

A 1D Var Retrieval of Relative Humidity Using the ERA5 Dataset for the Assimilation of Raman Lidar Measurements

S. MAHAGAMMULLA GAMAGE,^a R. J. SICA,^{a,b} G. MARTUCCI,^b AND A. HAEFELE^{b,a}

^a *Department of Physics and Astronomy, University of Western Ontario, London, Ontario, Canada;*

^b *Federal Office of Meteorology and Climatology, Payerne, Switzerland*

(Manuscript received 3 October 2019, in final form 27 August 2020)

ABSTRACT: We present a one-dimensional variational (1D Var) retrieval of fifth-generation European Centre for Medium-Range Weather Forecast reanalysis (ERA5) temperature and relative humidity profiles above Payerne, Switzerland, assimilating raw backscatter measurements from the MeteoSwiss Raman Lidar for Meteorological Observations (RALMO). Our reanalysis is called ERA5-reRH. We use an optimal estimation method to perform the 1D Var data retrieval. The forward model combines the Raman lidar equation with the Hyland and Wexler expression for water vapor saturation pressure. The error covariance matrix of ERA5 was derived from the differences between ERA5 and a set of 50 special radiosoundings that have not been assimilated into ERA5. We validate ERA5-reRH, ERA5, and RALMO temperature and relative humidity profiles against the same set of special radiosoundings and found the best agreement was with our reanalysis, with a bias of less than 2% relative humidity with respect to water (%RH_w) and a spread of less than 8%RH_w below 8 km in terms of relative humidity. Improvements for temperature in our reanalysis are only found in the boundary layer, as ERA5 assimilates a large number of upper-air temperature observations. Our retrieval also provides a full uncertainty budget of the reanalyzed temperature and relative humidity including both random and systematic uncertainties.

KEYWORDS: Humidity; Temperature; Lidars/Lidar observations; Data assimilation; Reanalysis data

1. Introduction

Water vapor is of great importance to the atmosphere's radiative budget, chemistry, and dynamics. The global radiosonde network provides most of the relative humidity measurements for forecast models. The radiosonde network is widely spread over the world and provides relative humidity measurements with high vertical resolution. However, the temporal resolution of the routine sonde measurements is typically only two soundings per day. Also, it is well known that the radiosonde relative humidity measurements are often not reliable in the upper troposphere and above (Miloshevich et al. 2001; Noh et al. 2016; Ferreira et al. 2019).


There is a series of other techniques to measure atmospheric water vapor including active and passive remote sensing instruments deployed from surface and space as well as airborne in situ instruments. For an extensive overview, the reader is referred to Kämpfer (2012).

The Raman lidar is one of the best instruments for measurements of water vapor throughout the troposphere, with high vertical and temporal resolutions (Whiteman et al. 1992). For Raman lidars that possess temperature profiling capability exploiting pure rotational Raman scattering, the water vapor information can be combined with temperature to yield relative humidity. Mattis et al. (2002) report an uncertainty between 5% and 25% relative humidity with respect to water, hereafter referred to as %RH_w, with temperature being

the dominant source of uncertainty. Using the Hyland and Wexler formulation (1983) one can show that a 1–2 K temperature accuracy is required to reduce the relative uncertainty in the relative humidity to, on average, less than 10%RH_w.

In this study we apply a one-dimensional variational (1D Var) data assimilation scheme to reanalyze the fifth-generation European Centre for Medium-Range Weather Forecasts reanalysis (ERA5) relative humidity profiles above Payerne, Switzerland, assimilating measurements from a Raman lidar capable of measuring water vapor mixing ratio and temperature. The 1D Var scheme is based on the work in Gamage et al. (2019) and Sica and Haeefe (2016) and uses raw (level 0) measurements from the lidar, as opposed to computed water vapor mixing ratio or temperature profiles determined from the lidar. The 1D Var reanalyzed ERA5 profiles (ERA5-reRH) include a complete profile-by-profile uncertainty budget as well as the contribution of the measurements to the reanalysis product and vertical resolution. We have chosen to reanalyze relative humidity profiles in units of relative humidity with respect to water, since this allows a direct comparison with radiosondes which generally report RH_w (Dirksen et al. 2014; Miloshevich et al. 2009). We calculated RH_w from temperature and water vapor mixing ratio using the Hyland and Wexler formulation (see section 2a). In our method it is simple to convert to relative humidity with respect to ice as required.

The paper is organized as follows: Section 2 provides a description of the Raman lidar and ERA5 data we use. The forward model and the implementation of the 1D Var algorithm together with a characterization of the ERA5-reRH profiles are given in section 3. In sections 3f(1) and 3f(2) we have presented two case studies of day and nighttime

 Denotes content that is immediately available upon publication as open access.

Corresponding author: R. J. Sica, sica@uwo.ca

DOI: 10.1175/JTECH-D-19-0170.1

© 2020 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

ERA5-reRH retrievals. Section 4 contains a validation of the ERA5-reRH dataset compared to the measurements from radiosondes. A detailed discussion of the results, our conclusions and an outlook to future work are given in sections 5 and 6.

2. Measurements and data used in the ERA5-reRH

a. RALMO

For this study we use Raman lidar measurements from the Raman Lidar for Meteorological Observations (RALMO), located in Payerne (46°48'N, 6°56'E), and operated by MeteoSwiss. RALMO is a fully automated lidar, operating near continuously since 2008, with an average uptime of 50%, with the primary loss of measurements due to events of precipitation and low clouds. The transmitting system of RALMO consists of a frequency tripled, Q-switched neodymium-doped yttrium–aluminum–garnet (Nd:YAG) laser at 354.7 nm generating up to 400 mJ per shot at a 30 Hz repetition rate. The laser pulses are 8 ns in duration. The lidar telescope receiver consists of four 30-cm-diameter mirrors that are tightly arranged around a 15-times beam expander. The mirrors are fiber-optically coupled to the polychromators. A near range optical fiber, located off-axis on one of the four mirrors, improves the signal-to-noise ratio in the partial overlap region and allows water vapor and temperature measurements below 400 m altitude (Dinoyev et al. 2013). The RALMO detection system consists of two polychromators isolating the water vapor and nitrogen Raman return at 407 and 387 nm, respectively, as well as four portions of the pure rotational Raman spectrum including high and low quantum number lines in the Stokes and anti-Stokes branches. The detection system captures the light from the polychromators using photomultiplier tubes operating in both analog and digital modes. There is a total of eight channels. A detailed description of RALMO is given by Dinoyev et al. (2013), and the instrument's validation is given in Brocard et al. (2013).

b. The ERA5 data

ERA5 is the fifth generation of the European Centre for Medium-Range Weather Forecasts (ECMWF) atmospheric analysis of the global climate (Hennermann and Berrisford 2017). ERA5 was produced using 4D Var data assimilation. It provides hourly temperature, relative humidity (over water above freezing and over ice below freezing temperature), specific humidity, geopotential, and many other atmospheric parameters with an uncertainty estimate at 37 pressure levels between the surface and the stratopause from 1979 onwards. ERA5 is combined with measurements from satellites and in situ instruments worldwide to provide a complete and consistent dataset. The data assimilation is done twice per day using 12 h windows from 0900 to 2100 UTC and 2100 to 0900 UTC (the following day). Further details of ERA5 can be obtained from Hennermann and Berrisford (2017) and the ERA5 data can be accessed either from the Meteorological Archival and Retrieval System (MARS) archive or from Climate Data Store (CDS) cloud server that has comparatively fast access (Hersbach et al. 2019).

Humidity is reported both as specific and relative humidity. For temperatures above 0°C, relative humidity is given with

respect to water, for temperatures below −23°C it is given with respect to ice, and a mix of the two for temperatures between 0° and −23°C (ECMWF 2018). Since here we systematically use a unit of relative humidity with respect to water, we convert ERA5 specific humidity, temperature, and pressure data to relative humidity with respect to water for all temperatures, hereafter referred to as $RH_{w,ERA5}$. For this calculation, we use the Hyland and Wexler formulation (Hyland and Wexler 1983).

As mentioned earlier, ERA5 assimilates both satellites and in situ measurements. A list of all the measurements is presented in Hennermann and Berrisford (2017) and the data usage in ERA5 for the segment from 1979 is presented in Hersbach et al. (2019). ERA5 also provides uncertainty (ensemble spread) estimates from a 10-member ensemble data assimilation. In general, lower ensemble spreads in ERA5 indicate higher confidence in the data. As shown in Hersbach et al. (2019) the accuracy of the ERA5 temperature data is improved over the years due to the increase of the number of temperature observations that are assimilated into the model.

No validation studies of the ERA5 specific humidity are available. This motivated us to evaluate the ERA5 specific humidity ensemble spreads for the same date in every decade starting from 1980 to 2010, as an estimate of the accuracy of the ERA5 specific humidity data. We do not find significant improvements in the ERA5 specific humidity data over that time, even with the higher observational coverage available in the 2000s relative to the 1980s. However, the ERA5 relative humidity ensemble spreads starting from the surface to about 15 km are about 10%RHw, whereas the relative humidity standard uncertainty of RALMO is better than 5%RHw. Thus, we expect a significant impact on the ERA5-reRH relative humidity data from the RALMO measurements.

3. 1D Var retrieval of relative humidity from Raman lidar measurements and ERA5 (ERA5-reRH)

For the 1D Var data retrieval scheme we use the optimal estimation method (OEM) as described by Rodgers (2000). The OEM technique minimizes a cost function that estimates the degree of fit of the atmospheric state to the measurements, constrained by a priori information (Palmer et al. 2000). The cost function is given by

$$\text{cost} = [\mathbf{y} - F(\mathbf{x}, \mathbf{b})]^T \mathbf{S}_y^{-1} [\mathbf{y} - F(\mathbf{x}, \mathbf{b})] + [\mathbf{x} - \mathbf{x}_a]^T \mathbf{S}_a^{-1} [\mathbf{x} - \mathbf{x}_a], \quad (1)$$

where \mathbf{y} is the measurement vector, \mathbf{x} is the state vector, and F is the forward model that relates the measurements to the state variables (in data assimilation F is often referred to as an observation operator). The forward model parameters \mathbf{b} are ancillary parameters needed to evaluate the forward model. The forward model contains all the physics describing the measurements. \mathbf{S}_y and \mathbf{S}_a are the error covariance matrices of the measurement and the a priori state vector \mathbf{x}_a , respectively. In data assimilation \mathbf{x}_a is normally referred to as background.

