde measurements for the sites of Manus Island and Sodankylä, however the remote sensing and the in-situ data well identify whether the air is more humid or drier than given by the a priori assumption (represented by the yellow star). For Lindenberg we observe two outliers in the 7) and the GRUAN random errors as discussed in Appendix A and presented in Fig. A2. For the MI ensemble the  $\Delta_{\text{MDL}}$  and  $\sigma_{\text{MDL}}$  show similar amplitudes and vertical behavior, meaning that the expected and the observed scatter agree reasonably well. For the LI08 data ensemble  $\Delta_{\text{MDL}}$  and  $\sigma_{\text{MDL}}$  agree well only above 5 km altitude. At lower altitudes the actually observed scatter is significantly larger than the scatter expected from the estimated MUSICA IASI and GRUAN errors, which might indicate the observation of different airmasses by IASI and GRUAN an underestimation of the MUSICA IASI random errors at Lindenberg below 5 km altitude.~~

~~In the middle troposphere (middle row) The comparison suggests a weak dry bias in the MUSICA IASI data between 2 and 10 km at Manus Island and above 10 km at Lindenberg. The former could be explained by errors in the simulated line intensities and the latter by errors in the simulated line shapes (see discussion in the context of Fig. 14) the retrieval and the measurements remain in quasi-linear relation over nearly one order of magnitude at all reference sites. This indicates the good sensitivity and precision of the remote sensing data at this altitude region. B). However, given the small number of ensemble members we should be careful and avoid premature conclusions.~~

~~In the upper troposphere close to the tropopause (bottom row) the  $\text{H}_2\text{O}$  variability at Manus Island is rather small. In agreement to the GRUAN radiosonde, the retrieval estimates an  $\text{H}_2\text{O}$  increase~~

### 5.3.2 Retrieval using external temperature data

During summer 2007 EUMETSAT L2 data have not been available and the retrievals for the LI07 and SK07 ensembles have been executed using the atmospheric temperature measured by the GRUAN radiosondes as the a priori atmospheric temperatures (see discussion in Sect. 3.4). In order to avoid inconsistencies when comparing the different ensembles we simulate retrievals of the MI and LI08 ensembles that also use the GRUAN radiosonde temperatures instead of the EUMETSAT L2 temperatures as the a priori atmospheric temperatures. The simulated retrieval products are obtained by adding  $\mathbf{G}\mathbf{K}_T(\mathbf{T}_{\text{L2}} - \mathbf{T}_{\text{GRUAN}})$  to the standard MUSICA IASI retrieval products, where  $\mathbf{G}$  is the gain matrix,  $\mathbf{K}_T$  the Jacobian matrix for atmospheric temperature, and  $\mathbf{T}_{\text{L2}}$  and  $\mathbf{T}_{\text{GRUAN}}$  are the atmospheric temperature state vectors of the EUMETSAT L2 and GRUAN data, respectively. For the altitudes above the radiosonde we extend the  $\mathbf{T}_{\text{GRUAN}}$  vector with a zonally and monthly mean temperature climatology (Rees et al., 1990). For calculating the combined MUSICA IASI and GRUAN random error (i.e. the expected scatter  $\Delta_{\text{MDL}}$ ) we have to consider the uncertainties in the GRUAN temperatures instead of the uncertainties in the EUMETSAT L2 temperatures. Appendix B gives a brief overview on the GRUAN temperature uncertainties.

Figure 12 depicts the data agreement for all four ensembles when GRUAN radiosonde temperatures are used as the a priori atmospheric temperatures. For Manus Island and Lindenberg (ensembles MI and LI08) the scatter in MDL ( $\sigma_{\text{MDL}}$ ) is significantly reduced if compared to Fig. 11 (figure showing the data agreement for MUSICA IASI retrievals that use EUMETSAT L2 temperatures as the a priori assumption. It should also be noted that the concentrations at this altitude are rather low (~40 ppmv) compared to the concentrations near the ground (about 20 000 ppmv at 1.8 km atmospheric temperature). A similar reduction is also observed in the theoretically predicted scatter ( $\Delta_{\text{MDL}}$ ), because the GRUAN temperatures have a much smaller uncertainty (typically 0.1-0.3 K) than the EUMETSAT L2 temperatures (we assume 1-2 K). At Lindenberg and Sodankylä the  $\text{H}_2\text{O}$  variability is higher and well captured by the IASI retrieval, despite the small humidity levels of below 100 ppmv (ensemble LI08)  $\sigma_{\text{MDL}}$  and  $\Delta_{\text{MDL}}$  agree much better for the retrieval products obtained by using GRUAN temperature as a priori than for the MUSICA IASI standard retrieval products (obtained by using EUMETSAT L2 temperature as a priori). This suggests that for Lindenberg and the year 2008 our uncertainty assumptions for the EUMETSAT L2 atmospheric temperatures (see Table 2) are probably too optimistic.

The error bars on the diagonal of the plots of bottom panels in Fig. 14 indicate the typical GRUAN errors ( $\hat{x}_{e,\text{GRUAN}}$  as detailed in Appendix A) and the root square sum of the typical leading statistical IASI errors ( $x_e$  from Eq. 7), whereby we have considered measurement noise and uncertainties in surface emissivity, surface skin temperatures and atmospheric temperatures. 12 show the data agreement for the LI07 and SK07 ensembles. These ensembles are exclusively representative for summer observations. We observe that the  $\sigma_{\text{MDL}}$  values are generally larger than the  $\Delta_{\text{MDL}}$  values, meaning that we probably underestimate the MUSICA IASI random errors. In addition we find a wet bias of up to 30% below 2 km altitude and a dry bias of about 20% at around 14 km.

### 5.3.3 Bias and scatter

For a better statistical quantification of the deviations of the remote sensing data from the GRUAN reference data, we introduce a skill score DL describing the difference of the logarithmic values of the respective water vapour concentrations:-

$$\begin{aligned} \text{DL} &= \frac{\ln([\text{H}_2\text{O}]_{\text{retrieval}}) - \ln([\text{H}_2\text{O}]_{\text{GRUAN}})}{[\text{H}_2\text{O}]_{\text{GRUAN}}} \\ &\approx \frac{[\text{H}_2\text{O}]_{\text{retrieval}} - [\text{H}_2\text{O}]_{\text{GRUAN}}}{[\text{H}_2\text{O}]_{\text{GRUAN}}}, \end{aligned}$$

where  $[\text{H}_2\text{O}]_{\text{GRUAN}}$  is the regridded and smoothed radiosonde  $\text{H}_2\text{O}$  data (i.e.  $\hat{x}_{\text{GRUAN}}$  from Eq. 13) An upper tropospheric dry bias is consistently observed in the analysis of the LI08, LI07 and  $[\text{H}_2\text{O}]_{\text{retrieval}}$  is the retrieved IASI  $\text{H}_2\text{O}$  data. The so defined skill score DL is a good measure for the relative difference between the GRUAN and IASI data:-

As a good measure for the mean relative difference between GRUAN and IASI we can use the mean difference of logarithmic values (MDL):-

$$\begin{aligned} \text{MDL} &= \frac{1}{N} \sum_{i=1}^N \text{DL}_i \equiv \frac{1}{N} \sum_{i=1}^N [\ln([\text{H}_2\text{O}]_{\text{retrieval}}) - \ln([\text{H}_2\text{O}]_{\text{GRUAN}})]_i \\ &\approx \frac{1}{N} \sum_{i=1}^N \left( \frac{[\text{H}_2\text{O}]_{\text{retrieval}} - [\text{H}_2\text{O}]_{\text{GRUAN}}}{[\text{H}_2\text{O}]_{\text{GRUAN}}} \right)_i. \end{aligned}$$

- 5 Similarly, we can use the standard deviation of the logarithmic differences as a measure for the relative scatter between GRUAN and IASI and introduce  $\sigma_{\text{MDL}}$  as-

$$\sigma_{\text{MDL}} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\text{DL}_i - \text{MDL})^2}.$$

For illustrating the variation of the atmospheric state we introduce  $\sigma_{\text{GRUAN}}$  as-

$$\sigma_{\text{GRUAN}} = \sqrt{\frac{1}{N} \sum_{i=1}^N [\ln([\text{H}_2\text{O}]_{\text{GRUAN}})_i - \overline{\ln([\text{H}_2\text{O}]_{\text{GRUAN}})}]^2}.$$

- 10 SK07 ensembles, but not seen in the analysis of the MI ensemble. A systematic uncertainty source that affects upper tropospheric  $\text{H}_2\text{O}$  at Lindenberg and Sodankylä but not at Manus Island is the shape of the water vapour lines (see discussion in the context of Fig. B). So deficits in simulating the line shapes might explain this upper tropospheric dry bias. In the near surface atmosphere we observe a wet bias at the two continental sites Lindenberg and Sodankylä, but only for the ensembles that are limited to the summer season (LI07 and SK07). Our error estimation study suggests that small uncertainties in the emissivity can cause
- 15 large errors at these continental sites. So an uncertainty in the used IREMIS emissivity is a candidate for explaining the surface near wet bias; however, the  $\text{H}_2\text{O}$  retrieval response for a  $-1\%$  uncertainty in the emissivity differs between observations and can be positive or negative (see Fig. 9). This means that emissivity uncertainties can only explain the bias if the sign of the emissivity uncertainty is correlated with the atmospheric state (e.g. the uncertainty in the used monthly IREMIS surface emissivity is typically positive for dry atmospheric conditions and typically negative for humid atmospheric conditions) or
- 20 surface conditions (e.g. the uncertainty in the IREMIS data is typically positive/negative for a surface with high/low emissivity or high/low skin temperatures).

Figure ??

#### 5.4 Global overview on data agreement

- Figure 13 depicts the vertical profiles of the ~~aforementioned skill scores calculated~~ data agreement skill score parameters for all coinciding observations without separating the different ~~sites and time periods~~. The MDL value is within  $\pm 0.12$  for all altitudes, indicating that the retrieval is in very good agreement to ensembles. This analysis is based on 100 individual comparisons.

Below 12 km altitude the MDL value oscillates between  $-10\%$  and  $+11\%$  and at around 14 km altitude it reaches  $-21\%$ . The scatter in MDL ( $\sigma_{\text{MDL}}$ ) is almost 29% close to the surface but generally smaller than 20% above 1 km altitude. Above 5 km altitude this observed scatter is only slightly larger than the scatter predicted from the estimated errors ( $\Delta_{\text{MDL}}$ ). At lower altitudes the ~~smoothed radiosonde data~~ predicted scatter is clearly smaller than the observed scatter. An explanation of the observed upper tropospheric bias and increased scatter at low altitudes is given in the previous section: the dry bias in the upper troposphere might have its origin in an incorrect modeling of the spectroscopic line shapes and the increased scatter near the surface might be due to uncertainties in the IREMIS emissivities.

The observed scatter between GRUAN and IASI ( $\sigma_{\text{MDL}}$ ) ~~decreases slightly linearly with altitude starting at  $\pm 0.3$  at the ground and reaching  $\pm 0.2$  at 9 km. In contrast the  $\sigma_{\text{MDL}}$  is significantly smaller than the  $1\sigma$  variation of the smoothed radiosonde data ;  $\sigma_{\text{GRUAN}}$ , reaches more than  $\pm 0.5$  in the lower troposphere ( $\sigma_{\hat{q}}$ ), which reaches about 50% near the surface and more than  $\pm 1.0$  100% in the middle and upper troposphere. This reflects the large variation in the atmospheric water vapour concentration data we use for our evaluation study (see also Fig. 2).~~

~~The small values of MDL and the small variations of  $\sigma_{\text{MDL}}$  throughout the atmosphere prove that the retrieval is able to detect the~~ The MUSICA IASI data product does well capture most of these variations. For demonstrating this capability Fig. 14 illustrates correlations between the MUSICA IASI retrieval products and the smoothed GRUAN data for selected altitudes. The respective altitudes are highlighted in Figs. 3 and 5, which documents that at all sites the MUSICA IASI product for 1.8, 3.6, 6.4, and 9.8 km are independent and well sensitive to real atmospheric variations. Near the surface the sensitivity is generally limited and at 13.6 km only the Manus Island data are reasonably sensitive to the actual atmospheric variations.

