Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **482,** 206–218 (2019)
Advance Access publication 2018 October 1

doi:10.1093/mnras/sty2668

Forecasting water vapour above the sites of ESO's Very Large Telescope (VLT) and the Large Binocular Telescope (LBT)

Alessio Turchi ,[1]★ Elena Masciadri,[1] Florian Kerber[2] and Gianluca Martelloni[1,3]

[1]*INAF – Osservatorio Astrofisico di Arcetri, L.go E. Fermi 5, I-50125 Florence, Italy*
[2]*European Southern Observatory, Karl-Schwarzschild-Str.2, D-85748 Garching, Germany*
[3]*INSTM, Via della Lastruccia 3-13, I-50019 Sesto Fiorentino (Florence), Italy*

Accepted 2018 September 27. Received 2018 September 11; in original form 2018 August 10

ABSTRACT

Water vapour in the atmosphere is the main source of the atmospheric opacity in the infrared and sub-millimetric regimes and its value plays a critical role in observations done with instruments working at these wavelengths on ground-based telescopes. The scheduling of scientific observational programmes with instruments such as the VLT Imager and Spectrometer for mid-Infrared at Cerro Paranal and the Large Binocular Telescope Interferometer (LBTI) at Mount Graham would definitely benefit from the ability to forecast the atmospheric water vapour content. In this contribution, we present a study aiming at validating the performance of the non-hydrostatic mesoscale Meso-NH model in reliably predicting precipitable water vapour (PWV) above the two sites. For the VLT case we use, as a reference, measurements done with a Low Humidity and Temperature PROfiling radiometer (LHATPRO) that, since a few years, is operating routinely at the VLT. LHATPRO has been extensively validated on previous studies. We obtain excellent performances on forecasts performed with this model, including for the extremely low values of the PWV ($\leq$1 mm). For the LBTI case, we compare one solar year predictions obtained with the Meso-NH model with satellite estimates again obtaining an excellent agreement. This study represents a further step in validating outputs of atmospheric parameters forecasts from the ALTA Center, an operational and automatic forecast system conceived to support observations at LBT and LBTI.

Key words: site testing – atmospheric effects – methods: data analysis – methods: numerical.

1 INTRODUCTION

Astronomical observations in the infrared (IR) part of the spectrum, from ground-based telescopes, are mainly affected by the water vapour content[1] of the atmosphere, which is the main source of opacity in IR and sub-millimetre domains. Ground-based observations at these wavelengths can be done only in transparent windows of the spectrum without water vapour absorption lines. Atmospheric transmission depends on the distribution of atmospheric components, such as H_2O, O_2, CO_2, and other molecules. Water vapour is highly variable in the atmosphere both in altitude, time and geographic location. Atmospheric water content stratification is usually expressed in absolute humidity (AH), typically in g m^{-3}, which represents the total mass of water per volume of air. For astronomic purposes, the parameter of interest is the integral of AH over the whole vertical height, which is the precipitable water vapour (PWV), usually expressed in mm, which represents the total mass of water vapour in a column of unit cross-sectional area extending from the surface to the top of the atmosphere.

In order to identify sites with the minimum water vapour content in the atmosphere and preferably those with the minimum variability, a lot of effort has been spent in the past in searching for atmospherically dry sites for ground-based telescopes, typically at high elevations (Giovanelli et al. 2001; Saunders et al. 2009; Otarola et al. 2010; Sims et al. 2012; Tremblin et al. 2012). An accurate forecast of PWV conditions, performed with numerical models, can allow observatories using service mode operations (Primas et al. 2016) to plan observations that require very low PWV conditions while optimizing the use of available observing time. This will allow the possibility to plan in advance service-mode observations and match the science targets of a specific observing programme with the corresponding optimal weather conditions, thus maximizing the scientific output of an observatory. A few studies already tried to apply mesoscale models to PWV forecast on astronomical sites (Giordano et al. 2013; Pozo et al. 2016; Perez-Jordan et al 2015).

The main goal of this paper is to validate and quantify the reliability of the mesoscale atmospherical model Meso-NH (Lafore et al.

★ E-mail: aturchi@arcetri.astro.it
[1]In some cases OH lines are more critical.

1998; Lac et al. 2018) in forecasting the PWV above Cerro Paranal, site of the Very Large Telescope (VLT) and Mount Graham, site of the Large Binocular Telescope (LBT).

VLT Imager and Spectrometer for mid-Infrared (VISIR; Lagage et al. 2004) is the instrument (imager and spectrometer) at VLT which operates in the mid-infrared regime (5–20 μm) and it is strongly dependent on the PWV. Besides that, a few other instruments, running at shorter wavelengths (near-infrared or visible) benefit, at different levels, from the low levels of PWV and, in general, from knowing in advance the quantity of water present in the atmosphere. Among these, CRIRES (Kaeufl et al. 2004), X-SHOOTER (Vernet et al. 2011), MUSE (Henault et al. 2003) and at VLTI GRAVITY (Eisenhauer et al. 2008). In the latter case, the PWV fluctuations between the different telescopes, represent an important deterioration in interferometric observations.

At LBT, the LBTI instrument (Hinz et al. 2014, i.e. Large Binocular Telescope Interferometer) is running in the 2.9–13 μm range and it is highly affected by the PWV too. To appreciate the importance to forecast the exact value of the PWV content in an operational configuration, we note that it has been observed that the VISIR sensitivity in Q band (17–25 μm) can allow to detect targets up to 3 mag higher in extremely low PWV conditions (0.1–0.2 mm; Smette, Horst & Navarrete 2008). It is straightforward to conclude that a good managing of the PWV forecast may open new scientific opportunities both for imaging and spectroscopy from ground-based telescopes, as described e.g. in Kerber et al. (2014).

The interest for validating PWV provided by Meso-NH is twofold. From one side, we have installed an operational and automatic system (ALTA Center[2]) conceived for the forecast of a set of atmospheric parameters relevant for the ground-based astronomy and astroclimatic parameters above Mount Graham to support observations of LBT and LBTI (Masciadri, Vernin & Bougeault 1999; Masciadri et al. 2017). ALTA Center uses the Meso-NH and the Astro-Meso-NH codes. The first code is useful to predict atmospheric parameters, the latter to predict the optical turbulence (C_N^2) and the integrated astroclimatic parameters (seeing, isoplanatic angle and wavefront coherence time). PWV is part of the atmospheric parameters and we intend to prove the model performances in forecasting the PWV that is extremely useful for LBTI. On the other side, the implementation of a similar operational and automatic system for Cerro Paranal, site of the VLT is on-going. Previous studies performed on Cerro Paranal and Mount Graham have already proven excellent performances of the Meso-NH model in forecasting atmospheric parameters such as wind speed and direction, temperature and relative humidity at the ground level (Lascaux, Masciadri & Fini 2013, 2015; Turchi, Masciadri & Fini 2017) and above Cerro Paranal excellent performances in reconstructing the vertical profiles of the same atmospherical parameters on around 20 km (Masciadri, Lascaux & Fini 2013). This study can provide information on the model ability in forecasting the PWV that is critical for instruments such as VISIR, CRIRES, X-SHOOTER, MUSE, GRAVITY.

Cerro Paranal is characterized by extremely dry conditions, with ground relative humidity typically in the range 5 per cent–20 per cent and a median value of PWV $\sim$ 2.4 mm (Kerber et al. 2012). Starting from 2011, it has been installed at the VLT, a Low Humidity and Temperature Profiling microwave radiometer (LHATPRO) that routinely monitors the PWV in this location. LHATPRO has been previously extensively validated by Kerber et al. (2012) and it has been specifically engineered for monitoring dry sites. We have no equivalent instrument above Mount Graham. The strategy of the paper is therefore to prove that the model is able to reconstruct reliable PWV above Cerro Paranal using LHATPRO measurements as a reference. This will be done by comparing observations with model outputs on two rich statistical samples in 2013 and 2017. Once the model has been validated for Cerro Paranal, we apply the same model to Mount Graham. This can be done because the forecast of the PWV does not require a model calibration such as that for the optical turbulence in which some specific parameters of the turbulent energy scheme are tuned to optimize the model behaviour for a specific site. In Section 5, we will discuss why this assumption is justified.

In Section 2, we present the model configurations used for this study. In Section 3, we describe the LHATPRO instruments used for measuring PWV and AH and the criteria used for the samples selection. In Section 4, we describe the results of the model validation performed at Cerro Paranal (VLT site) on different years and conditions in statistical terms. In Section 4.2, we provide a comparison between predictions obtained with Meso-NH model and the ECMWF General Circulation Model. In Section 4.3, we compare the vertical distribution of the AH observed and reconstructed by the Meso-NH model. In Section 4.4, we present a test case done on an extremely low PWV event in 2017. In Section 5, we perform a climatological comparison done on Mount Graham (LBT site) between PWV estimated from Meso-NH and PWV measured with GOES satellites measurements (Carrasco et al. 2017). Finally, in Section 6 we present our conclusions.

