Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **466,** 1925–1943 (2017) doi:10.1093/mnras/stw2863

Forecasting surface-layer atmospheric parameters at the Large Binocular Telescope site

Alessio Turchi,★ Elena Masciadri★ and Luca Fini
INAF - Osservatorio Astrofisico di Arcetri, L.go E. Fermi 5, I-50125 Florence, Italy

Accepted 2016 November 3. Received 2016 October 14; in original form 2016 September 2

ABSTRACT

In this paper, we quantify the performance of an automated weather forecast system implemented on the Large Binocular Telescope (LBT) site at Mt Graham (Arizona) in forecasting the main atmospheric parameters close to the ground. The system employs a mesoscale non-hydrostatic numerical model (Meso-Nh). To validate the model, we compare the forecasts of wind speed, wind direction, temperature and relative humidity close to the ground with the respective values measured by instrumentation installed on the telescope dome. The study is performed over a large sample of nights uniformly distributed over 2 yr. The quantitative analysis is done using classical statistical operators [bias, root-mean-square error (RMSE) and σ] and contingency tables, which allows us to extract complementary key information, such as the percentage of correct detections (PC) and the probability of obtaining a correct detection within a defined interval of values (POD). The results of our study indicate that the model performance in forecasting the atmospheric parameters we have just cited are very good, in some cases excellent: RMSE for temperature is below 1°C, for relative humidity it is 14 per cent and for the wind speed it is around 2.5 m s^{-1}. The relative error of the RMSE for wind direction varies from 9 to 17 per cent depending on the wind speed conditions. This work is performed in the context of the ALTA (Advanced LBT Turbulence and Atmosphere) Center project, whose final goal is to provide forecasts of all the atmospheric parameters and the optical turbulence to support LBT observations, adaptive optics facilities and interferometric facilities.

Key words: turbulence – atmospheric effects – instrumentation: adaptive optics – methods: data analysis – methods: numerical – site testing.

1 INTRODUCTION

The study presented in this paper was developed in the context of the Advanced LBT Turbulence and Atmosphere (ALTA Center Project 2016; Masciadri, Vernin & Bougeault, 1999; Masciadri et al. 2016) project, which aims to implement an automated forecast system for the Large Binocular Telescope (LBT) using non-hydrostatic mesoscale atmospheric models. The purpose of ALTA is to forecast classical atmospheric parameters (wind speed and direction, temperature and relative humidity) that are relevant for ground-based astronomy and astroclimatic parameters (C_N^2 profiles, seeing ϵ, isoplanatic angle θ_0 and wavefront coherence time τ_0) to support ground-based observations by LBT. The forecasts of all these parameters are crucial for the telescope operations and are relevant for adaptive optics (AO) applications.

In the context of this extended project, the goal of this paper is to validate the forecasts of atmospheric parameters [temperature, relative humidity (RH), wind speed and direction] close to the ground above Mount Graham (Arizona), the site of the LBT, by comparing the outputs of the model (i.e. the forecast system) with measurements of the same parameters (stored in the telescope telemetry) taken by instruments placed on the telescope. More precisely, this paper quantifies the confidence level of the model predictions of the parameters just cited.

Knowing in advance the value of these parameters close to the ground is crucial for maximizing the efficiency of telescope operations and scheduling the planned observations. We refer the reader to section 2 of Masciadri, Lascaux & Fini (2013), which extensively explains how the atmospheric parameters close to the ground play a fundamental role in optimizing the ground-based observations of telescope facilities, particularly if they are supported by AO systems. Here we summarize the main arguments. The dome seeing, one of the main contribution to the total seeing, is proportional to the temperature gradient between the primary mirror and the dome temperature ΔT_M and between the dome temperature and the external temperature ΔT_I. Therefore, knowing in advance the temperature close to the ground is fundamental for minimizing the thermal gradients and, as a consequence, the dome seeing. The wind speed close to the ground is the main cause of vibrations of

★ E-mail: aturchi@arcetri.astro.it (AT); masciadri@arcetri.astro.it (EM)

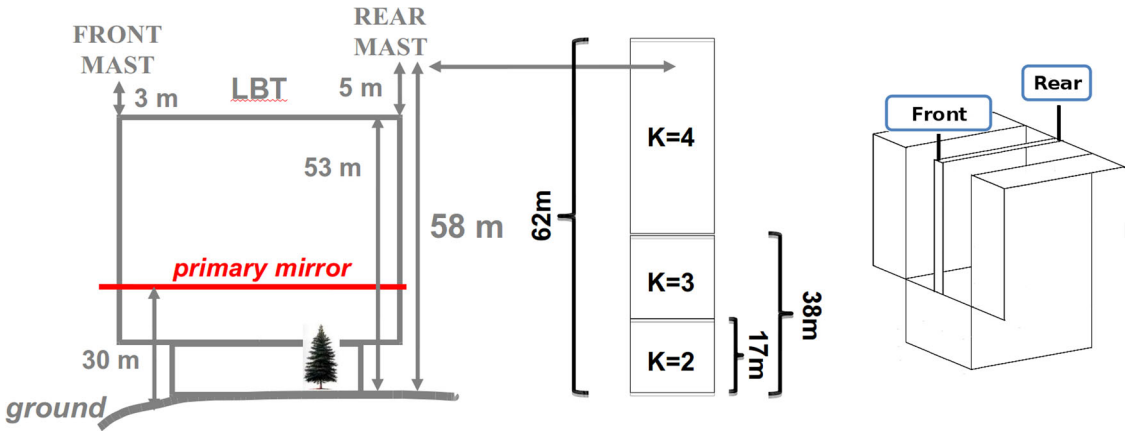

Figure 1. Large Binocular Telescope dome scheme with weather mast locations. We report also the model levels $K = 2$, 3 and 4 with the corresponding heights.

the telescope mirrors. The noise produced by wind bursts is one of the main causes of error introduced in AO systems. This effect is mostly visible when the wind hits the mirrors frontally, while it is minimal when hitting them laterally. Knowing precisely in advance the wind direction helps in selecting the suitable part of the sky for observations to minimize the impact of strong winds on AO systems. Also, the wind direction is known to be correlated well to seeing conditions. The RH forecast is very useful for closing the dome when the RH reaches values larger than a fixed threshold. Of course, all these elements, jointly with a forecast of the optical turbulence (not analysed in this paper), will contribute to optimizing the scheduling of telescope observations and the telescope scientific outputs.

The ALTA Center is an integral part of the new strategy conceived by the LBT Observatory (LBTO) to optimize the science operations of the LBT telescope in the near future (Veillet et al. 2016). Commissioning of the ALTA project is articulated in different phases. By the term 'commissioning', we mean validating the model forecasts or, equivalently, estimating the model performance in predicting specific parameters. The results in this paper are basically a certification of how good or bad the model performances are with respect to the atmospheric parameters close to the ground. The users and staff responsible for LBT scheduling can now take advantage of forecasts of the atmospheric parameters by knowing how the model performs.

The numerical models used in the ALTA project to forecast the atmospheric parameters are the Meso-Nh model (Lafore et al. 1998) developed by the Centre National des Recherches Météorologiques (CNRM) and Laboratoire d'Aérologie (LA) and the Astro-Meso-Nh module (Masciadri et al. 1999), which is used to provide forecasts of optical turbulence parameters. We limit this paper to the former.

Among the first studies on using non-hydrostatic mesoscale models for forecasting atmospheric parameters close to the surface for astronomical applications, we highlight Masciadri, Vernin & Bougeault (2001) and Masciadri (2003). At that epoch, the employment of sub-kilometric horizontal resolutions was particularly relevant in the simulations (in the context of astronomical applications). More recently, further attempts followed with the same mesoscale model, Meso-Nh (Lascaux et al. 2009), above the Antarctic plateau of Dome C and with the WRF model (Giordano et al. 2013)[1] above Roque de los Muchchos. More recently, forecasting atmospheric parameters was part of a large validation campaign conducted within the MOSE project, commissioned by the European Southern Observatory (ESO) and applied to Cerro Paranal and Armazones in Chile using the Meso-Nh model (Masciadri et al. 2013). In that context, a detailed study was carried out on the performance of the Meso-Nh model in forecasting the most important atmospheric parameters close to the ground (Lascaux, Masciadri & Fini 2013, 2015). For the first time, an exhaustive statistical analysis including statistical operators such as the percentage of correct detections and the probability of detection in a specific range of values has been presented, providing evidence for excellent model performance (to our knowledge, the best ever achieved in this field) for basically all the parameters.

