Earth System
Open Access Science
Data Discussions

Water vapor Raman-lidar observations from multiple sites in the framework of WaLiNeAs

Frédéric Laly[1,2], Patrick Chazette[1], Julien Totems[1], Jérémy Lagarrigue[1,3], Laurent Forges[1], Cyrille Flamant[4]

[1]LSCE/IPSL, CNRS-CEA-UVSQ, University Paris–Saclay, CEA Saclay, Gif sur Yvette, France
[2]ADDAIR Company, 78530, Buc, France
[3]Now at Division Technique INSU (DT INSU), UPS 855, Meudon, France
[4]LATMOS/IPSL, Université Paris–Saclay, UVSQ, Guyancourt, France

Correspondence to: Frédéric Laly (Frederic.laly@lsce.ipsl.fr)

Abstract. During the Water Vapor Lidar Network Assimilation (WaLiNeAs) campaign, 8 lidars specifically designed to measure water vapor mixing ratio (WVMR) profiles were deployed on the western Mediterranean coast. The main objectives were to investigate the water vapor content during case studies of heavy precipitation events in the coastal Western Mediterranean and assess the impact of high spatio-temporal WVMR data on numerical weather prediction forecasts by means of state–of–the–art assimilation techniques. Given the increasing occurrence of extreme events due to climate change, WaLiNeAs is the first program in Europe to provide network–like, simultaneous and continuous water vapor profile measurements. This paper focuses on the WVMR profiling datasets obtained from three of the lidars ~~managed~~ [run] by the French component of the WaLiNeAs team. These lidars were deployed in the ~~towns~~ [cities] of Coursan, Grau du Roi and Cannes. This measurement setup enabled monitoring of the water vapor content within the low troposphere ~~along~~ [over] a period of three months over autumn – winter 2022 and four months in summer 2023. The lidars measured the WVMR profiles from the surface up to approximately 6–10 km at night, and 1–2 km during daytime. [They had] ~~with~~ a vertical resolution of 100 m and a time ~~sampling~~ [resolution] between 15 – 30 min [and they were] selected to meet the needs of weather forecasting with an uncertainty lower than 0.4 g kg[-1]. The paper presents details about the instruments, the experimental strategy, as well as the datasets ~~given in NETcdf format~~ [are provided.] The final dataset is divided in two [sub-]datasets; the first with a time resolution of 15 min, which contains a total of 26 423 WVMR vertical profiles and the second with a time resolution of 30 min to improve the signal to noise ratio and signal altitude range.

1 Introduction

The Mediterranean Basin has been identified as a hotspot of climate change for the years to come, as its population is expected to increase to 500 million inhabitants within the next 15 years (Giorgi and Lionello, 2008). In the context of global warming, this area has increasingly been subjected to heavy precipitation events (HPEs) that produce flash floods and landslides during autumn (e.g. Ricard et al., 2012). The scientific community noted that the frequency of HPEs alarmingly increased over the last 30 years (e.g. Flamant et al., 2021). Autum HPEs occur when the temperature difference between the sea surface and the

* an updated projection and citation is needed.

atmosphere is greatest, after Western Mediterranean waters have warmed all summer. Such temperature conditions favour water evaporation, which brings latent energy in the atmosphere, leading to deep convection processes and formation of mesoscale convective systems (MCSs) (e.g. Ducrocq et al., 2008; Duffourg et al., 2016; Chazette et al., 2016). The water vapor mixing ratio (WVMR) is therefore a crucial parameter to study for the energy balance of the troposphere (e.g. Held and Soden,

35   2000; IPCC, 2022).

Humid air masses from the Atlantic Ocean and Saharan regions are advected over the *western* Mediterranean Sea and reach the coast of southern France, which leads to HPEs (Duffourg and Ducrocq, 2011). It has already been established experimentally that before HPEs, the atmosphere is moister, with an increase in water vapor content in the first kilometres above ground level (Flamant et al., 2021). For instance, Chazette et al. (2016) used data acquired by a ground-based lidar in the Balearic Islands

40   along with satellites data to study the formation of MCSs which impacted the Cevennes–Vivarais area as they lead to HPEs. They highlighted *also* the fact that these MCSs were formed over the Mediterranean basin and were moistened as they passed over the sea, leading to an increase in water vapor content up to 5 km in the free troposphere.

Due to global warming, the intensity of HPEs increases, threatening human lives and leading to important economic and environmental costs (IPCC, 2022). Therefore, the ~~sampling~~ *monitoring* and forecasting of these events, which are two fundamental

45   components of a decision–making tool for local governments, are major but challenging objectives for meteorologists. Indeed, current measurement methods lack the temporal and vertical resolutions to correctly study the water vapor content initiating deep convection in the low troposphere (Flamant et al., 2021), where the spatio-temporal variability of the moisture field is greatest.

In response to these climatic threats, the international scientific community implemented the 10-year Hydrological Cycle

50   Experiment in the Mediterranean program (HyMeX, Drobinski et al., 2014). This program deployed a suite of instruments within the Special Observing Period to measure meteorological parameters over the western Mediterranean area at the surface and in the lower troposphere (Ducrocq et al., 2014; Duffourg et al., 2018). As part of the instrumental set-up, two Raman lidar systems (Chazette et al., 2014; Di Girolamo et al., 2020) were dedicated to measure atmospheric water vapor. *profiles.* These instruments provided the constraints needed not only to validate airborne and drifting balloon (Chazette et al., 2016)

55   measurements, but also to test the impact of their assimilation in the Application of Research to Operations at Mesoscale (AROME) model (Seity et al., 2011; Fourrié et al., 2019). The Raman lidar system used over the Balearic Islands also provided the opportunity for a validation campaign of the Infrared Atmospheric Sounding Interferometer (IASI) onboard the MetOp (meteorological operational) platform (Chazette et al., 2014), a key component for assimilation in numerical weather prediction models (Hilton et al., 2009; Guidard et al., 2011).

60   Improving the forecasting of HPEs over the *W* Western Mediterranean Basin by using what was learnt during HyMex was the main motivation of the new French initiative Water Vapor Lidar Network Assimilation (WaLiNeAs, Flamant et al., 2021). The main field campaign associated to WaLiNeAs took place between October 2022 and January 2023. It was followed by a lidar calibration validation† campaign at the Météo–France site in Toulouse from June to September 2023. That campaign also

† normaly the validation campaign is performed before the main campaign. explain WHY.! or comment on it.

Earth System
Science
Data
Open Access
Discussions

presented the opportunity to sample heavy rainstorms, as well as the severe heat wave of August 2023, which affected all of
65   southern Europe.

The goal of this paper is to provide an overview of the Raman lidar measurements performed from the French ground–based stations during WaLiNeAs and the complementary campaign carried out in Toulouse. The lidar profiles are now available to the national scientific community on the AERIS database (https://doi.org/10.25326/537). This represents a total of 26 423 lidar profiles, averaged over 15 min with a vertical resolution of 100 m. The experimental strategy is presented in Section 2, along
70   with the main objectives of the campaign, the site locations, the descriptions of the instruments used and the operating time periods. Section 3 describes the data processing methodology and also the algorithms for the assessment of uncertainties, computed with an end–to–end approach. Section 4 presents the results of data processing, after applying the methodology described in section 3. Section 5 details the final database structure, as well as the procedure for the reader to access the database, and defines the flags for data quality. A conclusion is presented in section 6.

75   **2   The ground–based experiment**

2.1 Main objectives

The main objective of the WaLiNeAs campaign is to improve the prediction of HPEs and the understanding of the initial conditions that generate these events by assimilating WVMR lidar datasets into mesoscale models, as represented in Fig. 1. Data acquired during the WaLiNeAs campaign in autumn and winter 2022 − 2023 will be assimilated in mesoscale models
80   such as the AROME mesoscale model, at the horizontal resolution of 1.3 km developed by Météo–France (Fourrié et al., 2019). These data will serve as constraints for the model to improve the precision in forecasting precipitation events. Current means of measurement providing data in the AROME mesoscale model have limited temporal and vertical resolution. As discussed in Chazette et al. (2014), IASI satellite data offer vertical resolution on the order of one kilometre in the lower troposphere and the weighting functions of the spectral channels use peak over 2 km above ground level (a.g.l.). Consequently, they lack the
85   necessary vertical precision to accurately measure water vapor in the altitude range of the atmospheric boundary layer, which contains the majority of water vapor content. This limitation can potentially result in errors and inaccuracies when predicting both the intensity and the location of HPEs. Radiosoundings are well resolved in altitude, but measurements are too punctual, with an average sampling frequency of two radiosoundings per day. Ground–based weather stations provide continuous data over time, but each of their measurement is given for a precise point in space and in altitude, and moreover, the correlation
90   between ground-based level measurements and the atmosphere above is frequently low (Chazette et al., 2017). On the other hand, Raman lidar data provide continuous and well resolved WVMR profile in altitude (e.g. Whiteman et al., 1992; Ansmann et al., 1992; Mattis et al., 2002). During WaLiNeAs, ground–based Raman lidars measure the WVMR up to 1.5–2.5 km during daytime and over 6 km during nighttime, with a vertical resolution of 100 m, as discussed in section 2.3. These performances let us sample the majority of vapor content in the troposphere, with sufficient resolution to identify the various processes that
95   may lead to HPEs.

Water vapor

Lidar data acquired during the WaLiNeAs and Toulouse campaign ~~are~~ *will* also be available on the AERIS database (https://metclim-lidars.aeris-data.fr/). This database will also serve for future case studies involving intercomparisons and validations with other measuring methods from aircrafts or satellites, as well as with mesoscale models (Fig.1).
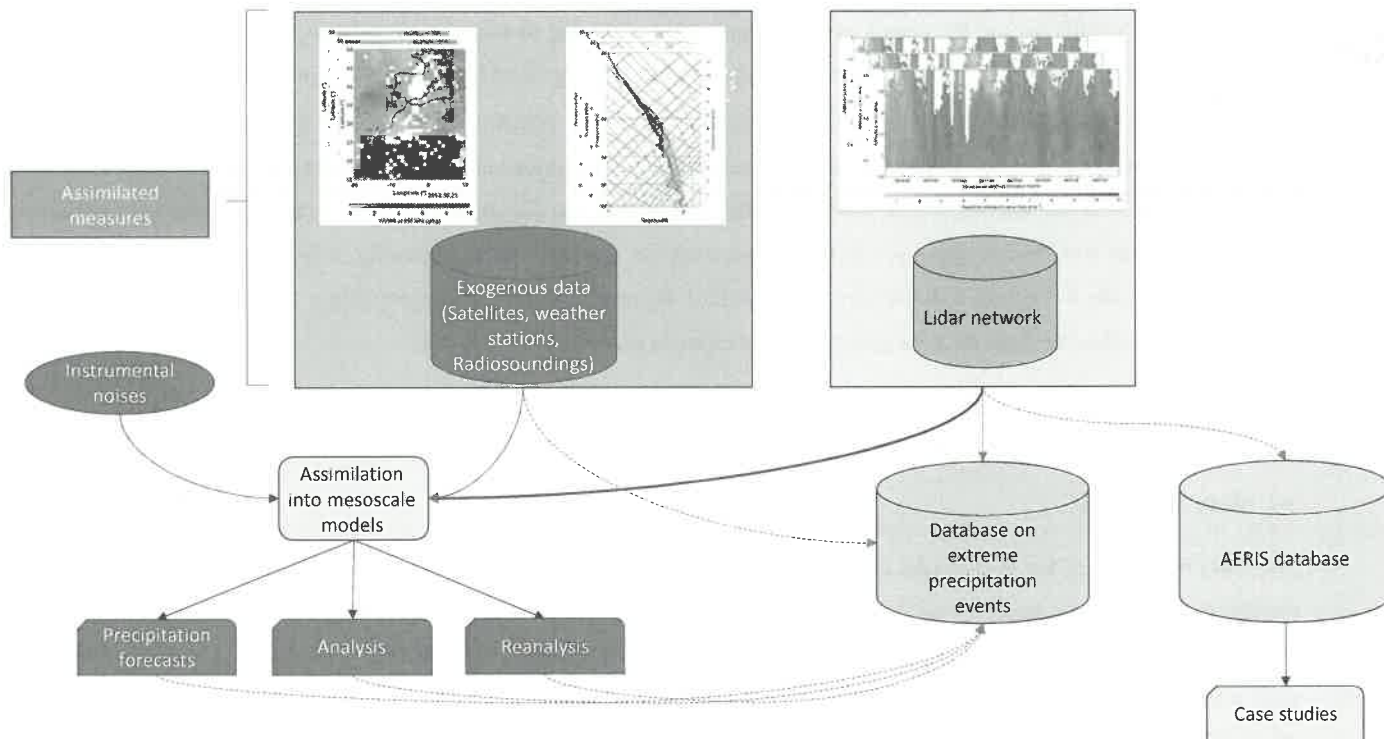
Figure 1. Flowchart of the use of lidar data acquired during the WaLiNeAs campaign. Lidar vertical profiles acquired using the lidar network (orange cylindrical box) have been added to the AERIS database represented by the grey cylindrical box. These data will therefore serve directly for case studies, shown as a green box at the bottom right. The primary goal of lidar measurements is to be assimilated into mesoscale meteorological models (green box on the left), which already assimilates other data from exogenous measurements, shown as a blue box in the top left. Together, instrument measurements, lidar data and models' outputs form a complete database on HPEs, represented by the yellow cylindrical box.