The Opack software package, developed by Eriksson et al. (2005), provides the OEM solver used in this study. The OEM solver uses the Marquardt–Levenberg method to

minimize the cost function given in Eq. (1) and yields the solution \hat{x} (hereafter called the retrieval) and its error covariance matrix \mathbf{S}_x . For the uncertainty budget of the retrieval we consider the following sources of uncertainty: measurement uncertainty described by \mathbf{S}_y and the forward model parameter uncertainties described by the forward model parameter error covariance matrix \mathbf{S}_b (see section 3d). The uncertainty due to the forward model itself is neglected since we use a sophisticated forward model as detailed in section 3c which accounts for all geophysical and instrument effects to within the measurement uncertainty. The different contributions to the retrieval error $\hat{\mathbf{S}}_x$ are given by the following expressions.

a. Measurement uncertainty

The error due to measurement uncertainty is referred to as the statistical uncertainty, since it accounts for the measurement noise:

$$\mathbf{S}_e = \mathbf{G}\mathbf{S}_y\mathbf{G}^T. \quad (2)$$

b. Forward model parameter uncertainties

The error due to the forward model parameter uncertainties is referred to as the forward model uncertainty:

$$\mathbf{S}_F = \mathbf{G}\mathbf{K}_b\mathbf{S}_b\mathbf{K}_b^T\mathbf{G}^T, \quad (3)$$

where \mathbf{K}_b is the forward model Jacobian computed with respect to \mathbf{b} . The gain matrix \mathbf{G} gives the sensitivity of the retrieval to the measurement:

$$\mathbf{G} = \frac{\partial \hat{x}}{\partial y}. \quad (4)$$

For more details of the OEM the book by Rodgers (2000), in particular section 5.7, in addition to Sica and Haeefele (2015, 2016), Farhani et al. (2019), and Gamage et al. (2019). These references provide more details into our application of the OEM using lidar measurements.

c. Forward model

The forward model for the lidar measurements is based on the Raman lidar equation and relates the measured back-scattered photocounts of the i th channel $N_{\text{obs},i}$ to the instrument and atmosphere as follows:

$$N_{\text{obs},i}(z) = \frac{C_i}{z^2} O_i(z) n_i(z) \frac{d\sigma_i(\pi)}{d\Omega} \exp\left\{-\int_0^z [\alpha_{\lambda_0}(z) + \alpha_{\lambda_i}(z)] dz\right\} + B_i(z). \quad (5)$$

Here C_i is the lidar constant for channel i that depends on the number of transmitted photons, detector efficiency and area of the telescope. The geometrical overlap $O_i(z)$ is a dimensionless parameter that describes the overlap between the transmitted laser beam and the field of view of the telescopes. The number density of the scattering molecule is $n_i(z)$ and $B_i(z)$ is the background of the observed signal. $d\sigma_i(\pi)/d\Omega$ is the differential Raman backscatter cross section where σ is the cross section and Ω is the solid angle. Finally, $\alpha_{\lambda_0}(z)$ and $\alpha_{\lambda_i}(z)$ are the extinction coefficients at the

transmitted wavelength and at the wavelength of receiver channel i .

For the digital channels, the lidar equation given in Eq. (5) presumes that the count rates depend on the number of received photons in a linear manner, in our case for count rates below 10 MHz. For higher count rates, the true and observed counts, N_{tr} and N_{obs} , respectively, are related as follows for a nonparalyzable counting system, as is appropriate for RALMO:

$$N_{\text{obs}} = \frac{N_{\text{tr}}}{1 + N_{\text{tr}}\gamma} \quad (6)$$

(Kovalev and Eichinger 2004). The dead time, γ , characterizes the response speed of the digital acquisition system. Equations (5) and (6), where (6) is valid only for digital channels, are evaluated eight times to produce the four digital and four analog signals corresponding to rotational-vibrational Raman scattering of water vapor ($i = \text{Wd}, \text{Wa}$) and nitrogen ($i = \text{Nd}, \text{Na}$) and pure rotational Raman (PRR) scattering of high ($i = \text{JHd}, \text{JHa}$) and low ($i = \text{JLd}, \text{JLa}$) quantum numbers.

For the PRR channels the number densities are equal to the air number density, $n_{\text{air}} = n_{\text{JHd}} = n_{\text{JHa}} = n_{\text{JLd}} = n_{\text{JLa}}$, which is replaced by pressure and temperature assuming the hydrostatic equilibrium and the ideal gas law (Behrendt 2005).

The saturation vapor pressure, $e_{w,s}$, which is needed to convert relative humidity to water vapor number density, $n_{\text{wv}} = n_{\text{wd}} = n_{\text{wa}}$, is expressed using the Hyland and Wexler formulation (Hyland and Wexler 1983).

The extinction coefficients $\alpha_{\lambda_0}(z)$ and $\alpha_{\lambda_i}(z)$ in Eq. (5) including both molecules and particles are given by

$$\alpha_{\lambda_0}(z) = \alpha_{\lambda_0,\text{mol}}(z) + \alpha_{\lambda_0,\text{par}}(z), \quad (7)$$

where $\alpha_{\lambda_0,\text{mol}}(z)$ is the molecular extinction coefficient at the emitted wavelength and $\alpha_{\lambda_0,\text{par}}(z)$ is the particle extinction coefficient at the emitted wavelength. Similarly, $\alpha_{\lambda_i}(z)$ can be represented as a summation of molecular and particle extinction. For PRR channels ($i = \text{JHd}, \text{JHa}, \text{JLd}, \text{JLa}$), $\alpha_{\lambda_0}(z)$ is assumed to be equal to $\alpha_{\lambda_i}(z)$, as the emitted and the received wavelengths are very close to each other.

The extinction $\alpha_{\lambda_0,\text{mol}}(z)$ given in Eq. (7) can be expressed using the Rayleigh cross section as $\sigma_{\lambda_0,\text{Ray}}$ and air density $n_{\text{air}}(z)$:

$$\alpha_{\lambda_0,\text{mol}}(z) = \sigma_{\lambda_0,\text{Ray}} n_{\text{air}}(z), \quad (8)$$

where $\sigma_{\lambda_0,\text{Ray}}$ is calculated for each wavelength using the expressions given by Nicolet (1984). In our work, the air density term in Eq. (8) is considered as a model parameter and is therefore estimated using the pressure and temperature either from sonde measurements or reanalysis data. All the particle extinction coefficients are expressed at their respective return wavelength using the following expression:

$$\alpha_{\lambda_i,\text{par}}(z) = \alpha_{\lambda_0,\text{par}}(z) \left(\frac{\lambda_i}{\lambda_0}\right)^{-\hat{a}(z)}, \quad (9)$$

where $\hat{a}(z)$ is the Ångström exponent as a function of altitude (Ansmann and Müller 2005).

TABLE 1. Values and associated uncertainties for the retrieval and forward model parameters. All ERA5 quantities used correspond to the location of the MeteoSwiss field station in Payerne.

Parameter	Value	Standard deviation
Measurements		
Digital (JL, JH, water vapor, nitrogen)	Measured	Poisson statistics
Analog (JL, JH, water vapor, nitrogen)	Measured	Autocovariance method
Retrieval parameters (a priori)		
Temperature	ERA5	Standard deviation of ERA5 and coincident sonde temperature measurements
Relative humidity	Recalculated using ERA5 reanalysis temperature, pressure, and specific humidity	Standard deviation of ERA5 and coincident sonde relative humidity measurements
Overlap functions	Estimated using the forward model and measurements	50% below and at transition height
Particle extinction	Estimated using measurements	10 ⁻³ above transition height 10 ⁻⁶ km ⁻¹ below and at transition height 50% above transition height
Lidar constants (analog/digital)	Estimated using the forward model	100%
Digital background noise	Mean above 50 km	Standard deviation above 50 km
Analog background noise	Mean above 50 km	Nighttime: Standard deviation above 50 km Daytime: Normalized standard deviation above 50 km
Dead time	Empirical fitting	10%
Forward model parameters		
Coupling constants (analog/digital)	Estimated measurements and sonde temperature	Standard deviation of the linear fitting
Air density	Radiosonde	1%
Seed pressure	HSEQ: Pressure at the heights point of the retrieval grid	10%
Ångström exponent	1 for all heights	10%

We now introduce calibration factors R to eliminate the four lidar constants from each channel, digital water vapor (C_{Wd}), analog water vapor (C_{Wa}), digital high quantum PRR (C_{JHd}) and analog high quantum PRR (C_{JHa}), from the forward model:

$$R_{WVd} = \frac{C_{Wd}}{C_{Nd}}, \quad (10a)$$

$$R_{WVa} = \frac{C_{Wa}}{C_{Na}}, \quad (10b)$$

$$R_{PRRd} = \frac{C_{JHd}}{C_{JLd}}, \quad (10c)$$

$$R_{PRRa} = \frac{C_{JHa}}{C_{HLa}}. \quad (10d)$$

These calibration factors are determined outside the 1D Var process using collocated radiosonde profiles. For details concerning the calibration see [Sica and Haeefe \(2016\)](#) and [Gamage et al. \(2019\)](#).

d. Error covariance matrices

In [section 2](#) we have explained the method of constructing the error covariance matrices for both retrieval and model parameters using the uncertainty values given in [Table 1](#).

1) MEASUREMENT NOISE

The error covariance matrices of the two sets of Raman lidar measurements, analog and digital, are diagonal assuming no

correlation of noise between channels. For the analog channels and the digital measurements that are not in the linear range (count rate is >10 MHz), the variances are estimated using the autocovariance function method given by [Lenschow et al. \(2000\)](#). The measurements from the digital channels that are in the linear range follow Poisson statistics, where the variance is equal to the signal.

2) A PRIORI (BACKGROUND) RELATIVE HUMIDITY AND TEMPERATURE

The a priori or background error covariance matrices of temperature and relative humidity are key parameters in the 1D Var process, since they directly control the weight that is given to the background, which here is ERA5. As was mentioned in [section 1](#), the ERA5 data are produced assimilating an extensive set of global observations. Thus, the accuracy of the reanalysis is typically higher than for forecasts. To construct the a priori error covariance matrices for temperature and relative humidity, we use a tent function to parameterize the off-diagonal elements, which decay linearly from the variance on the diagonal with an e -folding distance called the correlation length, as discussed by [Eriksson et al. \(2005\)](#), with negative values set to zero.