In this work, we aim to perform a model validation of the surface atmospheric parameters calculated on a sample of 144 nights uniformly distributed between 2014 and 2015. We intend to use the same strategy used by Lascaux et al. (2015) but on a different site. This is the first study of the performance of a non-hydrostatic mesoscale model applied to an astronomical site in which the model is already running in an operational system. Besides, we present a characterization of the surface-layer atmospheric parameters above the site of LBT. To our knowledge, there are no published results in the literature and this is a fundamental first step for our analysis.

In Section 2, we give an overview of the LBT site, measurement characteristics, sample selection criteria and model setup. In Section 3, we present a synthesis of the climatological analysis of the surface atmospheric parameters: temperature, RH, wind speed and direction. In Section 4, we describe the strategy for the statistical analysis performed in this study. In Section 5, we show the results of the model validation, in terms of statistical operators and contingency tables. In Section 6, we draw the conclusions and perspectives.

2 NUMERICAL ANALYSIS OVERVIEW

2.1 Observations

Measurements used as a reference are obtained from the weather stations positioned on two masts placed above the telescope dome and afterwards stored in the telemetry of LBT. Fig. 1 shows the telescope and the associated weather station instruments. The LBT dome has a flat roof that is 53 m above the ground. The mast labelled 'front' is positioned on the front side of the dome, facing the telescope line-of-sight direction. This mast has only one anemometer measuring wind speed and wind direction, placed at a height

[1] This study used only a kilometric horizontal resolution.

Table 1. Selected nights (in UT, yyyy/mm/dd) for preliminary model validation.

2014/01/01	2014/01/06	2014/01/11	2014/01/16	2014/01/21	2014/01/26	2014/02/01	2014/02/05
2014/02/11	2014/02/16	2014/02/21	2014/02/26	2014/03/01	2014/03/06	2014/03/11	2014/03/16
2014/03/21	2014/03/26	2014/04/01	2014/04/06	2014/04/11	2014/04/16	2014/04/21	2014/04/26
2014/05/01	2014/05/06	2014/05/11	2014/05/15	2014/05/21	2014/05/26	2014/06/01	2014/06/06
2014/06/11	2014/06/16	2014/06/21	2014/06/26	2014/07/01	2014/07/06	2014/07/10	2014/07/16
2014/07/22	2014/07/26	2014/08/01	2014/08/06	2014/08/11	2014/08/16	2014/08/19	2014/08/28
2014/09/01	2014/09/06	2014/09/11	2014/09/16	2014/09/21	2014/09/26	2014/10/01	2014/10/06
2014/10/11	2014/10/16	2014/10/21	2014/10/26	2014/11/01	2014/11/06	2014/11/11	2014/11/16
2014/11/21	2014/11/26	2014/12/01	2014/12/06	2014/12/11	2014/12/16	2014/12/21	2014/12/26
2015/01/01	2015/01/06	2015/01/10	2015/01/16	2015/01/21	2015/01/26	2015/02/04	2015/02/06
2015/02/11	2015/02/16	2015/02/21	2015/02/26	2015/03/06	2015/03/07	2015/03/11	2015/03/16
2015/03/21	2015/03/26	2015/04/01	2015/04/06	2015/04/11	2015/04/16	2015/04/21	2015/04/28
2015/05/06	2015/05/07	2015/05/11	2015/05/18	2015/05/21	2015/05/26	2015/06/01	2015/06/06
2015/06/11	2015/06/15	2015/06/21	2015/06/26	2015/07/01	2015/07/06	2015/07/11	2015/07/16
2015/07/21	2015/07/26	2015/08/14	2015/08/16	2015/08/17	2015/08/18	2015/08/21	2015/08/24
2015/09/02	2015/09/06	2015/09/12	2015/09/16	2015/09/21	2015/09/26	2015/10/10	2015/10/16
2015/10/24	2015/10/25	2015/10/26	2015/10/28	2015/11/03	2015/11/08	2015/11/10	2015/11/14
2015/11/24	2015/11/25	2015/12/05	2015/12/11	2015/12/14	2015/12/19	2015/12/21	2015/12/30

Table 2. Total number of nights used to validate each atmospheric parameter, together with the nights excluded from the total sample of 144 nights from Table 1.

Parameter	Number of nights	Excluded nights
Temperature	144	–
Relative humidity	143	2015/09/21
Wind speed	139	2014/02/01 2014/03/01 2014/11/16 2014/12/26 2015/01/01
Wind direction	136	2014/02/01 2014/03/01 2014/07/16 2014/07/22 2014/11/16 2014/12/26 2015/01/01 2015/03/26

of 3 m above the roof (i.e. 56 m above the ground). The second mast labelled 'rear' is placed on the rear side of the dome and is equipped with an anemometer, 5 m above the roof (i.e. 58 m above the ground), and a set of sensors measuring temperature, RH and pressure, 2.5 m above the roof (i.e. 55.5 m above the ground). The weather stations send data to the telemetry streams approximately each second, which are stored in the LBT archive. We checked that the rear and front anemometers provide the same measurements for the wind direction, thus for this specific parameter, we used measurements from the rear anemometer for the analysis presented in this paper. For the wind speed, we detected discrepancies between the two sensors, depending on the direction of the incoming wind. We conclude that this was due to the position of the anemometers relative to the LBT roof and we found a method to disentangle the biased measurements and to discard them. We have been forced, therefore, to use both front and rear sensors to reconstruct a reliable measure of this parameter. In Section 5.3, we discuss the actual procedure for selecting the wind speed measurements.

2.2 Sample selection

To validate the model, we selected a rich sample of 144 nights uniformly distributed between 2014 and 2015. This sample is rich enough to be statistically significant and permits us to perform a reasonable number of simulations with a homogeneous model configuration. The selection criterion was, therefore, the following: starting from 2014 January 1, we selected a date every roughly 5 d with the full night data available. When the telemetry data were missing, we selected the closest night with full night data available. This criterion permitted us to have around six nights for each month. The selected dates are reported in Table 1. We observe that the telemetry archive lacks some measurements, mainly due to temporary failures of the sensors. Telemetry streams have flags that signal if the corresponding instrument is working and online; however, when the night sample was already selected, we noticed that there were cases in which the instrument was clearly malfunctioning while the corresponding flag was signalling that it was working properly. Specifically, we had to remove one night from the RH measurement sample, five nights from the wind speed sample and eight nights from the wind direction sample. Refer to Table 2 for the number of nights used to validate each atmospheric parameter considered in this analysis.

Besides, we noticed that the RH values extracted from the telemetry were distributed between a minimum of 4 per cent to a maximum of 104 per cent (exact numbers). Values larger than 100 per cent are obviously not realistic. We do not know the cause of this problem. We, therefore, considered two cases in our analysis. In case A, we assume that the sensor has a bias of +4 per cent and we subtract 4 per cent from all measurements. In case B, we simply disregard all measurements between 100 per cent and 104 per cent. In both cases, measurements are included in the interval [0 per cent, 100 per cent].[2]

2.3 Model configuration

The numerical model used to produce the forecasts of the aforementioned parameters is Meso-Nh,[3] which is an atmospheric non-hydrostatic mesoscale model that simulates the time evolution of

[2] For completeness, besides cases A and B, we may assume a case C in which the problem causing a RH larger than 100 per cent may affect also the other measurements in an unknown way. However, this should be equivalent to assuming that no measurements are reliable. This does not seem to be the case, as we will see later.

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

Table 3. Horizontal resolution of each Meso-Nh imbricated domain.

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
Domain 4	0.1	100 × 100	10 × 10

weather parameters in a three-dimensional volume over a finite geographical area. The coordinate system is based on a Mercator projection, which is most suitable at low latitudes as in the LBT case, while the vertical levels use the Gal-Chen and Sommerville coordinate systems (Gal-Chen & Sommerville 1975). The model filters acoustic waves thanks to an anelastic formulation of the hydrodynamic equations. Simulations are made using a one-dimensional mixing length proposed by Bougeault and Lacarrere (1989) with a one-dimensional 1.5 closure scheme (Cuxart, Bougeault & Redelsperger 2000). The exchange between the surface and the atmosphere is computed with the Interaction Soil Biosphere Atmosphere scheme (Noilhan & Planton 1989).