2.2 Experimental strategy

2.2.1 Meteorological context

Duffourg and Ducrocq (2011) highlight that humid air masses have various remote origins before reaching France. On average, two days before an event, the majority of humid air masses reaching France come with the southerly flow from Africa, which may bring water vapor from the tropical Atlantic (Winschall et al., 2012), and the westerly flow originating from the Atlantic Ocean, in connection to extratropical cyclones (Dettinger, 2011; Flaounas et al., 2014; Pfahl et al., 2014). Note that Duffourg and Ducrocq (2011) also point to a contribution from the eastern Mediterranean area.

Once the precipitating system reaches the western Mediterranean region, it ~~can~~ *may* follow two main paths before reaching France's Mediterranean coasts: one along the southern Spanish coast before reaching the Balearic Islands, heading northward/north–

after passing

115 eastward and one from Tunisia ~~that runs~~ over Sardinia, heading northward/north–westward. The Mediterranean Sea acts as a heat and moisture source, and coastal orography (i.e. Massif Central, Pyrenees, Alps) induces mesoscale convergence and lift of moist air (Ricard et al., 2012). The interaction between the synoptic conditions, topography, and mesoscale features determine the location and intensity of precipitation. Lastly, the shape and position of mountain ranges may enhance rainfall in very specific areas, leading to destructive floods.

120 **2.2.2 Experimental set up**

To study HPEs over the western Mediterranean area within the framework of WaLiNeAs, 8 lidar sites managed by Spanish, French, Italian and German research teams (Flamant et al., 2021) were set up on the Mediterranean coasts of France and Spain. Their coordinated efforts made it possible to track air masses bringing water vapor content towards the south of France. A 9th site was also set up to complete the validation of lidar measurements, near the Météo-France radiosonde station in Toulouse

125 (South-west France). The locations of the sites involved during the WaLiNeAs campaign are shown in Fig. 2, with the different air masses impacting them on the western Mediterranean Basin (inspired from Flamant et al. (2021), Fig. 5). The coastal lidar sites were chosen to study the moisture in the low troposphere upstream the mountainous areas whose windward sides have been the most impacted by HPEs during the last decades, namely Languedoc–Roussillon, Cévennes–Vivarais, southern Alps, and Corsica (Ricard et al., 2012; Ducrocq et al., 2014; Duffourg et al., 2016, 2018).

run   *one*

130 Four lidar sites (Table 1) were ~~managed~~ by the French team during two seasons. The longest was during autumn 2022, where 3 sites were set up on the Mediterranean coast: Coursan, Grau du Roi and Cannes. Coursan is positioned upstream the Languedoc–Roussillon region. Low–level flows bringing precipitating systems are oriented easterly, usually due to a low–pressure area between the Balearic Islands and Corsica. This region is surrounded by the Pyrenees and Massif Central Mountain ranges, which create a venturi effect, bringing strong winds and humidity over lands and potentially leading to rainfall. Grau

135 du Roi is located upstream from the Rhone valley and the Cevennes mountains. These regions are also subjected to a southerly flow, veering slightly west near the coast, and among the most impacted by HPEs in the Mediterranean Basin. The orographic situation is similar to the one for Coursan, as the Rhone Valley is surrounded by the Massif Central and the Alps mountains. Finally, Cannes is located at the foothill of the Maritime Alps, a region which is also impacted by HPEs due to the Alps topography. During autumn, this region is also subjected to a southerly/south–westerly flow, which can bring elevated dust

140 plumes originating from the Sahara Desert. Those 3 sites were followed by an additional one in Toulouse, so as to validate the calibration process. The campaign has offered the opportunity to sample extreme weather events in southwestern France. Toulouse is located in the Midi–Pyrénées* region, which is also considered a climatic hotspot for the years to come, as the intense heatwaves and violent storms occurring during summer cause significant environmental and economic damages, and

frequently,

threaten human lives. At the end of summer and autumn, tropical air masses are advected over the western Mediterranean Sea

145 and bring important amount of moisture over southern France, sometimes crossing the Pyrenees mountains. Air masses are thus subjected to the Foehn effect, warming them in the leeward of the mountains. Besides, Toulouse lies in the path of Atlantic air masses and Autan winds blowing from the southeast, which can generate storms and HPEs over the Midi–Pyrénées region.

** follow the same orthographic all over the paper (Pyrénées)*

The 4 French lidar sites were equipped with i) the H_2O Raman Ultraviolet Sounder second generation (HORUS–2) at Coursan and Toulouse, ii) the H_2O Raman Ultraviolet Sounder first generation (HORUS–1) at Grau du Roi, and iii) the Water Vapor and Aerosol Lidar (WALI) at Cannes. The sites are indicated by a red cross in Fig. 2 and their geographical coordinates are given in Table 1, together with the altitude of the site above mean sea level (a.m.s.l.). The other lidar sites managed by the different European teams are also shown in Fig. 2, but as yellow crosses. Note that lidar instruments deployed across all sites are described by Flamant et al. (2021). WALI (Chazette et al., 2014; Totems et al., 2021) is embedded in the Mobile Atmospheric Station (MAS) (e.g. Raut and Chazette, 2009) shown in Fig.3a. The HORUS lidars have been developed for the purpose of the WaLiNeAs campaign at Laboratoire des Sciences du Climat et de l'Environnement (LSCE). As a new generation of compact and autonomous systems (Fig.3b), they were conceived specifically to measure water vapor content in the low tropposphere.
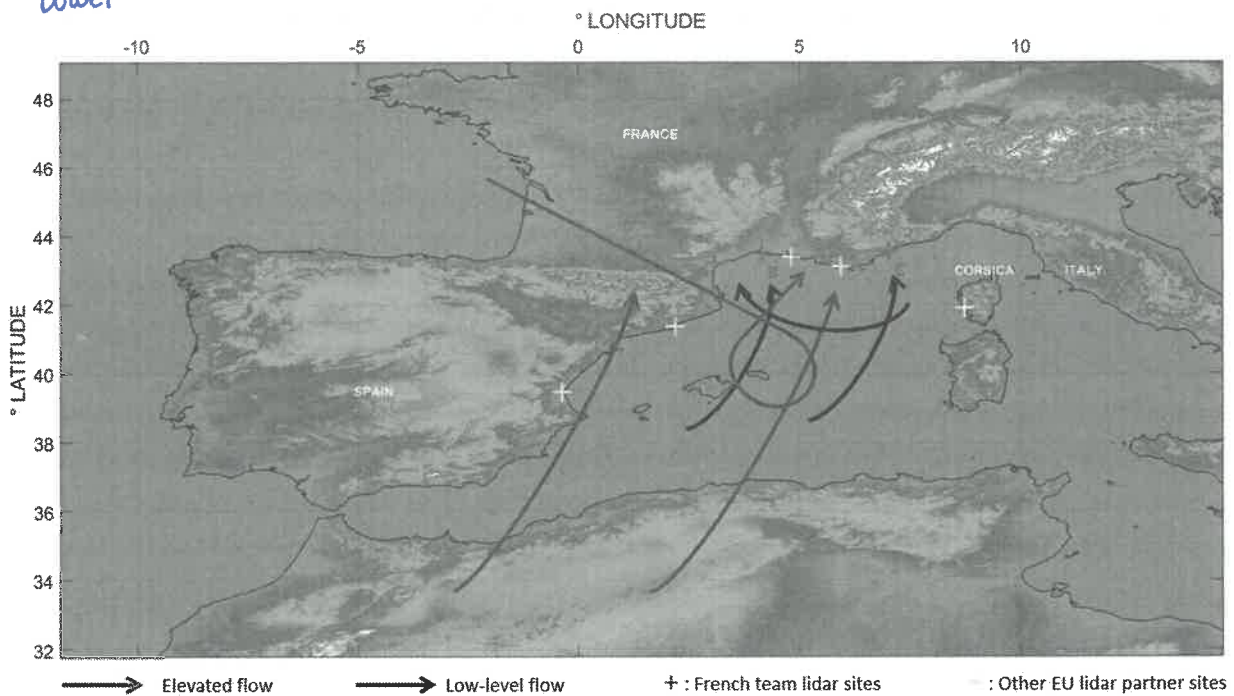
Figure 2. Map of the WaLiNeAs campaign lidar sites and of the main flow patterns in the low levels (black arrows) and in altitude (between 2 and 4 km, brown arrows) described in Flamant et al. (2021). Red crosses represent the French team lidar sites and yellow crosses those of the other European teams. See Table 1 for letters signification.
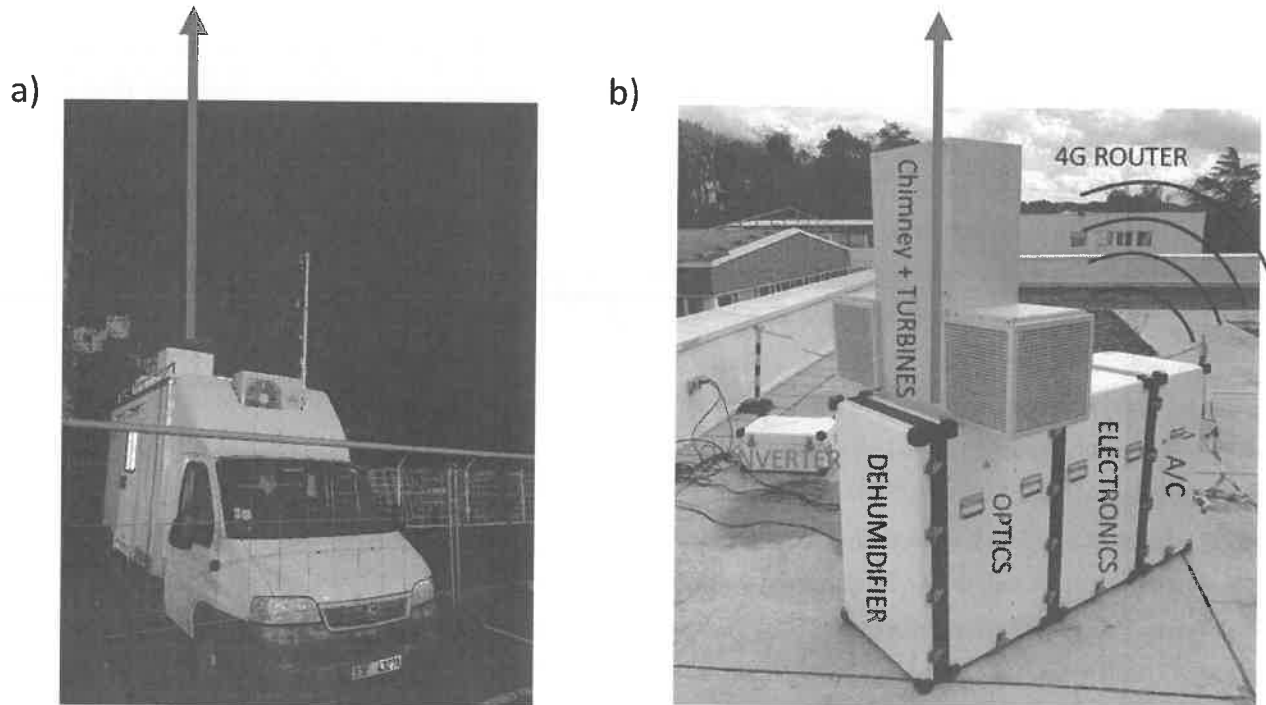
Figure 3. Picture of a) the MAS truck station containing WALI and b) HORUS-1 or -2 composed of several enclosures containing the air conditioning (A/C), electronic components (Electronics), optical components (Optics) and the dehumidifier. HORUS has a chimney to limit the sky background and avoid direct sunlight, as well as turbines to keep the emission window clean. HORUS is connected to an inverter to prevent power outages and a 4G router is present to access the lidar remotely. Lidar emission beams are represented by the purple arrows.

Table 1. Coordinates and altitude a.m.s.l. (above mean sea level) of the French WaLiNeAs lidar sites

Identification	Place	Latitude Longitude	Altitude a.m.s.l.
A	Coursan	43°14'5''N 3°3'49''E	4 m
B	Grau du Roi	43°31'14''N 4°7'39''E	7 m
C	Cannes	43°32'29''N 6°57'30''E	4 m
D	Toulouse	43°34'28" N 1°22' 25" E	157 m

Earth System
Science
Data

Open Access

Discussions

2.3 Lidar characteristics

170 The main characteristics of WALI and HORUS 1&2, the three lidars deployed during the WaLiNeAs campaign, are summarized in Table 2. HORUS is composed of 3 modules to create a compact and autonomous instrument (Fig. 3b). The electronics module supplies power to the other two modules and contains all the electronics and the optical spectral analysers, which consist of two rack-mounted fiber optic polychromators. The optics module contains the laser transmitter and the two

175 reception telescopes. Each receiving telescope acquires a N_2–Raman channel and a H_2O–Raman channel to improve the signal to noise ratio. An AC module maintains the internal temperature of the lidar, which is crucial for the correct functioning of the laser and optical detection. Above the optics module, turbines remove water and any particles that can be deposited on the lidar windows during precipitation events. The chimney prevents direct sunlight from entering the lidar, limiting damage due to focused light as well as the impact of sky background on the signals. HORUS is inverter-powered to prevent power cuts

180 affecting the campaign. Finally, a 4G router is embedded in each lidar to control it remotely. Inside this field–proof enclosure, the optical architecture of HORUS is almost identical to that of WALI (Totems et al., 2021), and presented in Fig.4a. WALI was developed at LSCE (Chazette et al., 2014) to simultaneously study the aerosol content in the atmosphere, with elastic reception channels, as well as the temperature (measured but not logged in the database) and water vapor profiles, with rotational and vibrational Raman channels respectively. A schematic representation of the WALI system components is given

185 in Fig. 4b. It is embedded in the MAS van (Raut and Chazette, 2009), offering a mobile temperature–controlled work environment.