2 MODEL CONFIGURATION

We use the atmospheric model called Meso-NH[3] (hereafter MNH) to forecast the PWV (Lafore et al. 1998; Lac et al. 2018). This is a non-hydrostatic mesoscale model that computes the evolution of weather parameters in a three-dimensional volume over a finite geographical area. It uses a forward in time (FIT) numerical scheme to compute the hydrodynamic equations. The coordinate system is based on Mercator projection, while the vertical levels use the Gal-Chen and Sommerville coordinate systems (Gal-Chen & Sommerville 1975). We consider wave-radiation open boundary conditions with Sommerfeld equation for the normal velocity components (Carpenter 1982). The model itself is based on an elastic formulation of hydrodynamic equations, which allows for the filtering of acoustic waves. The simulations make use of a one-dimensional mixing length proposed by Bougeault & Lacarrère (1989) with a one-dimensional 1.5 closure scheme (Cuxart, Bougeault & Redelsperger 2000). We take into account the interaction between surface and atmosphere parameters with the ISBA (Interaction Soil Biosphere Atmosphere) scheme (Noilhan & Planton 1989). We use a Kessler microphysical scheme (Kessler 1969) for the water. The radiation scheme used is the ECMWF one (Hogan et al. 2016). The long wave radiation is computed following the Rapid Radiation Transfer Model (Mlawer et al. 1997). In the short wave, the ECMWF version of RRTM (Morcrette et al. 1999) is used.

The VLT site is located at Cerro Paranal (24.62528 S, 70.40222 W) at a height of 2635 m above sea level, while LBT is located at Mount Graham (32.70131 N, 109.88906 W) at a height of 3191 m above sea level.

[2]http://alta.arcetri.astro.it/

[3]http://mesonh.aero.obs-mip.fr/mesonh/

Table 1. Horizontal resolution of each MNH imbricated domain at Cerro Paranal (VLT).

Domain	ΔX (km)	Grid points	Domain size (km)
Domain 1	10	80 × 80	800 × 800
Domain 2	2.5	64 × 64	160 × 160
Domain 3	0.5	150 × 100	75 × 50

Table 2. Horizontal resolution of each MNH imbricated domain at Mount Graham (LBT).

Domain	ΔX (km)	Grid points	Domain size (km)
Domain 1	10	80 × 80	800 × 800
Domain 2	2.5	64 × 64	160 × 160
Domain 3	0.5	120 × 120	60 × 60

MNH simulations are fed with the initialization data provided by the European Centre for Medium Weather Forecasts (ECMWF), calculated with their General Circulation Model (GCM) extend on the whole globe.

Simulations cover the night time at the specific site, which is relevant for astronomical observations on each site. The date of each simulation in this paper is identified by the UT day 'J' in which the night starts. For the Cerro Paranal case, in this study we simulate 15 h initializing the model at 18:00 UT of the day 'J', forcing the model each 6 h with data coming from the GCM of the ECMWF and we treat/analyse results in the interval [00:00–09:00] UT of day 'J + 1'. This interval permits to fit the nighttime during the whole solar year.

For the Mount Graham case instead, we initialize the model at 00:00 UT of day 'J' (nighttime is always enclosed in the same UT day) with the same forcing scheme at intervals of 6 h. As will be described and explained in Section 5, in the case of Mount Graham model outputs will not be compared to measurements obtained *in situ* at specific dates but we compare model outputs with satellite outputs in climatologic terms. We evaluate therefore the model outputs as they are calculated in the ALTA Center during a full solar year.

We use a grid-nesting technique (Stein et al. 2000) that consists of using a set of different imbricated domains, described in tTbles 1 and 2, with a digital elevation model (DEM, i.e. orography) extended on smaller and smaller surfaces having a progressively higher horizontal resolution. In this way, using the same vertical grid resolution, we achieve the highest horizontal resolution on the innermost domain extended on a limited surface around the summit to provide the best possible prediction at the specific site. Each domain is centred on the telescope coordinates.

The DEM used for domains 1 and 2, on both sites, is the GTOPO,[4] with an intrinsic resolution of 1 km. In domain 3 of Cerro Paranal, we use the ISTAR,[5] with an intrinsic resolution of 500 m (16 arcsec), while in the case of Mount Graham we use the SRTM90[6] (Jarvis et al. 2008), with an intrinsic resolution of approximately 90 m (3 arcsec). Even if in principle the SRTM DEM is available also above Cerro Paranal, in previous studies we observed that model outputs are better correlated to measurements using the ISTAR DEM. We therefore preferred to implement this solution in our study.

In our configuration, the grid-nesting allows a two-way interaction between each couple of father and son domains i.e. each couple of contiguous domains. Under these conditions, the atmospheric flow in the inner domains is the most realistic because the atmospheric flow inside each domain is in a constant thermodynamic equilibrium with the outer domain's flow.

In the Paranal case, we use 62 vertical levels on each domain, starting from 5 m above ground level (a.g.l.), while in the Mount Graham case we use 54 vertical levels on each domain, with the first grid point equal to 20 m a.g.l. In both cases, the levels have a logarithmic stretching of 20 per cent up to 3.5 km a.g.l. From this point onward, the model uses an almost constant vertical grid size of $\sim$ 600 m up to $\sim$23 km, which is the top level of our domain. The grid mesh deforms uniformly to adapt to the orography, so the actual size of the vertical levels can stretch in order to accommodate for the different ground level at each horizontal grid point. The different size of the first grid point (5 and 20 m) is due to the fact that instruments providing atmospheric measurements (wind, temperature, relative humidity and pressure) and observations of the optical turbulence are located at different heights above the ground in the two sites (for example Lascaux et al. 2013, 2015; Masciadri et al. 2017; Turchi et al. 2017). This same model is used for the PWV predictions as well as for the forecast of all the parameters that we have just mentioned.

The PWV value is provided by the MNH model with a time sampling of two minutes of simulated time and it is calculated in the innermost domain having a horizontal resolution of 500 m.

3 MEASUREMENTS AND INSTRUMENTATION

The instrument used as a reference in this paper is the LHATPRO that has been installed at Cerro Paranal in 2011 (Kerber et al. 2012) and since then runs continuously 24/7 providing measurements stored in the ESO archive.[7] This instrument is completely automated and is manufactured by Radiometer Physics GmbH. It uses multiple microwave channels in the frequency bands of 183 GHz (H_2O) and 51–58 GHz (O_2) in order to retrieve, among others, the humidity and temperature profiles up to 10 km of altitude above the ground level. Measurements are taken on 39 vertical levels with a resolution that varies from 10 m at the ground level up to 1 km at the topmost height. For more detailed description, see Rose et al. (2005). As explained in Kerber et al. (2012), the 183 GHz line is extremely important because it allows to resolve the extremely low levels of PWV present on a dry site such as Paranal (median value is around 2.4 mm).

LHATPRO was validated for astronomical use in 2011 (Kerber et al. 2012) against radiosoundings, showing a good correlation with measurements, an accuracy of 0.1 mm, and a precision of 0.03 mm for the PWV. The instrument starts to saturate for PWV values above 20 mm, which are however very rare bad weather events at Paranal (typically below 15 mm). Absolute humidity (AH) vertical profiles were however never considered in previous ESO studies. The vertical distribution of the humidity is not particularly relevant for astronomical applications. However, we investigate also this parameter in this paper mainly to better understand the model

[4] https://lta.cr.usgs.gov/GTOPO30

[5] This was bought by the ESO from the ISTAR Company (Sophia Antipolis, Nice, France). The method is based on two stereoscopic images of the same location taken at different angles, obtained by SPOT satellites.

[6] http://www.cgiar-csi.org/data/srtm-90m-digital-elevation-database-v4-1

[7] http://www.eso.org/asm/ui/publicLog?name = Paranal

behaviour and quantifying its performances and eventually the space for improvements in the calculation of the PWV.

To investigate both PWV and AH, we first selected a sample of 120 nights in 2013. ESO gave us access to the vertical distribution of AH in the period [2012–2014]. The 120 nights are uniformly distributed on the year 2013 (approximately one night every 3 days, 10 days on each month, blindly selected). The PWV retrieved from LHATPRO is sampled every 5 s, while AH profiles are sampled with a frequency of 1 min. MNH outputs are sampled every 2 min (both PWV and AH). Similarly to what has been already done in other similar studies (Lascaux et al. 2015; Turchi et al. 2017), in order to efficiently compare model outputs and observations for an operational application, we performed a moving average of data with a window of 1 h on both data sets (forecasts and measurements) to remove the high frequencies. After this procedure, data are resampled on 20 min and then compared. This permits us to put in evidence the trend of measurements and model outputs. In the rare case that no measurement is available we discard the simulated values.

Besides this analysis performed on 2013 data, we decided to test the model also on a statistical sample related to a more recent period, successive to 2016 March, i.e. the time in which the ECMWF GCM has been upgraded to a horizontal resolution of 9 km instead of 16 km. This test has the goal to investigate if the MNH model performances might improve using more accurate initialization data. We selected therefore another sample of 120 nights uniformly distributed in 2017 (approximately one night every 3 days, blindly selected) and we performed a separate comparison observations versus mesoscale model. LHATPRO PWV data in ESO archive were missing from most of the dates in the period 2017/06/11–2017/07/13, with some other sporadic case of missing data; however, we were still able to select at least 10 days in each month with usable data.

Finally, as we will see in Section 4, to investigate the most challenging case, i.e. the model performances in forecasting the extremely low PWV values (PWV $\leq$ 1 mm) we selected a third sample. The criterium we used is to select all nights of the solar year 2017 in which the PWV is lower than 1 millimetre for at least 30 per cent of the night. We counted 35 nights (see Table B1). This was necessary because we want to maximize the number of cases in which PWV is weaker than 1 mm. In the previous samples, the criterium of uniformity selection over a whole solar year resulted in a scarcity of events with PWV $\leq$ 1 mm. The investigation of this sample permits us to have a more solid estimate of the model performances in reconstructing the lowest conditions of PWV. Looking at the cumulative distribution of the PWV on the observations on the full year of 2013 and 2017 (Fig. A1 – bold line), we retrieve that the percentage of PWV $\leq$ 1 mm is of the order of 12 per cent. Even if the percentage is relatively small, it is important for ESO to optimize the use of this portion of time to retrieve the best advantage for VISIR.