The model is initialized with forecast data provided by the European Centre for Medium Weather Forecasts (ECMWF), calculated with their hydrostatic General Circulation Model (HRES) extended over the whole globe, with a horizontal resolution of 16 km.[4] Initialization of the Meso-Nh model starts each night at 00:00 UT (i.e. 17:00 MST) and the Meso-Nh model is forced every 6 h with new data from the ECMWF. We simulate a total of 15 h up to 15:00 UT (i.e. 08:00 MST) to cover completely the night-time period of Mt Graham. This is exactly the procedure used in the operational configuration.

The Mt Graham site of LBT is located at coordinates [32.70131, −109.88906], at a height of 3221 m above sea level. We used a grid-nesting technique (Stein et al. 2000), which consists of using different imbricated domains, described in Table 3, with a digital elevation model (DEM, i.e. orography) extended on smaller and smaller surfaces having a progressively higher horizontal resolution. This procedure lets us achieve a high horizontal resolution, using the same vertical grid resolution, on a sufficiently small region around the summit of interest to provide better model predictions at the specific site. Each domain is centred on the LBT coordinates. A graphic representation of the model domains is given in Fig. 2. Lascaux et al. (2013) proved that a horizontal resolution of 100 m is necessary at Cerro Paranal and Cerro Armazones to reconstruct the wind speed close to the ground when the wind is strong. In our study, we compare the results obtained with a 500-m resolution with those obtained with a 100-m resolution for the wind speed to verify if similar conditions are found also above Mt Graham. For the other atmospheric variables (temperature, RH and wind direction), we used a resolution of 500 m, which is sufficient to reconstruct reliable values of these parameters.

For the orography of domains 1 and 2, we used the GTOPO[5] DEM, with an intrinsic resolution of 1 km. In domains 3 and 4, we utilized the SRTM90[6] DEM (Jarvis et al. 2008), with an intrinsic resolution of approximately 90 m (3 arcsec). The resolution is obviously defined by the number of grid points of the atmospheric model, therefore it is 500 and 100 m in our case.

For all the couples of imbricated domains, we use a two-way interacting grid-nesting, in which the interface between the outer and inner domains has a bidirectional interaction. This allows the atmospheric flow between each domain to be in constant thermodynamic equilibrium with the outer one, allowing for the propagation of gravity waves through the whole area mapped by the simulation independently of the specific domain.

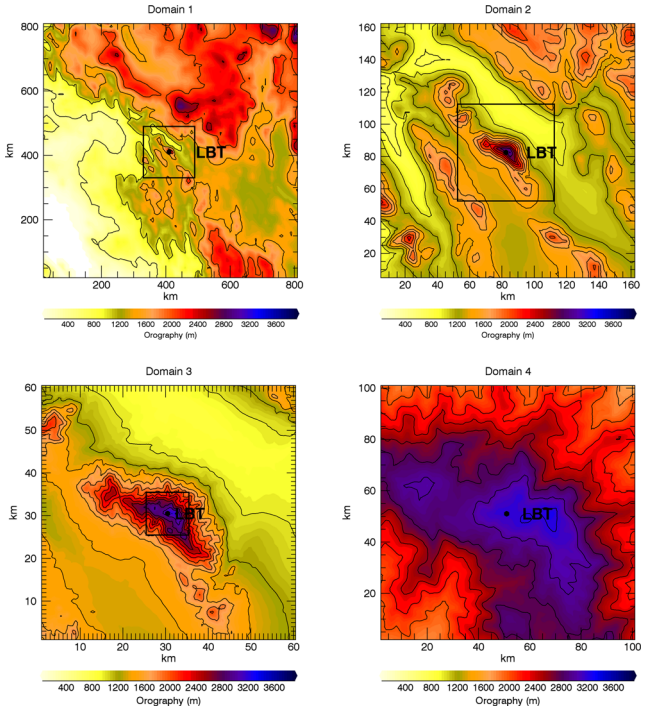

Figure 2. Horizontal maps of model nested domains, from the outer to the innermost domain, as reported in Table 3. The black dot in the centre of each domain corresponds to the position of LBT, while the colour scale represents the orography elevation.

The simulation has 54 physical vertical levels on each domain, with the first grid point set to 20 m above ground level (a.g.l.), and a logarithmic stretching of 20 per cent up to 3.5 km a.g.l. following:

$$\frac{\Delta z(k+1)}{\Delta z(k)} = 1 + \frac{20}{100}. \quad (1)$$

From this point onwards, the model uses an almost constant vertical grid size of ∼600 m up to 23.57 km, which is the top level of our domain.

As reported in Section 2.1, LBT weather stations are positioned ∼55–58 m above ground, and this corresponds to the third physical Meso-Nh level ($K = 4$, Fig. 1) spanning the interval [38–62] m at the LBT site coordinates. The model output for each level from the innermost model is representative of the whole interval spanned by the level itself.

The model has been configured to give access to the temporal evolution of the surface parameters calculated on the summit (LBT location) with a temporal frequency equal to the time step (order of a few seconds) of the innermost domain.

3 CLIMATOLOGIC CHARACTERIZATION OF THE SURFACE-LAYER ATMOSPHERIC PARAMETERS

We report in this section the climatology distributions of the atmospheric parameters (temperature, wind speed and direction, and RH) in the surface layer at the summit of Mt Graham. To our

[4] From 2016 March, the resolution increased to roughly 9 km.
[5] https://lta.cr.usgs.gov/GTOPO30
[6] http://www.cgiar-csi.org/data/srtm-90m-digital-elevation-database-v4-1

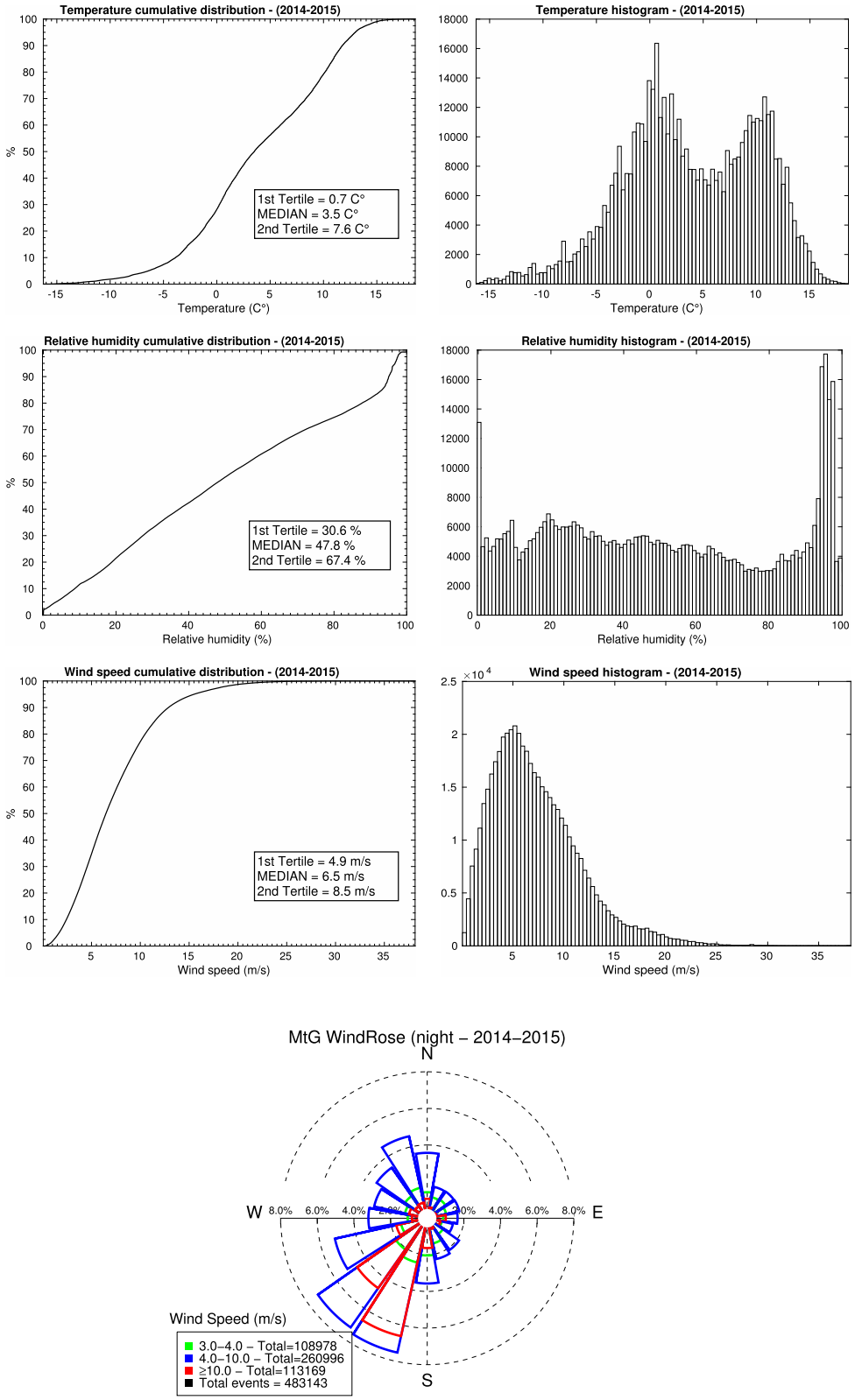

Figure 3. Cumulative distributions, histograms and wind rose computed for the whole years 2014 and 2015, considering measurements between sunrise and sunset.