The three lidar systems use pulsed frequency–tripled Nd:YAG lasers manufactured by Lumibird Quantel, with an emission wavelength of 354.7 nm. Laser beam expanders allowed to meet eye–safety standards (EN 60825–1) at the chimney exit. The UV pulse energy is respectively 30 mJ and 100 mJ for HORUS lidars and WALI, whereas the pulse repetition rate is 20 Hz

190 for WALI and HORUS–1, and 100 Hz for HORUS–2, enabling a better vertical range. The reception systems are 150–mm Newtonian telescopes, feeding filter–based spectral analysers (called polychromators on Fig.4) via an optical fiber. The acquisition system, employing PXI (PCI eXtensions for Instrumentation) technology, incorporates 12–bit digitizers manufactured by National Instrument® (https://www.ni.com/, last access 7 February 2023). These digitizers operate at a speed of 200 MHz, allowing for post–digitization photon counting. Table 2 provides an overview of the system's key characteristics

195 for each lidar. Full overlap, which represents the overlap between the transmitted beam and reception field of view, is reached 200 m above the lidar, as shown in Fig.5.

Table 2. Characteristics of the three lidars during the WaLiNeAs campaign

	WALI *Reference for HORUS*	HORUS–1	HORUS–2
Lidar type	Vibrational Raman for N_2 and H_2O		
Emission wavelength	354.7 nm		

Energy / Emission frequency	100 mJ / 20 Hz	30 mJ / 20 Hz	30 mJ / 100 Hz
Maximum daytime range	2000 m	1500 m	2500 m
Maximum nighttime range	10 km	7 km	12 km
Full overlap	200 m		
Minimum range	150 m		
Elastic channel	Yes	No	
Temperature channel	Yes	No	
Vertical resolution	*Raw* ~~Native~~: 0.75 m Final: 100 m		
Conditioning	MAS truck	ArtConcept®* Composite enclosures	
Time resolution	*Raw* ~~Real~~ ~~Native~~: 1 min Final: 15 min / 30 min (AC)		

* (https://www.art-concept.fr/, last access 7 February 2023)

[Handwritten annotations:]

Laser beam expansion (x 10⁻¹) factor   ×   ×   ×
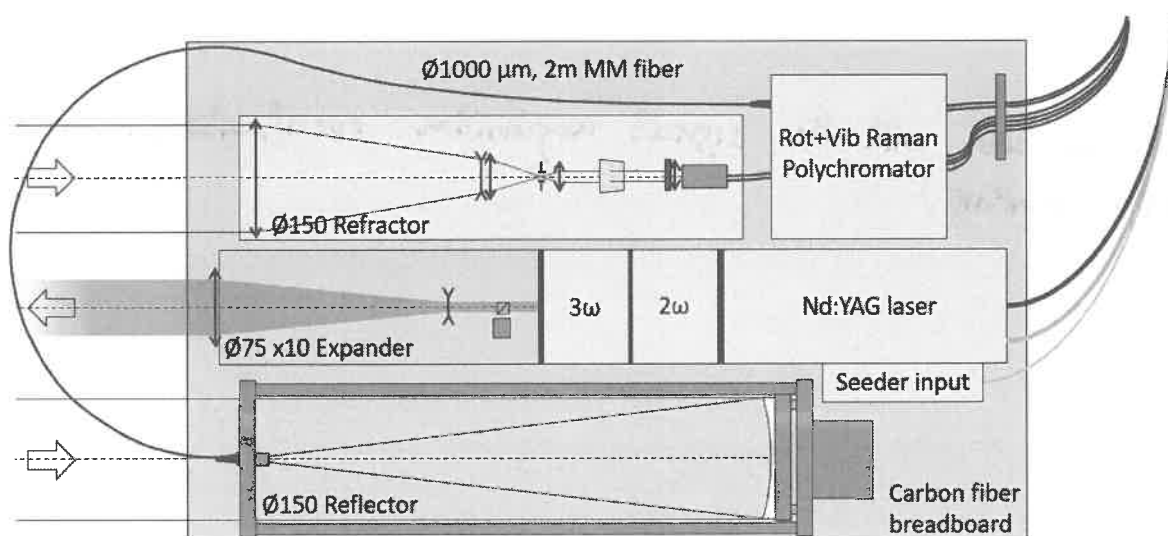
provide a line with the signal acquisition specifications (analog and photon counting)

a)
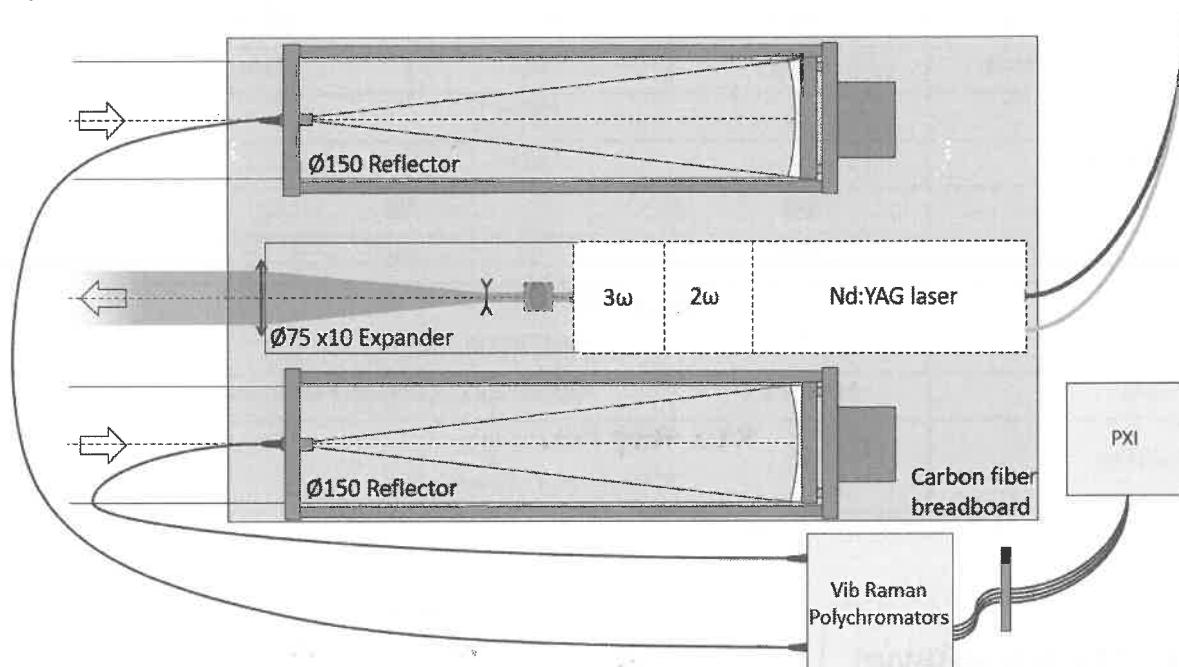
b)

200

Figure 4. Diagrams of emission and reception components. These diagrams are inspired by Fig. 2 in Totems et al. (2021) on WALI. a) For HORUS, the laser is shown in light grey to indicate the fact that it is on the other side of the optical table. The laser is cooled by water, which is fed through pipes shown in blue and black. It is equipped with a periscope, represented by a blue oval, which

205 carries the beam to the other side of the table before reaching the beam expander. Receiving telescopes ~~inject~~ fed the received signal into fibers directly connected to the Raman polychromators, enabling the signal to be processed by the PXI. b) For WALI, all components are present on the same side of the optical table. The blue cube letting the laser beam emission through represents a dichroic plate before the beam expander. The Raman Vibrational telescope is the same as HORUS, and the Rotational Raman reception system is described in Totems et al. (2021). WALI contains an injector represented by the yellow fiber to stabilise the emission wavelength, which is important for measuring temperature.
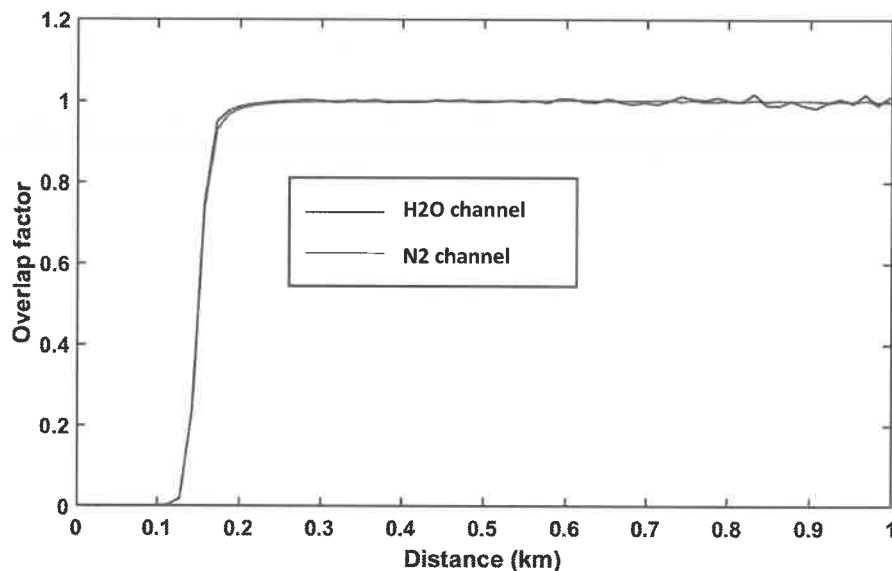
Figure 5. Lidar overlap factors for dinitrogen and water vapor Raman channels. The 3 lidars are built using the same telescope architecture, with identical overlap, factors.

2.4 Lidar operating time periods

The Periods during which the 3 French lidars were operational are summarized in Fig.6. Both WALI and HORUS–1 acquired
215 around two months of data. WALI had acquired data between 4 October 2022 and 12 January 2023. The lidar stopped several times over October and November, due to power drops which were not compensable by the inverter. It was then necessary to reboot the lidar manually until a remote–controlled power distribution unit was installed after mid–December, allowing to restart the lidar remotely if necessary. A short downtime in mid–December was necessary for routine maintenance on the laser. HORUS–1 acquired data continuously between 26 October 2022 and 12 January 2023. The lidar was briefly switched off for
220 standard maintenance at the beginning of November and December, respectively.

During the WaLiNeAs campaign, HORUS–2 acquired data between 6 October 2022 and 4 November 2022. Unfortunately, the lidar was unable to acquire data after a manufacturing defect induced a laser failure. Due to other laser failures probably caused by the same manufacturing defect, the lidar was also down several times during the month of October. After laser repair, HORUS–2 was redeployed in Toulouse between 31 May 2023 and 25 September 2023. The lidar was only off once
225 during that period, between the evening of 31 June 2023 and the morning of 3 July 2023, after a thunderstorm caused a power outage on the Météo–France site. The lidar was also briefly stopped on 25 July 2023 for maintenance. The performances of