At the altitudes where the MUSICA IASI product show very good sensitivity (1.8, 3.6, 6.4, and 9.8 km) we observe a very good correlation and can demonstrate that the MUSICA IASI product can correctly capture the large variations that are present in atmospheric water vapour ~~concentration in all climate regions~~. For instance, at 3.6 km the mixing ratios range from below 200 ppmv to almost 20000 ppmv and at 9.8 km from 10 ppmv to more than 1000 ppmv. Please note that these large variations are reliably reproduced by the MUSICA IASI processor, although the retrieval works with a single humidity a priori value that is used at all sites and during all seasons (indicated as the yellow star in Fig. 14). Near the surface the correlation is a bit weaker than at higher altitudes, mainly due to the some outliers belonging to the LI07 and SK07 ensembles (the ensembles representing the summer season over land). At 13.6 km altitude we observe a good correlation, which demonstrates the possibility of detecting  $\text{H}_2\text{O}$  at Manus Island. However, at Lindenberg and Sodankylä these variations are strongly driven by actual atmospheric variations that take place at lower altitudes (see magenta lines in Figs. 3 and 5).

For our theoretical error analyses in Sect. 4 we assume that the relative errors have a component that is mostly random (Fig.7) and another component that is mostly systematic (Fig.B). For the comparison study we proceed similarly and examine bias and scatter, which means that we describe the variances in the MUSICA IASI data by the variances in the GRUAN data ( $\sigma_{\hat{q}}^2$ ) and the variance in the difference between MUSICA IASI and GRUAN ( $\sigma_{\text{MDL}}^2$ ). Using this description we can calculate the  $R^2$  value that represents the portion of the MUSICA IASI variance that is in full agreement (fully correlated) with the

GRUAN variance:

$$R^2 = \frac{\sigma_{\hat{g}}^2}{\sigma_{\hat{g}}^2 + \sigma_{\text{MDL}}^2}. \quad (20)$$

Each panel of Fig. 14 contains the  $R^2$  value calculated for the respective altitude.

The error blue bars on the diagonal of the plots of Fig. 14 indicate the typical GRUAN errors ( $\Delta\hat{g}$  as detailed in Appendix A) and the root-square-sum of the typical leading MUSICA IASI random errors ( $\Delta x$ ), whereby we have considered measurement noise, uncertainties in surface emissivity and uncertainties in the GRUAN temperatures. The MDL and  $\sigma_{\text{MDL}}$  values are also written in each panel as bias (b) and scatter (s) value, respectively. They are the same as shown in Fig. 13.

Figure 15 resumes the capability of the MUSICA IASI retrieval product for capturing real atmospheric  $\text{H}_2\text{O}$  variations at different altitudes by showing vertical profiles of the  $R^2$  values calculated according to Eq. (20) for the different ensembles individually and when considering all 100 individual comparisons together. Between 1 km and 12.5 km altitude (and when considering all comparisons together the MUSICA IASI products) detects more than 90% of the atmospheric variations in agreement with GRUAN.

## 6 Summary and outlook

In this paper, we compare water vapour profiles retrieved from IASI spectra by the MUSICA MetOp/IASI retrieval with in situ measurements from GRUAN radiosondes at three different reference sites representative for of three different climate zones (tropics, midlatitudes-mid-latitudes and polar region). Additionally-In addition, we provide an extensive theoretical error estimation of the retrieval's water vapour product for the respective reference sites considering a-large-number-of-many different uncertainty sources.

The error estimations of the MUSICA MetOp/IASI water vapour profiles at the different reference sites reveal that for the lowermost 3 km the errors can be as large as 4030%. The most important uncertainty sources are unrecognized clouds, and uncertainties in lower tropospheric temperature ,in-surface-skin-temperature-, and in surface emissivity. Between 3 and 6 km the error can be as large as 20%, mainly due to middle atmospheric temperature uncertainties and unrecognized high cirrus clouds. Above 6 km the errors are typically smaller than 20% and mainly caused by uncertainties in upper tropospheric temperatures and uncertainties in spectroscopic pressure broadening parameters.

For the empirical validation study the remote sensing MUSICA MetOp/IASI  $\text{H}_2\text{O}$  profiles have been compared to 100 different Vaisala RS92 radiosonde measurements that have been processed by the GRUAN lead centre. The scatter found for the difference between GRUAN and IASI is smaller than 25%-between-1.5-and-10-km-21% above 1.8km altitude. It is slightly higher near the groundand-close-to-the-tropopause. This is in good agreement with errors as given for the GRUAN data and the errors as estimated for the MUSICA IASI product. It is important to note that the coincidences correspond to five different years and represent three different climate zones, giving the here-presented-study-study here presented a good global representativeness. We demonstrate that the MUSICA MetOp/IASI retrieval is able to correctly capture variations in  $\text{H}_2\text{O}$  profiles between 1.5 km-1 km above ground up to the upper troposphere.

The comparison indicates to a dry bias of the remote sensing data of 20% in the upper troposphere of the middle and high latitude sites, but not at the tropical site. We find that deficits in spectroscopic line shape modeling could explain such behavior. For the current MUSICA IASI retrieval a Voigt line shape model is assumed and HITRAN 2016 pressure broadening parameters are used. It would be interesting to investigate if the usage of more sophisticated line shape models (e.g. speed-dependent Voigt line shape model) could reduce the upper tropospheric bias and improve the agreement between the MUSICA IASI remote sensing and GRUAN in-situ data. For the continental sites (Lindenberg and Sodankylä) and during summer we observe a wet bias in the MUSICA IASI data with respect to GRUAN. Uncertainties in land surface emissivity being correlated to atmospheric or surface conditions (e.g. negative/positive emissivity uncertainties occurring in line with very dry/humid atmospheric conditions or hot/cold skin temperatures) could explain this behavior. It would be interesting to test if the usage of a daily surface emissivity product instead of the monthly mean IREMIS data (which have been used for the here presented retrievals) improves the agreement between MUSICA IASI and GRUAN.

## 7 ~~Data-availability~~

~~The here presented MUSICA MetOp/IASI data are available on the MUSICA website . The structure of the data will be explained in a manuscript that is currently prepared for submission to Earth System Science Data (ESSD) . In the meanwhile please-~~

## 7 Data availability

The MUSICA IASI data presented here are available on the MUSICA website <http://www.imk-asf.kit.edu/english/musica.php>. Please contact M. Schneider for more details. The GRUAN data are available at the GRUAN website: <https://www.gruan.org/data/data-products/gdp/rs92-gdp-2/>.



## Appendix A: Uncertainties of GRUAN water vapour volume mixing ratios

In order to perform a valid comparison between remote sensing data and in situ measurements, the uncertainties of the in situ data have to be considered.

- 5 GRUAN provides uncertainties for the relative humidity ( $\Delta\rho$ ), for the temperature ( $\Delta T$ ), and for the pressure ( $\Delta p$ ). The water vapour volume mixing ratio (WVMR) is defined as

$$\text{WVMR} = \frac{\rho E(T)}{p - \rho E(T)} \approx \frac{\rho E(T)}{p}, \quad (\text{A1})$$

where  $E$  is the water vapour saturation pressure. The GRUAN WVMR error for each individual radiosonde ~~ea~~can be calculated as

$$10 \text{ WVMR}_e = \sqrt{\left(\frac{\Delta E(T)}{E(T)}\right)^2 + \left(\frac{\Delta \rho}{\rho}\right)^2} \times \text{WVMR}. \quad (\text{A2})$$

Uncertainties in atmospheric pressure  $p$  can be neglected if compared to the uncertainties of  $E(T)$  and  $\rho$ . For the calculation of the water vapour saturation pressure we use the same formula as GRUAN from Hyland and Wexler (1983). Since  $E(T)$  is a highly non-linear function, we estimate the uncertainty of  $E$  by

$$\Delta E = \max\{|E(T + \Delta T) - E(T)|; |E(T - \Delta T) - E(T)|\}. \quad (\text{A3})$$

- 15 According to Dirksen et al. (2014) there are correlated and uncorrelated errors. We investigate both separately. Figure A1 depicts ~~for~~ the correlated and uncorrelated GRUAN WVMR errors ( $\text{WVMR}_e$ ) in the top and bottom panels, respectively. Black lines indicate the data ensembles that cover all seasons (Manus Island and Lindenberg 2008) and red lines the ensembles that are only representative for the summer season (Lindenberg 2007 and Sodankylä).

- For a reasonable comparison the vertically ~~high~~highly resolved GRUAN profiles have to be adjusted to the vertical resolution of the remote sensing profiles (see Sect. 5.1). This means a significant reduction of the vertical resolution and the uncorrelated errors will cancel out. The regridding and smoothing of the ~~uncorrelated~~correlated errors is accomplished as follows: First, the errors  $\text{WVMR}_e$  are added to the measured WVMR data. Second, for  $\text{WVMR} + \text{WVMR}_e$  we perform the regridding as described in Sect. 5.1, i.e. we calculate the regridded version of the erroneous GRUAN WVMR profile. The difference between the erroneous and the original profiles (of the regridded versions) give the regridded GRUAN WVMR ~~error~~( $\epsilon_{\text{GRUAN}}$ ):  
 25 uncertainty profile ( $\Delta g$ ). Above the radiosonde (where we set  $g$  equal to the retrievals a priori) we set the uncertainty to 100%. Then we calculate an uncertainty covariance matrix  $\mathbf{S}_g$  using the values from the uncertainty profile  $\Delta g$  and a large correlation length of 30 km individually for the two blocks representing the data measured by the radiosonde and the data above the radiosonde. Third, in analogy to Eq. (13) we apply the averaging kernels to  ~~$\epsilon_{\text{GRUAN}}$~~  $\mathbf{S}_g$  and get the error ~~in~~covariance for the regridded and smoothed GRUAN profiles as:

$$30 \text{ } \underline{\epsilon_{\text{GRUAN}} \mathbf{S}_g} = (\mathbf{A}_{11} + \mathbf{A}_{12}) \underline{\epsilon_{\text{GRUAN}} \mathbf{S}_g} (\mathbf{A}_{11} + \mathbf{A}_{12})^T. \quad (\text{A4})$$

Figure A2 depicts ~~these error profiles~~( $\hat{x}_{e,\text{GRUAN}}$ ) the square root values of the diagonal of  $\mathbf{S}_g$  for the different ensembles. The uncertainties typically increase from 5% near the ground to ~~values~~5-20% at around 10 km altitude. For higher altitudes it decreases again due to the decaying sensitivity (see averaging kernel plots of Fig. 3).

## Appendix B: Uncertainties of GRUAN temperatures

The MUSICA IASI retrieval uses the EUMETSAT L2 temperatures as the a priori atmospheric temperatures. However, in summer 2007 EUMETSAT L2 data are not available and instead we use the GRUAN temperatures for the LI07 and SK07 retrievals. In addition, for the MI and LI08 ensembles we simulate retrievals that use the GRUAN temperatures instead of the EUMETSAT L2 temperatures as the a priori atmospheric temperatures. For all these retrievals the uncertainty in the GRUAN temperatures and not the uncertainty in the EUMETSAT L2 temperatures has to be considered for the error estimation.

Figure A3 depicts the correlated GRUAN temperature uncertainty profiles after regridding the data to the MUSICA retrieval grid points by using a triangle inverse-distance weighted averaging function (as for the first H<sub>2</sub>O regridding step, see Sect. 5.1). We assume that the uncorrelated uncertainties cancel out by this averaging. Below 20 km altitude the GRUAN temperature uncertainties are well within 0.3 K, i.e. they are much smaller than the uncertainties of 1-2 K we assume for the EUMETSAT L2 temperatures. The GRUAN nighttime temperature data have an uncertainty that is rather constant with altitude, whereas for the daytime data the uncertainty monotonically increases with altitude. Above the top altitude of the radiosonde we use a monthly and zonally averaged temperature climatology (COSPAR International Reference Atmosphere; Rees et al., 1990) and assume a temperature uncertainty of 5 K (this explains the instantaneous uncertainty increase that can be observed for some Manus Island and Lindenberg radiosondes).