At Mount Graham (LBT), we do not have the availability of accurate on-site measurements of the PWV. In this case, we compared the statistics obtained from the forecasts of the PWV produced by the ALTA project (i.e. the operational version of the model we are treating here – see Section 1) in the period of almost one solar year (from 2016/09/21 to 2017/06/08) with exception of the months of July and August in which the telescope LBT is closed because of monsoon season, with satellite data obtained by Carrasco et al. (2017) over the years 1993–1999, for a total of 58 months of observations.

While LHATPRO provides us both direct measure of PWV (in mm) and a profile of (AH) (in g m^{-3}), MNH model gives us the vertical profile of water vapour mixing ratio M (kg kg^{-1}), pressure P (Pascal), and temperature T (Kelvin degrees). We can however obtain PWV and AH from M with the appropriate relations 1 and 2.

$$\mathrm{PWV} = -\frac{1}{g\rho_{\mathrm{H_2O}}}\int_{P_0}^{P_{\mathrm{top}}} M\,\mathrm{d}P \tag{1}$$

$$\mathrm{AH} = 10^3 M\frac{P}{T R_d}, \tag{2}$$

PWV expressed in mm and AH in g m^{-3}. In the above equations $\rho_{\mathrm{H_2O}} = 10^3\,\mathrm{kg\,m^{-3}}$ is the water density, $g = 9.81\ \mathrm{m\,s^{-2}}$ is the standard gravity acceleration, and $R_d = 287.05\ \mathrm{J/(kg\times K)}$ is the specific gas constant for dry air. We integrate between the ground level pressure P_0 and the top level ($\sim$20 km a.g.l.) pressure P_{top}. We note that the water vapour scale height is in the range 1.5–2.5 km. Above the latter height the water content decrease drastically and is typically negligible above 10 km (Querel, Naylor & Kerber 2016).

4 MODEL VALIDATION ON CERRO PARANAL

Simulations done with the MNH model related to the three samples described in Section 3 have been performed and results have been compared to LHATPRO measurements.

4.1 Model performances in reconstructing the PWV

A statistical analysis has been performed on the three independent samples of nights using the classical statistical operators BIAS, RMSE, and σ, defined as

$$\mathrm{BIAS} = \sum_{i=1}^{N}\frac{(Y_i - X_i)}{N} \tag{3}$$

$$\mathrm{RMSE} = \sqrt{\sum_{i=1}^{N}\frac{(Y_i - X_i)^2}{N}}, \tag{4}$$

where X_i are the individual observations and Y_i are the individual simulations computed at the same time index i, with $1 \leq i \leq N$, N being the total sample size. As done in previous studies (Lascaux et al. 2013, 2015; Turchi et al. 2017), from the above quantities, we deduce the bias-corrected RMSE (σ):

$$\sigma = \sqrt{\mathrm{RMSE}^2 - \mathrm{BIAS}^2}. \tag{5}$$

The previously defined indicators, which provide us information on the statistical and systematic errors, were computed over two different data subsets. The first one is obtained by considering all the available measurements and forecasts while the second one is selected by considering only the [X_i, Y_i] couples corresponding to LHATPRO measurements with PWV $\leq$ 5 mm. This allows us to properly characterize the model performance in low PWV value ranges, which are useful for telescope operations.

We first analysed the 120 nights 2013 sample. In Fig. 1 (left column), we reported the corresponding scatterplots together with the computed regression lines (which must be compared to the dashed diagonal bisecting the plot). In Table 3 (first row), we reported the RMSE, BIAS, and σ computed in the previously specified ranges. The same calculation is performed on the sample of 120 nights on 2017 (Fig. 1, right column and Table 3, second row).

We notice that the statistical indicators computed on the 2013 and 2017 samples are consistent. Results are very convincing, since the

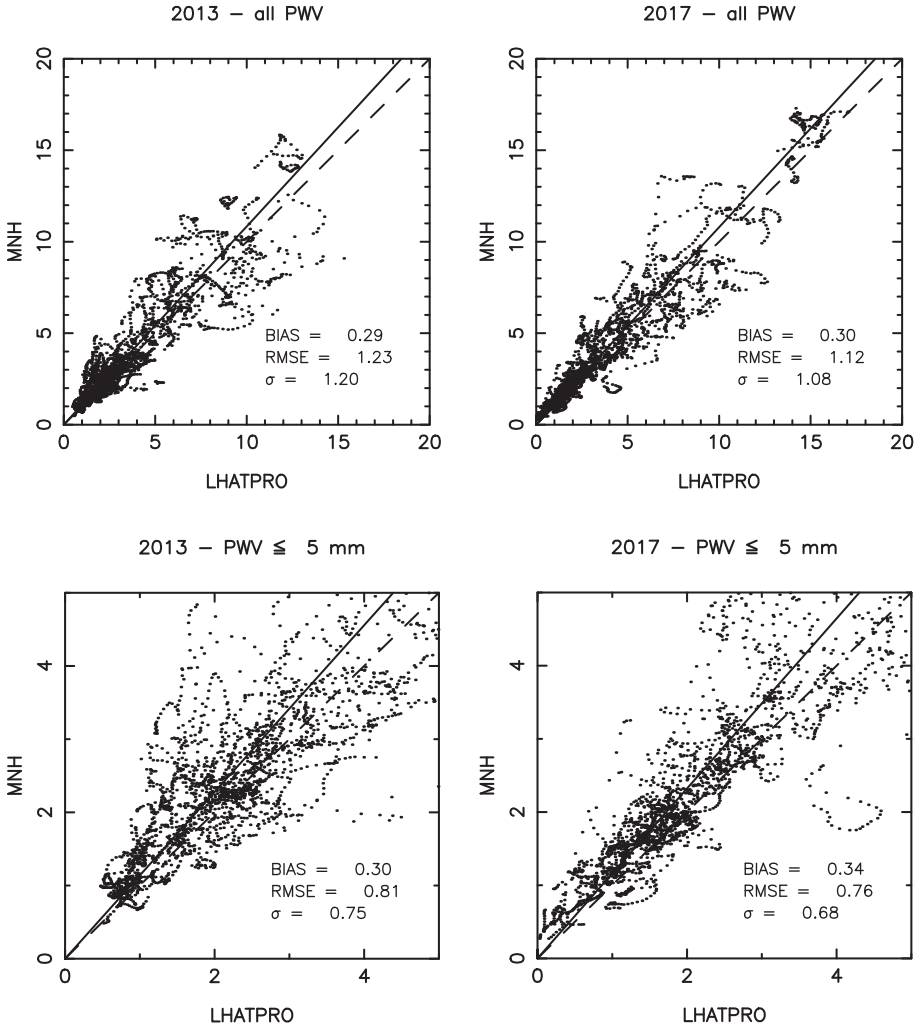

Figure 1. Cerro Paranal – scatterplot computed on the raw data set PWV_{raw}. The plots are computed over the 120 nights 2013 sample (left column) and the 120 nights 2017 sample (right column). The indicators are computed over the whole sample (first row) and on the sub-sample in which PWV measurements are ≤5 mm (second row). The dashed line corresponds to the 45° that should represent a perfect match between model and measurements. The straight line corresponds to the regression line computed on the data points.

Table 3. Cerro Paranal – in each column are the statistical indicators computed both on the raw model output and by applying correction in equation (6). The statistics are computed over the 120 nights 2013 sample (16 km resolution initialization data),and the 120 nights 2017 sample (9 km resolution initialization data) . The indicators are computed over the whole sample (all PWV) and on a sample filtered by selecting LHATPRO measurements with PWV ≤ 5 mm.

	RMSE (mm) (raw/corr)	BIAS (mm) (raw/corr)	σ (mm) (raw/corr)
2013 - 120n			
All PWV	1.23/1.17	0.29/-0.10	1.20/1.17
PWV ≤ 5 mm	0.81/0.73	0.30/-0.04	0.75/0.73
2017 - 120n			
All PWV	1.12/1.05	0.30/-0.11	1.08/1.05
PWV ≤ 5 mm	0.76/0.65	0.34/-0.01	0.68/0.65

dispersion of the data along the diagonal of the scatterplot is quite reduced. There is a cone effect with the dispersion increasing for large PWV values, confirmed by the statistical indicators computed on all the sample and on the PWV ≤ 5 mm range. In the latter case the RMSE is reduced by ∼1/3 in both samples, with respect to the full sample (from 1.23 to 0.81 mm in the 2013 sample and from 1.12 to 0.76 mm in the 2017 sample). The statistical operators RMSE and σ are slightly better on 2017 with respect to 2013. The increased horizontal resolution of the initialization data passing from 2013 to 2017 can be the possible cause of this improvement. In both years (2013 and 2017), the RMSE in the PWV ≤ 5 mm case is well below 1 mm. Results can be considered, therefore, very satisfactory. We also notice, in all cases, a residual BIAS of the order of ∼0.3 mm indicating that the model systematically slightly overestimates the PWV values measured by the LHATPRO. We refer to Section 4.3 for a discussion on this point. This bias effect can however be reduced or eliminated with a regression model obtained from the statistics shown in the previous analysis. We searched for a regression which minimizes errors on all the data sets and on all the range of values of PWV and we selected an optimal correction reported in equation (6):