HORUS–2 were of high quality, allowing us to measure water vapor content up to 2.5 km during daytime and up to the tropopause during nighttime. Among the three lidars, HORUS–2 is the most efficient field lidar for measuring water vapor content in the troposphere.
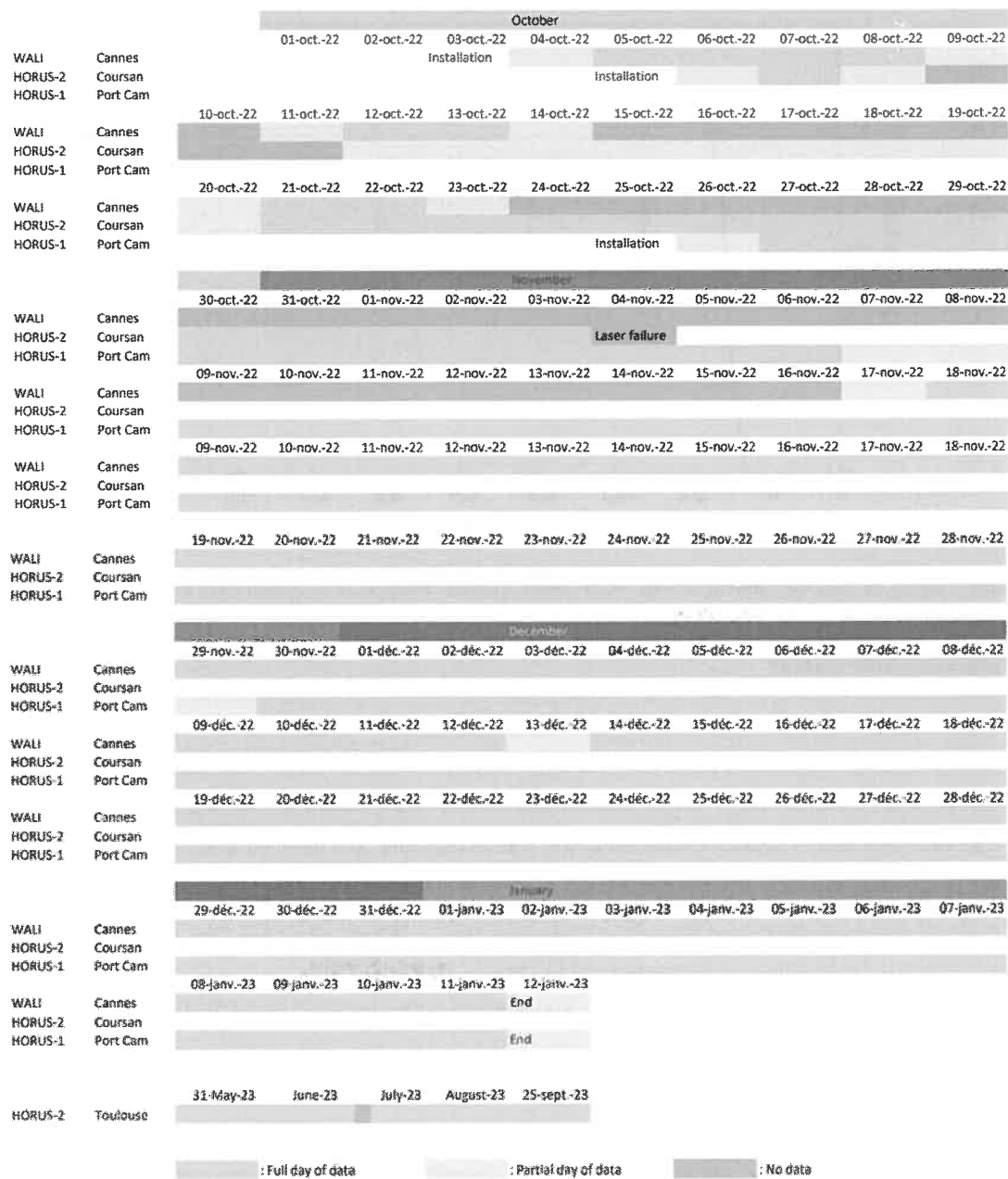
230

Figure 6. Daily lidar data availability for each ground–based lidar station.

3   Methodology

235   This section describes the method followed to invert the data from the raw lidar signals to the WVMR profiles. It also describes the method used to study the lidar instrumental error budget with an end–to–end model.

3.1 Basic Raman lidar equation

referring to analog and photon counting detector, respectively.

Vibrational Raman lidars acquire signals corresponding to the dinitrogen and water vapor backscattering in the atmosphere. Raw lidar profiles are expressed in millivolts *(mV)* and sampled at a rate of 200 MHz. As described in Totems et al. (2021) these

240   profiles are then corrected from both sky background radiance and detection solid angle. During the acquisition process, the lidar profiles are sampled with a ~~native~~ *raw* resolution of 0.75 m along the line of sight. A temporal averaging of 1000 profiles for WALI and HORUS–1 and over 5000 profiles for HORUS–2 translates to approximately one recording every 50 seconds over the campaign.

Typically,

The lidar acquires the range–corrected Raman signal S_i from ground level z_G at the altitude a.m.s.l. z of channel i (N_2 or H_2O),

245   of wavelength λ_i (386.6 nm for the dinitrogen channel and 407.5 nm for the water vapor channel) following the equation:

$$S_i(z) = K_i . g_i \cdot \beta_i(z) . O_i(z) . exp\left(- \int_{z_G}^{z} \left(1 + \eta_{i,m}\right) . \alpha_m(z') + \left(1 + \eta_{i,a}\right) . \alpha_a(z') . dz'\right) \tag{1}$$

K_i is the instrumental constant of channel i, which is a function of the lidar components, such as the laser emission energy, the transmission of reception optics, as well as the quantum efficiency of the photodetector. g_i is the photodetector gain, which depends on the level of high voltage (HV) applied to it. $\beta_i(z)$ is the volume backscattering coefficient defined as a function of the density profile N_i of gas i and the associated differential cross–section taken in backscatter condition (σ_i^π):

$$\beta_i(z) = N_i(z) . \sigma_i^\pi, \tag{2}$$

respectively

250   Spectral dependences for air molecules and aerosols are characterized by parameters $\eta_{i,m}$ and $\eta_{i,a}$ according to the relationships:

$$\begin{cases} \eta_{i,m} = \left(\frac{\lambda_i}{354.67}\right)^{-4.09} \\ \eta_{i,a} = \left(\frac{\lambda_i}{354.67}\right)^{-A} \end{cases}, \tag{3}$$

while

where A is the Ångström exponent of aerosol, α_m and α_a are ~~respectively~~ the molecular and aerosol extinction coefficients at 354.67 nm, respectively. The molecular extinction coefficient is determined following Nicolet (1984), using radiosoundings and climatological databases (Chazette et al., 2012a) .

here we an updated citation and formula.

255   ### 3.2 Inversion and atmospheric correction

3.2.1 WVMR retrieval

We aim to assess the WVMR (r_H) vertical profile, which is defined as the water vapor mass (m_{H_2O}) per dry air mass (m_a), expressed in g kg^{-1} at the altitude z:

$$r_H(z) = \frac{m_{H_2O}(z)}{m_a(z)},$$ (4)

This can also be written as:

$$r_H(z) = \frac{N_H(z)}{N_N(z)} \cdot \frac{M_H}{M_N} \cdot r_N,$$ (5)

260 Where N_i and M_i are ~~respectively~~ the density profile and the molar mass coefficient for molecule i, and r_N is the dinitrogen mixing ratio.

From equations (1) and (2), we can then calculate r_H from the lidar profiles according to the relationship:

$$r_H(z) = K_0 \cdot \underbrace{\frac{O_N(z)}{O_H(z)}}_{OR(z)} \cdot \frac{(S_H(z)/g_H)_M}{(S_N(z)/g_N)_M} \cdot C_m(z) \cdot C_a(z),$$ (6)

where K_0 is the calibration coefficient calculated for a reference HV of 940 V. The variables C_m and C_a are associated with atmospheric transmission corrections for molecules and aerosols, respectively. The water vapor (H) and dinitrogen (N) 265 channels are corrected for the detection gains g_H and g_N, respectively. The WVMR is calculated on a time–average ($\langle \ \rangle$) of M profiles for each altitude z with a vertical resolution of 100 m. This procedure is well established, as presented in Totems et al. (2021) or Chazette et al. (2014).

3.2.2 Molecular and aerosols transmission corrections

The method for retrieving the WVMR by Raman lidar measurement requires a correction of the atmospheric transmission at 270 the wavelengths used. Molecular transmission is a function of air density, and therefore of temperature and pressure, which are usually derived from climatological thermodynamic profiles, or radiosoundings when available. The corrective multiplicative term C_m is given by (Chazette et al., 2014):

$$C_m(z) = exp\left(-\left[\eta_{N,m} - \eta_{H,m}\right] \cdot \int_{z_G}^{z} \alpha_m(z') \cdot dz' \right)$$ (7)

Similar to the corrective multiplicative term of molecular transmission, the corrective multiplicative term C_a for aerosols transmission is written as follow (Chazette et al., 2014):

$$C_a(z) = exp\left(-\left[\eta_{N,a} - \eta_{H,a}\right] \cdot \int_{z_G}^{z} \alpha_a(z') \cdot dz' \right)$$ (8)

275 ### 3.3 Calibration of lidar-derived WVMR

The purpose of the calibration process is to find the constant K_0 in Eq. 6. Usually, this constant is found by comparing lidar–derived WVMR with coincident radiosounding profiles. During WaLiNeAs, because of the presence of no-fly zones, no radiosoundings were available close to the lidar sites. Hence, we use ground–based weather stations (PTU VAISALA® 303,

14

Open Access Earth System Science Data Discussions

https://www.vaisala.com/, last access 7 February 2023) calibrated just before the campaign by the VAISALA company to

280   perform the lidar calibration in terms of WVMR. For the meteorological probe, the absolute uncertainties on pressure, temperature and relative humidity are 0.25 hPa, 0.2 °C and 1 %, respectively. This leads to an error of 0.2 g kg^{-1} for the WVMR assessment. The weather stations were close to the lidars at ~2 m from ground level, just above the lidars. To retrieve K_0, we compared the WVMR derived from the meteorological probes with the one derived from the Raman lidar between 200 and 400 m, when the overlap factor is 1. Such comparison is reliable when the vertical gradient of r_H is close to 0, indicating a

285   well–mixed lower troposphere. It is worth noting that to calibrate the dual–telescope HORUS lidars, two distinct constant values must be employed for each of the two channels. This approach involves performing a cross calibration between the two telescopes (hereafter denoted as T1 and T2), while maintaining a constant ratio between the calibration constants associated with each of them. The results associated with the calibration process are presented in section 4.1.

3.4 Error budget calculation

290   As discussed in Chazette et al. (2012b, 2014), the determination of WVMR is affected by uncertainties stemming from three primary sources:

- The shot noise and the natural variability of the atmosphere, which are characterized by the signal–to–noise ratio (SNR_λ) of the lidar system,

- Bias due to calibration issues usually associated with in situ measurements, coincident with lidar profiles (subsection
295   3.3),

- Bias due to the contributions from molecular and aerosol (subsection 3.2.2) components in the atmosphere.

To the first order and considering all sources of uncertainty as independent, the total relative uncertainty ε_H on the WVMR (r_H) is given by the following equation (Chazette et al., 2014):

$$\varepsilon_H \approx \sqrt{\frac{1}{SNR_N^2} + \frac{1}{SNR_H^2} + (\varepsilon_m^2 + \varepsilon_a^2) + (\varepsilon_{K_0}^2 + \varepsilon_{HV}^2 + \varepsilon_O^2)} \qquad (9)$$

Where the relative bias associated with the spectral–dependent properties of the extinction for molecules and aerosols is given
300   by ε_m and ε_a. ε_{K_0} is the relative bias due to the calibration constant K_o, ε_O the relative bias due to the overlap factors and ε_{HV} the uncertainty resulting from the HV variations. High voltage (HV)

HV variations are mainly due to changes in sky background values during the day. The uncertainty is thus caused by the atmospheric variability during HV changes. The relative uncertainty resulting from the HV variation has been laboratory tested. Its contribution is ~1–2 % and may be higher with HV below 700 V (~3–4%). Note that below 600 V, the photomultipliers
305   may have a non-linear response. The relative bias on the overlap factor is negligible, as full overlap is reached above 200 m. In addition, a ratio between the two detection channels is performed to obtain the WVMR, which strongly limits overlap effects because the backscatter signals due to N_2 and H_2O follow the same optical path in the lidar architecture.

15

Note: HV variations are due to the electronic circuit which is controlling the HV of the PMTs. Power voltage variations (220V) induce the variations of the HV of the PMTs.

To simulate the contribution of shot noise, we employed a Monte Carlo approach similar to the one used by Royer et al. (2011) and Chazette et al. (2014). The schematic representation of the method is given in Fig. 7. First, we need to select reference

310   WVMR vertical profiles ($r_{H,ref}$) that are representative of the observations. To achieve this, lidar measurements averaged on 15 min are inverted to provide this reference dataset, which is then used as input to the end–to–end model. In a second step, we apply Eq. 1 to simulate the vertical lidar signals for each channel, relying on atmospheric parameters that also act as inputs to the model (Fig. 8). The molecular contribution is simulated through a climatological model as in Chazette (2003). The lidar instrumental constant is obtained by isolating a low–noise part of the actual signal, typically between 1000 and 1500 m, and

315   fitting it with the simulated signal.