knowledge, there are no published results on the characterization of these parameters above Mt Graham. This should be, therefore, the first in this respect. We report the cumulative distribution and histogram for each parameter, for the full years 2014 and 2015 (730 nights) considered in this analysis (Fig. 3), for the summer (April–September, Fig. 4) and for the winter (October–March, Fig. 5) of the same date sample. For each date, we extract the subsample of data included between sunset and sunrise. After

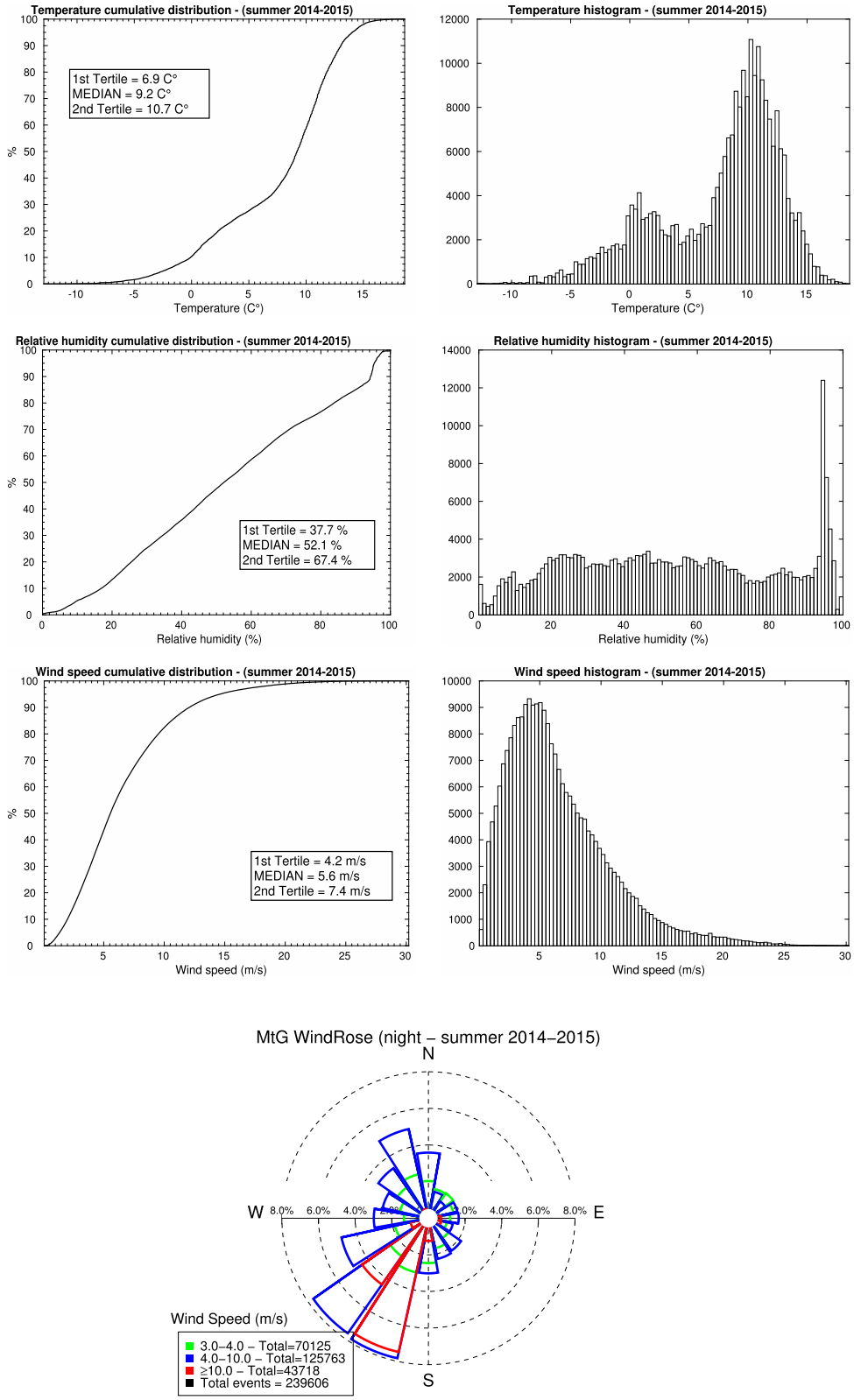

Figure 4. Cumulative distributions, histograms and wind rose computed for the summers (April–September) of 2014 and 2015, considering measurements between sunrise and sunset.

the above procedure, the distributions are computed. Tertiles and median values are calculated for each parameter.

In Fig. 3, we observe that temperature distribution has two distinct peaks at around ∼10°C and ∼1°C. By observing the partial temperature distributions in Figs 4 and 5 relative to summer and winter, we can easily detect that the two temperature peaks are related to the median temperature in summer and winter, respectively. However, in Fig. 5, it is evident that there is a secondary small