$$PWV_{corr} = \frac{PWV_{raw}}{1.04} - 0.25, \quad (6)$$

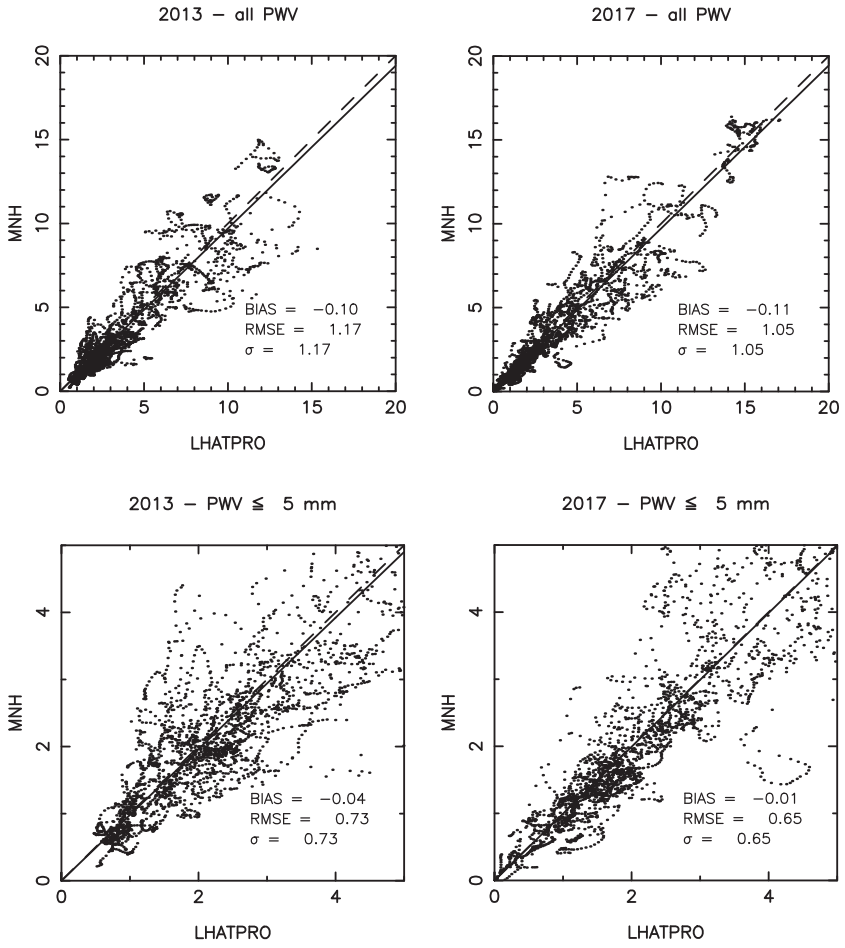

Figure 2. Cerro Paranal – corrected data set. Scatterplot computed on the corrected data set PWV_{corr} (with correction in equation 6). The plots are computed over the 120 nights 2013 sample (left column) and the 120 nights 2017 sample (right column) .The indicators are computed over the whole sample (first row) and on the sub-sample in which PWV measurements with PWV $\leq$ 5 mm (second row). The dashed line corresponds to the 45° that should represent a perfect match between model and measurements. The straight line corresponds to the regression line computed on the data points.

where PWV_{raw} is the uncorrected model output and PWV_{corr} is the one corrected by the optimal regression line.

Table 3 reports the PWV_{corr} values obtained through equation (6), with the corresponding scatterplots shown in Fig. 2.

We observe that the results are consistently improved on all the data sets, especially in the PWV $\leq$ 5 mm range. We are able to obtain a negligible BIAS on all the data sets and the RMSE is basically identical to σ, meaning that the residual error left is purely random, which confirms the validity of the applied correction. We also confirm that, concerning the PWV $\leq$ 5 mm data set which is the most relevant one on Cerro Paranal (which have a median PWV$\sim$2.4 mm), the error on the 2017 data set (RMSE $= \sigma =$ 0.65 mm) is lower than the one computed on the 2013 samples (RMSE $= \sigma =$ 0.73 mm). Since future initialization data from ECMWF will have a resolution which is equal or superior to the one available in 2017, we argue that we should consider the indicators computed on the 2017 sample as the reference for potential future applications. In this paper, we do not study the correlation coefficient because in previous studies (Masciadri et al. 2017 – Annex B) it has been shown that this parameter can be misleading. In many cases the model has an almost perfect agreement with measurements; however, it shows a low correlation coefficient (due to small random fluctuations along the same baseline). In other cases the model may show a large bias or large local discrepancies with measurements; however, it would have a good correlation coefficient. In other words, the correlation coefficient is not particularly relevant for the flexible-scheduling application.

Besides of the bias, RMSE, and σ, we validated the model through the contingency table method, in a similar way to the previous studies on Cerro Paranal from Lascaux et al. (2015), which provide useful informations that complement the previous statistical analysis. While referring the reader to Lascaux et al. (2015) for specific details on the method, here we report the main informations which are useful to understand the method. A contingency table is a method to analyse the relationship between variables in a categorical way. In practice, we pre-define value intervals and we count the number of times in which the couples measurement/simulation $[X_i, Y_i]$ both fall into the same interval. With such defined tables it is possible to evaluate the probability of success of the model using several different statistical operators. As in previous studies, we use the percentage of correct detection (PC), which represents the global probability of having both measurements and model to agree on the same interval. We also define the probability of detection (POD) in a specific range of values, which quantifies the specific agreement

Table 4. 3 × 3 contingency table – 2013 120 nights sample with applied correction in equation (6).

PWV (mm)		LHATPRO		
		PWV ≤ 1	1 <PWV ≤ 5	PWV> 5
MNH	PWV ≤ 1	350	104	0
	1 <PWV ≤ 5	78	1914	68
	PWV> 5	0	90	636

Sample size = 3240; PC = 89.5%; EBD = 0.0%
POD_1 = 81.8%; POD_2 = 90.8%; POD_3 = 90.3%

Table 5. 3 × 3 contingency table – 2017 120 nights sample with applied correction in equation (6).

PWV (mm)		LHATPRO		
		PWV ≤ 1	1 <PWV ≤ 5	PWV> 5
MNH	PWV ≤ 1	325	151	0
	1 <PWV ≤ 5	52	1805	95
	PWV> 5	0	92	698

Sample size = 3218; PC = 87.9%; EBD = 0.0%
POD_1 = 86.2%; POD_2 = 88.1%; POD_3 = 88.0%

probability for each defined range, and the extremely bad detection (EBD), which defines the probability of measurements and model to fall into distant categories (see definitions of PC, POD, and EBD in Lascaux et al. 2015 – equations 9–13). If we define three range values (3 × 3 table), in the case of a perfect model prediction we would have PC = PODs = 100 per cent and EBD = 0, while in the random prediction case we would have PC = PODs = 33 per cent and EBD = 22.5 per cent. It is also possible to define $N \times N$ tables (see Lascaux et al. 2015) if it is desired.

Here, we report the contingency tables computed on the full 2013 and 2017 samples after correction (see equation 6). We selected the following specified ranges: PWV ≤ 1 mm, 1 mm < PWV ≤ 5 mm, PWV > 5 mm because this set appeared the most interesting from an observational point of view. Also 5 mm is approximately the third quartile of the PWV distribution over Cerro Paranal, as shown in Fig. A1. In Tables 4 and 5, we see that model performance is satisfactory on both samples (2013 and 2017), with a PC = 89.5 per cent on the 2013 sample and a PC = 87.9 per cent on the 2017 sample. Specifically, we note that we have EBD = 0 per cent in both cases, meaning that the model never make dramatic errors. In each specified range, the performance is comparable to the global PC. We note that the most challenging probability of detection of PWV ≤ 1 mm (i.e. POD_1) presents excellent values for both 2013 (POD_1 = 81.8 per cent) and 2017 (POD_1 = 86.2 per cent). It is possible that the improved performance observed on 2017 with respect to 2013 is due to the higher horizontal resolution of the initialization data.

If we look at the cumulative distribution calculated for LHATPRO and MNH on the samples of 120 nights on 2013 and 2017 (Fig. A1, black dashed line and red line), it is possible to conclude that the samples are perfectly representative of the typical conditions of PWV at Cerro Paranal. Median values and first and third quartiles of the sub-sample of 120 nights and the correspondent sample reconstructed by the model are very well correlated also with the bold full line representing the sample of all the nights of 2013 and 2017. In particular, PWV is ≤ 1 mm for 13 per cent of time on 2013 and for 12 per cent of time on 2017. Very similar values to what has been observed on the whole years 2013 and 2017 (Fig. A1 – bold line).

As we anticipated in Section 3, even if the first two samples appear representative of the global distribution, to better investigate the most challenging region of PWV ≤ 1 mm, we analysed the third sample obtained selecting all the nights in 2017 having at least 30 per cent of the night in which the PWV ≤ 1mm as described in Section 3. We identified 35 nights (Table B1). On the third sample of 35 nights, we have ∼73 per cent of measurements below or equal to 1 mm, with almost all of the measurements below or equal to 2 mm.