*Author contributions.* C. Borger performed most calculations for this work during his master thesis at KIT IMK-ASF and prepared the manuscript together with M. Schneider and in collaboration with all coauthors. M. Schneider developed the IASI retrievals in the framework of the MUSICA project and B. Ertl supported these developments by making the processing chain more efficiently. F. Hase wrote the PROFIT and PRFFWD codes. O. García helped in reading and formatting the EUMETSAT IASI L2 data. M. Sommer provided the GRUAN radiosonde measurements in a very useful data format for the sites of Lindenberg and Sodankylä. M. Höpfner helped us with the KOPRA calculations used for estimating the effect of the scattering by cirrus and mineral dust particles. S. Tjemkes provided all necessary data for the site of Manus Island in a very useful data format in the framework of a planned exercise called “Intercomparison of Hyperspectral Retrieval Codes”. X. Calbet collected the IASI/GRUAN coincidences over Manus Island and helped us in the interpretation of the radiosonde’s measurement uncertainties.

*Acknowledgements.* This research has largely benefit from results of the projects MUSICA (funded by the European Research Council under the European Community’s Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement number 256961), MOTIV (funded by the Deutsche Forschungsgemeinschaft under GZ SCHN 1126/2-1) and INMENSE (funded by the Ministerio de Economía y Competitividad from Spain, CGL2016-80688-P). We acknowledge the support by the Deutsche Forschungsgemeinschaft and the Open Access Publishing Fund of the Karlsruhe Institute of Technology.

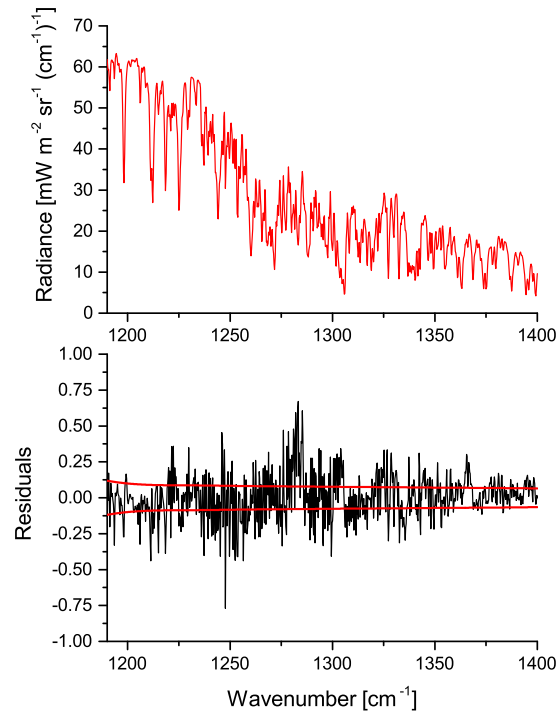


## References

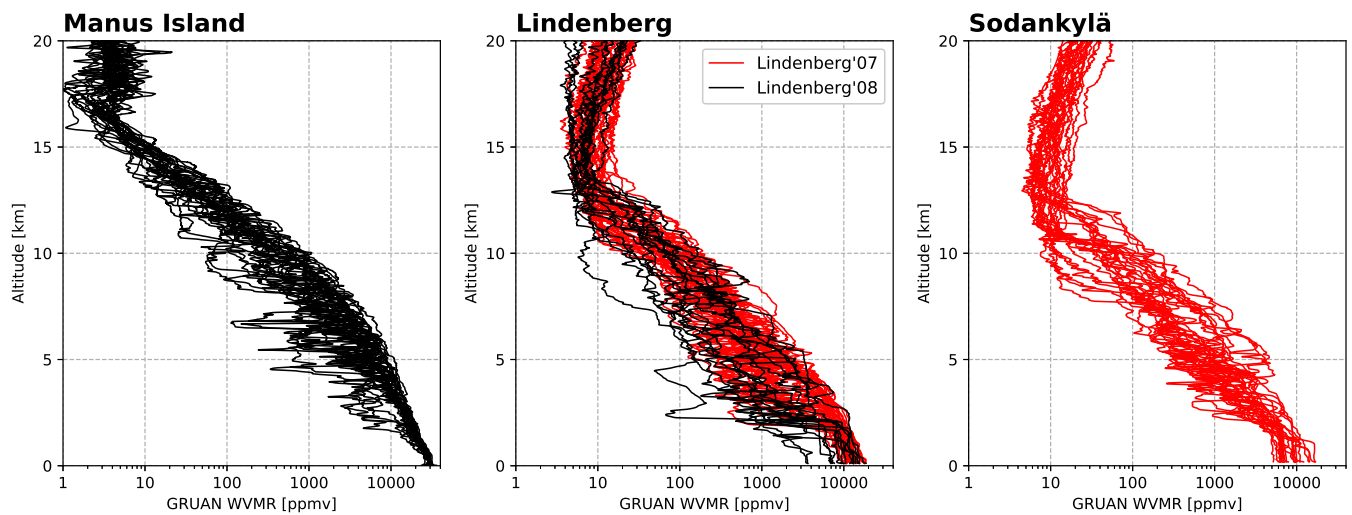
- August, T., Klaes, D., Schluessel, P., Hultberg, T., Crapeau, M., Arriaga, A., O'Carroll, A., Coppens, D., Munro, R., and Calbet, X.: IASI on Metop-A: Operational Level 2 retrievals after five years in orbit, 113, 1340–1371, 2012.
- 5 Barthlott, S., Schneider, M., Hase, F., Blumenstock, T., Kiel, M., Dubravica, D., García, O. E., Sepúlveda, E., Mengistu Tsidu, G., Takele Kenea, S., Grutter, M., Plaza-Medina, E. F., Stremme, W., Strong, K., Weaver, D., Palm, M., Warneke, T., Notholt, J., Mahieu, E., Servais, C., Jones, N., Griffith, D. W. T., Smale, D., and Robinson, J.: Tropospheric water vapour isotopologue data ( $\text{H}_2^{16}\text{O}$ ,  $\text{H}_2^{18}\text{O}$ , and  $\text{HD}^{16}\text{O}$ ) as obtained from NDACC/FTIR solar absorption spectra, *Earth System Science Data*, 9, 15–29, <https://doi.org/10.5194/essd-9-15-2017>, <http://www.earth-syst-sci-data.net/9/15/2017/>, 2017.
- 10 Calbet, X., Kivi, R., Tjemkes, S., Montagner, F., and Stuhlmann, R.: Matching radiative transfer models and radiosonde data from the EPS/Metop Sodankylä campaign to IASI measurements, *Atmospheric Measurement Techniques*, 4, 1177–1189, <https://doi.org/10.5194/amt-4-1177-2011>, <http://www.atmos-meas-tech.net/4/1177/2011/>, 2011.
- Calbet, X., Peinado-Galan, N., Rípodas, P., Trent, T., Dirksen, R., and Sommer, M.: Consistency between GRUAN sondes, LBLRTM and IASI, *Atmospheric Measurement Techniques*, 10, 2323–2335, <https://doi.org/10.5194/amt-10-2323-2017>, <https://www.atmos-meas-tech.net/10/2323/2017/>, 2017.
- 15 Clerbaux, C., Boynard, A., Clarisse, L., George, M., Hadji-Lazaro, J., Herbin, H., Hurtmans, D., Pommier, M., Razavi, A., Turquety, S., Wespes, C., and Coheur, P.-F.: Monitoring of atmospheric composition using the thermal infrared IASI/MetOp sounder, *Atmos. Chem. Phys.*, 9, 6041–6054, 2009.
- Delamere, J. S., Clough, S. A., Payne, V. H., Mlawer, E. J., Turner, D. D., and Gamache, R. R.: A far-infrared radiative closure study in the Arctic: Application to water vapor, *Journal of Geophysical Research: Atmospheres*, 115, n/a–n/a, <https://doi.org/10.1029/2009JD012968>, <http://dx.doi.org/10.1029/2009JD012968>, d17106, 2010.
- 20 Dirksen, R. J., Sommer, M., Immler, F. J., Hurst, D. F., Kivi, R., and Vömel, H.: Reference quality upper-air measurements: GRUAN data processing for the Vaisala RS92 radiosonde, *Atmospheric Measurement Techniques*, 7, 4463–4490, <https://doi.org/10.5194/amt-7-4463-2014>, <http://www.atmos-meas-tech.net/7/4463/2014/>, 2014.
- 25 Gao, B.-C. and Kaufman, Y. J.: Water vapor retrievals using Moderate Resolution Imaging Spectroradiometer (MODIS) near-infrared channels, *Journal of Geophysical Research: Atmospheres*, 108, n/a–n/a, <https://doi.org/10.1029/2002JD003023>, <http://dx.doi.org/10.1029/2002JD003023>, 4389, 2003.
- García, O. E., Sepúlveda, E., Schneider, M., Wiegeler, A., Borger, C., Hase, F., Barthlott, S., Blumenstock, T., and de Frutos, A. M.: Upper tropospheric  $\text{CH}_4$  and  $\text{N}_2\text{O}$  retrievals from MetOp/IASI within the project MUSICA, *Atmospheric Measurement Techniques Discussions*, 2017, 1–32, <https://doi.org/10.5194/amt-2016-326>, <http://www.atmos-meas-tech-discuss.net/amt-2016-326/>, 2017.
- 30 Gordon, I., Rothman, L., Hill, C., Kochanov, R., Tan, Y., Bernath, P., Birk, M., Boudon, V., Campargue, A., Chance, K., Drouin, B., Flaud, J.-M., Gamache, R., Hodges, J., Jacquemart, D., Perevalov, V., Perrin, A., Shine, K., Smith, M.-A., Tennyson, J., Toon, G., Tran, H., Tyuterev, V., Barbe, A., Császár, A., Devi, V., Furtenbacher, T., Harrison, J., Hartmann, J.-M., Jolly, A., Johnson, T., Karman, T., Kleiner, I., Kyuberis, A., Loos, J., Lyulin, O., Massie, S., Mikhailenko, S., Moazzen-Ahmadi, N., Müller, H., Naumenko, O., Nikitin, A., Polyansky, O., Rey, M., Rotger, M., Sharpe, S., Sung, K., Starikova, E., Tashkun, S., Auwera, J. V., Wagner, G., Wilzewski, J., Wcisło, P., Yu, S., and Zak, E.: The HITRAN2016 molecular spectroscopic database, *Journal of Quantitative Spectroscopy and Radiative Transfer*, <https://doi.org/https://doi.org/10.1016/j.jqsrt.2017.06.038>, <http://www.sciencedirect.com/science/article/pii/S0022407317301073>, 2017.