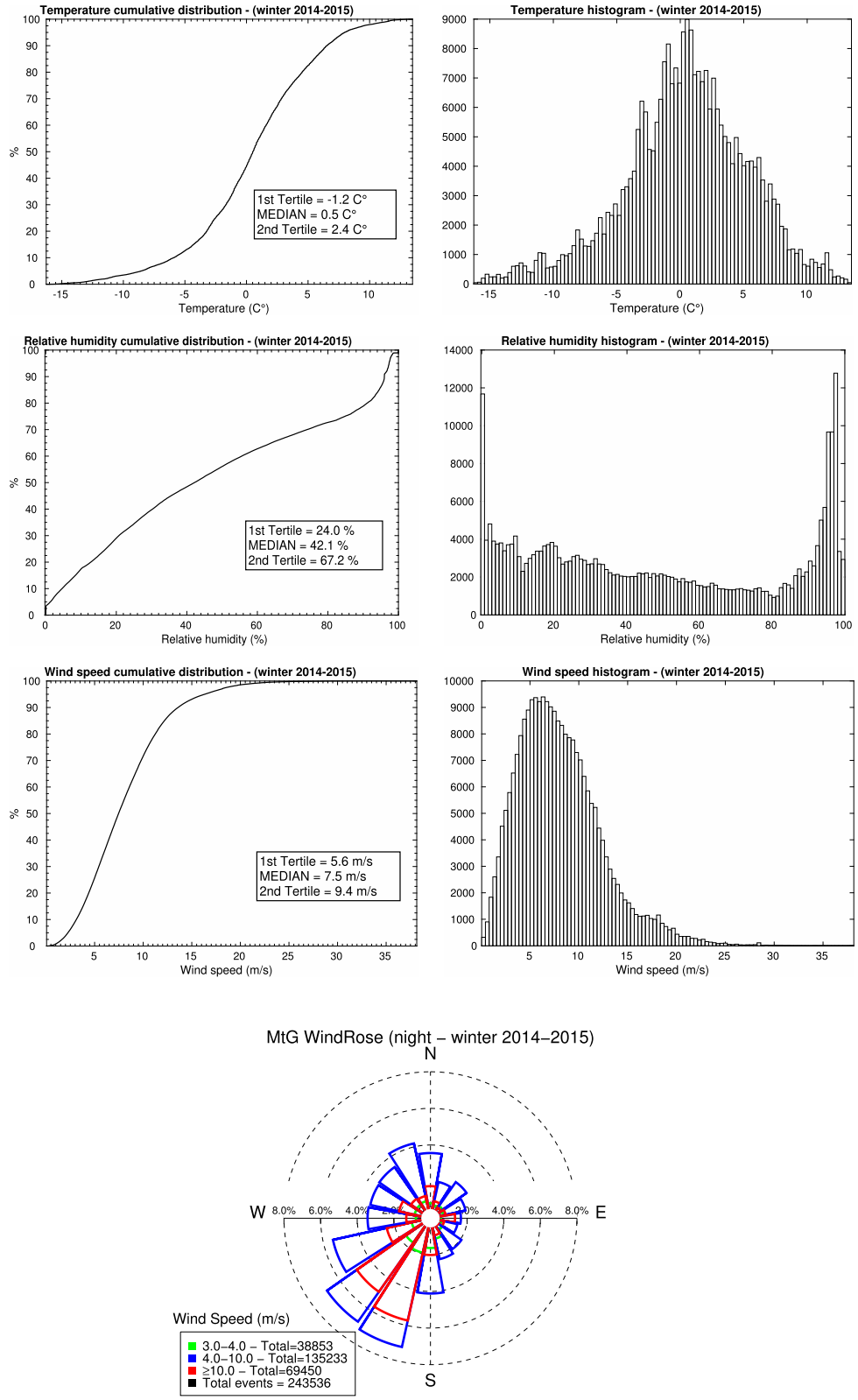

Figure 5. Cumulative distributions, histograms and wind rose computed for the winters (October–March) of 2014 and 2015, considering measurements between sunrise and sunset.

peak at around ∼1°C meaning that even in summer there are times when the outside temperature is near or below the freezing point. The RH distribution shows that there are numerous times, both in summer and winter, when the values are near ∼100 per cent, mostly related to rainy events. However, the driest season is winter, while the summer tends to have a flatter distribution of RH values. The median value of RH for the whole year is 47.8 per cent. In winter it is 42.1 per cent and in summer it is 52.1 per cent.

Table 4. Total sample size [couples of points ($p_{i,\mathrm{sim}}$, $p_{i,\mathrm{obs}}$)] for each atmospheric parameter related to the total 144 nights from Table 1. For the wind direction, we considered different samples because we filtered out values associated with a wind speed (WS) larger than 3, 5 and 10 m s^{-1}.

Parameter	Sample size
Temperature	5201
Relative humidity	5165
Wind speed	4996
Wind direction (WS $>$ 3 m s^{-1})	4423
Wind direction (WS $>$ 5 m s^{-1})	3626
Wind direction (WS $>$ 10 m s^{-1})	1311

Regarding the wind speed distribution, we observe that the median value for the whole year is 6.5 m s^{-1} while in winter it is $\sim$7.5 m s^{-1} and in summer it is $\sim$5.6 m s^{-1}. The percentage of times when the wind speed is larger than 15 m s^{-1} is 6 per cent (with no major differences between winter and summer) and when larger than 20 m s^{-1}, the percentage is 1 per cent.

The wind direction distribution does not show much seasonal variability. From its distribution, we observe that strong winds (above 10 m s^{-1}, those most critical for astronomers) are mainly from the south-west direction; however, there is always a chance of observing strong winds from any direction.

4 ANALYSIS STRATEGY

As already done in recent studies (Lascaux et al. 2013, 2015) on Cerro Paranal and Cerro Armazones,[7] we decided to perform a moving average over a 1-h time window, from 30 min before to 30 min after, both on measurements and numerical forecasts. This allows for the filtering of fast frequencies and enables us to estimate the model performance over the slower-moving trends, which are of interest for the telescope operation and planning. Data were then resampled over 20-min intervals. This value was selected because the time required to switch a beam from an instrument to another one (or a program to another one) is typically of the order of 20 min. It makes, therefore, no sense to use a higher frequency. We highlight that, even if we calculate a moving average that smooths out the high frequencies, it is extremely important to have a high-frequency output from the model (as in our case) instead of treating outputs with a temporal sampling of 1 h (as is the case in many studies in the literature), since this allows for a more reliable estimate of the atmospheric quantities over the selected time window. Finally, we selected values only between sunset and sunrise, computed with an ephemeris for each date.

In Table 4, we report the total sample size [number of couples of points ($p_{i,\mathrm{sim}}$, $p_{i,\mathrm{obs}}$)] considered for each atmospheric parameter, once the procedure described in this section was applied. For the wind direction, we considered different samples in which we filtered out values that correspond to wind speeds lower than 3, 5 and 10 m s^{-1} (see discussion in Section 5.4).

The validation procedure used in this paper, similarly to what has already been done for ESO telescope sites by (Lascaux et al. 2015), followed two different statistical approaches. First, we computed the classical statistical operators [bias, root-mean-square error (RMSE) and σ], defined as follows:

$$\mathrm{Bias} = \sum_{i=1}^{N} \frac{(Y_i - X_i)}{N} \tag{2}$$

and

$$\mathrm{RMSE} = \sqrt{\sum_{i=1}^{N} \frac{(Y_i - X_i)^2}{N}}, \tag{3}$$

where X_i are the individual observations and Y_i the individual numerical forecasts calculated at the same time index i. N is the total sample size.

From the above quantities, we deduce the bias-corrected RMSE (σ):

$$\sigma = \sqrt{\mathrm{RMSE}^2 - \mathrm{Bias}^2}. \tag{4}$$

The previously defined indicators provide global information on the statistical and systematic errors of the model.

To refine the statistical analysis and give a more practical estimate of the model score of success, we proceeded to perform also a different analysis based on contingency tables, which are able to provide complementary information, with respect to the above statistical operators, on the reliability of the model in a realistic use-case scenario. A contingency table is a method of studying the relationship between two or more categorical variables and it provides multiple statistical operators: the percentage of correct detections (PC), the probability of detecting a parameter within a specific range of values (POD$_i$) and the probability of an extremely bad detection (EBD). Refer to Lascaux et al. (2015) for an exhaustive discussion of contingency tables and a definition of all these statistical operators.

Table 5. Climatological distribution of atmospheric parameters at Mount Graham. Left column: first tertile (33 per cent), central column: median (50 per cent), right column: second tertile (66 per cent). Values are computed over the 2014 and 2015 distribution of observations from LBT telemetry.

Mount Graham	33%	Median (50%)	66%
Temperature (°C)	0.7	3.5	7.6
Relative humidity (%)	30.6	47.8	67.4
Wind speed (m s^{-1})	4.9	6.5	8.5

For temperature, wind speed and RH, we decided to use as categories the tertiles of the cumulative distribution of these parameters (3 $\times$ 3 contingency tables) in which the thresholds are the tertiles of a climatological analysis performed on measurements (see Table 5). An exception was made for the wind direction, which was divided into quadrants defined as follows: north = [315°, 45°], east = [45°, 135°], south = [135°, 225°] and west = [225°, 315°] (4 $\times$ 4 contingency table). As can be seen in Section 5, the selected sample of nights (144) on which we performed the model validation covers all the conditions revealed by the climatological analysis. In other words, there are no biased effects.

5 MODEL VALIDATION

In this section, we report the results obtained with the procedure described in Section 4. The validation is performed over the full sample of nights distributed over the whole two years 2014 and 2015, with actual sample sizes for each parameter reported in Tables 1 and 2. Here we report the RMSE, bias and σ obtained

[7] A decision taken in agreement with the ESO and LBTO staff.

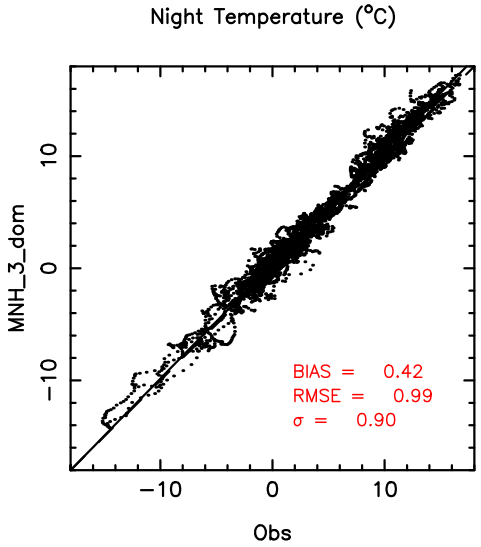

Figure 6. Scatter plot for temperature, comparing model outputs (MNH) and measurements (Obs). The full black line is the regression line passing by the origin, while the dashed line represents the reference diagonal line for unbiased results.