To determine the variances, we first calculate the mean and standard deviation of the temperature and relative humidity differences between sonde measurements (observations) and ERA5 data from 2004 to 2015. Since the routine soundings from Payerne made at 1100 and 2300 UT are assimilated in

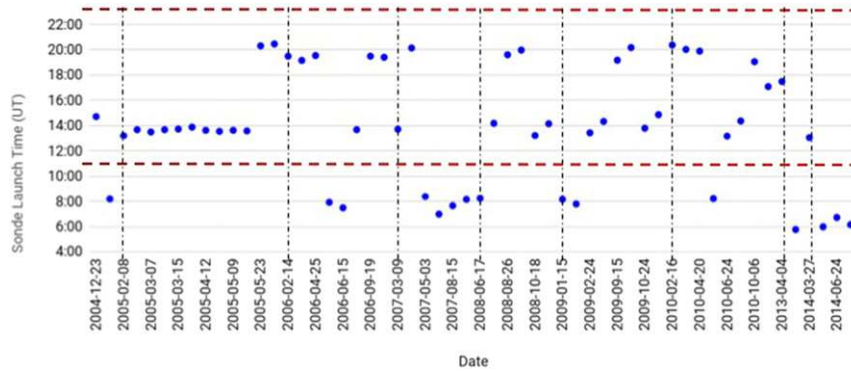


FIG. 1. Dates and times of the sondes launched from Payerne, Switzerland, that coincide with the ERA5 data used to estimate correlations of temperature and relative humidity. The two red dashed lines indicate the usual sonde launching times of 1100 and 2300 UT from Payerne.

ERA5, we only consider a set of special soundings that were made at times between 0600–0900, 1300–1500, and 1800–2100 UT which are not assimilated in ERA5. Figure 1 shows dates and times of the 56 special soundings considered in our calculation. The differences show a systematic bias in the ERA5 temperature and relative humidity with respect to the sonde measurements (Fig. 2). Up to about 12 km altitude, ERA5 shows an overall warm bias with a maximum of about 4.5 K at the surface and a secondary maximum of 3.6 K around 8 km. Above 12 km a 0.5 K cold bias exists. From the surface to about 1.5 km altitude the ERA5 relative humidity is lower relative to the sonde by as much as 20% RHw, while above 1.5 km ERA5 is lower by 0 to 15% RHw. We use these mean differences to obtain a bias-corrected ERA5 dataset for the Payerne site. The corrected ERA5 data are the a priori information used in our 1D Var processing.

To determine the correlation lengths of the temperature and relative humidity errors we first computed the correlation matrices of the differences between ERA5 and the coincident special soundings as shown in Fig. 3. The correlation lengths for temperature and relative humidity was then estimated to be 1 km and 750 m, respectively, throughout the entire troposphere.

3) PARTICLE EXTINCTION AND OVERLAP

For particle extinction we calculate an a priori or background profile from the backscatter ratio measured by the lidar. To convert the backscatter ratio to particle extinction we assume a lidar ratio and use the same molecular extinction profile as in the OEM retrieval [Eqs. (8) and (9)]. As in previous studies (Sica and Haeefe 2016; Gamage et al. 2019), we assume the lidar ratio for clear-sky conditions (backscatter ratio smaller than 2) to be 80 sr inside the boundary layer and 50 sr elsewhere. Inside clouds we assume the lidar ratio to be 20 sr below 6 km (liquid cloud), and 15 sr above (cirrus cloud; Ansmann et al. 1992; Pappalardo et al. 2004). We identify the presence of cloud when the value of backscatter ratio is greater than 2. The cloud layer base altitude is used to define the altitude where the overlap retrieval hands over to the retrieval of

particle extinction, since retrieving both simultaneously at the same altitude is not possible due to the high degree of linear dependence (for further details, see Gamage et al. 2019). In cloud-free conditions, this handover takes place at 6 km, where full overlap has been reached. For the particle extinction error covariance matrix we assume a standard deviation of 50% above and 10^{-6} km^{-1} below the handover altitude. For overlap a standard deviation of 50% is assumed below and 10^{-3} above the handover altitude. The off-diagonal elements are parameterized using a tent function with a correlation length of 100 m.

4) BACKGROUND, LIDAR CONSTANTS, AND DEAD TIMES

The a priori backgrounds and their variances for both analog and digital channels are determined by the mean and the variance of the measurements above 50 km altitude. The four a priori lidar constants (C_{JLd} , C_{JLa} , C_{Nd} , and C_{Na}) required for the 1D Var process are estimated by fitting the forward model in a specified region to the respective Raman lidar measurements. We consider 3.8 ns as the a priori dead times for the digital photon counting systems, values which were found by previous studies using RALMO, and also consistent with the values specified by the manufacturer (Sica and Haeefe 2015, 2016; Dineev et al. 2010; Gamage et al. 2019).

5) FORWARD MODEL PARAMETERS

The model (**b**) parameters used in the forward model are the calibration factors (for analog channels: R_{PRRa} , R_{wva} ; and for digital channels: R_{PRRd} , R_{wvd}), Ångström exponent, seed pressure, and air density for atmospheric transmission. The values and uncertainties of the other **b** parameters are given in Table 1.

e. Other 1D Var retrieval specifications

Prior to the assimilation, the raw lidar data are coadded to 30 m bins in altitude and 30 min in time (± 15 min around the ERA5 analysis time). The retrieval grid spans from 600 m above mean sea level (MSL) to 20 km MSL with a grid spacing of 90 m.

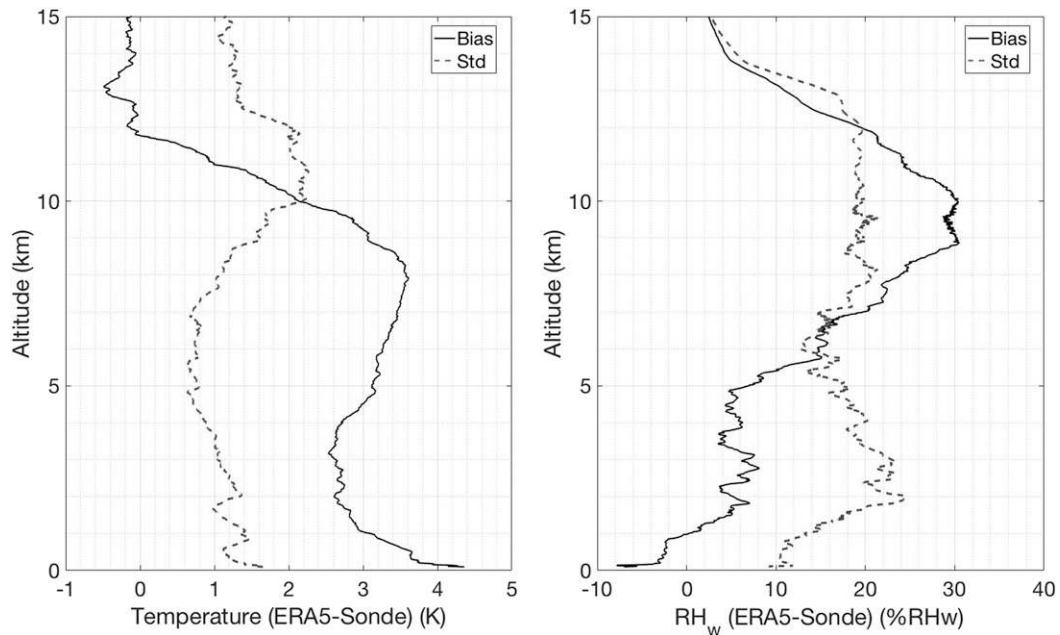


FIG. 2. (left) Profiles of ERA5 minus radiosonde mean temperature bias (solid black curve) and standard deviation of the temperature differences (dashed gray curve). (right) Profiles of ERA5 minus radiosonde mean relative humidity bias (solid black curve) and standard deviation of the relative humidity differences (dashed gray curve). The bias is calculated using 56 specific sonde measurements launched between the years 2004 and 2015 from Payerne, which are not included in the ERA5 calculations. Corresponding sonde launch date and time information is given in Fig. 1.

f. Characterization of ERA5-reRH: Two case studies

In this section we present two representative case studies from the new ERA5-reRH dataset corresponding to a day and a nighttime retrieval. To be able to demonstrate the benefit

of combining Raman lidar with ERA5, we repeated the same processing in the following figures using the U.S. standard atmospheric climatology instead of ERA5 for the a priori temperature and water vapor profiles. Apart from the a priori

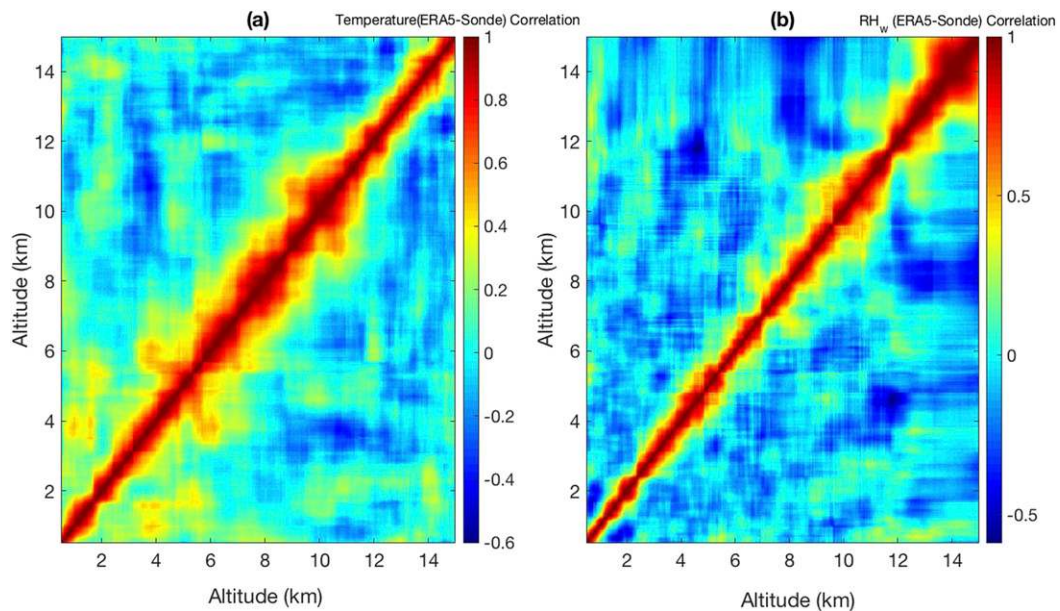


FIG. 3. (a) Temperature and (b) relative humidity a priori correlation matrices estimated using ERA5 and 56 special radiosounding measurements.