- Grossi, M., Valks, P., Loyola, D., Aberle, B., Slijkhuis, S., Wagner, T., Beirle, S., and Lang, R.: Total column water vapour measurements from GOME-2 MetOp-A and MetOp-B, *Atmospheric Measurement Techniques*, 8, 1111–1133, <https://doi.org/10.5194/amt-8-1111-2015>, <http://www.atmos-meas-tech.net/8/1111/2015/>, 2015.
- 5 Hase, F., Hannigan, J. W., Coffey, M. T., Goldman, A., Höpfner, M., Jones, N. B., Rinsland, C. P., and Wood, S.: Intercomparison of retrieval codes used for the analysis of high-resolution, *J. Quant. Spectrosc. Ra.*, 87, 25–52, 2004.
- Herbin, H., Hurtmans, D., Clerbaux, C., Clarisse, L., and Coheur, P.-F.:  $\text{H}_2^{16}\text{O}$  and HDO measurements with IASI/MetOp, *Atmos. Chem. Phys.*, 9, 9433–9447, 2009.
- Hess, M., Koepke, P., and Schult, I.: Optical Properties of Aerosols and Clouds: The Software Package OPAC, *Bulletin of the American Meteorological Society*, 79, 831–844, [https://doi.org/10.1175/1520-0477\(1998\)079<0831:OPOAAC>2.0.CO;2](https://doi.org/10.1175/1520-0477(1998)079<0831:OPOAAC>2.0.CO;2), [http://dx.doi.org/10.1175/1520-0477\(1998\)079<0831:OPOAAC>2.0.CO;2](http://dx.doi.org/10.1175/1520-0477(1998)079<0831:OPOAAC>2.0.CO;2), 1998.
- 10 Hyland, R. and Wexler, A.: Formulations for the thermodynamic properties of the saturated phases of  $\text{H}_2\text{O}$  from 173.15 K to 473.15 K, *ASHRAE transactions*, 89, 500–519, 1983.
- Koepke, P., Gasteiger, J., and Hess, M.: Technical Note: Optical properties of desert aerosol with non-spherical mineral particles: data incorporated to OPAC, *Atmospheric Chemistry and Physics*, 15, 5947–5956, <https://doi.org/10.5194/acp-15-5947-2015>, <http://www.atmos-chem-phys.net/15/5947/2015/>, 2015.
- 15 Masuda, K., Takashima, T., and Takayama, Y.: Emissivity of pure and sea waters for the model sea surface in the infrared window regions, *Remote Sensing of Environment*, 24, 313 – 329, [https://doi.org/http://dx.doi.org/10.1016/0034-4257\(88\)90032-6](https://doi.org/http://dx.doi.org/10.1016/0034-4257(88)90032-6), <http://www.sciencedirect.com/science/article/pii/0034425788900326>, 1988.
- 20 McMillin, L. M., Zhao, J., Rama Varma Raja, M. K., Gutman, S. I., and Yoe, J. G.: Radiosonde humidity corrections and potential Atmospheric Infrared Sounder moisture accuracy, *J. Geophys. Res.*, 112, D13S90, <https://doi.org/10.1029/2005JD006109>, 2007.
- Mlawer, E. J., Payne, V. H., Moncet, J.-L., Delamere, J. S., Alvarado, M. J., and Tobin, D. C.: Development and recent evaluation of the MT\_CKD model of continuum absorption, *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 370, 2520–2556, <https://doi.org/10.1098/rsta.2011.0295>, 2012.
- 25 Payne, V. H., Mlawer, E. J., Cady-Pereira, K. E., and Moncet, J. L.: Water Vapor Continuum Absorption in the Microwave, *IEEE Transactions on Geoscience and Remote Sensing*, 49, 2194–2208, <https://doi.org/10.1109/TGRS.2010.2091416>, 2011.
- Pequignot, E., Blumstein, D., and Larigauderie, C.: CNES Technical Note: IASI Noise Covariance Matrix, Tech. rep., IA-TN-0000-3271-CNE, 2008.
- Pougatchev, N., August, T., Calbet, X., Hultberg, T., Oduleye, O., Schlüssel, P., Stiller, B., Germain, K. S., and Bingham, G.: IASI temperature and water vapor retrievals – error assessment and validation, *Atmospheric Chemistry and Physics*, 9, 6453–6458, <https://doi.org/10.5194/acp-9-6453-2009>, <https://www.atmos-chem-phys.net/9/6453/2009/>, 2009.
- 30 Rees, D., Barnett, J., and Labitzke, K.: Cospar international reference atmosphere: 1986 part II. Middle atmosphere models., *Adv. Space Res.*, 10, 520, 1990.
- Risi, C., Noone, D., Worden, J., Frankenberg, C., Stiller, G., Kiefer, M., Funke, B., Walker, K., Bernath, P., Schneider, M., Wunch, D., Sherlock, V., Deutscher, N., Griffith, D., Wennberg, P., Strong, K., Barthlott, S., Hase, F., G.-O., Smale, D., Mahieu, E., Sayres, D., Bony, S., Lee, J., Brown, D., Uemura, R., and Sturm, C.: Process-evaluation of tropospheric humidity simulated by general circulation models using water vapor isotopic observations. Part 1: comparison between models and datasets, *J. Geophys. Res.*, 117, <https://doi.org/10.1029/2011JD016621>, 2012.
- 35 Rodgers, C.: *Inverse Methods for Atmospheric Sounding: Theory and Praxis*, World Scientific Publishing Co., Singapore, 2000.

- Rosenkranz, P. W.: Retrieval of water vapor from AMSU-A and AMSU-B measurements, in: IGARSS 2000. IEEE 2000 International Geoscience and Remote Sensing Symposium. Taking the Pulse of the Planet: The Role of Remote Sensing in Managing the Environment. Proceedings (Cat. No.00CH37120), vol. 7, pp. 3033–3035 vol.7, <https://doi.org/10.1109/IGARSS.2000.860326>, 2000.
- 5 Schneider, M. and Hase, F.: Improving spectroscopic line parameters by means of atmospheric spectra: Theory and example for water vapour and solar absorption spectra, *J. Quant. Spectrosc. Radiat. Transfer*, 110, 1825–1839, <https://doi.org/10.1016/j.jqsrt.2009.04.011>, 2009.
- Schneider, M. and Hase, F.: Optimal estimation of tropospheric H<sub>2</sub>O and  $\delta$ D with IASI/METOP, *Atmos. Chem. Phys.*, 11, 11 207–11 220, <https://doi.org/10.5194/acp-11-11207-2011>, 2011.
- Schneider, M., Wiegeler, A., Barthlott, S., González, Y., Christner, E., Dyroff, C., García, O. E., Hase, F., Blumenstock, T., Sepúlveda, E.,  
10 Mengistu Tsidu, G., Takele Kenea, S., Rodríguez, S., and Andrey, J.: Accomplishments of the MUSICA project to provide accurate, long-term, global and high-resolution observations of tropospheric {H<sub>2</sub>O, $\delta$ D} pairs – a review, *Atmospheric Measurement Techniques*, 9, 2845–2875, <https://doi.org/10.5194/amt-9-2845-2016>, <http://www.atmos-meas-tech.net/9/2845/2016/>, 2016.
- Schneider, M., Borger, C., Wiegeler, A., Hase, F., García, O. E., Sepúlveda, E., and Werner, M.: MUSICA MetOp/IASI {H<sub>2</sub>O, $\delta$ D} pair retrieval simulations for validating tropospheric moisture pathways in atmospheric models, *Atmospheric Measurement Techniques*, 10,  
15 507–525, <https://doi.org/10.5194/amt-10-507-2017>, 2017.
- Seemann, S. W., Borbas, E. E., Knuteson, R. O., Stephenson, G. R., and Huang, H.-L.: Development of a Global Infrared Land Surface Emissivity Database for Application to Clear Sky Sounding Retrievals from Multispectral Satellite Radiance Measurements, *Journal of Applied Meteorology and Climatology*, 47, 108–123, <https://doi.org/10.1175/2007JAMC1590.1>, <http://dx.doi.org/10.1175/2007JAMC1590.1>, 2008.
- 20 Sommer, M., Dirksen, R., and von Rohden, C.: Brief Description of the RS92 GRUAN Data Product (RS92-GDP), GRUAN-TD-4, Tech. rep., GRUAN Lead Centre, <https://www.gruan.org/documentation/gruan/td/gruan-td-4/>, version 2, 2016.
- Stevens, B. and Bony, S.: What Are Climate Models Missing?, *Science*, 340, 1053–1054, <https://doi.org/10.1126/science.1237554>, <http://www.sciencemag.org/content/340/6136/1053.short>, 2013.
- Stiller, G.: The Karlsruhe optimized and precise radiative transfer algorithm (KOPRA), Tech. rep., Karlsruhe, wissenschaftliche Berichte /  
25 Forschungszentrum Karlsruhe in der Helmholtz-Gemeinschaft, FZKA ; 6487, 2000.
- Susskind, J., Barnett, C. D., and Blaisdell, J. M.: Retrieval of atmospheric and surface parameters from AIRS/AMSU/HSB data in the presence of clouds, *IEEE Transactions on Geoscience and Remote Sensing*, 41, 390–409, <https://doi.org/10.1109/TGRS.2002.808236>, 2003.
- Wiegeler, A., Schneider, M., Hase, F., Barthlott, S., García, O. E., Sepúlveda, E., González, Y., Blumenstock, T., Raffalski, U., Gisi, M., and Kohlhepp, R.: The MUSICA MetOp/IASI H<sub>2</sub>O and  $\delta$ D products: characterisation and long-term comparison to NDACC/FTIR data, *Atmospheric Measurement Techniques*, 7, 2719–2732, <https://doi.org/10.5194/amt-7-2719-2014>, <http://www.atmos-meas-tech.net/7/2719/2014/>, 2014.
- 30 Worden, J., Kulawik, S., Frankenberg, C., Payne, V., Bowman, K., Cady-Peirara, K., Wecht, K., Lee, J.-E., and Noone, D.: Profiles of CH<sub>4</sub>, HDO, H<sub>2</sub>O, and N<sub>2</sub>O with improved lower tropospheric vertical resolution from Aura TES radiances, *Atmospheric Measurement Techniques*, 5, 397–411, <https://doi.org/10.5194/amt-5-397-2012>, <http://www.atmos-meas-tech.net/5/397/2012/>, 2012.
- 35 Zhang, J., Chen, H., Li, Z., Fan, X., Peng, L., Yu, Y., and Cribb, M.: Analysis of cloud layer structure in Shouxian, China using RS92 radiosonde aided by 95 GHz cloud radar, *Journal of Geophysical Research: Atmospheres*, 115, n/a–n/a, <https://doi.org/10.1029/2010JD014030>, <http://dx.doi.org/10.1029/2010JD014030>, d00K30, 2010.

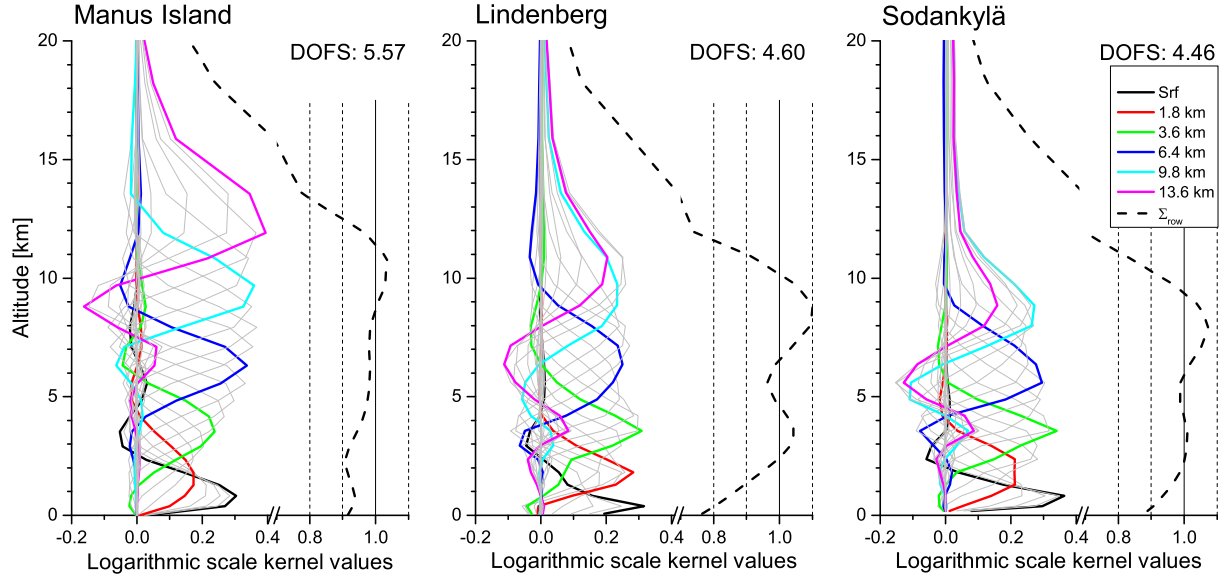


**Figure 1.** Example of an infrared ~~spectra~~-spectrum measured by IASI (upper panel) and residuals between satellite observation and radiative transfer simulation (bottom panel) at Manus Island (2012-10-15 11:46:26 UT, satellite zenith angle ~~10.16°~~10.2°, integrated water vapour ~~59.49 kg m<sup>-2</sup>~~48.0 mm). The red lines in the bottom panel indicate the typical IASI noise measurement level as given by the square root values of the diagonal elements of the IASI noise covariance matrix (Pequignot et al., 2008).