Table 6. 3 × 3 contingency table for the absolute temperature during the night, at 55.5 m a.g.l. at LBT, for the sample of 144 nights. We use the Meso-Nh $\Delta X = 500$ m configuration.

Temperature (°C)		Observations		
		$T < 0.7$	$0.7 < T < 7.6$	$T > 7.6$
Model	$T < 0.7$	1381	79	0
	$0.7 < T < 7.6$	267	1841	20
	$T > 7.6$	0	67	1546

Sample size: 5201; PC: 91.7 per cent; EBD: 0.0 per cent; POD_1: 83.8 per cent; POD_2: 92.7 per cent; POD_3: 98.7 per cent.

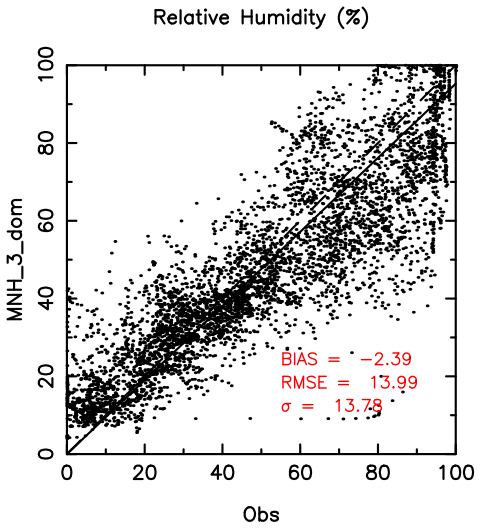

Figure 7. Scatter plot for RH, comparing model outputs (MNH) and measurements (Obs). The full black line is the regression line passing by the origin, while the dashed line represents the reference diagonal line for unbiased results.

Table 7. 3 × 3 contingency table for the RH during the night, at 55.5 m a.g.l. at LBT, for the sample of 143 nights. We use the Meso-Nh $\Delta X = 500$ m configuration.

Relative humidity (%)		Observations		
		RH < 30.6	30.6 < RH < 67.4	RH > 67.4
Model	RH < 30.6	1199	183	14
	30.6 < RH < 67.4	412	1475	649
	RH > 67.4	0	196	1037

Sample size: 5165; PC: 71.8 per cent; EBD: 0.3 per cent; POD_1: 74.4 per cent; POD_2: 79.6 per cent; POD_3: 61.0 per cent.

on the whole sample as well as the contingency tables with associated PC, POD and EBD values. The cumulative distributions of RMSE, bias and σ errors obtained considering each single night are reported in Appendix A. When not otherwise specified, model outputs are taken from the vertical level $K = 4$ (see Fig. 1), which is representative of the weather mast positions and in the grid point corresponding to the summit of Mt Graham in the innermost domain having a horizontal resolution of 500 m for temperature, RH and wind direction and a resolution of 100 m for the wind speed.

5.1 Temperature

In Fig. 6 is reported the scatter plot between observations and simulations calculated on the sample in Table 2. We can observe that all the statistical operators are well below 1°C (bias = 0.42°C, RMSE = 0.99°C and $\sigma = 0.9$°C) and this indicates excellent model performance. Table 6 is the contingency table obtained for the temperature parameter where the thresholds are the climatological tertiles defined in Table 5. We observe that the model has an impressive PC = 91.7 per cent and a zero EBD, as well as excellent POD_i (between 84 per cent and 99 per cent). These results confirm the excellent model performance as we found in a recent study carried out above Cerro Paranal on a sample of 129 nights (Lascaux et al. 2015).

If we consider a sample of bias, RMSE and σ calculated for each night and we calculate the cumulative distribution (see Appendix A), we observe that the median value of the bias, RMSE and σ of the temperature is well below 1°C too. This means that also if we look at the problem from a single night's perspective, the results are equivalently good.

If we consider the two separate subsamples of summer and winter, we find equivalent good model performance. Figs B1 and B2 (left panel) in Appendix B are scatter plots of temperature in summer and winter, while Tables B1 and B6 show the contingency tables calculated using, as thresholds, the tertiles associated with the respective subsamples. We can observe that the model performance remains good with no major differences observed in the two periods of the years. POD_i are well above 80 per cent with the unique exception of POD_2 in summer, which is 70.1 per cent, and in any case this is still very good.

5.2 Relative humidity

Relative humidity (RH) is the ratio (expressed in percentage) of the mixing ratio w to the saturation mixing ratio w_s with respect to water at the same temperature and pressure: $RH = (w/w_s) \times 100$. Fig. 7 reports the scatter plot of RH for the sample in Table 2, as done for the temperature. The bias = −2.4 per cent, RMSE = 14.0 per cent and $\sigma = 13.8$ per cent can be considered absolutely satisfactory. The points are well distributed along the regression line passing by the origin. We note that the dispersion of the cloud points increases with RH values. If we look at the contingency tables (Table 7), we observe a good global PC = 71.8 per cent and EBD = 0.3 per cent.

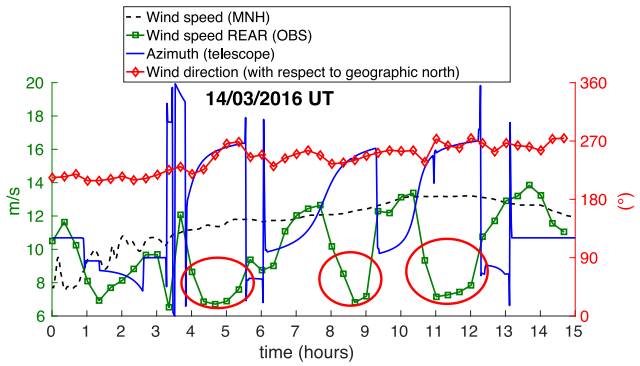

Figure 8. Time evolution of different parameters along the test night of 2016 March 14 UT. The full blue line represents the telescope azimuth (right *y*-axis showing the angles), the full red line with diamond dots represents the incoming observed wind direction, the full green line with square dots represents the wind speed measured on the rear anemometer, while the dashed black line represents the model output. The events in which the observed (rear) wind speed value drops drastically are highlighted with red circles. These events coincide with the telescope azimuth facing the incoming wind direction.

For POD_i, we find good values for $POD_1 = 74.4$ per cent and $POD_2 = 79.6$ per cent. POD_3 has decreased to 61 per cent. This matches with the increasing dispersion of the scatter plot for large RH (see Fig. 7). We note that POD_3 is the most interesting from an astronomical point of view. Astronomers are indeed interested in knowing when the RH is higher than the threshold for keeping the dome open for observations. Despite the more modest performance, 61 per cent is still well above the 33 per cent value of the random case. We highlight that these results have been obtained considering case A (see Section 2.2). We verified that case B provides negligible differences in the scatter plots and contingency table. The only interesting thing is that POD_3 is 56.2 per cent instead of 61 per cent. In conclusion, POD_3 is in the [56.2–61 per cent] range.

We observe that the rich sample of 144 night provides a much better result [56.2–61 per cent] with respect to what we had found in a preliminary study (Turchi, Masciadri & Fini 2016) in which POD_3 calculated on a small sample of 22 nights was 30 per cent.

If we consider a sample of bias, RMSE and σ calculated for each night and we calculate the cumulative distribution (see Appendix A), we observe that the median values of the bias, RMSE and σ are as good as those obtained on the whole sample.

Looking at the scatter plots of the subsamples of summer and winter (Figs B1 and B2, second panel), we do not observe any substantial difference between the periods. Also the contingency tables (Tables B2 and B7) for summer and winter confirm the substantially equivalent behaviour of the model with similar performance. POD_3 in summer is 62.8 per cent and in winter is 59.4 per cent (while POD_3 is 61 per cent for the total sample).

5.3 Wind speed

While studying wind speed measurements, we discovered that the front and rear anemometers tend to give significantly different measurements in some cases. The wind speed values tend to disagree when the wind is from specific directions: if the wind is from the front of the telescope, the rear anemometer tends to give a lower wind speed value with respect to the front one, while the opposite is true if the wind is from the back of the telescope. We report in Fig. 8 the observed wind speed behaviour for the test night of 2016 March 14 UT. It is evident that when the telescope orientation (azimuth) is coincident with the measured wind direction, so that the wind is from the front of the telescope, the wind speed measured on the rear anemometer suddenly drops more than 5 m s^{-1}. Also, the wind speed predicted by the model tends to agree with the anemometer that is facing the incoming wind direction. We interpreted this evidence as proof of a drag effect that slows down the wind passing over the dome.