We generated a total of $n = 400$ noise instances for each channel to ensure a normal noise distribution with at least one standard deviation. The noise level, represented by the standard deviation of the noise, is scaled based on real lidar profiles S_i averaged over a 15 min duration. Subsequently, we multiply the ratio of the two simulated channels by the calibration coefficient K_o to recover r_H, which is then compared to $r_{H,ref}$. Finally, we evaluate the error budget by calculating both the mean–bias (MB)

320   and the root mean square deviation (RMSD), according to the following relationships:

$$MB = \frac{1}{n} \cdot \sum_{k=1}^{n} \left[r_H(k) - r_{H,ref}(k) \right] \tag{10}$$

and

$$RMSD = \sqrt{\frac{1}{n} \cdot \sum_{k=1}^{n} \left[(r_H(k) - r_{H,ref}(k) \right]^2} \tag{11}$$

Different other bias sources and their impact are described in Totems et al. (2021). They are not considered here, as they are negligible compared to shot noise. The other sources of error due to calibration and atmospheric transmission have been computed in subsection 4.3, using the measurements performed during the field experiment.
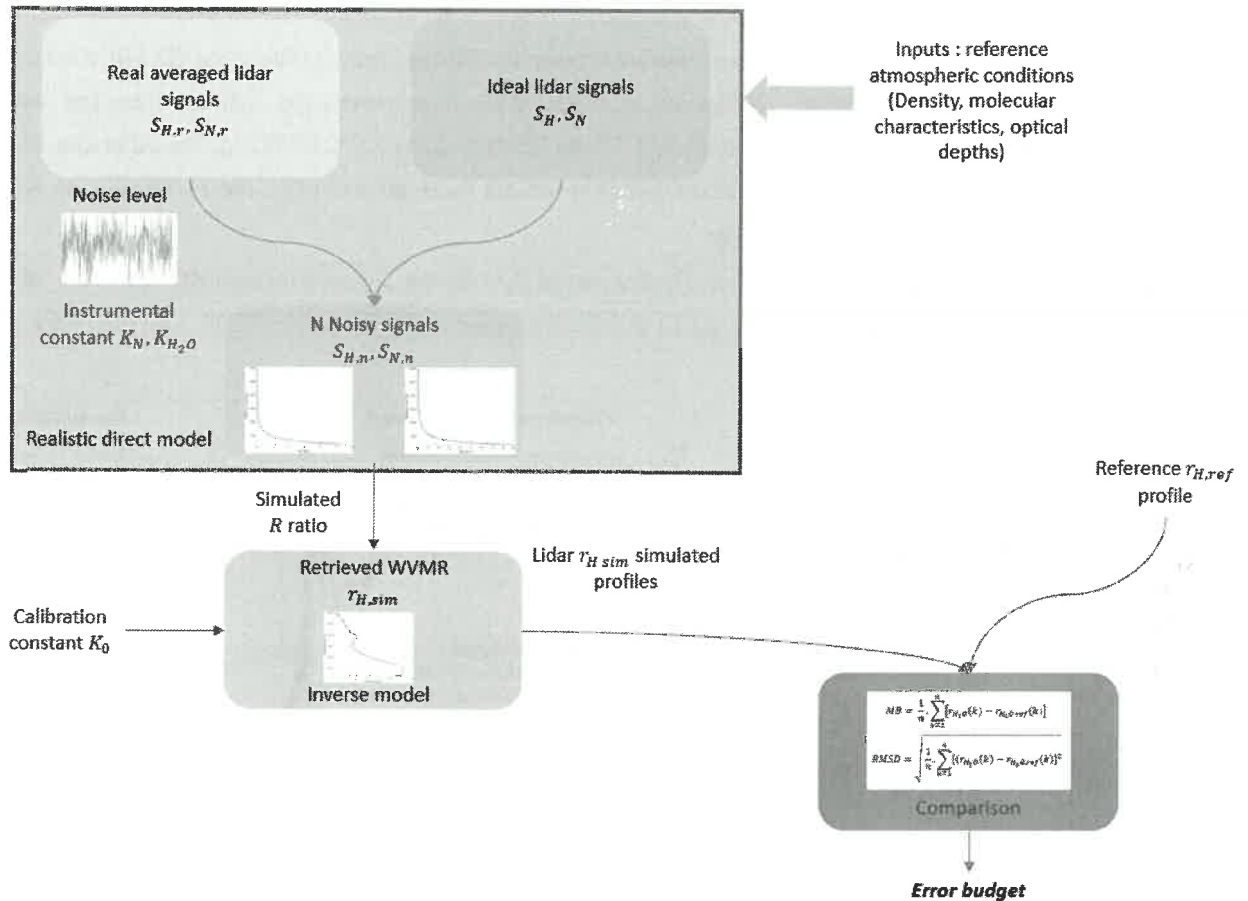
325

Figure 7. Diagram of the direct/inverse algorithm methodology for water vapor Raman lidar. The grey box includes several coloured boxes to describe the direct lidar model. Ideal lidar signals (S_H and S_N) are generated from the lidar equation considering atmospheric parameters (blue box) and compared to real lidar signals ($S_{H,r}$ and $S_{N,r}$) acquired with the lidars (yellow box) to estimate noise levels and system constants (K_{H2O} and K_N). This creates a certain number of noisy signals ($S_{H,n}$ and $S_{N,n}$) representing

330 **the lidar signals (green box). From water vapor and dinitrogen lidar signals, we estimate the WVMR ($r_{H,sim}$) using the ratio (R) of the two channels and applying the calibration constant Ko. This operation is the inverse model (orange box). The simulated WVMR is then compared to the reference WVMR ($r_{H,ref}$) to estimate the error budget (blue box).**

4   Results

4.1   Calibration

335   For each lidar, Fig. 8 (a, b, c) shows examples of periods during which the lidar-derived WVMR corresponds to the weather station-derived WVMR. Lidar measurements were extracted at 200 m a.g.l. The standard deviation on the data due to both instrumental noise and atmospheric variability is also indicated by red coloured areas. When time evolutions are close together, this corresponds to periods when dynamical vertical mixing homogenizes the lower troposphere. For these periods, the scatter

[handwritten note: Please, re-plot all vertical profiles accordingly by refering to above mean sea level (a.m.s.l) heights, to provide a standard for all profiles]

17

plots between the in situ and remote sensing measurements are shown in Fig. 8 (d, e, f) for each lidar site during the WaLiNeAs

340 campaign. These scatter plots show a good correlation between the lidar and the weather station WVMR when the atmosphere is well–mixed, with determination coefficients exceeding 0.90. The relative gap between lidar and weather station measurements is on average 4.4 % for HORUS–1, 2.7 % for HORUS–2 and 3.8 % for WALI. The calibration constants found for each lidar are 108 for WALI, 112 for HORUS–1 (T1) and 205 for HORUS–2 (T1). The T1/T2 ratios for HORUS–1 and HORUS–2 are 1.436 and 1.092, respectively.

345 Fig. 9 shows the scatter plots between the two N_2–Raman and H_2O–Raman channels for each HORUS lidars. All scatter plots highlight a linear relationship between T1 and T2 with determination coefficients (R^2) of 0.99. This shows that therefore, the cross–calibration method is relevant.
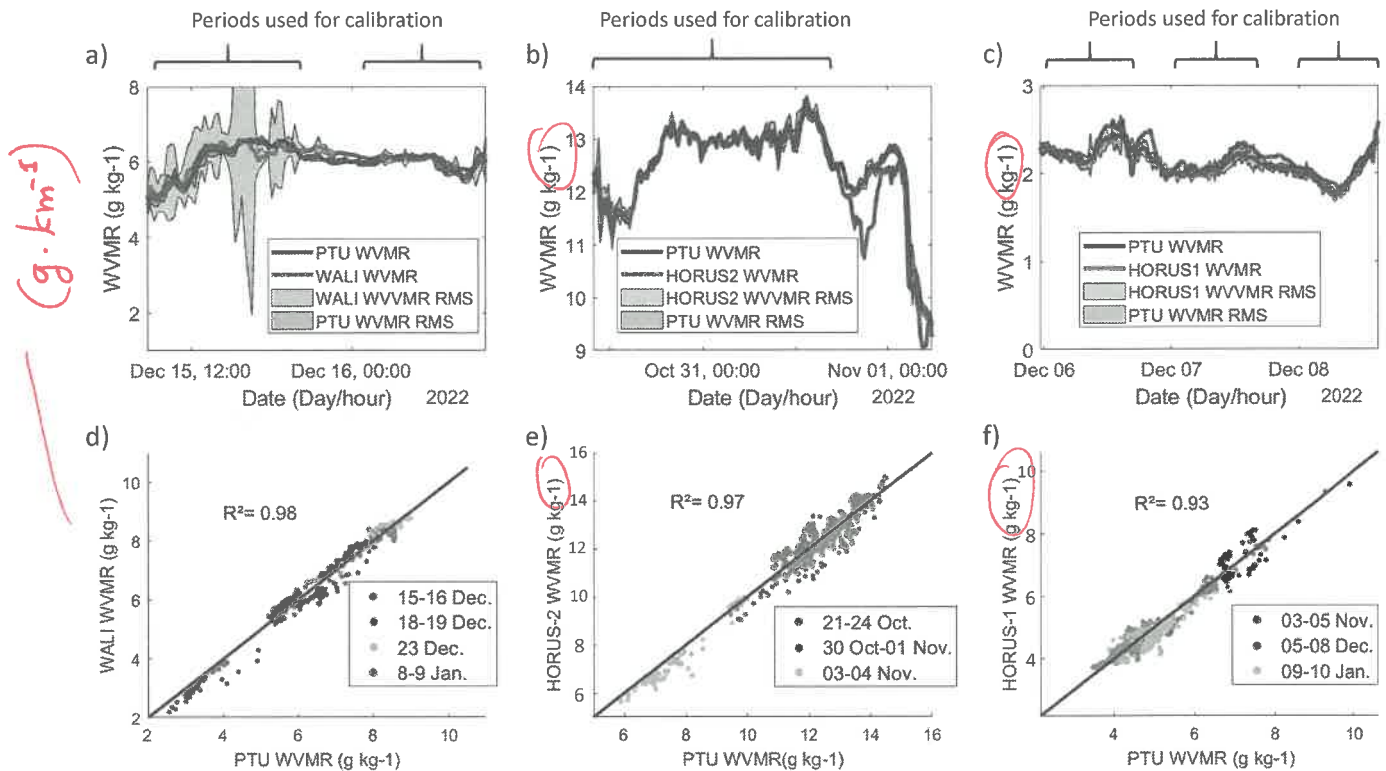
Figure 8. Examples of time series during which lidars and ground–based weather stations WVMR correspond to each other are given on figures a) for WALI, b) for HORUS-2 and c) for HORUS-1. Root mean square (RMS) deviations on WVMR are represented by the blue and red shaded areas. Scatter plots between the weather station and lidars WVMR for periods during which WVMR correspond to each other are shown in figures d) for WALI, e) for HORUS-2 and f) for HORUS-1. The different periods are represented by dots of different colours. Determination coefficients R² have been plotted on figures d, e, f to show the correlation between lidar and weather station observations. The temporal resolution is 15 min and lidar profiles measured around 200 m a.g.l.
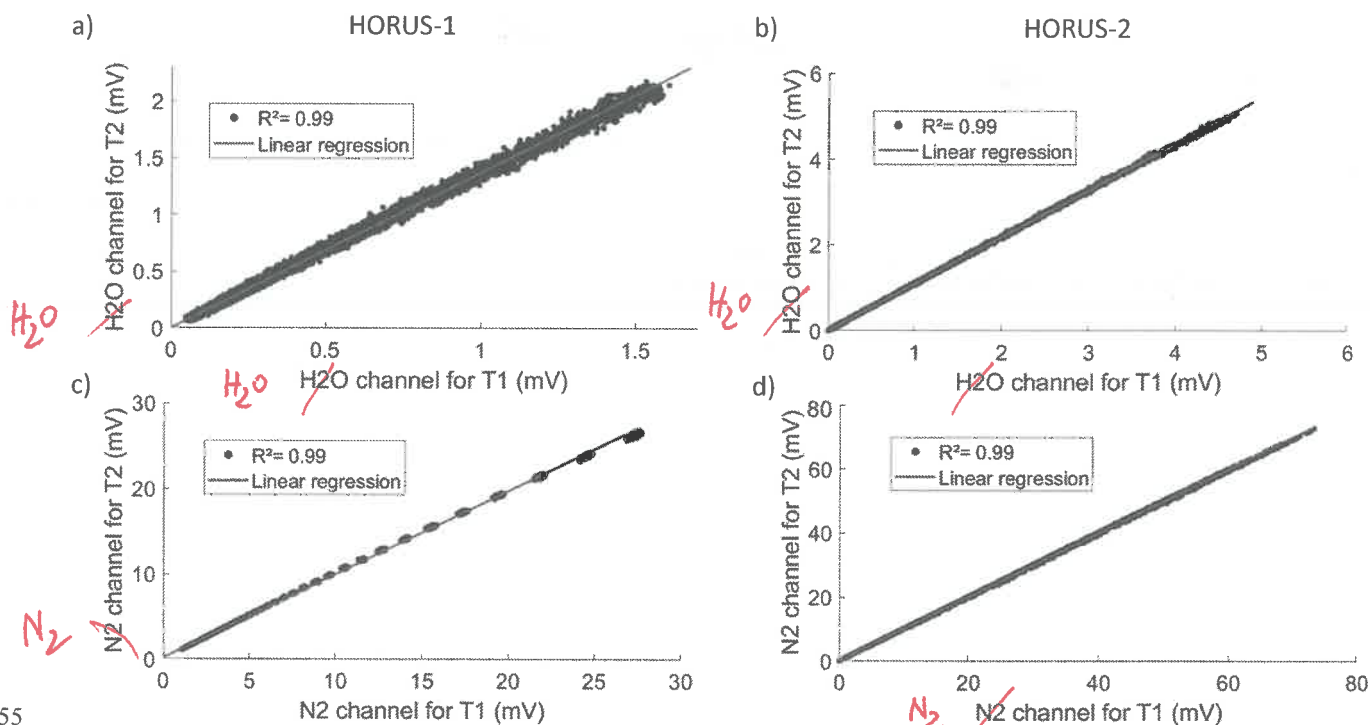
Figure 9. Dinitrogen (N_2) and Water vapor (H_2O) Raman channels scatter plots between the two telescopes (T1, T2) for a,c)HORUS–1 and b,d) HORUS–2 lidars. For HORUS–1, data were taken on 05 January 2023 between 200 m and 1000 m in altitude (a.g.l.) and between 00:00 UTC and 06:45 UTC. For HORUS–2, data were taken on 24 October 2022, between 200 m and 1000 m in altitude (a.g.l.) and between 00:00 UTC and 06:45 UTC. Initial signals corrected for gain and sky background with a time resolution of approximately 1 min and a vertical resolution of 15 m were used. The regression lines (red lines) and coefficients of determination (R^2) are plotted on each figure.