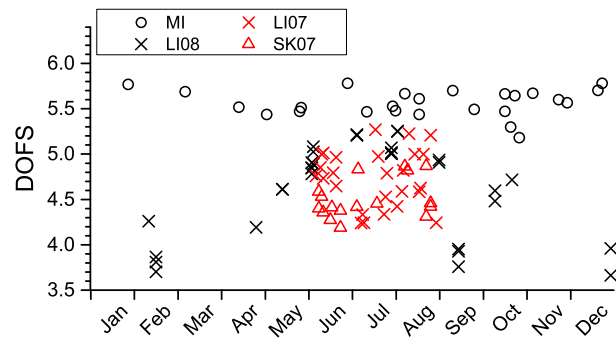


**Figure 2.** Vertical H<sub>2</sub>O profiles as measured by the 100 different GRUAN processed Vaisala RS92 radiosondes: from Manus Island, Lindenberg and Sodankylä used for our study. Black lines indicate radiosonde data ensembles that cover all seasons (Manus Island and Lindenberg 2008) and red lines indicate ensembles that cover the summer season only.

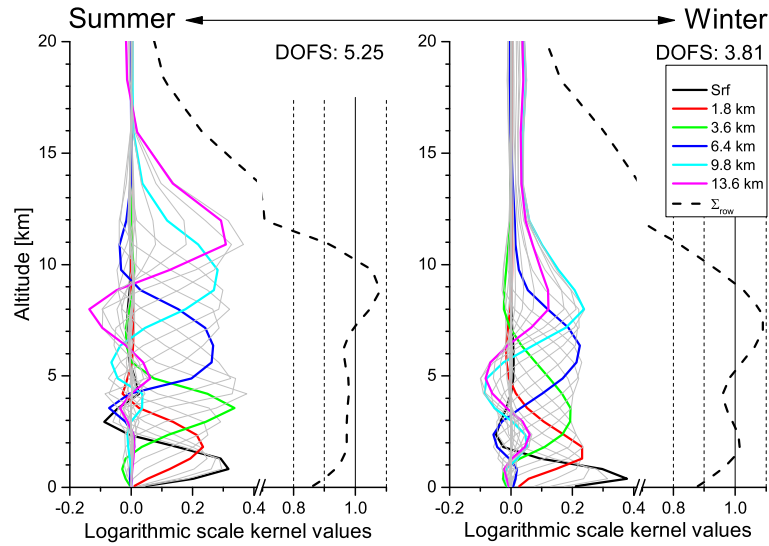




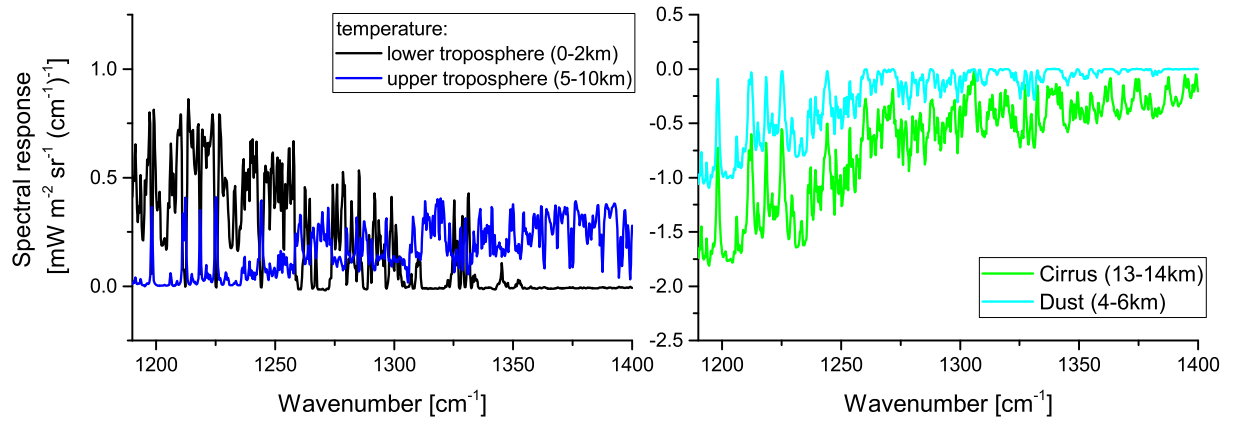
**Figure 3.** Example row kernels ( $\mathbf{A}_{11} + \mathbf{A}_{12}$ , see Eq. 11) for the three reference sites. Manus Island: 2013-11-28 11:39:37:1424 UT, satellite zenith angle  $12.44^\circ$   $12.4^\circ$ , ~~integrated~~ precipitable water vapour  $55.75 \text{ kg m}^{-2}$  46.4 mm; Lindenberg: ~~2008-06-03-08~~ 2008-10-08 20:5900:1438 UT,  $23.95^\circ$  16.3^\circ,  $20.05 \text{ kg m}^{-2}$  14.1 mm; Sodankylä: ~~2007-06-07-18~~ 2007-08-24 08:4031:3825 UT,  $17.70^\circ$  27.8^\circ,  $11.17 \text{ kg m}^{-2}$  13.7 mm. Numbers in the upper right corners of every panel indicate the respective degrees of freedom of signal (DOFS). Row kernels of selected altitudes are highlighted by thick coloured lines. The thick black dashed line represents the sum along the row of the averaging kernel matrix.



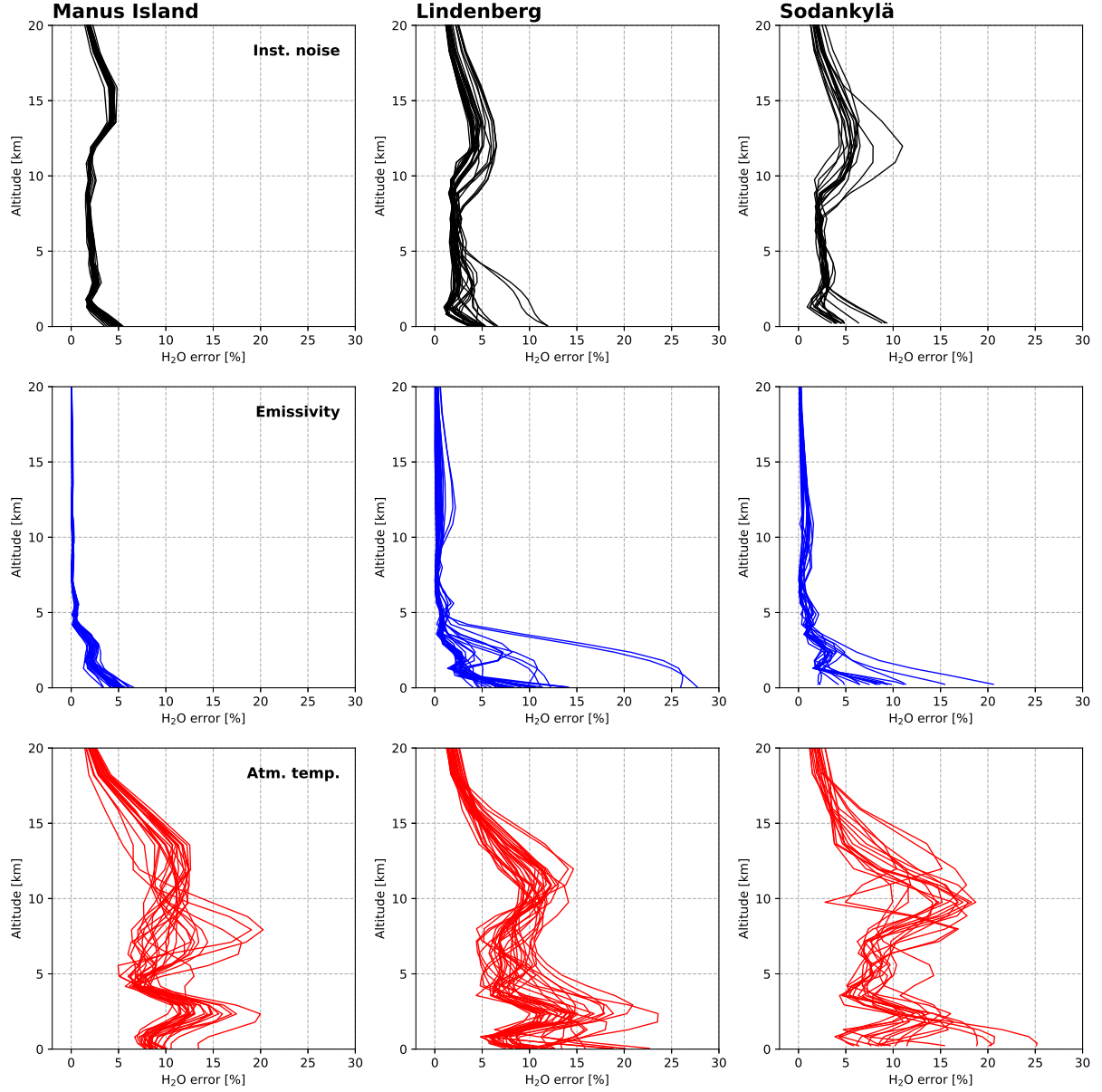
**Figure 4.** Variation of the DOFS values (degree-degrees of freedom for signal) for the three-reference-sitesfour different ensembles.



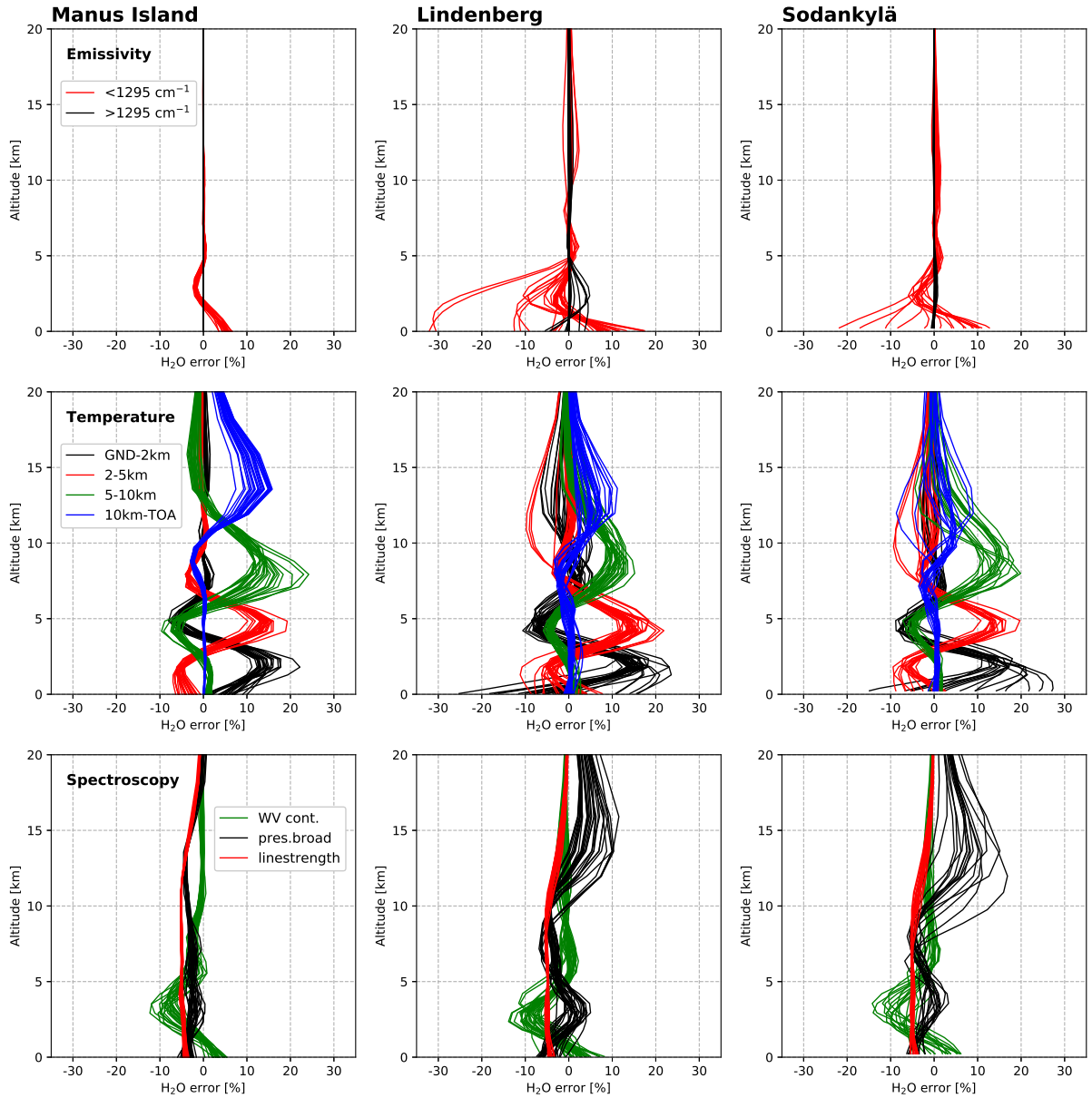
**Figure 5.** Spectral responses of uncertainty sources for a typical situation at Manus Island (same situation Same as for the kernel in Fig. 3). The left panel shows examples for the influence of uncertainties in temperatures (surface skin  $\Delta T = +2$  K, lower tropospheric  $\Delta T = +2$  K and upper tropospheric  $\Delta T = +1$  K). The right panel illustrates examples but for the influence of clouds (dust layer (4-6 km) and cirrus cloud (13-14 km a typical winter and 50% cloud fraction)) on the spectrum summer observation above Lindenberg. Please note the different y-axis scales Summer observation : 2008-08-01 08:39:01 UT, i.e. the positive response for positive temperature uncertainties and the negative response for unrecognized clouds satellite zenith angle 43.7°, precipitable water vapour 28.8 mm; Winter observation: 2008-02-15 09:54:30 UT, 41.2°, 3.2 mm.