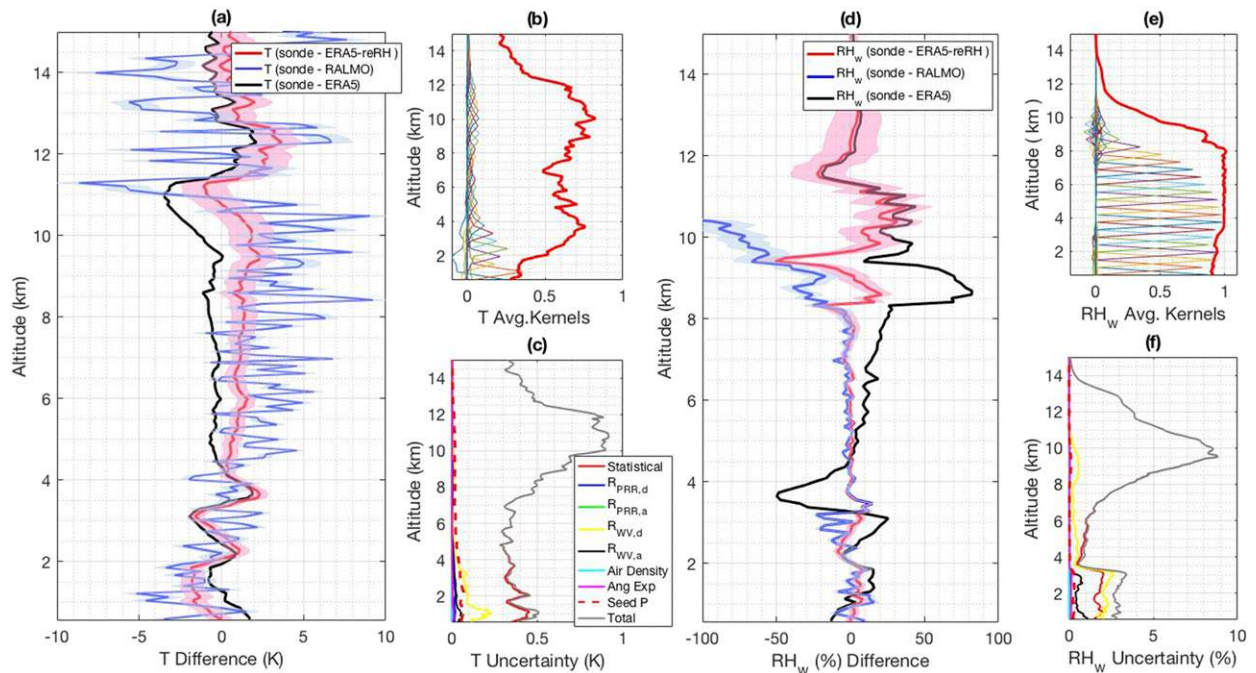


FIG. 4. (a) The difference between sonde and ERA5-reRH temperature profiles (red curve) and statistical uncertainty (shaded area) retrieved using RALMO nighttime measurements on 28 Aug 2012 with 30 min temporal and 90 m vertical resolutions. The sonde was launched at 2241 UT. The blue curve is the difference between sonde temperatures and the RALMO OEM temperature retrievals. The black curve is the difference between sonde and ERA5 bias-corrected a priori temperature profiles. (b) Averaging kernels for the temperature retrieval. The red curves show the response function. For clarity averaging kernels for every fifth altitude of the retrieval grid are shown. (c) Random and systematic uncertainties due to the forward model parameters for the temperature retrieval. Total uncertainty (gray curve), statistical uncertainty (red curve), $R_{PRR,d}$ digital coupling constant for PRR (blue curve), $R_{PRR,a}$ analog coupling constant for PRR (green curve), $R_{WV,d}$ digital coupling constant for WV/N2 (yellow curve), $R_{WV,a}$ analog coupling constant for WV/N2 (black curve), air density (cyan curve), Ångström exponent (purple curve), and seed pressure (dashed red curve). (d) The difference between sonde and ERA5-reRH relative humidity profiles (red curve) and the statistical uncertainty (shaded area). The blue curve is the difference between sonde and the RALMO OEM relative humidity profiles. The black curve is the difference between sonde and ERA5 bias-corrected a priori relative humidity profiles. (e) Averaging kernels for the relative humidity retrieval. (f) The total uncertainty budget for the relative humidity retrievals with the same color explanatory mapping as in (c).

profiles and the a priori error covariance matrix, all parameters are kept the same and this dataset is referred to as RALMO, indicating that this is essentially a pure lidar measurement. For RALMO retrievals we use an a priori temperature error covariance matrix with a standard deviation of 35 K and off-diagonal elements parameterized using a tent function with 1 km correlation length. The U.S. Standard Atmosphere model temperature serves as the a priori profile. The a priori relative humidity profile is constant in altitude with a value of 50%RH_w. The error covariance matrix is constructed in the same way as for temperature with a standard deviation of 100%RH_w and a correlation length of 1 km. For comparison, the ERA5-reRH reanalysis is compared to coincident sonde measurements and RALMO retrievals. The purpose of this section is to show that the ERA5-reRH improves the comparison of ERA5 with the radiosonde, and to quantify the impact of the lidar measurements. In this study we have only used measurements from Vaisala RS92 radiosondes. A statistical validation of ERA5-reRH is given in section 4.

1) CASE 1: NIGHTTIME, THIN CIRRUS CLOUD, 2241–2311 UT 28 AUGUST 2012

We consider 30 min of measurements starting from the launch time of the coincident sonde from Payerne. The calibration coupling constants required for our OEM were estimated using the 30 min raw lidar measurements and coincident sonde measurements as detailed in Gamage et al. (2019). During the time of the lidar measurement a cirrus cloud was present between 8 and 10 km.

Figure 4 shows the ERA5-reRH, RALMO retrievals, and ERA5 data in comparison with coincident sonde measurements. Figure 4a, shows the temperature difference between coincident sonde and ERA5-reRH (red curve) with the statistical uncertainty (shaded area), temperature difference between coincident sonde temperature and RALMO retrievals (blue curve), temperature difference between coincident sonde temperature and ERA5 temperature (black curve). The ERA5-reRH temperatures agree with the coincident sonde measurements within $\sim \pm 2$ K for all heights except between 11.5 and 12.5 km where the agreement is a maximum of ± 3 K. The RALMO retrieved temperatures contain more

noise compared to ERA5-reRH, as it uses a large error a priori covariances for temperature. The bias-corrected ERA5 temperatures also agree well (± 2 K) with the coincident sonde measurements, except in the region of 10–11.5 km.

The relative humidity difference between the sonde and ERA5-reRH (red curve) shows that the two profiles agree within $\pm 10\%$ for heights up to about 8 km (Fig. 4d). The RALMO relative humidity retrievals (blue curve) also closely follow the ERA5-reRH relative humidity retrievals up to 8 km. However, the RALMO relative humidity retrievals show more variations due to the requirement of a large error covariance. The comparison of the ERA5 and coincident sonde relative humidity products (black curve) show large deviations at height ranges ~ 2 –4 km and above 5 km ($\sim \pm 50\%$).

The measurement response, which varies from 0 to 1, is the sum of the averaging kernels and indicates the contribution of the measurement to the retrieval. It is unity when 100% of the retrieval is due to the measurements. When a retrieval is fully dependent on the a priori profile, the response function is equal to zero. The measurement response (red curve) for temperature (Fig. 4b), shows the lidar measurements contribution between 50% and 60% up to 12 km. Above 12 km the lidar impact drops quickly and ERA5-reRH becomes identical to ERA5.

Unlike for temperature, the measurement response for relative humidity (red curve in Fig. 4e), is greater than 90% up to about 8 km. The difference in lidar impact on temperature and humidity is related to the following 2 factors. First, ERA5 assimilates many temperature datasets while there are fewer humidity datasets available. Second, relative humidity is more variable in time and space than temperature. Hence, our confidence in ERA5 for temperature is higher than for relative humidity, which reduces the impact of the lidar data on temperature.

One of the main features of ERA5-reRH is the full uncertainty budget on a profile-by-profile basis that contains both random and systematic uncertainties. The full uncertainty budget is determined from the measurement and model parameter covariance matrices which are propagated through the retrieval using Gaussian error propagation. The list of uncertainty sources includes statistical uncertainty, uncertainty due to coupling constants, and seed pressure uncertainty [as mentioned in section 3d(5)]. Figure 4c shows the uncertainty budget for the temperature retrievals. Measurement noise is the dominant source of uncertainty up to 8 km in altitude and maximizes around 8–12 km to a value of 0.9 K. The second most important contribution is the uncertainty of the analog coupling constant for WV/N2 channels. Below 2 km, the uncertainty due to analog coupling constant for WV/N2 channels is on the order of 0.3 K decreasing to about 0.02 K above 4 km. Contribution from each of the other forward model parameters such as seed pressure, digital and analog coupling constants from PRR channels, and Ångström exponent to the temperature uncertainty is less than 0.1 K.

The full relative humidity uncertainty budget is shown in Fig. 4f. For all altitudes, the total uncertainty is on the order of 5% RHw with a maximum of 5.15% RHw at around 9 km. Below 3 km, the contribution from the analog coupling

constant for WV/N2 channels is about 4% RHw. The statistical uncertainty dominates most in the total uncertainty for altitudes above 3 km.

2) CASE 2: DAYTIME, CLEAR SKY, 1010–1040 UT 10 SEPTEMBER 2011

For the second case study, measurements from a coincident radiosonde flight from Payerne launched at 1010 UT are used to compare with our ERA5-reRH. During the time of the measurements, sky conditions remained clear, but the signal-to-noise ratio of the RALMO daytime water vapor measurements dropped below 1 at above 5.5 km due to the large background.