4.2 Example of WVMR temporal series

Examples of WVMR temporal series for each lidar are given in Fig.10. These profiles were obtained after processing the raw lidar signals as described in Section 3 with a vertical resolution of 100 m and a temporal resolution of 30 min.

HORUS-1 and WALI operated simultaneously for several days during the campaign. This allows to compare their sampling of the water vapor column. WVMR retrieved from HORUS–1 (Fig. 11a) and WALI (Fig. 11b) contain similarities due to the geographical location of the lidar sites (Fig. 2). However, HORUS–1 was in the Rhône delta, in the Camargue region, and sampled air masses that are influenced by the Mistral wind flowing down the Rhône valley. These can recirculate over the Mediterranean Sea to reach the Bay of Cannes. During daytime, both sites are subject to sea breezes, which can travel dozens of kilometres inland along the Rhône delta and even along the Durance River. In the case of the Cannes site, this breeze will help to carry humid air masses aloft over the mountains bordering the coast. It is worth noting that there is a significant contrast between day and night in the lower layers, below 2 km a.m.s.l., linked to the breeze cycle. Note that the range limitation of the lidar profiles during the night of 21–22 November 2022 corresponds to the presence of clouds.

19

Fig. 10c shows the evolution of WVMR [*vertical*] profiles over Toulouse from 20 to 25 August 2023. That period appears very humid,

375   with r_H values often exceeding 10 g kg^{-1} in the atmospheric [*planetary*] boundary layer [*(PBL)*]. Such values could be encountered in tropical latitudes. It should be noted that [*also*] the period studied corresponds to heatwave conditions, with daytime temperatures reaching 43°C at the Toulouse site. This shows the value of this dataset for studying not only extreme precipitation, but also extreme temperatures. These two types of extreme meteorological situations are among the main threats posed by climate change [*(citation)*]

[handwritten marginal note: ✳ provide 2 relevant citations.]

20
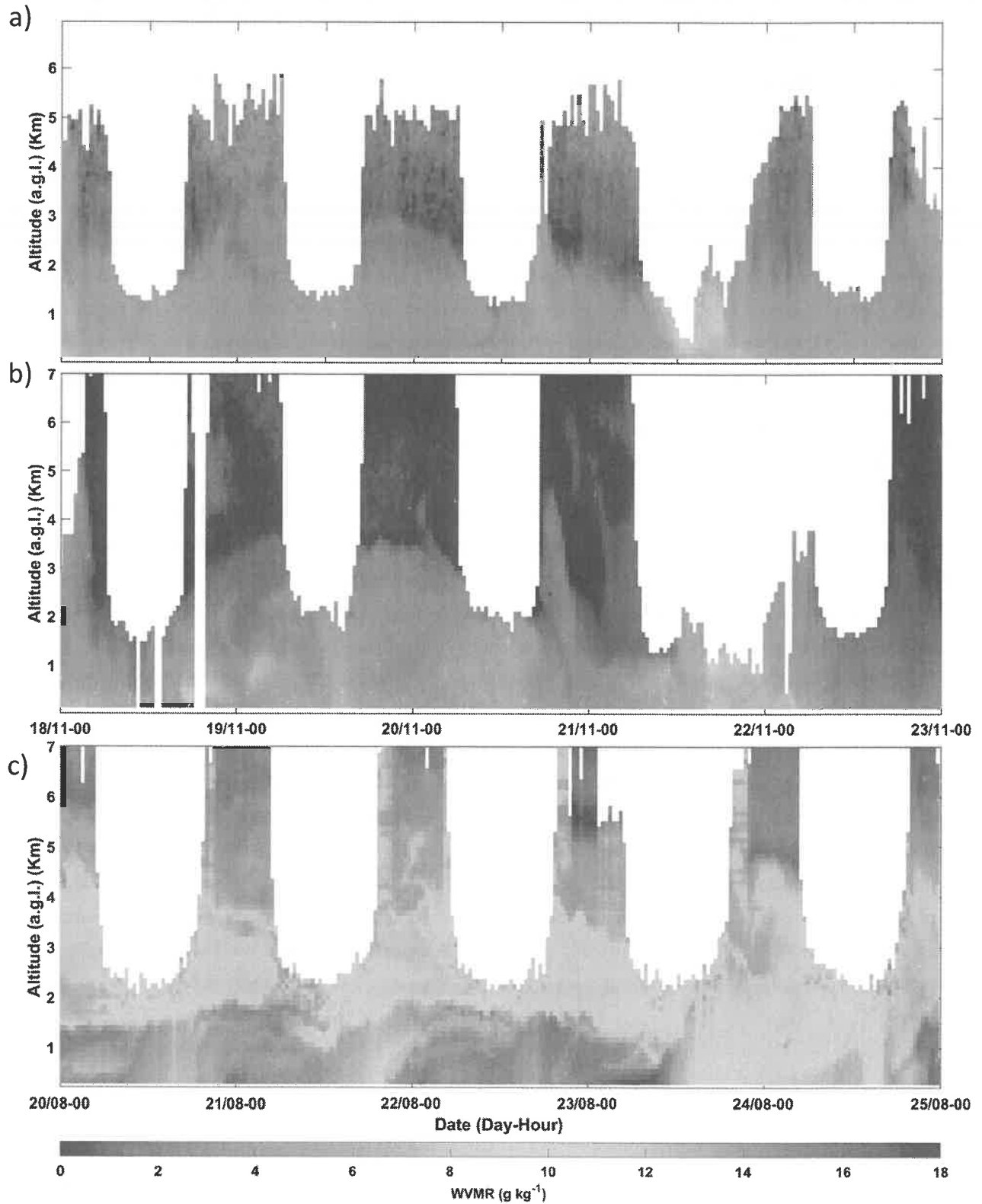
380   **Figure 10. Temporal evolution of vertical profiles of the WVMR derived from the Raman lidar as a function of altitude a.g.l. for a) HORUS–1, b) WALI and c) HORUS–2. Vertical resolution is 100 m and time resolution is 30 min. The white area corresponds to low quality WVMR retrieval.**

4.3 Errors on the lidar-derived WVMR profiles

4.3.1 Shot noise contribution

385   To estimate the shot noise contribution to lidar measurement, we apply the Monte Carlo approach explained in subsection 3.4. To characterize this uncertainty properly, we first need to estimate the Signal to Noise Ratio (SNR). This is made easier during nighttime, when the photon counting mode is activated. As explained in Measures (1984), the standard deviation (Eq. 9) is then equal to the square root of the returned lidar signal. Using the Monte Carlo approach, the SNR has been estimated for each lidar with profiles averaged over 15 min and a vertical resolution of 100 m as detailed in Section 4. Given the lidars

390   characteristics in Table 2, 15 min represents a total of 90000 laser shots averaged for HORUS–2, and 18000 for WALI and HORUS–1. The SNR is thus proportional to the square root of the total number of shots. During daytime we have also assessed the shot noise contribution to the error by estimating the SNR, which also takes solar luminance into account as in Measures (1984). Moreover, unlike the night detection, the day detection scheme is performed in analogue mode and we must account for the statistical variation in the detector gains.

395   Given that the signal level of the N_2–Raman channel is about 50 times greater than that of the H_2O–Raman channel, we can conside that the uncertainty on the WVMR (Eq 9) is inversely proportional to the SNR of the H_2O–Raman channel. This rough approximation assumes that signal noise is dominant over other noise sources, which is indeed the case. The evolution of the error ε_H in g kg^{-1} as a function of the water vapor channel signal to noise ratio SNR_H is then linear in a logarithm scale as shown in Fig. 11a and can be used to directly determine which SNR_H corresponds to the relative uncertainty ε_H.

400   The uncertainty on the WVMR due to the shot noise is plotted during nighttime in Fig. 11b–d–f and daytime in Fig. 11c–e–g, with the reference water vapor profile used in the model represented as black solid lines. These profiles are derived from measurements taken at contrasting periods: during the day of 20 November 2022 (00:00 UTC for nighttime profiles and 10:00 UTC for daytime profiles) for WALI and HORUS–1, and during the day of 2 August 2023 (00:00 UTC for nighttime profile and 08:00 for daytime profile) for HORUS–2. As expected, the RMSD values are higher for HORUS–1 due to its lower laser

405   emission energy. The values of the signal noise contribution to the total error are shown in Table 4 for each lidar system. Note that calculation of the standard deviation of vertical WVMR profiles over a time interval includes both signal noise and natural atmospheric variability. It is worth noting that atmospheric variability is a natural process and not a form of noise derived from the instrument. This natural variability is strongly influenced by the thermal stability of the troposphere due to convection and air mass advections. Generally, it varies more during daytime, but this may depend on the geographical

410   location.
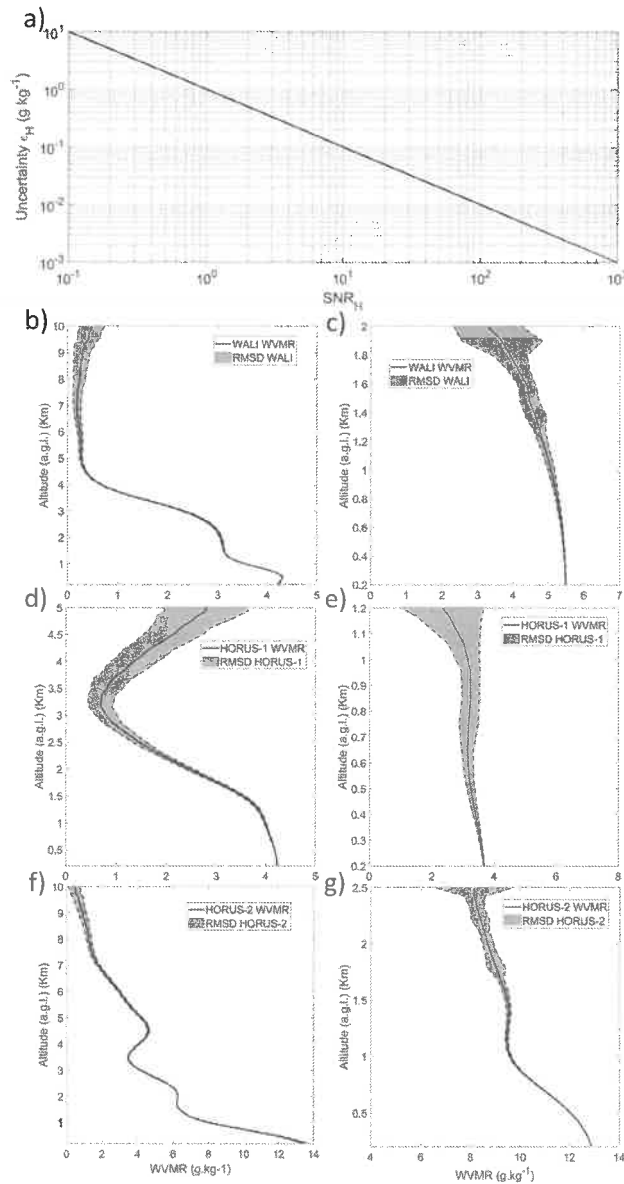
Figure 11. Uncertainty variation (ε_H) as a function of the water vapor channel Signal to Noise Ratio (SNR_H) is plotted in Fig. a). Lidar WVMR profiles as a function of altitude are plotted as solid black lines and their associated RMSD are represented by the grey shaded area during nighttime for b) WALI, d) HORUS-2, f) HORUS-1 and daytime for c) WALI, e) HORUS-1, g) HORUS-2. HORUS–1 and WALI nighttime profiles were taken on the night of 20 November 2022 at 00:00 UTC over Grau du Roi and Cannes, respectively. Daytime profiles were taken during the day of 20 November 2022 at 10:00 UTC. HORUS–2 nighttime profile was taken on the night of 2 August 2023 at 00:00 UTC over Toulouse. Daytime profile was taken during the day of 2 August 2023 at 8:00 UTC. The vertical resolution of these profiles is 100 m, and each profile has been averaged over 15 min.

23

Earth System
Science
Data

4.3.2 Relevance of the calibration

To calibrate the lidars, we used a ground-based weather station as described in subsection 3.3. To ensure that calibration is consistent with a conventional radiosonde calibration approach, we were able to compare HORUS–1 and HORUS–2 vertical profiles with specific radiosoundings. In Fig. 12a, a cross–comparison has been performed between HORUS–1 and radiosounding measurements from Nimes, 45 km north/north-east of the lidar's location. On the profiles shown in Fig. 12a, we are limited in altitude by the lidars' SNR, impacted by the presence of clouds above 4.5 km a.g.l. The cross–comparison carried out on the radiosounding of 23:15 UTC shows a similar behaviour against the altitude with mean differences of 0.58 g kg^{-1} on the entire profile. This is slightly higher than what we would have expected from the previous uncertainty study (cf. section 4.3.1) which suggested a mean RMSD of 0.15 g kg^{-1}. This can be explained by the natural variability of the atmosphere between the two sites used for the comparison, and the fact that the radiosonde drifts over several tens of kilometres between the ground and 4.5 km altitude. As shown in Fig. 12b, the differences between radiosounding and lidar data for HORUS–1 are significantly higher than those for HORUS–2. Indeed, lidar measurements obtained during the Toulouse campaign were compared with a spatiotemporal coincident radiosounding performed by Météo–France on 2 August 2023, 00:00 UTC. The two types of measurements match between ground level and 10 km a.g.l., close to the tropopause. The cross–comparison gives differences of 0.48 g kg^{-1} below 3 km a.g.l and of 0.28 g kg^{-1} above. Note that standard deviation for radiosounding WVMR has been estimated according to Di Girolamo et al. (2020) and reported in Fig. 12a–b in blue area. VAISALA® (https://www.vaisala.com/fr/, last access 13 February 2023) manufacturer information on the uncertainty affecting radiosoundings humidity measurements and translated into WVMR is specified to not exceed 0.20–0.25 g.kg^{-1} for temperatures higher than -40 °C.