**Figure 6.** Spectral responses of uncertainty sources for a typical situation at Manus Island (same situation as for the kernel in Fig. 3). The left panel shows examples for the influence of uncertainties in temperatures (surface skin  $\Delta T = +2$  K, lower tropospheric  $\Delta T = +2$  K and upper tropospheric  $\Delta T = +1$  K). The right panel illustrates examples for the influence of clouds (dust layer (4-6 km) and cirrus cloud (13-14 km and 50% cloud fraction)) on the spectrum. Note the different y-axis scales, i.e. the positive response for positive temperature uncertainties and the negative response for unrecognized clouds.



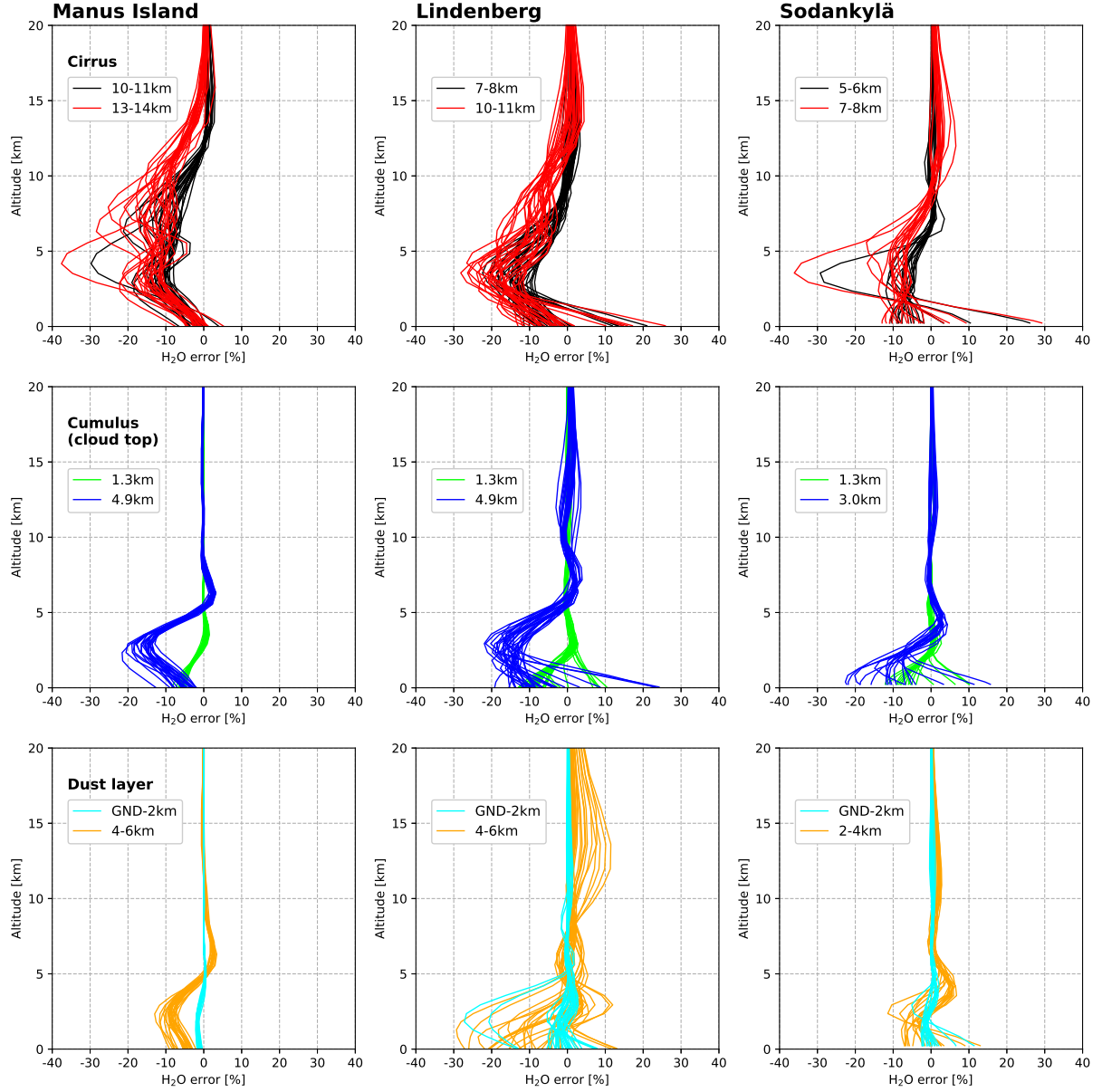
**Figure 7.**  $\text{H}_2\text{O}$  error profiles caused by derived from the random uncertainty sources: instrument noise, emissivity and atmospheric temperatures. Shown are the square root value values of the diagonal of the matrix  $\mathbf{S}_{\mathbf{e}}$  matrices  $\mathbf{S}_{\mathbf{x},\text{noise}}$  (for instrument noise) and  $\mathbf{S}_{\mathbf{x},\text{b}}$  (for emissivity and atmospheric temperatures) according to EqEqs. (9) and (8), respectively. Shown The data are the error profiles depicted for every situation examined in for all members of the validation study MI, LI08, and SK07 ensembles.



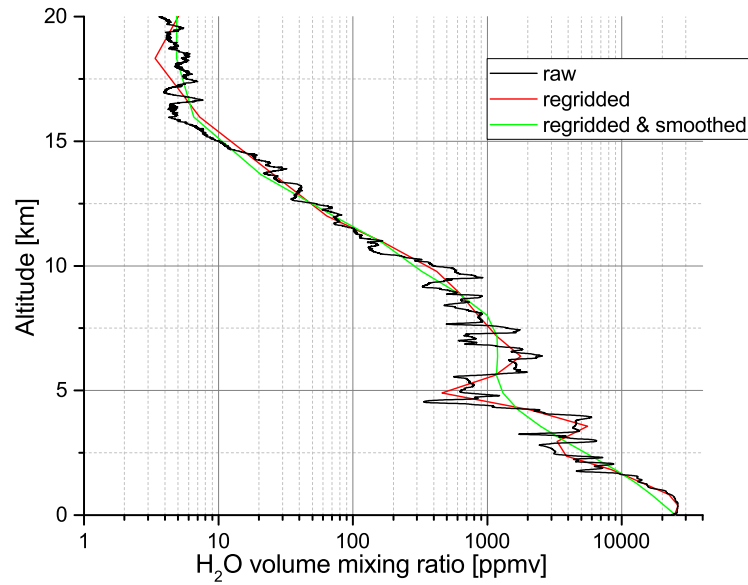
Same as Fig. ??, but for errors due to uncertainties and 1 K in the  $\text{H}_2\text{O}$  spectroscopic parameters (other layers), and spectroscopy (linestrength line strength, +5%, and pressure broadening, +5%) and the water vapour continuum (assuming a 10% underestimation of the “MT\_CKD” model). Shown are the errors in the atmospheric state vector  $\Delta\hat{x}$  according to Eqs. (7). The data are depicted for for all members of the MI, LI08, and SK07 ensembles.

Same as Fig. ??, but for errors due to uncertainties and 1 K in the  $\text{H}_2\text{O}$  spectroscopic parameters (other layers), and spectroscopy (linestrength line strength, +5%, and pressure broadening, +5%) and the water vapour continuum (assuming a 10% underestimation of the “MT\_CKD” model). Shown are the errors in the atmospheric state vector  $\Delta\hat{x}$  according to Eqs. (7). The data are depicted for for all members of the MI, LI08, and SK07 ensembles.

**Figure 8.**  $\text{H}_2\text{O}$  error pattern profiles  $\hat{x}_e$  according to Eq. due to uncertainties in surface properties (skin temperature and derived from the systematic uncertainty sources: emissivity (-1% in two different wavenumber regions) and, atmospheric air temperature temperatures (lower-tropospheric, mid-tropospheric, upper-tropospheric 2 K between surface and upper-atmospheric) at the 3 different reference sites 2 km



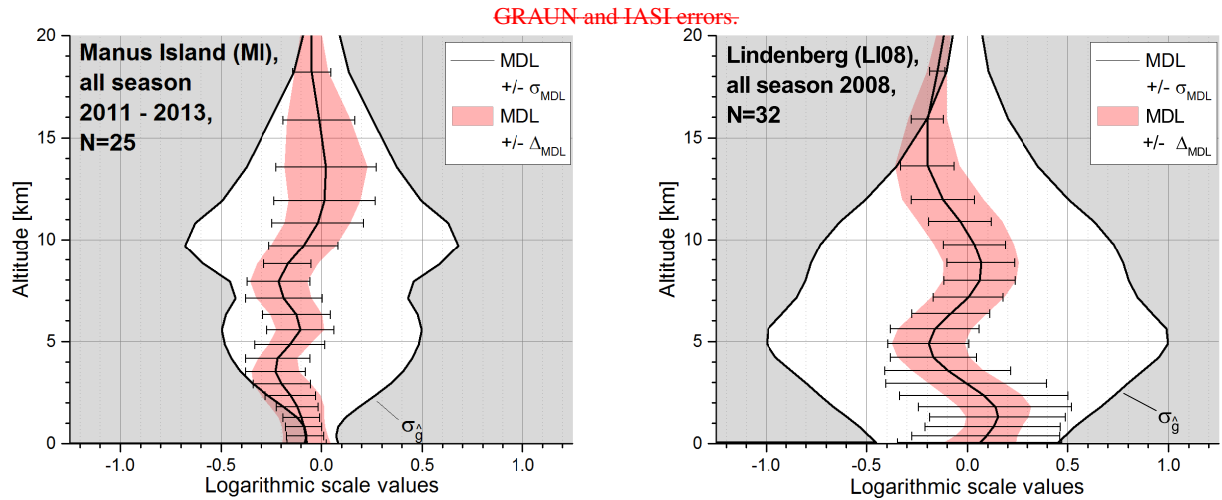
**Figure 9.** Same as Fig. ??B, but for errors due to unrecognized clouds (cirrus, mineral dust, cumulus). Top panels: cirrus clouds with (50% fractional coverage) located at typical middle/upper tropospheric altitudes (location of cloud layers see legend); Middle panels: cumulus clouds with (10% fractional coverage with cloud top altitudes as given in the legends; Bottom panel: mineral dust (homogeneous dust clouds with layers as given in the legends and with the composition according to OPAC "Desert").