The temperature difference between coincident sonde and ERA5-reRH temperature (red curve, Fig. 5a) shows that the two profiles agree within ± 3 K. The sonde and bias-corrected ERA5 temperature difference (black curve) also shows that the two profiles are in good agreement with each other, except in the region 10.5–12 km, where there ERA5 shows a cold bias of ~ 3 K to the sonde temperatures. Similar to the nighttime case study, the temperature difference between sonde and RALMO retrievals (black curve) contains more noise. Moreover, the daytime measurement response for temperature (red curve) in Fig. 5b drops below 0.5 at 7 km. Up to 6 km, the ERA5-reRH temperature retrieval depends about 70% on the lidar measurements.

The relative humidity difference between sonde and ERA5-reRH (red curve) and RALMO (blue curve) both show that the relative humidity from ERA5-reRH and RALMO profiles agree with the sonde measurements within $\pm 10\%$ in the region below 6 km (Fig. 5d). Below 8 km altitude, the difference between coincident sonde and ERA5 relative humidity (black curve) is about 45%.

The measurement response for relative humidity (red curve in Fig. 5e) shows the retrieved relative humidity depends more than 90% on the lidar measurements up to 5 km. As the lidar water vapor signal gets weaker, the relative humidity retrievals start to rely more on the a priori relative humidity profile. Above 5 km the ERA5-reRH retrieved relative humidity becomes identical to ERA5.

The temperature and the relative humidity uncertainty budgets are shown in Figs. 5c and 5f, respectively. The total temperature uncertainty is on the order of ~ 0.5 K for most altitudes and it is a maximum of ~ 0.9 K from 10 to 12 km. Uncertainty due to the water vapor calibration factors (R_{PRRa} and R_{PRRd}) dominates the temperature uncertainty below 1 km (~ 0.5 K). Elsewhere the statistical uncertainty dominates the total uncertainty. Uncertainty from other model parameters is on the order of ~ 0.1 K each.

The total relative humidity uncertainty for all altitudes is less than 7% RHw. The maximum value of $\sim 7\%$ RHw is below 1 km and relative humidity uncertainty peaks around 6%–6.5% RHw between 5.5 and 6 km in altitude. Uncertainty due to the digital WV/N2 coupling constant (R_{WVd}) is dominant below 3 km ($< \sim 5\%$ RHw), while the statistical uncertainty dominates above. Uncertainty due to other model parameters is on the order of $< 1\%$ RHw for each parameter for all altitudes.

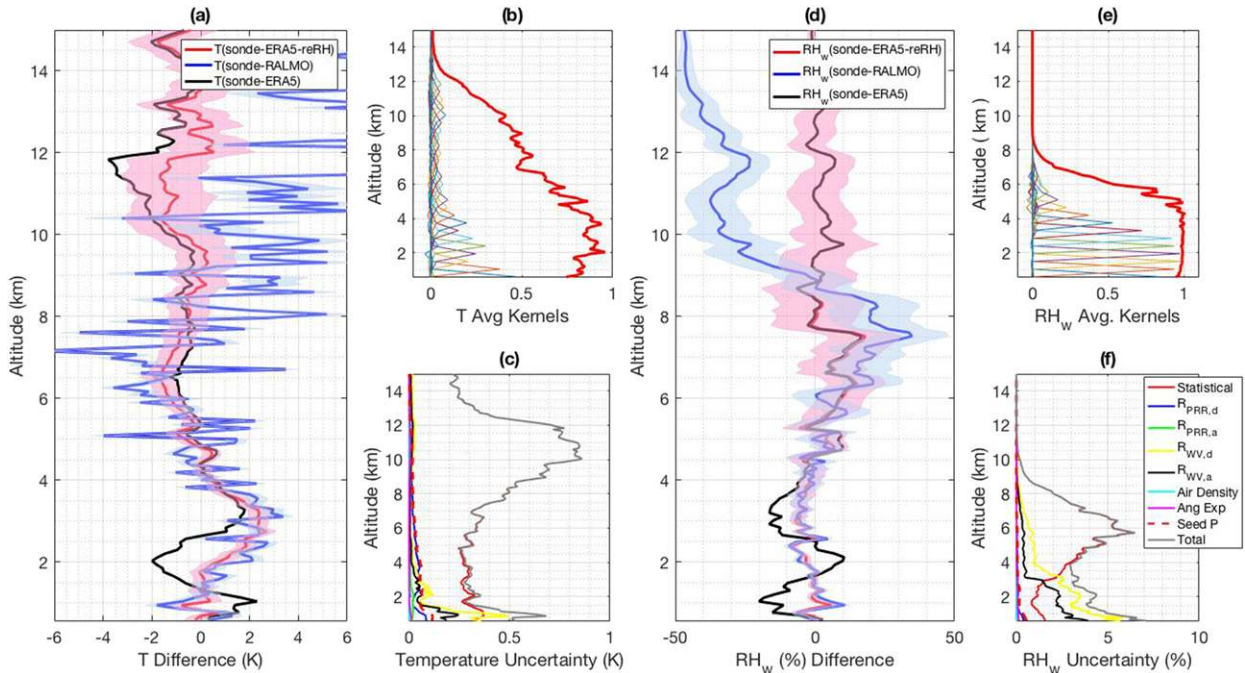


FIG. 5. (a) The difference between sonde and ERA5-reRH temperature profiles (red curve) and statistical uncertainty (shaded area) from RALMO daytime measurements on 10 Sep 2011 with 30 min temporal and 90 m vertical resolutions. The sonde was launched at 1010 UT. The blue curve is the difference between sonde temperatures and the RALMO OEM temperature retrievals. The black curve is the difference between sonde and ERA5 bias-corrected a priori temperature profiles. (b) Averaging kernels for the temperature retrieval. The red curves show the response functions. For clarity averaging kernels for every fifth altitude bin the retrieval grid are shown. (c) Random and systematic uncertainties due to the forward model parameters for the temperature retrieval. Total uncertainty (gray curve), statistical uncertainty (red curve), $R_{PRR,d}$ digital coupling constant for PRR (blue curve), $R_{PRR,a}$ analog coupling constant for PRR (green curve), $R_{WV,d}$ digital coupling constant for WV/N2 (yellow curve), $R_{WV,a}$ analog coupling constant for WV/N2 (black curve), air density (cyan curve), Ångström exponent (purple curve), and seed pressure (dashed red curve). (d) The difference between sonde and ERA5-reRH relative humidity profile (red curve) and the statistical uncertainty (shaded area). The blue curve is the difference between sonde and the RALMO OEM relative humidity profiles. The black curve is the difference between sonde and ERA5 bias-corrected a priori relative humidity profiles. (e) Averaging kernels for relative humidity retrieval. (f) The total uncertainty budget for the relative humidity retrievals with the same color explanatory mapping as in (c).

4. Results

Validation of the reanalysis against radiosonde measurements for an ensemble of 20 days

In this section we provide a comparison of ERA5, RALMO, and ERA5-reRH temperature and relative humidity profiles with coincident sonde measurements to evaluate the improvements in ERA5-reRH. The comparison includes 14 nighttime and 6 daytime profiles from 2011 to 2015. The dates that are used in the comparison are not affected by precipitation, thick cloudy conditions, or missing data. Calibration of the lidar is performed with respect to coincident sonde measurements for all 20 profiles to estimate coupling constants for temperature (Gamage et al. 2019).

1) NIGHTTIME

Figure 6 shows the differences (black curves) between ERA5, RALMO, and ERA5-reRH with respect to coincident nighttime sonde measurements in terms of temperature and relative humidity for the 14 soundings. The red curve is the mean of the 14 differences and the green shaded

area is the standard deviation. We define the mean of the differences between ERA5/RALMO/ERA5-reRH and the sonde as the bias, and the standard deviation of the differences of ERA5/RALMO/ERA5-reRH and the sonde as the spread.

In comparison with ERA5 and ERA5-reRH (Figs. 6a and 6c), RALMO temperatures show considerably more scatter. There is a significant warm bias in ERA5 between 10 and 12 km altitude that is not apparent in the RALMO and ERA5-reRH datasets. For a quantitative comparison we have shown bias and spread of the three datasets (red curves and green shaded areas shown in Figs. 6a,b,c) in Fig. 7.

Figures 7a and 7b show the nighttime temperature bias and spread of ERA5 (red curve), RALMO (blue curve), and ERA5-reRH (green curve). Below 4 km the temperature bias of RALMO and ERA5-reRH follow the same trend. Both the RALMO and ERA5-reRH temperature retrievals rely on lidar measurements below 4 km, but the spreads are considerably different. The spread of RALMO is in the range of 1–7 K while the spread of the ERA5-reRH is in the range 0.5–1.5 K below 4 km. This large difference in spread is due to