24
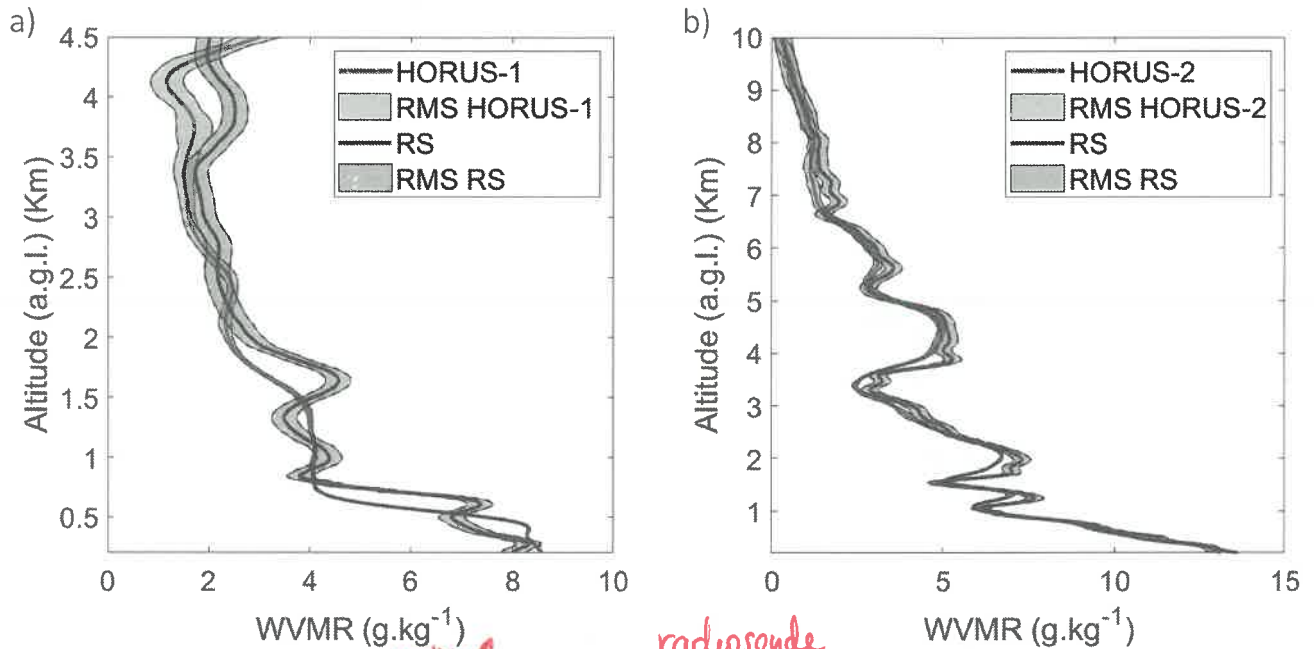
Figure 12. Cross-comparison of WVMR profiles derived from a) Horus–1 and a radiosonde (RS) during the night of 12/11/2022 (23:15 UTC). The lidar site was located in Grau du Roi and the Radiosonde in Nîmes. and b) Horus–2 and a RS during the night of 08/02/2023 (00:00 UTC) over the Météo-France site at Toulouse. Lidars and RS WVMR profiles are plotted in red lines and blue lines, respectively. The vertical resolution of profiles is 15 m. The coloured area gives the standard deviation around the mean value.

4.3.3 Molecules and aerosols contribution

The molecules and aerosols contributions on the uncertainty are low compared to the other error sources. The molecular contribution has been corrected in the final datasets using the outputs of the European Centre for Medium–Range Weather Forecasts (ECMWF) reanalysis (http://www.ecmwf.int/, last access 5 January 2024) ERA5. The residual WVMR uncertainty is less than 0.01 g kg^{-1}.

Aerosols contribution remains low, although higher than that linked to molecular transmission. The three lidars were set up near the sea (Fig.2) during the first part of WaLiNeAs campaign without major pollution or biomass burning aerosol events. The aerosol optical thickness (AOT) is lower than 0.15 at 355 nm except during two Saharan dust events in October 2022 (AERONET site of Toulon, https://aeronet.gsfc.nasa.gov/, last access 5 January 2024). The majority of aerosols present in the atmosphere are of marine origin with an Ångström exponent of ~1 in the UV spectral domain. Applying Eq. 8, the aerosol correction then changes r_H by only 0.7 % compared to around 5 % for the molecular transmission if it is not corrected. In the case of desert aerosol events mixed with marine aerosols, the Ångström exponent is ~0.9 and lower, so even with an AOT of the order of 0.35, they induce a relative bias on r_H of less than 1.6% (less than 0.1 g kg^{-1} in the dust layer). During the experiment over Toulouse, the AERONET station located in the site of Météo–France (https://aeronet.gsfc.nasa.gov/, last access 5 January 2024) highlights high values of AOT or Ångström exponent. Fig. 13 shows the bidimensional histogram of the AOT at 387 nm

460   and Ångström exponent between 387 and 440 nm enabling us to identify which type of aerosols were present in the atmospheric column. The most probable cases give an aerosol contribution to the lidar signal equal to 0.6 % in the case of polluted dust aerosols ($AOT = 0.15$ and $A = 0.8$); 2.3 % in the case of dust aerosols ($AOT = 1.2$ and $A = 0.4$) and 2.1 % for pollution aerosols ($AOT = 0.3$ and $A = 1.5$). The impact of the last two cases may be considered, but it should be noted that the temporal occurrence of these cases is less than 5% and induced an uncertainty lower than 0.12 g kg^{-1} on the WVMR.
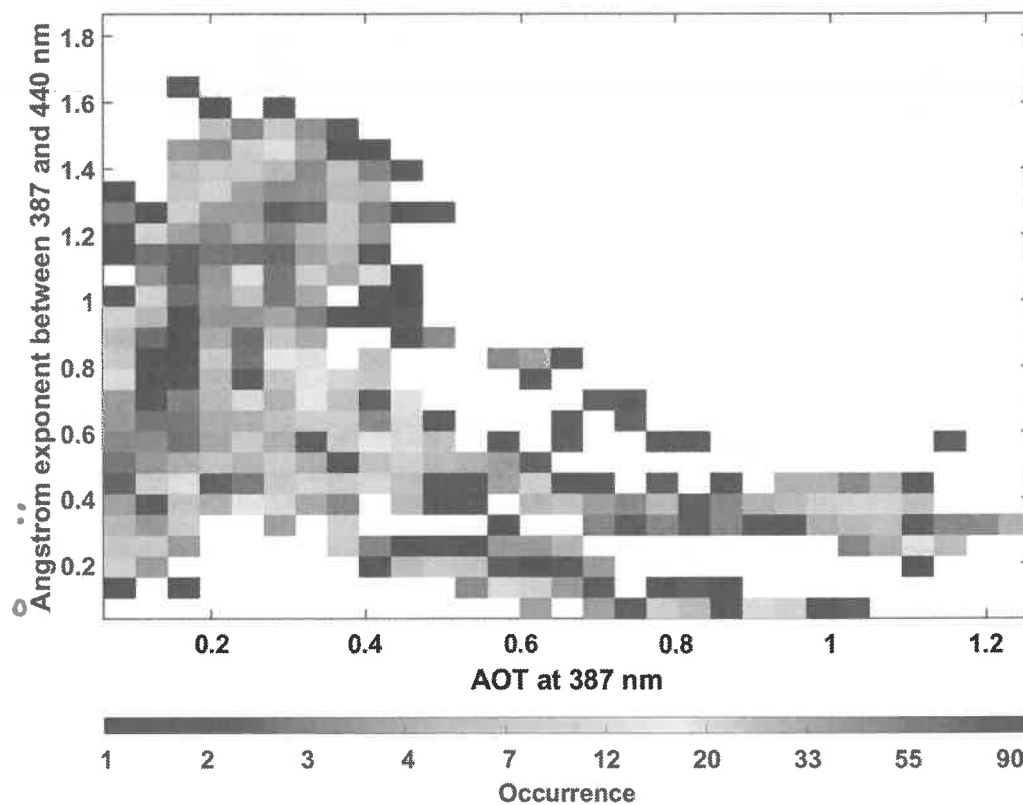
465   **Figure 13. Occurrence of both Ångström exponent and aerosol optical thickness (AOT) given by the AERONET photometer network in Toulouse. Data have been taken between 31 May and 25 September 2023.**

4.3.4 Error sources synthesis

For all three lidars, the contributions of the main bias and uncertainties sources are shown in Table 3 and Table 4, respectively. The bias that has the greatest impact on the signal is that of calibration, which depends mainly on both the HV variations and 470   the uncertainty linked to the meteorological probe. As expected, the higher RMSD are encountered during daytime and limit the altitude range of lidars. The higher the energy per laser shot, the better the precision is. It is also worth noting that the statistical uncertainties (RMSD) may vary based on the presence of more or less moist air masses in the lower/ middle

troposphere and are higher during daytime. However, they can be mitigated by extending the integration time to create a database of mean profiles.

475 **Table 3. Review of the biases impacting lidar measurements.**

Bias source		Bias value
Molecular contribution		< 0.1 %
Aerosols contribution		< 0.7 %
High voltage variation		1 − 2 %
Meteorological probe uncertainty		0.2 g kg⁻¹
Calibration	WALI	3.8 %
	HORUS−1	4.4 %
	HORUS−2	2.7 %

Table 4. Typical RMSD due to i) the shot noise and ii) the shot noise and the atmospheric variability (Total) during both nighttime and daytime. Uncertainties are given for different altitude ranges and for ach lidar (WALI, HORUS−1 and HORUS−2) accounting for the specific meteorology of each ground-based station during WaLiNeAs. Vertical and temporal resolution of lidar profiles
480 **considered are 100 m and 15 min, respectively.**

Lidar / RMSD		WALI	HORUS−1	HORUS−2
Nighttime	Shot noise	~ 0.01 − 0.03 g kg⁻¹ (0 − 2 km) ~ 0.03 − 0.05 g kg⁻¹ (2 − 5 km) ~ 0.1 − 0.3 g kg⁻¹ (5 − 10 km)	~ 0 − 0.07 g kg⁻¹ (0 −2 km) ~ 0.07 − 0.4 g kg⁻¹ (2 − 4 km) ~ 0.4 − 1 g kg⁻¹ (4 − 5.5 km)	~ 0 − 0.05 g kg⁻¹ (0 − 5 km) ~ 0.05 − 0.2 g kg⁻¹ (5 − 10 km)
	Total	~ 0.03 − 0.05 g kg⁻¹ (0 − 2 km) ~ 0.05 g kg⁻¹ (2 − 5 km) ~ 0.1 − 0.3 g kg⁻¹ (5 − 10 km)	~ 0 − 0.1 g kg⁻¹ (0 −2 km) ~ 0.1 − 0.4 g kg⁻¹ (2 − 4 km) ~ 0.4 − 1 g kg⁻¹ (4 − 5.5 km)	~ 0 − 0.1 g kg⁻¹ (0 − 5 km) ~ 0.1 − 0.2 g kg⁻¹ (5 − 10 km)
Daytime	Shot noise	~ 0 − 0.2 g kg⁻¹ (0 −1.5 km) ~ 0.2 − 1 g kg⁻¹ (1.5 − 2 km)	~ 0 − 0.3 g kg⁻¹ (0 −1 km) ~ 0.3 − 1 g kg⁻¹ (1 − 1.5 km)	~ 0 − 0.1 g kg⁻¹ (0 −1.6 km) ~ 0.1 − 0.3 g kg⁻¹ (1.6 − 2.5 km)
	Total	~ 0 − 0.4 g kg⁻¹ (0 −1.5 km) ~ 0.4 − 1 g kg⁻¹ (1.5 − 2 km)	~ 0 − 0.4 g kg⁻¹ (0 −1 km) ~ 0.4 − 1 g kg⁻¹ (1 − 1.5 km)	~ 0 − 0.4 g kg⁻¹ (0 −1.6 km) ~ 0.4 − 2 g kg⁻¹ (1.6 − 2.5 km)

Open Access Earth System Science Data Discussions

5 Data format and quality flag

5.1 Data Format

For each lidar site, lidar and weather stations data are available within the AERIS database as NetCDF files (version 4) via https://metclim-lidars.aeris-data.fr/ (last access 6 January 2024). For each site, two NetCDF files have been created

485 corresponding to time resolution of 30 min or 15 min. The vertical resolution of WVMR profiles is 100 m for all lidar profiles. Daily lidar data availabilities are given in Fig. 6 and measurement configurations of each lidar are described in Table 1. Additional general information is given in Table 5 and Appendix A describes Parameters available in each NETcdf file. The datasets published on the AERIS database (https://en.aeris-data.fr/, last access 6 January 2024) are freely available. The digital object identifier (DOI) for all data is https://doi.org/10.25326/537. The typical sizes for different NetCDF files are between

490 3.5 Mo and 20 Mo for files with a time resolution of 30 min, and between 7 Mo and 40 Mo for files with a time resolution of 15 min.