In Fig. 3 and Table 6, we report the statistical indicators computed on the 35 nights sample. Looking at the results after correction, we obtain the excellent values of RMSE = 0.27 mm and a σ = 0.25 mm. This value is comparable to the level of accuracy obtained by different instrumentation (VISIR, CRIRES, UVES, X-SHOOTER) that is of the order of 0.1–0.2 mm (Kerber et al. 2010) – Table 4 and (Kerber et al. 2014) – Table 1. We can conclude therefore that the level of the model performances is very satisfactory. If we put all together the 35 nights of 2017 with the sample of 120 nights of 2017[8] and we calculate the contingency table (see Table 7), we obtain a POD_1 = 84.3 per cent that is very similar to the POD_1 = 86.2 per cent (Table 5). This is consistent with the fact that we had proven that the sample of 120 nights of 2017 is very representative of the typical conditions above Cerro Paranal.

Finally, Table 8 shows the POD for PWV ≤ 2 mm, 1 mm, and 0.5 mm obtained with the sample of 135 nights. In the last column, we consider the POD calculated assuming an uncertainly of 0.2 mm. As we will see in Section 4.4, this corresponds to the typical uncertainty of instruments measuring the PWV. Besides to LHATPRO, there are also other instruments: VISIR, XSHOOTER, UVES, CRIRES). This means that we consider the MNH forecast as a hit if it falls within 0.2 mm of the corresponding measurement. We conclude that the MNH model has a probability of correct detection for PWV ≤ 2 mm of 97 per cent, a probability of correct detection for PWV ≤ 1 mm of 93 per cent, and a probability of correct detection for PWV ≤ 0.5 mm of 79 per cent.

4.2 Meso-NH model versus ECMWF

In this section, we compare the performances of the MNH model with the General Circulation Model of the European Center for Medium Range Weather Forecasts (ECMWF). We extract ECMWF historical data from the MARS catalogue.[9] We consider the sequence of forecasts calculated at 00:00 and 12:00 h UT with a time sampling of 1 h (i.e. a data point each hour). We pickup the ECWMF forecasts as if we are in an operational configuration, i.e. we consider that data are made available by ECMWF approximately 6 h after the calculation time. For each date J, extracted data corresponding to hours 6:00–17:00 UT are forecasts calculated at 00:00 UT, while data corresponding to 18:00–05:00 UT (of day after) are calculated at 12:00 UT of the same day J, following the 6 h delay of the ECMWF data delivery. This corresponds to the forecast system delivered by ECMWF to ESO in the 2017 period.

In the comparison between ECMWF and MNH forecasts, data have been resampled every 20 min in order to compute a statistical analysis in identical conditions to the previous studies. We performed the comparison on the largest possible sample, i.e. the

[8]Note that the total sample is done by 135 nights (and not 155) because part of the 35 nights is already included in the 120 nights. The two sub-samples overlap in part.
[9]https://software.ecmwf.int/wiki/display/WEBAPI/MARS + service

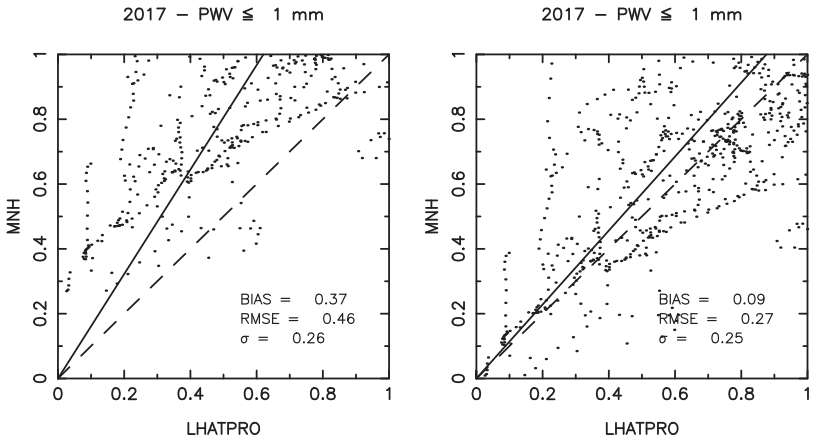

Figure 3. Cerro Paranal – scatterplot computed on PWV. The plots are computed over the 35 nights 2017 sample obtained by selecting all nights on 2017 with PWV $\leq$ 1 mm for, at least, 30 per cent of the night. In the left-hand panel, we show the uncorrected (raw) scatterplot, while in the right-hand panel we show the scatterplot obtained with correction in equation (6). The indicators are computed by filtering the sample selecting LHATPRO measurements with PWV $\leq$ 1 mm. The dashed line corresponds to the 45° that should represent a perfect match between model and measurements. The straight line corresponds to the regression line computed on the data points.

Table 6. Cerro Paranal – in each column are the statistical indicators computed both on the raw model output and by applying correction in equation (6). We did no special optimization for the low PWV values. The statistics are computed over the 35 nights 2017 sample obtained by selecting the nights with episodes of low PWV $\leq$ 1 mm. The indicators are computed by filtering the sample selecting LHATPRO measurements with PWV $\leq$ 1 mm.

	RMSE (mm) (raw/corr)	BIAS (mm) (raw/corr)	σ (mm) (raw/corr)
PWV$\leq$1 mm	0.46/0.27	0.37/0.09	0.27/0.25

Table 7. 3 × 3 contingency table – 2017 135 nights sample with applied correction in equation (6).

PWV (mm)		LHATPRO		
		PWV $\leq$ 1	1 <PWV $\leq$ 5	PWV> 5
MODEL	PWV $\leq$ 1	581	217	0
	1 <PWV $\leq$ 5	108	1832	95
	PWV> 5	0	92	698

Sample size = 3623; PC = 85.9%; EBD = 0.0%
POD_1 = 84.3%; POD_2 = 85.6%; POD_3 = 88.0%

Table 8. Cerro Paranal – POD values computed on the 135 nights 2017 sample for PWV $\leq$ 2 mm, 1 mm, and 0.5 mm. MNH values are corrected by equation (6). In the last column, we consider a 0.2 mm tolerance on the selected intervals in order to take into account the uncertainty on measurements. This means that we consider the MNH forecast as a hit if it falls within 0.2 mm of the corresponding measurements.

PWV range	POD	POD ($\pm$0.2 mm)
PWV$\leq$0.5 mm	64.8%	79.2%
PWV$\leq$1 mm	84.3%	93.5%
PWV$\leq$2 mm	95.9%	97.3%

135 nights sample (see Section 4.1). In Fig. 4, we show the scatterplots of the PWV in the three regimes (all values, PWV $\leq$ 5 mm, PWV $\leq$ 1 mm) between ECMWF (top row) and MNH (bottom line). In Table 9 are reported the corresponding statistical operators. The comparison ECMWF versus MNH is done on the raw measurements without the post-processing correction (see equation 6) because the important information, in this context, is the difference between the two models in the same conditions. The goal of this exercise is to establish if there is a gain or not in using MNH. All the statistical operators (BIAS, RMSE, and σ) are visibly larger for ECMW than MNH. The ECMWF BIAS is a factor between 2 and 4 larger than the MNH one, while ECMWF RMSE and σ are a factor 2 larger.

Table 9. Cerro Paranal – in each column are the statistical indicators computed by comparing LHATPRO measurements on the 135 nights 2017 sample either with the ECMWF GCM forecasts and the MNH forecasts. The indicators are computed over the whole sample (All PWV), on a sample filtered by selecting LHATPRO measurements with PWV $\leq$ 5 mm and finally on a sample filtered with PWV $\leq$ 1 mm.

PWV (mm)	RMSE (mm) (ECMWF/MNH)	BIAS (mm) (ECMWF/MNH)	σ (mm) (ECMWF/MNH)
All PWV	2.01/1.06	1.35/0.30	1.49/1.02
PWV$\leq$5	1.45/0.72	1.04/0.33	1.02/0.64
PWV$\leq$1	0.99/0.46	0.83/0.38	0.54/0.26

The most critical parameters are, however, the RMSE and even more σ that represents the pure statistical error. In all the regimes, RMSE and σ are visibly larger for the ECMWF case than for the MNH one. In the most challenging case, i.e. where the PWV $\leq$ 1 mm, we have a $\sigma = 0.54$ mm (ECMWF) and $\sigma = 0.26$ mm (MNH), i.e. a not negligible factor 2 in gain. For RMSE, the ECMWF case (0.99) is a factor 2.15 larger than MNH (0.46).

Such result proves that MNH can provide a not negligible improvement with respect to ECMWF in terms of scheduling of VLT instruments depending on PWV estimates.