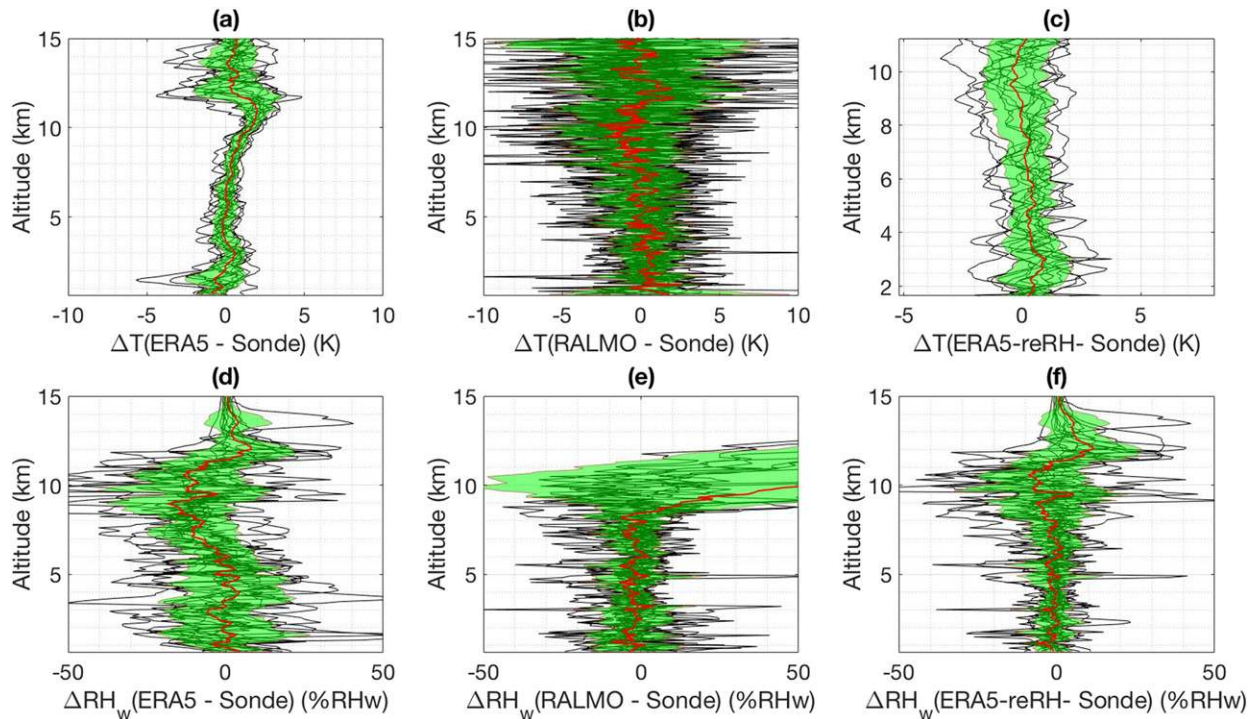


FIG. 6. (a) Nighttime temperature differences between bias-corrected ERA5 and sonde measurements in black curves for 14 nights. (b) Temperature difference between nighttime RALMO retrievals and sonde measurements. (c) Temperature difference between nighttime ERA5-reRH retrievals and sonde measurements. (d) Relative humidity differences between nighttime bias-corrected ERA5 and sonde measurements in gray curves for 14 nights. (e) Relative humidity difference between nighttime RALMO retrievals and sonde measurements. (f) Relative humidity difference between nighttime ERA5-reRH retrievals and sonde measurements. Red curves and green shaded areas show the mean and the standard deviations of the temperature and relative humidity differences of each panel.

the use of better-informed a priori temperature profiles and error covariances in ERA5-reRH compared to RALMO.

ERA5 shows a significant warm bias between 0.5 and 2 K in the altitudes above 8 km. In the same altitude range RALMO's bias varies from -3 to 2 K and the ERA5-reRH bias is in the range of -1 to 0.5 K. Overall, above 8 km ERA5-reRH temperature has the smallest bias compared to ERA5 and RALMO. Thus, the ERA5-reRH temperatures agrees best with the coincident sonde measurements.

The spread of RALMO temperatures increases with altitude (Fig. 7b). The spread of ERA5 is smaller than that of ERA5-reRH and RALMO, except in the 1–2 km and 11.5–12.5 km altitude ranges. In those two altitude regions, the spread of ERA5-reRH is the smallest. Thus, our 1D Var retrieval minimizes variations in ERA5, while the use of the lidar measurements allows the retrieval to determine an optimal temperature profile. Overall ERA5-reRH shows the smallest temperature biases, but the spread is slightly greater than ERA5 due to the higher statistical uncertainty of the lidar measurements in the upper troposphere.

We made the same comparison for the relative humidity products. Figures 6d, 6e, and 6f show the nighttime relative humidity differences (black curves) between ERA5, RALMO, and ERA5-reRH with respect to the sonde measurements. The spread of RALMO relative humidity in Fig. 6e is

comparatively smaller than the spread of ERA5. However, the RALMO relative humidity retrievals are restricted to an average altitude of about 11 km where the RALMO retrievals response function drops below 0.9. Above this altitude the RALMO retrievals begin to depend significantly on the a priori relative humidity profile, and the spread (green shaded area) increases significantly. The spread of ERA5-reRH shown in Fig. 6f is smaller compared to ERA5 and RALMO. Thus, by assimilating the lidar into ERA5 we have improved the relative humidity retrievals relative to the coincident sonde measurements. Above 11 km, the average cutoff altitude of the RALMO retrievals, the bias and spread of ERA5-reRH is identical to ERA5, as the lidar measurement impact on the retrieval is negligible.

Figures 7c and 7d show the bias and spread of ERA5 (red curve), RALMO (blue curve), and ERA5-reRH (green curve) for relative humidity. Below 6 km the bias of ERA5 varies between -6 and $+6$ RHw. The bias of RALMO is between -10 and 4 RHw and the bias of ERA5-reRH is in the range of -6 to 2 RHw. Also, below 6 km ERA5-reRH has the smallest spread, while ERA5 shows the largest spread. Therefore, below 6 km ERA5-reRH relative humidity agrees best with the coincident sonde measurements.

Figure 7c shows that, from 6 to 11.5 km, ERA5 has a dry bias with a maximum of 18 RHw at 9 km compared to the sonde.

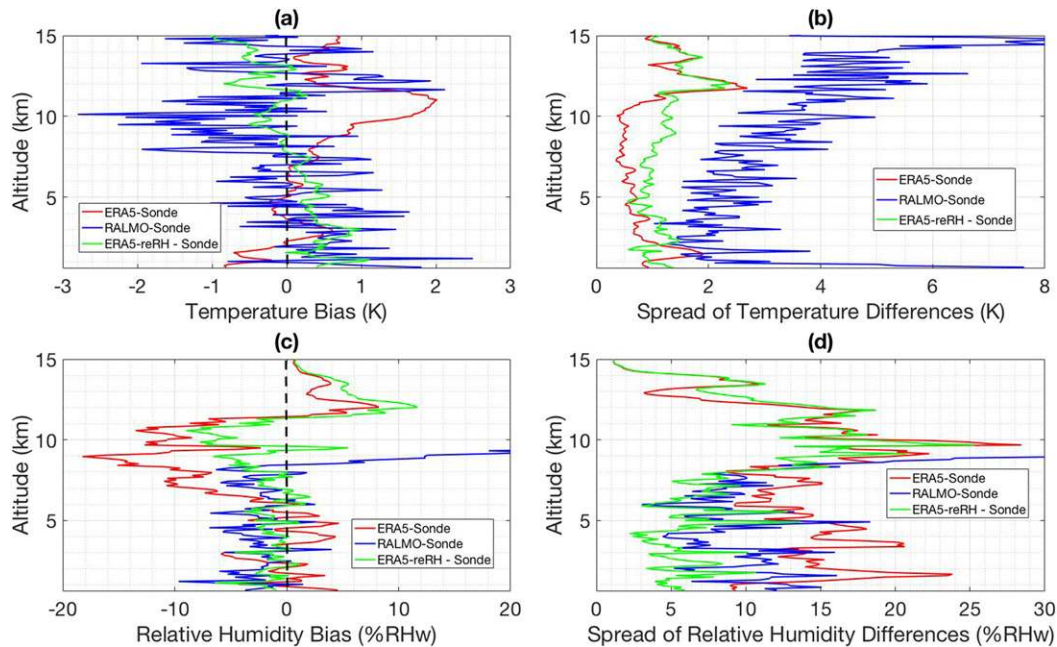


FIG. 7. (a) Nighttime temperature biases between ERA5 and sonde (red curve), RALMO and sonde (blue curve), and ERA5-reRH and sonde (green curve). (b) Spread of the temperature differences. (c) Nighttime relative humidity biases. (d) Spread of the relative humidity differences.

In the same altitude range, ERA5-reRH shows a smaller bias than ERA5 and at 9 km the ERA5-reRH bias is about 8%RHw. RALMO and ERA5-reRH biases between 6 and 9 km are about -6 to 2%RHw. However, above 9 km RALMO's bias increases significantly. In terms of spread, ERA5 shows the largest values up to about 9 km, followed by RALMO. Above 9 km ERA5 and ERA5-reRH both have almost the same bias and spread, indicating that the contribution of the lidar reduces quickly above this level. Overall, below 8 km, we find the best agreement for ERA5-reRH in terms of relative humidity with a bias smaller than 4%RHw and a spread smaller than 9%RHw as compared to ERA5 (bias and spread smaller than 10 and 20%RHw, respectively) and RALMO (bias and spread smaller than 6 and 14%RHw).

2) DAYTIME

Figures 8a–c show bias (red) and spread (green shaded area) of ERA5, RALMO, and ERA5-reRH for daytime in a similar format to the previous nighttime case. Interpretation requires caution as for daytime we have considered measurements only from six soundings which limits the statistical significance of the results. For comparison purposes, we have superimposed the temperature biases and the spreads of ERA5 (red curve), RALMO (blue curve), and ERA5-reRH (green curve) in Figs. 9a and 9b.

Daytime temperature differences between RALMO and sonde (black curves) shown in Fig. 8b have a larger spread than the temperature differences between ERA5 and sonde (Fig. 8a) and ERA5-reRH and sonde (Fig. 8c).

The ERA5-reRH temperature bias at 4–13 km is between -0.5 and 2 K, slightly larger than ERA5. Both the bias and spread of ERA5-reRH are identical to ERA5 above 13 km, indicating that the lidar measurements have no impact at these altitudes.

Figures 8d–f show the daytime relative humidity differences between ERA5 and sonde (red curve), RALMO and sonde (blue curve), and ERA5-reRH and sonde (green curve). The corresponding relative humidity biases and spreads are shown in Figs. 9c and 9d. Unlike for temperature, we found a large bias in ERA5 relative humidity at lower altitudes (Fig. 9c). Below 2.5 km, ERA5 has a wet bias with a maximum of 18%RHw while from 2.5 to 4 km it has a large dry bias, with a maximum of 30%RHw.

Bias and spread of ERA5-reRH below 5 km are within 10%RHw and 16%RHw, respectively, similar to RALMO as the retrieval is primarily informed by the lidar measurements. Figure 9d shows that below 5 km, the spread of ERA5 is the greatest. Above 5 km contribution of the lidar drops drastically and ERA5 and ERA5-reRH are nearly identical, indicating the ERA5-reRH relative humidity retrievals essentially depend on the a priori relative humidity.