Table 5. General data file description

NetCDF General information

Dataset name format:	WaLiNeAs_lidar-site_lidar-name_start-date_end-date_time-resolution_file-version	
DOI:	https://doi.org/10.25326/537	
Date created:	2023 – xx – xx	
Contact:	Patrick Chazette – LSCE – patrick.chazette@lsce.ipsl.fr	
Period:	Date begin:	yyyy – mm – dd
	Date end:	yyyy – mm – dd
Project:	WaLiNeAs	

5.2 Data quality

WVMR products include the two binary quality indicators Flags and GAB in the dataset to provide information on data relevance and quality. The first quality indicator Flags is coded with "1" and "0" over 4 bits. This indicator is defined in Table 6. For each altitude of WMVR profiles, the Flags indicates in which range the RMSD on WVMR lies. The different ranges, defined in Table 6, provide information on the statistical precision of the measurement. The minimum threshold is set at 0.4 g kg^{-1} to fulfil the World Meteorological Organization (WMO) requirements for atmospheric water vapor measurement accuracy. In order to simplify its re–reading by users, the indicator is converted into decimal numbers in NetCDF files. Before being used, it must be converted back to binary. For example, the decimal number 15 corresponds to the binary number "1111". The GAB parameter takes the value of either 1 or NaN (Not a Number) for each altitude level of the WVMR temporal evolution. The value 1 indicates data with a good signal, little noise, and little error while the value NaN accounts for noisy signals with a high error value constituting a poor quality signal. The threshold for poor quality data has been set empirically when the SNR for the water vapor channel is less than 1 and the RMSD for the WVMR is greater than 0.5.

Table 6. Flags and GAB quality indicators description. B1, B2, B3 and B4 are the parameter identification bits. Flags values are calculated from lidar profiles RMSD, named std_WVMR in the table and in the database. The value for poor data quality GAB and Flags 0000 is NaN.

Flags	B1	B2	B3	B4
NaN	0	0	0	0
Std_WVMR $\geq$ 2 g kg^{-1}	0	0	0	1
1 g.kg^{-1} < Std_WVMR < 2 g kg^{-1}	0	0	1	1
0.4 g.kg^{-1} < Std_WVMR $\leq$ 1 g kg^{-1}	0	1	1	1
Std_WVMR $\leq$ 0.4 g kg^{-1}	1	1	1	1

GAB	B1
Poor data quality	NaN
Good data quality	1

6 Conclusion

The WaLiNeAs project aims to predict extreme precipitation event by measuring WVMR at high spatio-temporal resolution in the low troposphere using Raman lidar technology and by investigating the impact of its variability in numerical weather prediction models' forecasts. It is the only instrument currently available to achieve the required vertical and temporal resolutions to improve meteorological forecasting performed by the new generation of mesoscale models such as AROME. As part of the main WaLiNeAs field measurement campaign, the three lidars which constituted the French component of the

project continuously measured WVMR profiles over southern France during fall and winter 2022-2023, i.e., the season most propitious to HPEs in the Western Mediterranean. A second campaign was conducted near Toulouse, France, between June and September 2023 during which WVMR variability associated with summer storms was documented with a single lidar. All

520   data have been processed to retrieve the vertical profiles of WVMR. The uncertainties have been quantified for various configurations of measurements, during nighttime and daytime, as well as for different meteorological situations. They agree with the recommendations given by the WMO, with an absolute accuracy on WVMR lesser than 0.4 g kg^{-1}. On cloudless nights, 15 and 30 min averages provide the accuracy required to constrain mesoscale modelling between the ground and the tropopause (~10 km). During the day, the range is greatly reduced, and the lidar used gives access to altitudes relative to the

525   ground of the order of 2 km above ground. Final datasets include WVMR profiles and parameters measured by the in situ weather stations associated with each lidar. Data quality assessment parameters are also provided. All datasets are available as NetCDF files and can be freely downloaded from the AERIS database (https://doi.org/10.25326/537, last access 13 February 2024). Vertical lidar profiles allow the measurement of the water vapor content in the atmosphere with sufficient spatio-temporal resolution to study the different processes that can occur in the air column, mainly in the atmospheric boundary layer.

530   Lidar-derived WVMR also allows to identify and study the initial conditions that can lead to extreme precipitation events. Given the temporal (15 and 30 min) and vertical (100 m) resolution of the lidar profiles, the assimilation of lidar data into mesoscale models such as AROME will improve the models' accuracy in predicting which areas will be affected by extreme weather phenomena. Indeed, ground-based lidar measurements fill a gap in observations of the lower troposphere, between the ground and ~2 km in altitude. They will also allow the study and understanding of different weather periods such as dust

535   events, heatwaves, or HPEs when different atmospheric processes occur in the atmosphere, resulting in high atmospheric water vapor content.

Data availability. Data are freely available from https://doi.org/10.25326/537 (Chazette et al., 2023, last access 13 February 2024).

540

Author contributions. FL took part in the field campaigns in autumn 2022 and summer 2023, calibrated the data, wrote the paper, and participated to the creation of the database; PC coordinated the two field campaigns, wrote the paper, and created the database; JT and JL took part in the field campaigns in autumn 2022; CF coordinated the WaLiNeAs project and took part in the field campaigns in autumn 2022. All authors contributed the preparation of the campaigns and to the proofreading of the

545   paper.

Competing interests. The authors declare that they have no conflict of interest.

References

Ansmann, A., Riebesell, M., Wandinger, U., Weitkamp, C., Voss, E., Lahmann, W., and Michaelis, W.: Combined raman elastic-backscatter LIDAR for vertical profiling of moisture, aerosol extinction, backscatter, and LIDAR ratio, Applied Physics B Photophysics and Laser Chemistry, 55, https://doi.org/10.1007/BF00348608, 1992.

Chazette, P.: The monsoon aerosol extinction properties at Goa during INDOEX as measured with lidar, Journal of Geophysical Research: Atmospheres, 108, https://doi.org/10.1029/2002jd002074, 2003.

Chazette, P., Bocquet, M., Royer, P., Winiarek, V., Raut, J. C., Labazuy, P., Gouhier, M., Lardier, M., and Cariou, J. P.: Eyjafjallajökull ash concentrations derived from both lidar and modeling, Journal of Geophysical Research Atmospheres, 117, https://doi.org/10.1029/2011JD015755, 2012a.

Chazette, P., Dabas, A., Sanak, J., Lardier, M., and Royer, P.: French airborne lidar measurements for Eyjafjallajökull ash plume survey, Atmos Chem Phys, 12, https://doi.org/10.5194/acp-12-7059-2012, 2012b.

Chazette, P., Marnas, F., Totems, J., and Shang, X.: Comparison of IASI water vapor retrieval with H2O-Raman lidar in the framework of the Mediterranean HyMeX and ChArMEx programs, Atmos Chem Phys, 14, https://doi.org/10.5194/acp-14-9583-2014, 2014a.

Chazette, P., Marnas, F., and Totems, J.: The mobile Water vapor Aerosol Raman LIdar and its implication in the framework of the HyMeX and ChArMEx programs: Application to a dust transport process, Atmos Meas Tech, 7, 1629–1647, https://doi.org/10.5194/amt-7-1629-2014, 2014b.

Chazette, P., Flamant, C., Raut, J. C., Totems, J., and Shang, X.: Tropical moisture enriched storm tracks over the Mediterranean and their link with intense rainfall in the Cevennes-Vivarais area during HyMeX, Quarterly Journal of the Royal Meteorological Society, 142, https://doi.org/10.1002/qj.2674, 2016.

Chazette, P., Totems, J., and Shang, X.: Atmospheric aerosol variability above the Paris Area during the 2015 heat wave - Comparison with the 2003 and 2006 heat waves, Atmos Environ, 170, https://doi.org/10.1016/j.atmosenv.2017.09.055, 2017.

Chazette, P., Laly, F., Totems, J. & Lagarrigue, J. (2023). WaLiNeAs France. [dataset]. Aeris. https://doi.org/10.25326/537

Dettinger, M.: Climate change, atmospheric rivers, and floods in California - a multimodel analysis of storm frequency and magnitude changes, J Am Water Resour Assoc, 47, https://doi.org/10.1111/j.1752-1688.2011.00546.x, 2011.

Drobinski, P., Ducrocq, V., Alpert, P., Anagnostou, E., Béranger, K., Borga, M., Braud, I., Chanzy, A., Davolio, S., Delrieu, G., Estournel, C., Filali Boubrahmi, N., Font, J., Grubišić, V., Gualdi, S., Homar, V., Ivančan-Picek, B., Kottmeier, C., Kotroni,

V., Lagouvardos, K., Lionello, P., Llasat, M. C., Ludwig, W., Lutoff, C., Mariotti, A., Richard, E., Romero, R., Rotunno, R., Roussot, O., Ruin, I., Somot, S., Taupier-Letage, I., Tintor, J., Uijlenhoet, R., and Wernli, H.: HYMEX: A 10-year

585   multidisciplinary program on the mediterranean water cycle, Bull Am Meteorol Soc, 95, https://doi.org/10.1175/BAMS-D-12-00242.1, 2014.

Ducrocq, V., Nuissier, O., Ricard, D., Lebeaupin, C., and Thouvenin, T.: A numerical study of three catastrophic precipitating events over southern France. II: Mesoscale triggering and stationarity factors, Quarterly Journal of the Royal Meteorological Society, 134, https://doi.org/10.1002/qj.199, 2008.

590   Ducrocq, V., Braud, I., Davolio, S., Ferretti, R., Flamant, C., Jansa, A., Kalthoff, N., Richard, E., Taupier-Letage, I., Ayral, P. A., Belamari, S., Berne, A., Borga, M., Boudevillain, B., Bock, O., Boichard, J. L., Bouin, M. N., Bousquet, O., Bouvier, C., Chiggiato, J., Cimini, D., Corsmeier, U., Coppola, L., Cocquerez, P., Defer, E., Delanoë, J., Di Girolamo, P., Doerenbecher, A., Drobinski, P., Dufournet, Y., Fourrié, N., Gourley, J. J., Labatut, L., Lambert, D., Le Coz, J., Marzano, F. S., Molinié, G., Montani, A., Nord, G., Nuret, M., Ramage, K., Rison, W., Roussot, O., Said, F., Schwarzenboeck, A., Testor, P., Van Baelen,

595   J., Vincendon, B., Aran, M., and Tamayo, J.: HyMeX-SOP1: The field campaign dedicated to heavy precipitation and flash flooding in the northwestern mediterranean, Bull Am Meteorol Soc, 95, https://doi.org/10.1175/BAMS-D-12-00244.1, 2014.

Duffourg, F. and Ducrocq, V.: Origin of the moisture feeding the heavy precipitating systems over southeastern France, Natural Hazards and Earth System Science, 11, https://doi.org/10.5194/nhess-11-1163-2011, 2011.

Duffourg, F., Nuissier, O., Ducrocq, V., Flamant, C., Chazette, P., Delanoë, J., Doerenbecher, A., Fourrié, N., Di Girolamo,

600   P., Lac, C., Legain, D., Martinet, M., Saïd, F., and Bock, O.: Offshore deep convection initiation and maintenance during the HyMeX IOP 16a heavy precipitation event, Quarterly Journal of the Royal Meteorological Society, 142, https://doi.org/10.1002/qj.2725, 2016.

Duffourg, F., Lee, K. O., Ducrocq, V., Flamant, C., Chazette, P., and Di Girolamo, P.: Role of moisture patterns in the backbuilding formation of HyMeX IOP13 heavy precipitation systems, Quarterly Journal of the Royal Meteorological Society,

605   144, https://doi.org/10.1002/qj.3201, 2018.

Flamant, C., Chazette, P., Caumont, O., Di Girolamo, P., Behrendt, A., Sicard, M., Totems, J., Lange, D., Fourrié, N., Brousseau, P., Augros, C., Baron, A., Cacciani, M., Comerón, A., De Rosa, B., Ducrocq, V., Genau, P., Labatut, L., Muñoz-Porcar, C., Rodríguez-Gómez, A., Summa, D., Thundathil, R., and Wulfmeyer, V.: A network of water vapor Raman lidars for improving heavy precipitation forecasting in southern France: introducing the WaLiNeAs initiative, Bulletin of

610   Atmospheric Science and Technology, 2, https://doi.org/10.1007/s42865-021-00037-6, 2021.

Flaounas, E., Kotroni, V., Lagouvardos, K., and Flaounas, I.: CycloTRACK (v1.0)-tracking winter extratropical cyclones based on relative vorticity: Sensitivity to data filtering and other relevant parameters, Geosci Model Dev, 7, https://doi.org/10.5194/gmd-7-1841-2014, 2014.

Fourrié, N., Nuret, M., Brousseau, P., Caumont, O., Doerenbecher, A., Wattrelot, E., Moll, P., Bénichou, H., Puech, D., Bock,

615   O., Bosser, P., Chazette, P., Flamant, C., Di Girolamo, P., Richard, E., and Saïd, F.: The AROME-WMED reanalyses of the