4.3 Model performances in reconstructing the absolute humidity vertical profiles AH

While not being crucial for astronomical observations, we study here the absolute humidity (AH) vertical profiles as measured by LHATPRO and reconstructed by MNH because this can provide us informations on where there is space for improving model performances. We used the sample 120 nights of 2013 because it is the

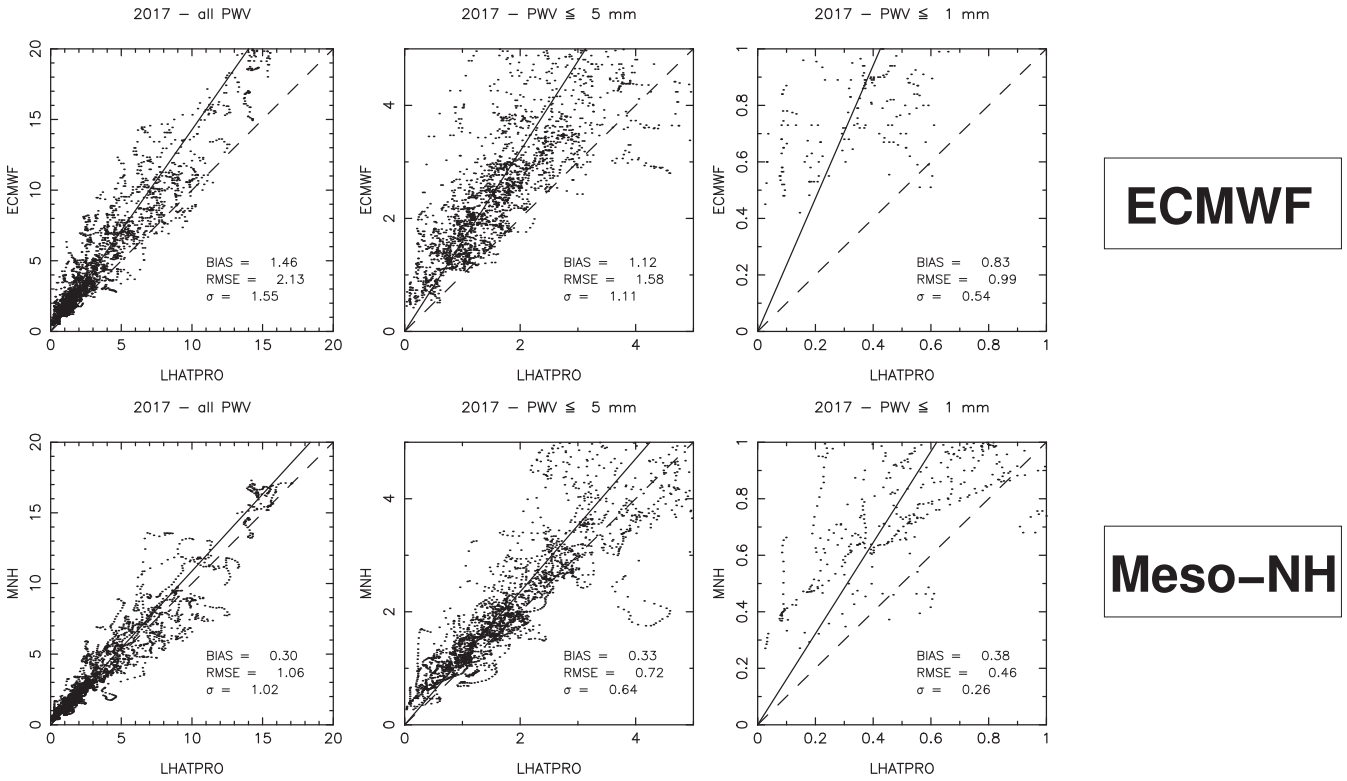

Figure 4. Cerro Paranal – scatterplot computed on the 135 nights 2017 sample comparing LHATPRO measurements either with ECMWF GCM forecasts (*top row*) and MNH forecasts (*bottom row*). The plots are relative to the whole sample (first column), to a sample filtered by selecting LHATPRO measurements with PWV $\leq$ 5 mm (second column) and PWV $\leq$ 1 mm (third column). The dashed line corresponds to the 45° that should represent a perfect match between model and measurements. The straight line corresponds to the regression line computed on the data points.

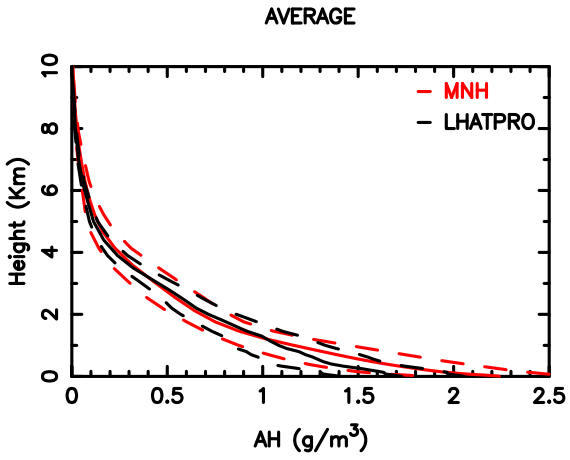

Figure 5. Cerro Paranal – average vertical AH profiles computed over the 120 nights 2013 sample by the MNH model (red) and by LHATPRO (black line). Dash lines indicate the standard deviation for the model (red) and LHATPRO (black).

unique one for which we have the vertical stratification of measurements on the about 20 km above the ground. We compared therefore the AH profiles measured by LHATPRO with those reconstructed by the model obtained with equation (2). In Fig. 5 is shown the average AH profile obtained by both LHATPRO and MNH, with the variability over the whole sample expressed as standard deviation and represented by the dashed lines. Above 10 km a.g.l., the water content in the atmosphere is almost negligible and does not contribute in a sensible way to the total PWV value. The agreement is meaningful because it shows that the model is correctly reproducing the highly variable distribution of water vapour across the atmosphere. Only below 1 km the model tends to overestimate, even if within the standard deviation margins. This finding is consistent with what observed in previous sections in the statistical analysis of PWV in which it was evident a residual BIAS (very small in reality) that we have corrected in a post-processing phase and can be introduced in the operational configuration. This result tells us that there is still some space for further improvement of the model in the very low part of the atmosphere where it is dominant in the soil–atmosphere interaction. This might be a topic for future investigations. However, we have shown in the previous sections that even at present we can overcome this shortcoming efficiently through a statistical approach in a post-processing phase. This element, however, has basically no impact on the forecast model performances because we have shown that it can be statistically corrected in the post-processing phase in an efficient way.

4.4 Forecast of an episode of extremely low PWV

We show here results obtained in relation to a single extremely low PWV event registered at Cerro Paranal in 2017. PWV show extremely low values in one night in a period of three consecutive nights [2017/07/15–2017/07/17] UT[10] in which the PWV was

[10]Our convention on dates is that yyyy/mm/dd UT indicates the start of the local night.

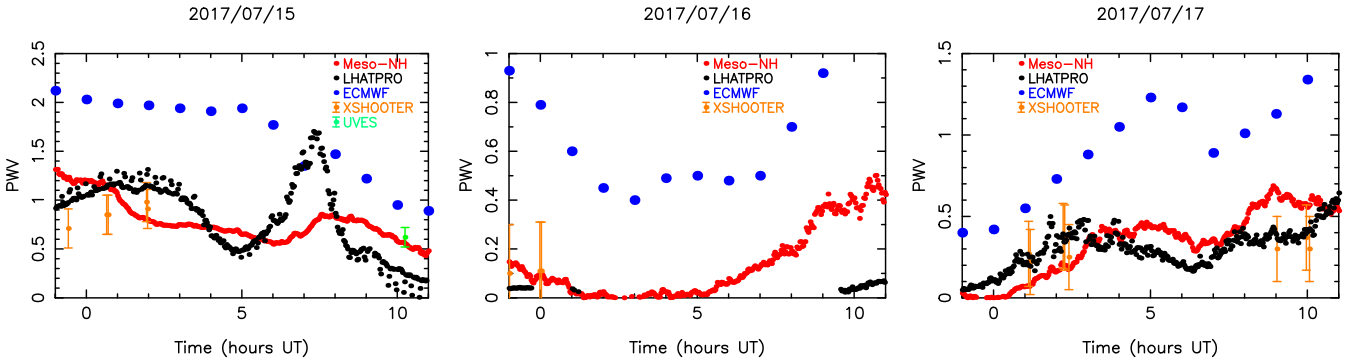

Figure 6. Time evolution of PWV values over three nights centred around the extremely low PWV event of 2017/07/17. Black dots correspond to LHATPRO measurements, red dots correspond to the MNH forecast (2-min sampling time), the blue dots correspond to the ECMWF forecast (1-h sampling time), the orange dots correspond to X-SHOOTER measurements and green dots correspond to UVES measurements. According to Table 6, we estimate the MNH model error to be around 0.25 mm.

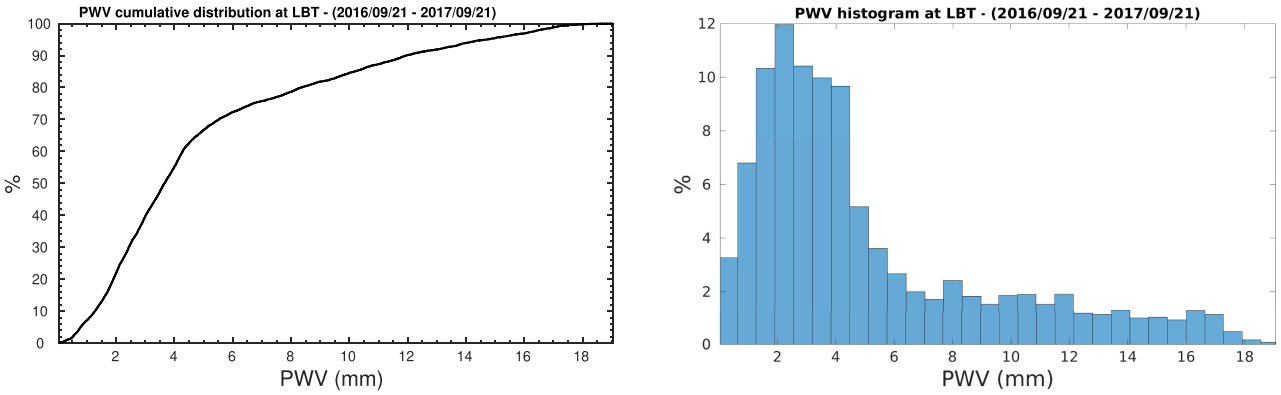

Figure 7. Cumulative distribution (left) and histogram (right) of PWV values on Mount Graham forecasted by MNH model as provided by the ALTA Center. Data refer to 283 nights between 2016/09/21 and 2017/09/21.

lower than 1 mm reaching values smaller than 0.025 mm[11] on 2017/07/16. Fig. 6 shows the time evolution of the PWV during the three nights. Predictions of the MNH model are shown together with measurements from the radiometer LHATPRO and from other VLT instruments. During these three nights we have the availability of PWV measurements from UVES and X-SHOOTER instruments. X-SHOOTER is a multiwavelength medium resolution spectrograph and operates on wavelength between 300 and 2500 nm (Vernet et al. 2011). We refer to Kerber et al. (2014) for more details on the instruments X-SHOOTER and UVES.