Table 8. 3 × 3 contingency table for the wind speed during the night, at 58 m a.g.l. at LBT, for the sample of 139 nights. We use the Meso-Nh $\Delta X = 500$ m configuration.

Wind speed (m s^{-1})		Observations		
500-m resolution		WS < 4.9	4.9 < WS < 8.5	WS > 8.5
Model	WS < 4.9	847	388	24
	4.9 < WS < 8.5	368	1036	515
	WS > 8.5	46	365	1407

Sample size: 4996; PC: 65.9 per cent; EBD: 1.4 per cent; POD_1: 67.2 per cent; POD_2: 57.9 per cent; POD_3: 72.3 per cent.

Thanks to the above finding, we looked for a criterion to select wind speed measurements to be used in this validation study: the idea is that wind speed is taken from the front anemometer if the wind is within a specific incident angle α (measured by the telescope anemometers) with respect to the telescope orientation (azimuth), otherwise the measurement is taken from the rear anemometer (see Fig. 9, left side). To identify the width of this specific angle, we selected different angle apertures from which to select front anemometer measurements and we looked for when the RMSE was a minimum (see Fig. 9, right side). We observed that there is a saturation effect between an aperture of $\alpha = 60^\circ$ ($\pm 30^\circ$ from the telescope azimuth) and $\alpha = 140^\circ$ ($\pm 70^\circ$ from the telescope azimuth) where the RMSE, with respect to the model outputs, is a minimum. We also observed that selecting any value of α within the range $[60^\circ, 140^\circ]$ does not significantly change the values of the other statistical indicators used in this paper. For this reason, we decided to select front wind speed measurements if the incident angle of the wind is within $\alpha = 60^\circ$ with respect to the telescope azimuth, otherwise we select the rear measurements.

While the other parameters were all computed at 500-m horizontal resolution (see Table 3), in this section we report the results of the wind speed (measured with the criterion explained above) computed with two different horizontal resolutions for the sample reported in Table 2. In the left-hand panel of Fig. 10, we report the scatter plot computed with 500-m horizontal resolution and in the right-hand panel the plot computed with 100-m horizontal resolution. We observe that the overall result is slightly better at 500-m resolution, with negligible bias $= -0.2$ m s^{-1}, RMSE $= 2.4$ m s^{-1} and $\sigma = 2.4$ m s^{-1}, while at 100-m resolution, the bias $= 0.7$ m s^{-1} is slightly larger, yielding an RMSE $= 2.7$ m s^{-1} and $\sigma = 2.6$ m s^{-1}. However, as explained in Section 2.3 and in Lascaux et al. (2015), it is evident that the results at 500-m resolution tend to underestimate strong wind speeds (larger than 10 m s^{-1}).

The above result is confirmed in the contingency Tables 8 and 9, which report the results obtained with 500-m and 100-m resolution, respectively. In both cases, we obtain a good global PC $\simeq$ 65 per cent. At 500-m resolution, we have higher performance at low to medium wind speeds, with $POD_1 = 67.2$ per cent and $POD_2 = 57.9$ per cent, while at 100-m resolution the performance is $POD_1 = 61.0$ per cent and $POD_2 = 42.3$ per cent. The result is reversed in

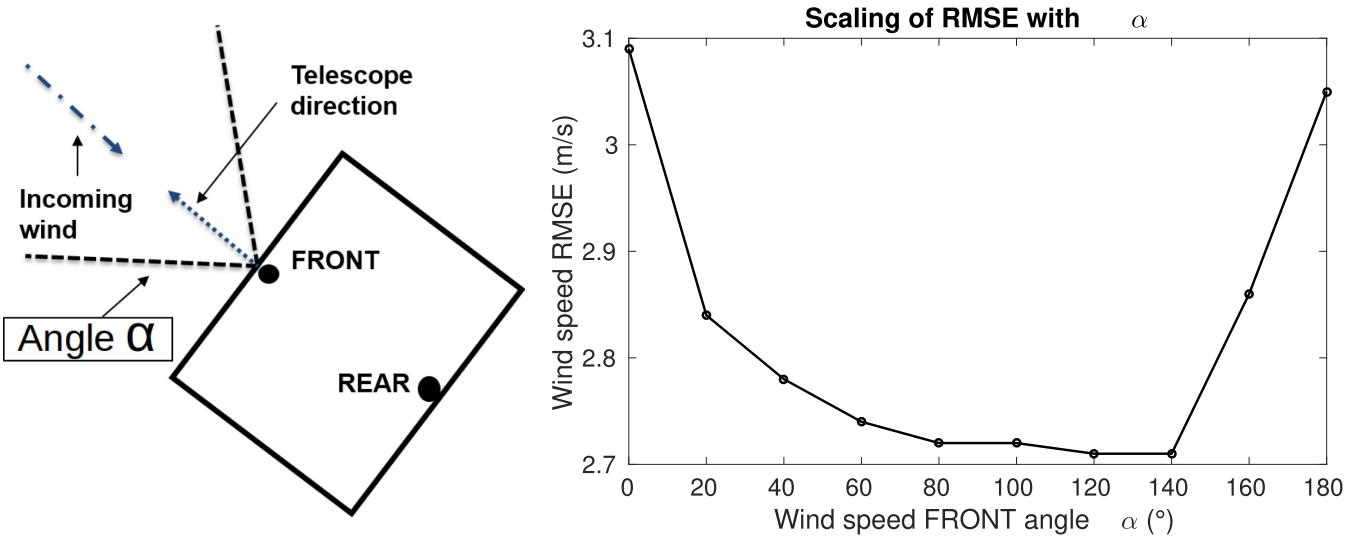

Figure 9. Left: Wind speed selection scheme. Wind speed is taken from the front anemometer if the wind is within incident angle α with respect to the telescope orientation (azimuth), otherwise the measurement is taken from the rear anemometer. Right: RMSE between model outputs and wind speed measurements with respect to the total angle α.

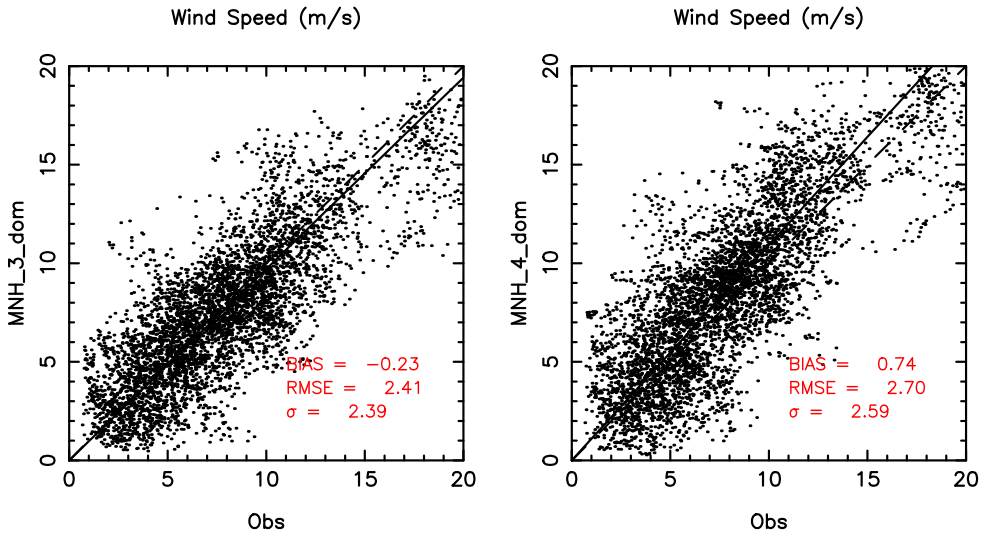

Figure 10. Scatter plot for wind speed, comparing model outputs (MNH) and measurements (Obs). The full black line is the regression line passing by the origin, while the dashed line represents the reference diagonal line for unbiased results. Left: Results obtained with the model output at 500-m horizontal resolution (domain 3). Right: Results obtained with the model output at 100-m horizontal resolution (domain 4).

Table 9. 3 × 3 contingency table for the wind speed during the night, at 58 m a.g.l. at LBT, for the sample of 139 nights. We use the Meso-Nh $\Delta X = 100$ m configuration.