5. Discussion

We have combined Raman lidar measurements with ERA5 data, using a 1D Var data assimilation approach based on the optimal estimation method to retrieve temperature and humidity from an initial subset of the RALMO database. The raw lidar measurements, i.e., backscatter profiles from rotational and vibrational Raman scattering, without any data preprocessing such as photocount corrections, background subtraction, overlap, or gluing are used to combine with ERA5. The output of the 1D Var data process retrieve relative humidity with respect to water and temperature. The dataset comes along with a full characterization of uncertainty and vertical resolution on a profile-by-profile basis. Prior to assimilation, ERA5 temperature and relative humidity have been bias-corrected using a set of special radiosoundings

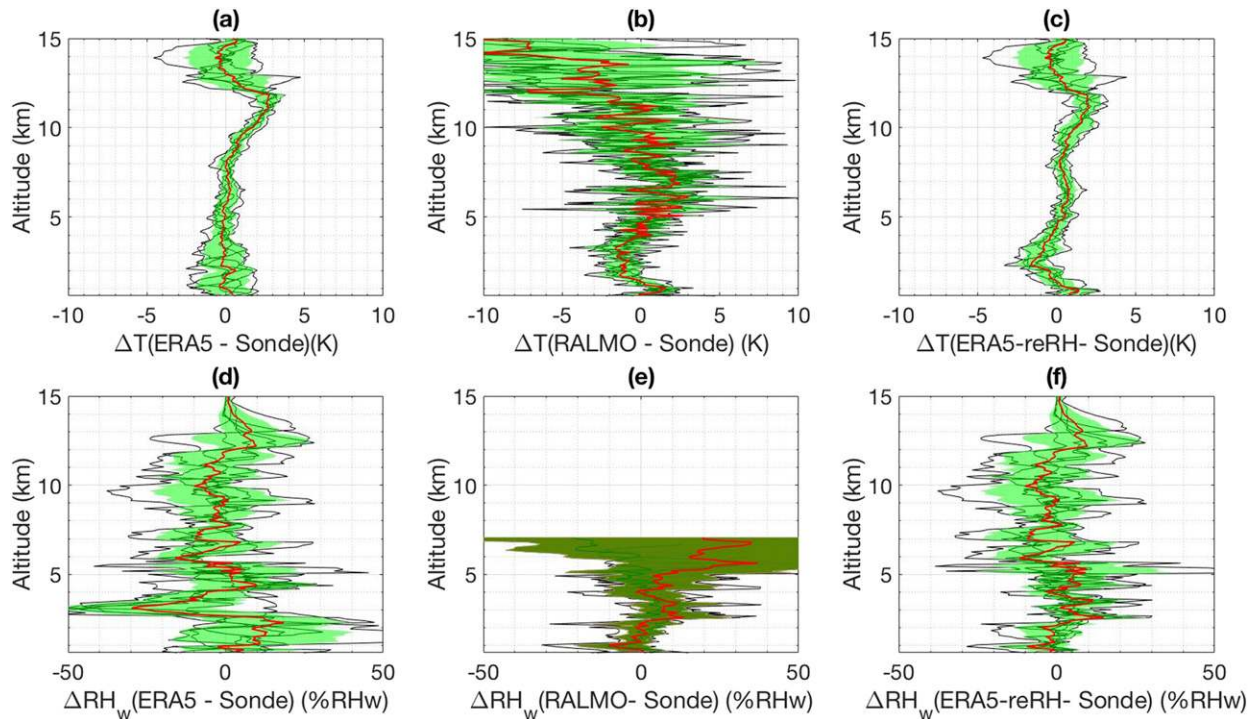


FIG. 8. (a) Daytime temperature differences between bias-corrected ERA5 and sonde measurements in black curves for six day times. (b) Temperature difference between daytime RALMO retrievals and sonde measurements. (c) Temperature difference between daytime ERA5-reRH retrievals and sonde measurements. (d) Relative humidity differences between daytime bias-corrected ERA5 and sonde measurements in gray curves for six day times. (e) Relative humidity difference between daytime RALMO retrievals and sonde measurements. (f) Relative humidity difference between daytime ERA5-reRH retrievals and sonde measurements. Red curves and green shaded areas show the mean and the spreads of the temperature and relative humidity differences of each panel.

which have not been assimilated into ERA5. The same set of sonde measurements has been used to determine the ERA5 background (a priori) error covariance matrix.

The comparison of ERA5, ERA5-reRH, and RALMO temperature and relative humidity profiles with coincident sonde measurements is given in section 4 and reveals that ERA5-reRH relative humidity profiles are significantly improved compared to both ERA5 and RALMO (the lidar-only retrieval). For temperature, only a modest improvement is found in terms of bias (but not in terms of standard deviation), mostly in the boundary layer and the upper tropospheric and lower stratospheric (UTLS) region. The impact of the lidar measurement is largest where the raw signals have a large signal-to-noise ratio, i.e., below 5 km during daytime and below 11 km during nighttime. The larger improvement seen for relative humidity than for temperature is expected since ERA5 assimilates relatively fewer humidity datasets relative to temperature and also because water vapor is highly variable both spatially and temporally. The ERA5 temperature data have been continually improved over the last few decades (Hersbach et al. 2019) and the temperature uncertainty is on the same order as the standard uncertainty of radiosondes. In contrast, the ERA5 relative humidity data for the last few decades do not show any significant improvement.

Full uncertainty budgets showed the total ERA5-reRH temperature uncertainties for both day and night are less than 1 K for all altitudes. The ERA5-reRH relative humidity uncertainty is less than 4%RHw for nighttime and for daytime up to 7% RHw (Figs. 4, 5). Statistical and calibration uncertainties introduce most of the relative humidity uncertainty. For comparison, the ERA5 relative humidity uncertainty is greater than 10% RHw up to 10 km, reaching a maximum of 25%RHw at 2 km (Fig. 2). Our ERA5-reRH relative humidity uncertainties for both day and nighttime retrievals are less than 10%RHw with a temperature accuracy of 1 K. Thus, our findings show that the relative humidity uncertainty is on the order of 10% with a temperature accuracy of 1 K, as claimed by Mattis et al. (2002).

We have chosen to not attempt to estimate smoothing errors for the retrievals. Estimation of the smoothing error is not straightforward, as the true state of the retrieval and the covariance matrix of a real ensemble of states are not normally well known. Rodgers (2000) and von Clarmann (2014) discuss in detail how in practice smoothing error may not be accurately quantified, in part as smoothing error propagation between the measurement and retrieval grids does not follow Gaussian error propagation. In Sica and Haeefe (2015) smoothing error was calculated for Rayleigh lidar temperature retrievals, but for their lidar water vapor retrieval Sica and Haeefe (2016) chose not to include it, both due to the implications of the von

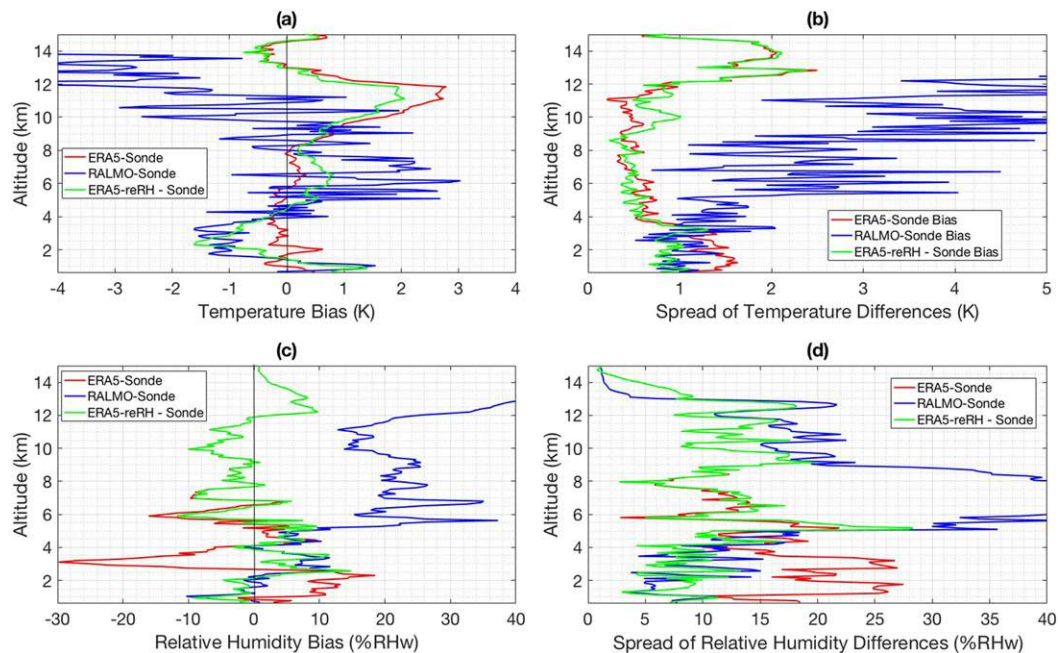


FIG. 9. (a) Daytime temperature biases between ERA5 and sonde (red curve), RALMO and sonde (blue curve), and ERA5-reRH and sonde (green curve). (b) Spread of the relative temperature differences. (c) Daytime relative humidity biases. (d) Spread of the relative humidity differences.

Clarmann (2014) result, and because the radiosonde and lidar measurements had similar vertical resolutions, which is the case in this retrieval as well.

Our ERA5-reRH relative humidity can also be compared to a previous determination of relative humidity using a combination of 2 Raman lidars, one which measured rotational temperature and the other, water vapor mixing ratio (Wang et al. 2011). These measurements were then combined into a relative humidity profile which was compared to a radiosonde measurement. A relative humidity statistical uncertainty of less than 10%RHw up to an altitude of 2 km was reported.

6. Conclusions

We have successfully assimilated Raman lidar measurements into ERA5 to generate an excellent quality dataset of relative humidity. Our ERA5-reRH reanalysis is an optimal combination of Raman lidar measurements and the ERA5 data that improves the determination of temperature and relative humidity compared to the lidar or ERA5 alone. ERA5-reRH overcomes some limitations of previous datasets based on lidar data, such as lack of coincident humidity and temperature measurements or excessive measurement noise in addition to providing a better characterization of the systematic uncertainties. Both daytime and nighttime ERA5-reRH retrieved temperature and relative humidity profiles that are in excellent agreement with coincident radiosonde measurements.

Our study demonstrates the potential benefits to improve numerical weather prediction (NWP) through assimilation of temperature and humidity information from Raman lidar

measurements. MeteoSwiss plans to assimilate the Payerne Raman lidar measurements with ERA5 on an operational basis in the near future.