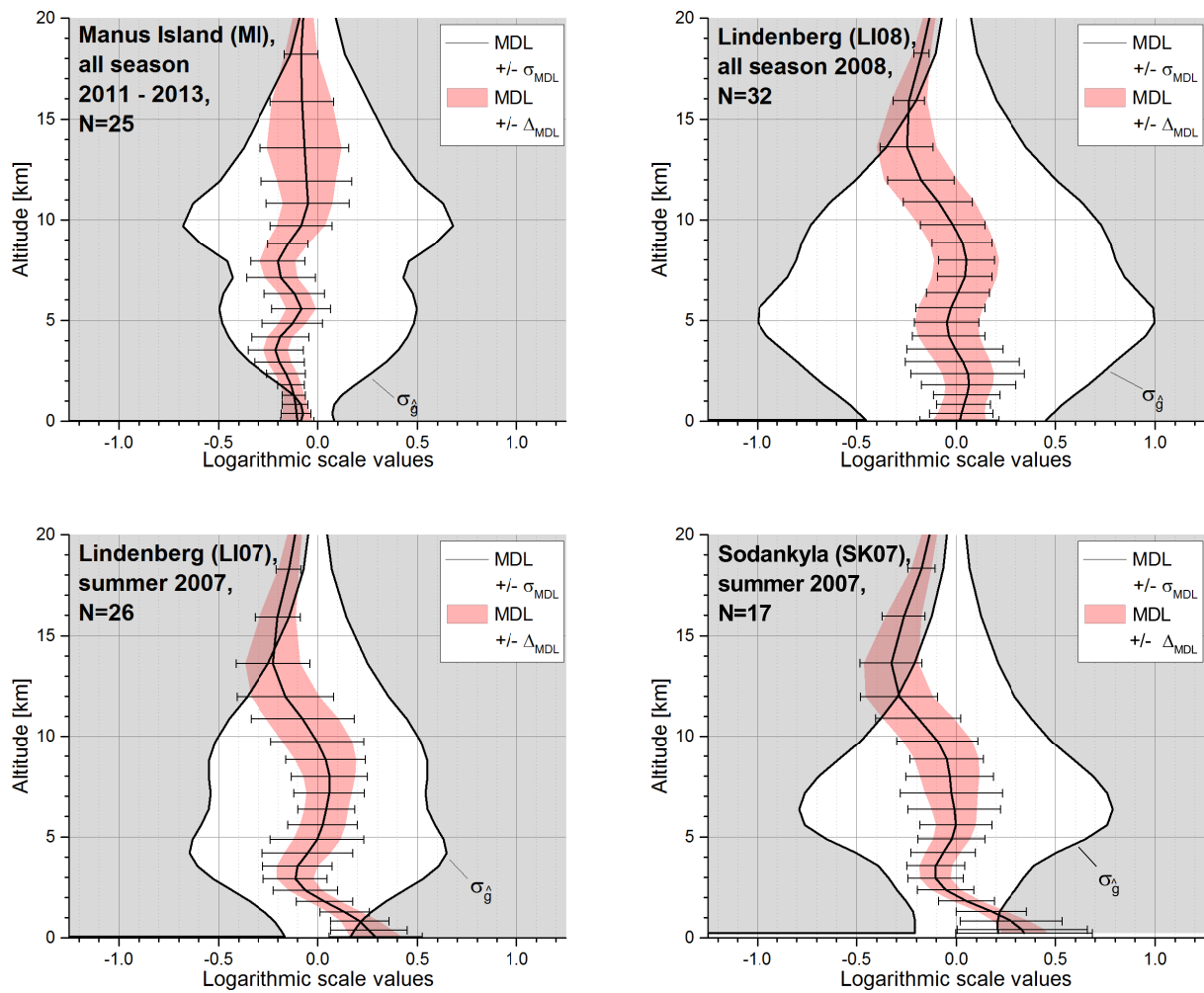
**Figure 10.** Example for the regridding and smoothing of the GRUAN data required before validating the MUSICA ~~MetOp~~/IASI retrieval of  $\text{H}_2\text{O}$  profiles. Black line: raw GRUAN data; Red line: regridded GRUAN data ( $\hat{x}_{\text{GRUAN}}^{\text{red}}$ ); Green line: regridded and smoothed radiosonde data ( $\hat{x}_{\text{GRUAN}}^{\text{green}}$ , according to Eq. 13).



Correlation between GRUAN (along y-axes) and MUSICA MetOp/IASI data (along x-axes) at the 3 different reference sites for 3 different atmospheric levels (lower, middle and upper troposphere). The yellow star represents the a priori assumption for the respective retrieval level (the retrieval uses globally the same a priori). The respective retrieval level altitudes are given in the individual scatter plots. Red and black colour distinguish the remote sensing data ensembles that use different input data (MI and LI08, on the one hand, and LI07 and SK07, on the other hand; see Sect. 3.1 and Tab. 1). The blue dotted line represents the 1-to-1 diagonal and the blue error bars indicate the typical



**Figure 11.** Vertical profiles of retrieval skill scores calculated according to Eqs. (16)-and using data from all ensembles - (19) for the MI ; LI07, LI08 and SK07) LI08 ensembles. MDL (red line) is and error bars represent the mean difference and the 1σ scatter between IASI and smoothed GRUAN data ; ±σ<sub>MDL</sub> (blue lines MDL and blue ±σ<sub>MDL</sub>). The red shaded area around MDL is the 1σ scatter between expected due to MUSICA IASI and smoothed GRUAN data ; ±σ<sub>GRUAN</sub> errors (black lines ±Δ<sub>MDL</sub>) is. The grey shaded area represents the area beyond the 1σ variability of smoothed GRUAN data (area beyond ±σ<sub>g</sub>).



**Figure 12.** Same as Fig.11, but for MUSICA IASI retrievals that use the GRUAN temperature profiles as the a priori atmospheric temperatures and all four ensembles (MI, LI08, LI07 and SO).

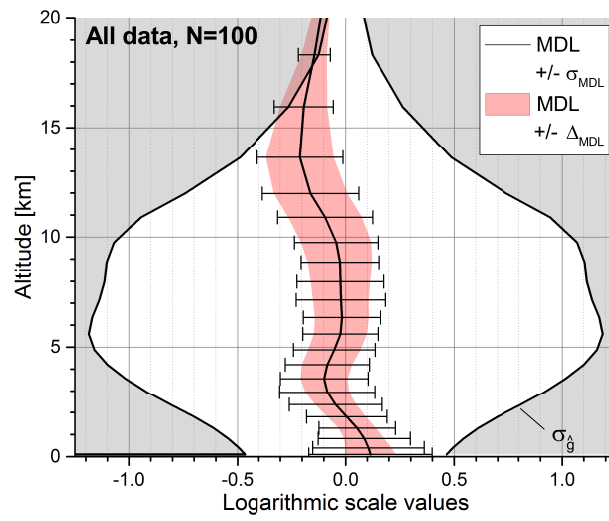
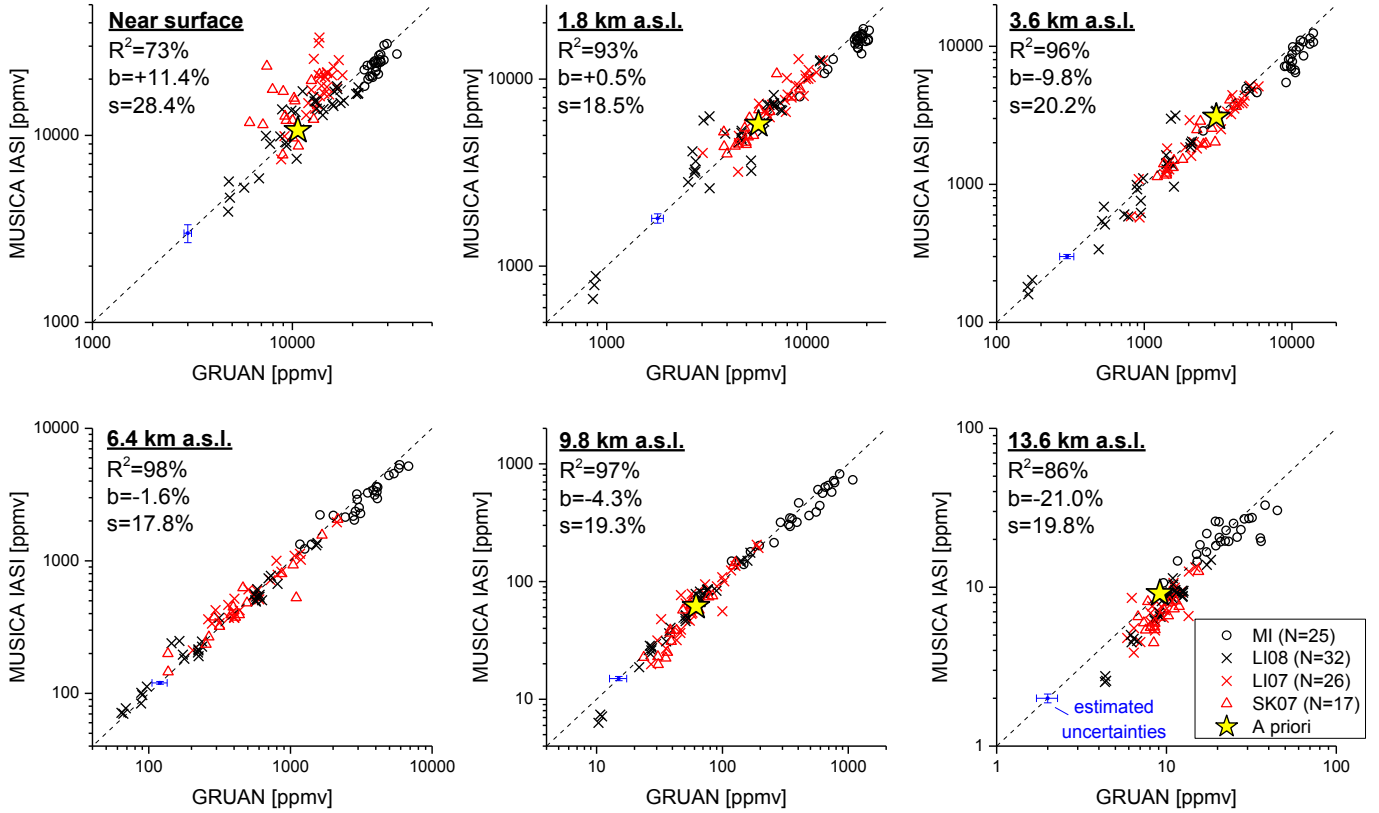
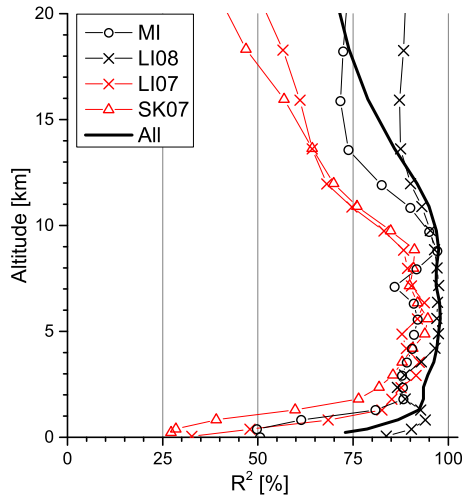