In Fig. 6, we have therefore a sampling of 5 s for LHATPRO and 120 s for MNH model. We considered an error bar of 0.20 mm for X-SHOOTER and 0.1 mm for UVES as reported by Kerber et al. (2014). For this special case, we run longer simulations that cover a 18 h time interval between 18 UT to 12 UT of the day after. We plotted all the available measurements from LHATPRO (black dots), X-SHOOTER (orange dots), and UVES (green dots) over the selected nights. The model shows a very good correlation with measurements done with LHATPRO and other VLT instruments. In almost all cases, the discrepancy between the model and the measurements is within the error bar of the VLT instruments. In this case, the PWV measurements are in the lowest range of values (PWV$\leq$1 mm), so according to Table 6, we estimate the model error to be around 0.25 mm. ECMWF forecasts (blue dots), extracted with the same procedure explained in Section 4.2, are unable to represent the extremely low PWV values of this episode with the same accuracy of the mesoscale MNH model, and this behaviour is particularly evident in the 2017/07/16 night. During this night, the number of LHATPRO PWV measurements that appear on the figure is small. The apparent missing measurements correspond in reality to a PWV equal to zero because PWV retrieval from brightness temperature fails if the PWV value is too low. This means that where we do not see the black points (LHATPRO PWV measurements) we can consider PWV equal to zero. LHATPRO measurements are consistent with both X-SHOOTER and MNH up to an exceptional degree. From this simple test, we do expect that the MNH forecast can be indeed able to predict in advance such rare events with excellent levels of accuracy and support infrared observations at VLT. Also MNH shows definitely better performances with respect to the ECMWF predictions. We also highlight the fact that, besides

[11] when the measurement is not visible in the figure it means that PWV is basically equal to zero as it will be explained later on.

the evident better performances of MNH with respect to ECMWF, the temporal sampling of the former (2 min) is definitely better than the latter (1 h).

5 MODEL VALIDATION ON MOUNT GRAHAM

In the case of LBT Observatory (Mount Graham), we do not have an instrument similar to LHATPRO deployed on the site. We cannot therefore have access to accurate measurements of PWV and the vertical AH stratification. However, the excellent results obtained in our analysis performed over Cerro Paranal make us confident that the model can perform reasonably well also on Mount Graham, since they are similar sites (mountain tops) and no specific model calibration as that required for the optical turbulence (Masciadri et al. 2017) is done before running the simulations. We assume that the coefficient of correction calculated a posteriori (equation 6) are valid not only for Cerro Paranal but also for other sites. The PWV is indeed weakly dependent on the topography. The fact that, as we will see, model outputs are consistent with measurements at Mount Graham from a climatologic point of view, indicates that this assumption is reasonable. It is in any case obvious that, the presence of measurements *in situ*, might permit a more sophisticated validation at Mount Graham similar to that done for Cerro Paranal. To provide at least a climatological verification, we compare here the climatologic statistical values of operators characterizing the PWV as reconstructed by the MNH model from the ALTA Center project and as measured from GOES satellites measurements by Carrasco et al. (2017) on a total of 58 months between 1993 and 1999. ALTA Center is also characterized by a three domains structure as is the case described here for Cerro Paranal having the same horizontal resolution of 10, 2.5, and 0.5 km.[12]

In Carrasco et al. (2017), the PWV values are reconstructed from the brightness temperature at 6.7 μm, measured by satellites, through a semi-empirical method developed by Soden & Bertherton (1993). In section 3.3.1 of the Carrasco paper, the technique was validated on Mount Graham against radiometer measurements provided by the Submillimeter Telescope Observatory (SMTO). The comparison was done by considering only radiometer values less than 7 mm (which ensures the validity of the comparison) and the found agreement was good.

Thanks to the ALTA system already running on the telescope site since last year, we have the availability of 283 simulated nights from 2016/09/21 (date in which the PWV forecast was initially implemented in ALTA) to 2017/06/30 (UT dates). This is basically one solar year if we exclude the July–August summer shutdown period of the telescope.

In Fig. 7 is shown the distribution of PWV values produced by the ALTA system over the selected 1 yr period, corrected by equation (6) that was obtained by the analysis on Cerro Paranal. We assume that the correction can be applied everywhere because in principle we did not adjust the model parameters with a site-dependent calibration. From the statistics computed over the simulated period, if we consider only values of PWV ≤ 7 mm as indicated by Carrasco et al. (2017) to respect the validation conditions, we obtain first quartile, median, and third quartile of the distribution reported in Table 10. The agreement with Carrasco's values is very good and confirms the validity of ALTA forecasts on Mount Graham. We note that, considering the nature of the GOES satellites estimates that we are treating, we cannot perform more accurate analyses. It has been already noted that GOES satellites is preferably used for statistical analysis since the scatter on individual nights is very high (Kerber et al. 2010). In part because we cannot perform a comparison on the same specific nights but we can only perform a comparison in climatological sense as described here. While this kind of validation is not as detailed as in the Paranal case and it can mainly achieve a climatologic estimate, the results we obtained in Table 10 indicate that MNH provides very consistent estimate for the PWV above Mount Graham.

[12] A fourth domain with higher horizontal resolution is included but it is used only for the wind speed.

Table 10. Comparison between the values (first quartile, median, and third quartile) as obtained by the MNH model (from ALTA Center project) on one solar year and the values as retrieved by Carrasco et al. (2017) on a sample of 58 months between 1993 and 1999. See the text for more details.

	First quartile (mm)	Median (mm)	Third quartile (mm)
ALTA	1.9	2.9	4.1
Carrasco et. al (2017)	2.0	2.9	4.3

6 CONCLUSIONS

In this paper, we validated the MNH model in forecasting the PWV above Cerro Paranal, site of the VLT and Mount Graham, site of the LBT. A detailed analysis has been performed at Cerro Paranal with two rich statistical sample of nights (120 nights) in two different years (2013 and 2017) showing excellent performances with the best on 2017 that corresponds to the best initialization data. To better investigate the portion of PWV values below 1 mm we also considered a slighter richer sample on 2017 (135 nights) including all the night of 2017 having a PWV ≤ 1 mm for at least 30 per cent of the night.

The statistical operators (BIAS, RMSE, and σ) applied on the sample of all PWV values, on the PWV ≤ 5 mm, and on the PWV ≤ 1 mm are all consistently extremely good with the RMSE and σ that decrease when the extension of the interval of analysis (PWV $\leq$ X mm) is limited to smaller X values reaching the smallest values of RMSE $= 0.27$ mm and $\sigma = 0.25$ mm in the most challenging region of PWV ≤ 1 mm.

We demonstrated that the MNH model provides a significant improvement with respect to the ECMWF forecast (roughly a factor 2) for the statistical operators (RMSE and σ) passing, respectively, from 1.58 to 0.72 mm (for PWV ≤ 5 mm) and from 0.99 to 0.46 mm (for PWV ≤ 1 mm) for RMSE and from 1.11 to 0.54 mm (for PWV ≤ 5 mm) and from 0.54 to 0.26 mm (for PWV ≤ 1 mm) for σ.

We also calculated 3×3 contingency tables with the correspondent percentage of corrected prediction (PC) and probability of detection in each individual sectors: PWV ≤ 1 mm, 1 mm $<$ PWV ≤ 5 mm, and PWV > 5 mm. We investigated finally the probability of detection for the most challenging sectors: ≤ 2 mm, ≤ 1 mm, and ≤ 0.5 mm. The probability to detect the PWV ≤ 2 mm is 97 per cent, the probability to detect the PWV ≤ 1 mm is 93 per cent and the probability to detect the PWV ≤ 0.5 mm is 79 per cent.

Besides, we also studied, in detail way, a case of extremely low PWV observed at the VLT lasted three days (with values smaller than 0.025 mm on one night). We showed that MNH can reconstruct the trend and the values observed by LHATPRO and by further two

instruments (X-SHOOTER and UVES) in this period in a much more precise way than ECWMF forecasts. The discrepancy between the model and the observations is almost comparable to the dispersion among the instruments and in many cases also with the declared accuracy of the instruments.

By studying the vertical stratification of the absolute humidity as measured by LHATPRO and reconstructed by the model we can conclude that the MNH model provides a vertical median profile very well correlated to measurements. Only very close to the surface, it is evident a slight overestimation of the model that is, anyway, within the standard deviation. This tells us that in this region of the atmosphere we might have some small margin of improvements for the model.