We also plan to use this methodology and the RALMO database to characterize ice supersaturation layers (ISS) in the upper troposphere. Accurate relative humidity retrievals and uncertainties are essential to detect ISS. Previous studies (Comstock et al. 2004; Immler et al. 2008) used Raman lidar water vapor measurements combined with radiosonde temperature measurements to determine RH_i and thus to detect ISS layers. The combination of a temperature and humidity Raman lidar with ERA5 provides a relative humidity dataset with unprecedented quality, together with a profile-by-profile uncertainty budget, in which to better characterize ISS layers.

Acknowledgments. We thank Dr. Ghazal Farhani for her helpful comments and suggestions that were extremely useful to us. We also thank the Western writing support center and Patricia Sica for their assistance in editing and proofreading this paper. This project has been funded in part by the National Science and Engineering Research Council of Canada and by the Canadian Space Agency under the Arctic Validation and Training for Atmospheric Research in Science (AVATARS) program.

REFERENCES

- Ansmann, A., and D. Müller, 2005: Lidar and atmospheric aerosol particles. *Lidar*, C. Weitkamp, Ed., Springer Series in Optical Sciences, Vol. 102, Springer, 105–141.
- , U. Wandinger, M. Riebesell, C. Weitkamp, and W. Michaelis, 1992: Independent measurement of extinction and backscatter

- profiles in cirrus clouds by using a combined Raman elastic-backscatter lidar. *Appl. Opt.*, **31**, 7113–7131, <https://doi.org/10.1364/AO.31.007113>.
- Behrendt, A., 2005: Temperature measurements with lidar. *Lidar*, C. Weitkamp, Ed., Springer Series in Optical Sciences, Vol. 102, Springer, 273–305.
- Brocard, E., R. Philipona, A. Haeefe, G. Romanens, A. Mueller, D. Ruffieux, V. Simeonov, and B. Calpini, 2013: Raman Lidar for Meteorological Observations, RALMO—Part 2: Validation of water vapor measurements. *Atmos. Meas. Tech.*, **6**, 1347–1358, <https://doi.org/10.5194/amt-6-1347-2013>.
- Comstock, J. M., T. P. Ackerman, and D. D. Turner, 2004: Evidence of high ice supersaturation in cirrus clouds using ARM Raman lidar measurements. *Geophys. Res. Lett.*, **31**, L11106, <https://doi.org/10.1029/2004GL019705>.
- Diniov, T., V. Simeonov, B. Calpini, and M. Parlange, 2010: Monitoring of Eyjafjallajökull ash layer evolution over Payerne Switzerland with a Raman lidar. *WMO Conf. on Meteorological and Environmental Instruments and Methods of Observation 2010*, Helsinki, Finland, World Meteorological Organization.
- , —, Y. Arshinov, S. Bobrovnikov, P. Ristori, B. Calpini, M. Parlange, and H. Bergh, 2013: Raman Lidar for Meteorological Observations, RALMO—Part 1: Instrument description. *Atmos. Meas. Tech.*, **6**, 1329–1346, <https://doi.org/10.5194/amt-6-1329-2013>.
- Dirksen, R., M. Sommer, F. Immler, D. Hurst, R. Kivi, and H. Vömel, 2014: Reference quality upper-air measurements: GRUAN data processing for the Vaisala RS92 radiosonde. *Atmos. Meas. Tech.*, **7**, 4463–4490, <https://doi.org/10.5194/amt-7-4463-2014>.
- ECMWF, 2018: IFS documentation—CY37R2: Part IV: Physical processes. ECMWF Doc., 174 pp., <https://www.ecmwf.int/en/elibrary/9239-part-iv-physical-processes>.
- Eriksson, P., C. Jiménez, and S. A. Buehler, 2005: Qpack, a general tool for instrument simulation and retrieval work. *J. Quant. Spectrosc. Radiat. Transfer*, **91**, 47–64, <https://doi.org/10.1016/j.jqsrt.2004.05.050>.
- Farhani, G., R. J. Sica, S. Godin-Beekmann, and A. Haeefe, 2019: Optimal estimation method retrievals of stratospheric ozone profiles from a DIAL. *Atmos. Meas. Tech.*, **12**, 2097–2111, <https://doi.org/10.5194/amt-12-2097-2019>.
- Ferreira, A. P., R. Nieto, and L. Gimeno, 2019: Completeness of radiosonde humidity observations based on the integrated global radiosonde archive. *Earth Syst. Sci. Data*, **11**, 603–627, <https://doi.org/10.5194/essd-11-603-2019>.
- Gamage, S. M., R. J. Sica, G. Martucci, and A. Haeefe, 2019: Retrieval of temperature from a multiple channel pure rotational Raman backscatter lidar using an optimal estimation method. *Atmos. Meas. Tech.*, **12**, 5801–5816, <https://doi.org/10.5194/amt-12-5801-2019>.
- Hennermann, K., and P. Berrisford, 2017: ERA5 data documentation. ECMWF, <https://confluence.ecmwf.int/display/CKB/ERA5%3A+data+documentation>.
- Hersbach, H., and Coauthors, 2019: Global reanalysis: Goodbye Era-Interim, hello ERA5. *ECMWF Newsletter*, No. 159, ECMWF, Reading, United Kingdom, 17–24, <https://www.ecmwf.int/node/19027>.
- Hyland, R., and A. Wexler, 1983: Formulations for the thermodynamic properties of the saturated phases of H₂O from 173.15 to 473.15 K. *ASHRAE Trans.*, **89**, 500–519.
- Immler, F., R. Treffeisen, D. Engelbart, K. Krüger, and O. Schrems, 2008: Cirrus, contrails, and ice supersaturated regions in high pressure systems at northern mid latitudes. *Atmos. Chem. Phys.*, **8**, 1689–1699, <https://doi.org/10.5194/acp-8-1689-2008>.
- Kämpfer, N., 2012: *Monitoring Atmospheric Water Vapour: Ground-Based Remote Sensing and In-Situ Methods*. ISSI Scientific Report Series, Vol. 10, Springer, 325 pp.
- Kovalev, V. A., and W. E. Eichinger, 2004: *Elastic Lidar: Theory, Practice, and Analysis Methods*. John Wiley and Sons, 640 p.
- Lenschow, D. H., V. Wulfmeyer, and C. Senff, 2000: Measuring second-through fourth-order moments in noisy data. *J. Atmos. Oceanic Technol.*, **17**, 1330–1347, [https://doi.org/10.1175/1520-0426\(2000\)017<1330:MSTFOM>2.0.CO;2](https://doi.org/10.1175/1520-0426(2000)017<1330:MSTFOM>2.0.CO;2).
- Mattis, I., and Coauthors, 2002: Relative-humidity profiling in the troposphere with a Raman lidar. *Appl. Opt.*, **41**, 6451–6462, <https://doi.org/10.1364/AO.41.006451>.
- Miloshevich, L. M., H. Vömel, A. Paukkunen, A. J. Heymsfield, and S. J. Oltmans, 2001: Characterization and correction of relative humidity measurements from Vaisala RS80—A radiosondes at cold temperatures. *J. Atmos. Oceanic Technol.*, **18**, 135–156, [https://doi.org/10.1175/1520-0426\(2001\)018<0135:CACORH>2.0.CO;2](https://doi.org/10.1175/1520-0426(2001)018<0135:CACORH>2.0.CO;2).
- , —, D. N. Whiteman, and T. Leblanc, 2009: Accuracy assessment and correction of Vaisala RS92 radiosonde water vapor measurements. *J. Geophys. Res.*, **114**, D11305, <https://doi.org/10.1029/2008JD011565>.
- Nicolet, M., 1984: On the molecular scattering in the terrestrial atmosphere: An empirical formula for its calculation in the atmosphere. *Planet. Space Sci.*, **32**, 1467–1468, [https://doi.org/10.1016/0032-0633\(84\)90089-8](https://doi.org/10.1016/0032-0633(84)90089-8).
- Noh, Y.-C., B.-J. Sohn, Y. Kim, S. Joo, and W. Bell, 2016: Evaluation of temperature and humidity profiles of Unified Model and ECMWF analyses using GRUAN radiosonde observations. *Atmosphere*, **7**, 94, <https://doi.org/10.3390/atmos7070094>.
- Palmer, P. I., J. Barnett, J. Eyre, and S. Healy, 2000: A nonlinear optimal estimation inverse method for radio occultation measurements of temperature, humidity, and surface pressure. *J. Geophys. Res.*, **105**, 17 513–17 526, <https://doi.org/10.1029/2000JD900151>.
- Pappalardo, G., and Coauthors, 2004: Aerosol lidar intercomparison in the framework of the EARLINET project 3 Raman lidar algorithm for aerosol extinction, backscatter, and lidar ratio. *Appl. Opt.*, **43**, 5370–5385, <https://doi.org/10.1364/AO.43.005370>.
- Rodgers, C. D., 2000: *Inverse Methods for Atmospheric Sounding: Theory and Practice*. Vol. 2. World Scientific, 256 pp.
- Sica, R., and A. Haeefe, 2015: Retrieval of temperature from a multiple-channel Rayleigh-scatter lidar using an optimal estimation method. *Appl. Opt.*, **54**, 1872–1889, <https://doi.org/10.1364/AO.54.001872>.
- , and —, 2016: Retrieval of water vapor mixing ratio from a multiple channel Raman-scatter lidar using an optimal estimation method. *Appl. Opt.*, **55**, 763–777, <https://doi.org/10.1364/AO.55.000763>.
- von Clarmann, T., 2014: Smoothing error pitfalls. *Atmos. Meas. Tech.*, **7**, 3023–3034, <https://doi.org/10.5194/amt-7-3023-2014>.
- Wang, Y., D. Hua, J. Mao, L. Wang, and Y. Xue, 2011: A detection of atmospheric relative humidity profile by UV Raman lidar. *J. Quant. Spectrosc. Radiat. Transfer*, **112**, 214–219, <https://doi.org/10.1016/j.jqsrt.2010.05.008>.
- Whiteman, D., S. Melfi, and R. Ferrare, 1992: Raman lidar system for the measurement of water vapor and aerosols in the Earth's atmosphere. *Appl. Opt.*, **31**, 3068–3082, <https://doi.org/10.1364/AO.31.003068>.