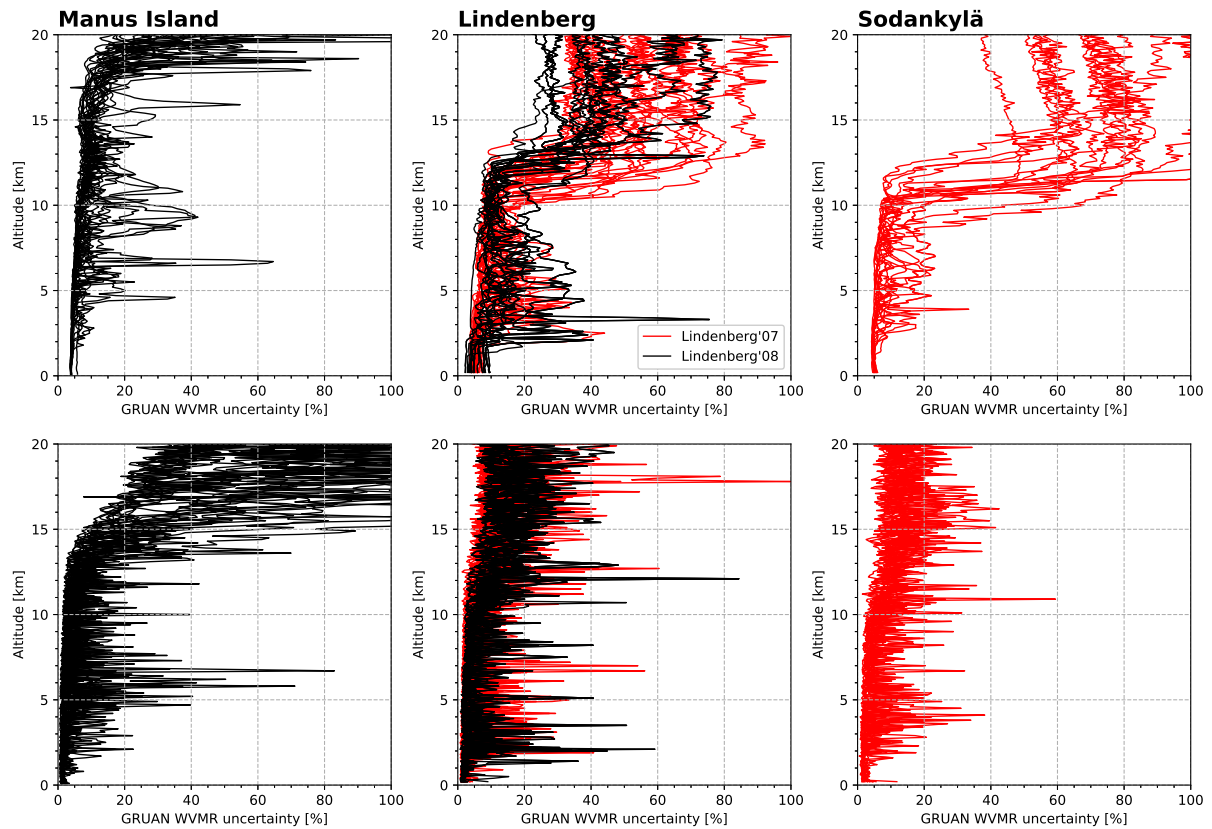
Figure 13. [Same as Fig.12, but considering all four ensembles as a single data set.](#)



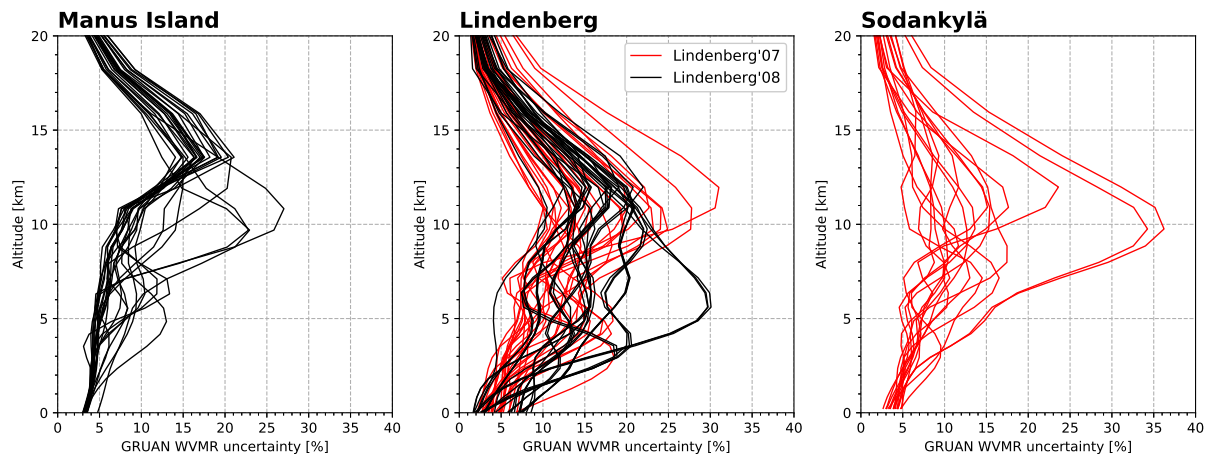
**Figure 14.** Correlation between GRUAN (along x-axes) and MUSICA IASI data (along y-axes) for the 6 different atmospheric altitudes that are highlighted in Figs. 3 and 5. All the presented data are for MUSICA IASI retrievals that use the GRUAN temperature profiles as the a priori atmospheric temperatures. The retrieval altitudes are given in the individual scatter plots, together with correlation coefficient ( $R^2$ ), bias (b) and scatter (s). Data belonging to the different ensembles can be identified by the symbols and colours as described in the legend (bottom right). The yellow stars represent the a priori value (the retrieval uses globally the same  $H_2O$  a priori) and the blue error bars indicate the typical GRAUN and IASI errors. The dotted line represents is the 1-to-1 diagonal.



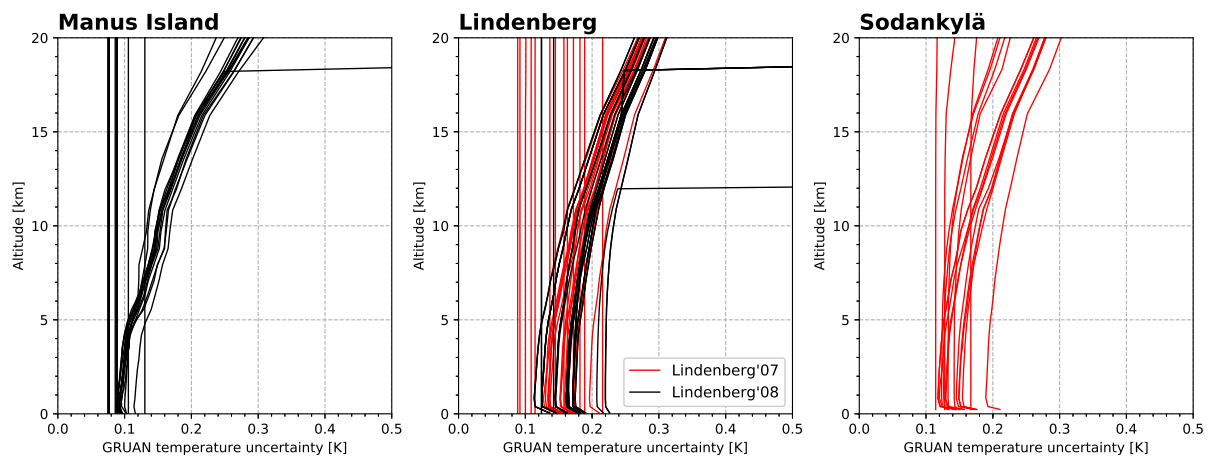
**Figure 15.** Profiles of correlation coefficients ( $R^2$ ) for comparison between GRUAN and MUSICA IASI (for retrievals that use the GRUAN temperature profiles as the a priori atmospheric temperatures). Different line colours and symbols for the different ensembles and thick black solid line for considering all four ensembles as a single data set (the  $R^2$  values for the latter are also written in the panels of Fig. 14).



**Figure A1.** Profiles of the WVMR errors of the GRUAN radiosondes: The top panels represent the correlated errors and the bottom panels the uncorrelated errors. The colours distinguish the different ensembles of the retrieval setup: black for MI and LI08, red for LI07 and SK07.



**Figure A2.** Same as top panels of Fig. A1, but for the correlated errors in the regridded and smoothed GRUAN radiosonde data.



**Figure A3.** Profiles of the correlated temperature errors of the GRUAN radiosondes. Above the radiosondes top height we use a zonally and monthly mean temperature climatology and assume an uncertainty of 5 K.



**Table 1.** Overview of on the different ensembles of GRUAN reference and the available retrieval input datasets.

	Manus Island	Lindenberg 2008	Lindenberg 2007	Sodankylä 2007
<del>time</del> -Acronym	<u>MI</u>	<u>LI08</u>	<u>LI07</u>	<u>SK07</u>
<u>Time</u> period	2011-2013	2008 (all months)	2007 ( <del>June, — July,</del> <del>August</del> <u>June-August</u> )	2007 ( <del>June, — July,</del> <del>August</del> <u>June-August</u> )
<del>ground</del> - <del>Number</del> of <u>independent</u> GRUAN <u>sondes</u>	<u>25</u>	<u>32</u>	<u>26</u>	<u>17</u>
<u>Ground</u> level	<u>EUMETSAT</u> IASI L2 (GTOPO30)	<u>EUMETSAT</u> IASI L2 (GTOPO30)	GTOPO30	GTOPO30
<del>emissivity</del> - <u>Emissivity</u>	<u>EUMETSAT</u> IASI L2 (Masuda et al., 1988)	<u>EUMETSAT</u> IASI L2 (IREMIS)	IREMIS	IREMIS
<del>cloud</del> - <u>Cloud</u> identifica- tion	<u>EUMETSAT</u> IASI L2, + visual inspection	<u>EUMETSAT</u> IASI L2	Zhang et al. (2010)	Zhang et al. (2010)
<del>a</del> - <u>A</u> priori for atmo- spheric and surface skin temperature	<u>EUMETSAT</u> IASI L2 (PPF v5)	<u>EUMETSAT</u> IASI L2 (PPF v4)	<del>Radiosonde</del> - <del>Radiosonde</del> <del>number</del> — <del>of</del> — <del>GRUAN</del> <del>sondes</del> — <del>25</del> — <del>32</del> — <del>26</del> <u>GRUAN sonde</u>	<del>17</del> - <u>GRUAN sonde</u>

**Table 2.** List of uncertainty assumptions used for the error estimation of the MUSICA ~~MetOp~~/IASI water vapour product (for emissivity and atmospheric temperature we assume random and systematic uncertainties).

<del>uncertainty source</del> <u>Source</u>	<del>uncertainty value</del> <u>Type</u>	<u>Value</u>
<u>Instrumental noise</u>	<u>Random (Gaussian pdf)</u>	<u>Noise covariance according to Pequignot et al. (2008)</u>
Surface emissivity	<del>2% Surface skin temperature</del> <u>Random (Gaussian pdf) + Systematic</u>	<del>2 K</del> <u>Random: 1% with spectral frequency correlation length of <math>100\text{ cm}^{-1}</math>; Systematic: -1% for <math>&lt;1295\text{ cm}^{-1}</math> and -1% for <math>&gt;1295\text{ cm}^{-1}</math></u>
<del>Temperature in lower troposphere (0-2 km)</del> <u>EUMETSAT L2 atmos. temp.</u>	<del>2 K</del> <u>Temperature in middle troposphere (Random (Gaussian pdf) + Systematic</u>	<u>Random: 2 K from ground-2 km and 1 K above 2 km altitude with correlation length increasing from 2 km at ground to 10 km in the stratosphere; Systematic: 2 K for ground-2 km, and 1 K for 2-5 km)</u> <del>+ K</del> <u>Temperature in upper troposphere (5-10 km) + K</u> <del>Temperature in upper atmosphere (10 km-) + K</del> <u>and 10 km-TOA</u>
Water vapour continuum	<u>Systematic</u>	10% of model MT_CKD v2.5.2
Line intensity $\text{H}_2\text{O}$ <u>and HDO</u>	<u>Systematic</u>	5%
Pressure-broadening $\text{H}_2\text{O}$ <u>and HDO</u>	<del>5% Line intensity HDO 5%</del> <u>Pressure-broadening HDO Systematic</u>	5%
<del><math>\text{CH}_4</math> concentrations 2% <math>\text{N}_2\text{O}</math> concentrations 2%</del> Opaque cumulus cloud	<u>Random (unknown pdf) but systematic sign</u>	10% fractional cover with cloud top at 1.3, 3.0 and 4.9 km
Cirrus cloud	<del>particle</del> <u>Random (unknown pdf) but systematic sign</u>	<u>Particle</u> properties according to OPAC "Cirrus 3", 1 km thickness, 50% fractional cover with cloud top at 6, 8, 11 and 14 km
Mineral dust cloud	<del>particle</del> <u>Random (unknown pdf) but systematic sign</u>	<u>Particle</u> properties according to OPAC "Desert", homogeneous coverage for layers: ground-2 km, 2-4 km and 4-6 km