Once the model has been validated in an extremely detailed way above Cerro Paranal (VLT), we compared measurements coming from GOES satellites (Carrasco et al. 2017) and simulations performed by MNH. This comparison has been done in climatological terms by comparing model estimates on one solar year with satellites measurements related to 58 months in the period 1993–1999. Comparison provided an excellent agreement on the median value (2.9 mm for both measurements and MNH) and first (1.9 and 2.0 mm) and third (4.1 and 4.3 mm) quartiles of the distribution for PWV ≤ 7 mm that, as proved by Carrasco et al. (2017), is the range of validity of the method.

The model MNH appears therefore an extremely efficient tool to forecast the PWV above Cerro Paranal (VLT), Mount Graham (LBT), and similar astronomical sites. The investigation of model performances on longer time-scales (larger than 15 h) is not a priority at the moment. On the other side, we are working on techniques/methods to improve the model behaviour on short time-scale (a few hours), a goal that is very critical for the flexible scheduling. This study represents an important step towards the set-up of a system for the operational forecast of various atmospheric parameters, among which the PWV, which is in progress by a few of us at INAF and it is conceived for the VLT.

ACKNOWLEDGEMENTS

ALTA Center project is funded by the Large Binocular Telescope Corporation. The authors thank Christian Veillet, Director of the Large Binocular Telescope, for his continued and valuable support given to this research activity. Authors also thank the LBTO staff for their technical support and collaboration. The authors thank the MOSE ESO Board and the Paranal Science Operation Team for their constant support. Part of the numerical simulations has been run on the HPCF cluster of the European Center for Medium Range Weather Forecasts (ECMWF) using resources from the Project SPITFOT.

REFERENCES

Bougeault P., Lacarrère P., 1989, Mon. Weather Rev., 117, 1872
Carpenter K. M., 1982, Q. J. R. Meteorol. Soc., 110, 717
Carrasco E., Avila R., Erasmus A., Djorgovski S. G., Walker A. R., Blum R., 2017, PASP, 129, 973
Cuxart J., Bougeault P., Redelsperger J.-L., 2000, Q. J. R. Meteorol. Soc., 126, 1
Eisenhauer F. et al., 2008, ESO Astrophys. Symp., 41, 431
Gal-Chen T., Sommerville R. C. J., 1975, J. Comput. Phys., 17, 209
Giordano C., Vernin J., Vazquez Ramio H., Munoz-Tunon C., Varela A. M., Trinquet H., 2013, MNRAS, 430, 3102
Giovanelli R. et al., 2001, PASP, 113, 803
Henault F. et al., 2003, Proc. SPIE, 4841, Instrument Design and Performance for Optical/Infrared Ground-based Telescopes, 1096
Hinz P. et al., 2014, Proc. SPIE, 9146, Optical and infrared Interferometry IV, 91460T
Hogan R. J., Bozzo A., 2016, ECMWF Technical Memorandum, Technical Report, ECRAD: A new radiation scheme for the IFS, 787, ECMWF
Jarvis A., Reuter H. I., Nelson A., Guevara E., 2008, Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database
Kaeufl H. U. et al., 2004, Proc. SPIE, 5492, Ground-based Instrumentation for Astronomy, 1218
Kerber F. et al., 2010, Proc. SPIE , Ground-based and Airborne Telescopes III, 7733, 77331M
Kerber F. et al., 2012, Proc. SPIE , Ground-based and Airborne Instrumentation for Astronomy IV, 8448, 84463N
Kerber F. et al., 2014, MNRAS, 439, 247
Kessler E., 1969, Meteor. Monog., 10, 84
Lac C. et al., 2018, Geosci. Model Dev., 11, 1929
Lafore J.-P. et al., 1998, Ann. Geophys., 16, 90
Lagage P. O. et al., 2004, The Messenger, 117, 12
Lascaux F., Masciadri E., Fini L., 2013, MNRAS, 436, 3147
Lascaux F., Masciadri E., Fini L., 2015, MNRAS, 449, 1664
Masciadri E., Vernin J., Bougeault P., 1999, A&AS, 137, 185
Masciadri E., Lascaux F., Fini L., 2013, MNRAS, 436, 1968
Masciadri E., Lascaux F., Turchi A., Fini L., 2017, MNRAS, 466, 520
Mlawer E. J., Taubman S. J., Brown P. D., Iacono M. J., Clough S. A., 1997, J. Geophys. Res., 102, 16663
Morcrette J. J., Barker H. W., Cole J. N. S., Iacono M. J., Pincus R., 2008, Mon. Weather Rev., 136, 4773
Noilhan J., Planton S., 1989, Mon. Weather Rev., 117, 536
Otarola A. et al., 2010, PASP, 122, 470
Perez-Jordan G., Castro-Almazan J. A., Munoz-Tunon C., Codina B., Vernin J., 2015, MNRAS, 452, 1992
Pozo D., Marin C., Illanes L., Cure M., Rabanus D., 2016, MNRAS, 459, 419
Primas F. et al., 2016, Proc. SPIE, Observatory Operations: Strategies, Processes, and Systems VI, 9910, 991002
Querel R. R., Naylor D. A., Kerber F., 2016, PASP, 123, 222
Rose T., Crewell S., Loehnert U., Simmer C., 2005, Atmos. Res., 75, 183
Saunders W. et al., 2009, PASP, 121, 976
Sims G. et al., 2012, PASP, 124, 74
Smette A., Horst H., Navarrete J., 2008, in Kaufer A., Kerber F., eds, ESO Astrophysics Symposia, The 2007 ESO Calibration Workshop. Springer-Verlag, Berlin, p. 433
Soden B. J., Bretherton F. P., 1993, J. Geophys. Res., 96, 16669
Stein J., Richard E., Lafore J. P., Pinty J. P., Asencio N., Cosma S., 2000, Meteorol. Atmos. Phys., 72, 203
Tremblin P., Schneider N., Minier V., Durand G. Al., Urban J., 2012, A&A, 548, A65
Turchi A., Masciadri E., Fini L., 2017, MNRAS, 466, 1925
Vernet J. et al., 2011, A&A, 536, A105

APPENDIX A: CERRO PARANAL CLIMATOLOGY

We report here the cumulative distributions of PWV obtained in 2013 and 2017 on the full year (all available LHATPRO measurements) and the 120 nights samples used in the analysis. We also report the cumulative distributions obtained by MNH simulations on the same samples, with correction in equation (6).

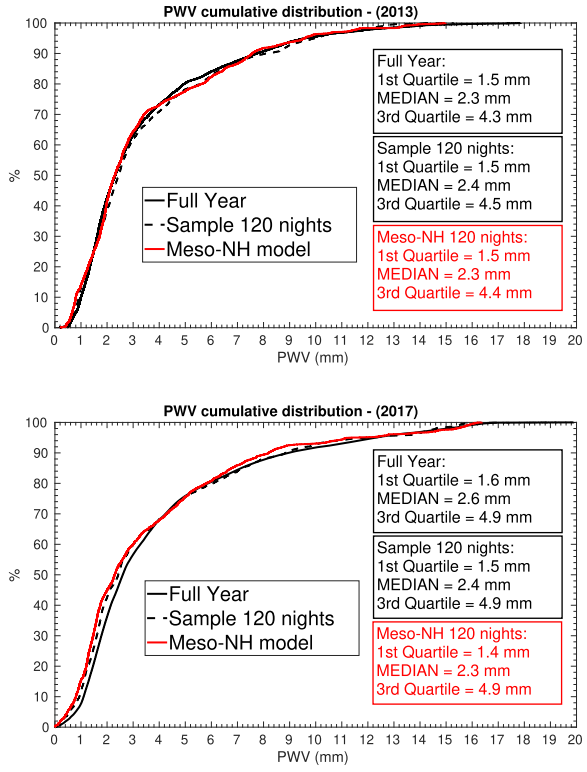

Figure A1. Cerro Paranal – cumulative distributions of PWV values in the years 2013 (top) and 2017 (bottom). We report the cumulative distribution computed over the LHATPRO measuraments on the whole year (full black line), the LHATPRO measurements over the 120 nights sample selected in each year (dashed black line) and the MNH forecasts on the same sample (full red line), which were used to compute the statistical indicators in Section 4.1. In the boxes in each figure we report the median, first quartile and third quartile values computed over the respective distributions.

APPENDIX B: SAMPLE 35 NIGHTS 2017

In this annex, we report the 35 nights selected in 2017, with at least 30 per cent of measured PWV $\leq$ 1 mm between 00:00 and 09:00 UT, whose analysis is reported in Fig. 3. The list is presented for reference in case of further studies on the selected sample of low PWV nights. The occurence of PWV $\leq$ 1 mm is more likely during the (southern) winter period; however, it may happen along the whole year. Since the uneven distribution of this phenomenon through the year we selected the present sample separately from the rest.

Table B1. Listof the 35 nights in 2017 yr (start date of the night in UT time, yyyy/mm/dd) with at least 30% of measured PWV$\leq$1 mm between 00:00 and 09:00 UT.

2017/04/28	2017/05/20	2017/05/26	2017/05/30
2017/07/15	2017/07/16	2017/07/17	2017/08/01
2017/08/02	2017/08/12	2017/08/26	2017/08/27
2017/09/04	2017/09/05	2017/09/06	2017/09/07
2017/09/10	2017/09/11	2017/09/12	2017/09/13
2017/09/14	2017/09/15	2017/09/20	2017/09/21
2017/09/25	2017/09/26	2017/10/01	2017/10/07
2017/10/09	2017/11/17	2017/11/18	2017/11/22
2017/11/29	2017/12/07	2017/12/08	

This paper has been typeset from a TEX/LATEX file prepared by the author.