Wind speed (m s^{-1}) 100-m resolution		Observations WS < 4.9	4.9 < WS < 8.5	WS > 8.5
Model	WS < 4.9	769	382	9
	4.9 < WS < 8.5	402	756	235
	WS > 8.5	90	651	1702

Sample size: 4996; PC: 64.6 per cent; EBD: 2.0 per cent; POD_1: 61.0 per cent; POD_2: 42.3 per cent; POD_3: 87.5 per cent.

the strong wind case, with a good $POD_3 = 72.3$ per cent at 500-m resolution that increases to an excellent $POD_3 = 87.5$ per cent at 100-m resolution. This is a logic result because it tells us that the higher resolution starts to play an important role when the wind speed is strong. We must also notice that in both cases the weaker performance reported in the POD_2 is caused by the fact that the interval between the first and second tertile is 3.6 m s^{-1}, which is slightly larger than the RMSE values.

Since the most interesting application, from an astronomical point of view, is to predict correctly ground-layer strong winds that would affect telescope operations, we definitely need 100-m horizontal resolution to obtain the best performance. However, since low wind speeds are better modelled by a 500-m horizontal resolution model, in the operative setup we will end up using a mixed forecast strategy, which will display results obtained with the 100-m resolution only for those nights characterized by strong wind speed. A test done on the sample we analysed in this paper tells us that a good strategy should be to take as a threshold 8.5 m s^{-1} (the second tertile of the wind speed distribution in Fig. 3). When the wind speed for the night is larger than this

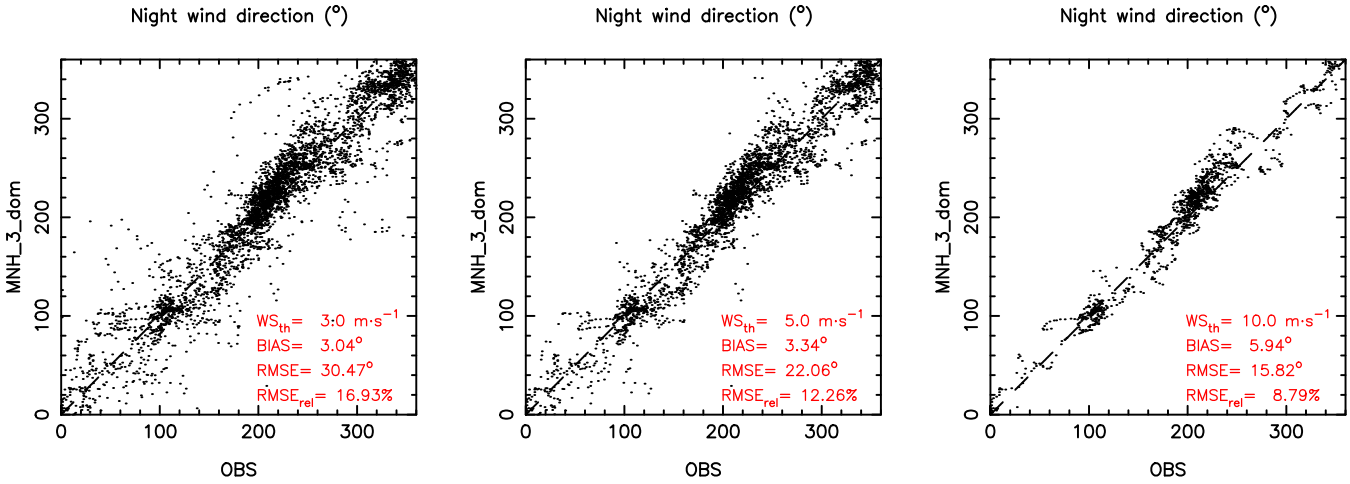

Figure 11. Scatter plots for wind direction, comparing model outputs (MNH) and measurements (OBS). The dashed line represents the reference diagonal line for unbiased results. Wind directions are filtered by the corresponding value of the wind speed. Left: Minimum wind speed of 3 m s^{-1}. Centre: Minimum wind speed of 5 m s^{-1}. Right: Minimum wind speed of 10 m s^{-1}.

threshold, we take the result of the model run at 100-m resolution, when the wind speed is lower we take results obtained by the model at 500-m resolution. In the operational procedure, we run, indeed, both models in sequence and we display the results from the 100-m or 500-m resolution whether the wind speed from the 100-m resolution is larger or weaker than 8.5 m s^{-1}. With the threshold at 8.5 m s^{-1}, we obtain POD$_1$ = 65.7 per cent, POD$_2$ = 45.0 per cent and POD$_3$ = 84.8 per cent, which provide an excellent POD$_3$. Also for the wind speed, we confirm the good model performance already observed above Cerro Paranal (Lascaux et al. 2015).

If we consider a sample of bias, RMSE and σ calculated for each night and we calculate the cumulative distribution (see Appendix A), we observe that the median values of the above parameters are lower than the the ones obtained on the whole sample. We obtain a median bias = 0.2 m s^{-1}, a median RMSE = 2.0 m s^{-1} and a median σ = 1.6 m s^{-1}. From the distributions, we observe that the nights in which the RMSE is larger than 4 m s^{-1} are fewer than 5 per cent of the total, with rare events (less than 1 per cent) in which the RMSE is larger than 6 m s^{-1}.

Looking at the scatter plots of the subsample of summer and winter (Figs B1 and B2, third panel) we do not observe any significant seasonal difference. Also the contingency tables (Tables B3, B4, B8 and B9) for summer and winter confirm an equivalent good behaviour of the model with similar performance.

5.4 Wind direction

In this analysis, we use the wind direction angle convention with respect to the geographical north, counting clockwise (0° = north, 90° = east). We compared the observed and simulated wind direction on different samples from which we filtered out all those measurements below a certain threshold, to compare model performance under different conditions. It is known indeed, that, when the wind speed is weak, the dispersion of the wind direction is high and it is harder to reconstruct the right wind direction. If we think about the realistic scenarios, from an astronomical point of view we are more interested in knowing the accuracy of the model in forecasting wind directions when the wind speed is high, because this impacts on telescope operations and instrument performance.

Table 10. 4 × 4 contingency table for the wind direction during the night, at 58 m a.g.l. at LBT. We use the Meso-Nh ΔX = 500 m configuration. Wind speed (WS) threshold is 3 m s^{-1}.

Wind direction WS > 3 m s^{-1}		Observations North	East	South	West
Model	North	737	53	20	145
	East	63	517	125	5
	South	21	63	1177	74
	West	85	1	434	903

Sample size: 4423; PC: 75.4 per cent; EBD: 1.1 per cent; POD$_N$: 81.3 per cent; POD$_E$: 81.5 per cent; POD$_S$: 67.0 per cent; POD$_W$: 80.1 per cent.

In this paper we decided to analyse the model performance with three different wind speed thresholds: 3, 5 and 10 m s^{-1}. The first threshold corresponds to negligible wind speed values, the second one is more or less close to the median value of the wind speed (6.5 m s^{-1}) while the last threshold corresponds to moderately high wind speed values. The total sample size in nights for the wind direction is reported in Table 2, while Table 4 shows the total sample size in data points, once the wind speed filter is applied.

As introduced in Lascaux et al. (2015), we calculated another statistical quantity RMSE$_{rel}$ = (RMSE/180°) × 100, i.e. the relative error of the RMSE calculated with respect to the worst possible prediction (reversed by 180°).

In Fig. 11, we report the scatter plots obtained for the three different wind speed filters. In all cases, the results are extremely good with a maximum RMSE$_{rel}$ = 16.9 per cent. Besides, we observed that, as expected, the larger the threshold on the wind speed, the better is the model performance. If we consider the threshold of 10 m s^{-1}, we find an excellent RMSE$_{rel}$ = 8.8 per cent.

Table 10 reports the contingency table computed for the first threshold (wind speed less than 3 m s^{-1}, the same used for Cerro Paranal), which permits us to discard just the very low wind speed values: the performance level is extremely good, with a global PC = 75.4 per cent and PODs in the four wind quadrants ranging from

Table 11. Contingency table statistical indicators for the wind direction during the night computed with different wind speed filters, at 58 m a.g.l. at LBT. We use the Meso-Nh $\Delta X = 500$ m configuration.

Operator	Wind speed filters		
	WS > 3 m s^{-1}(%)	WS > 5 m s^{-1}(%)	WS > 10 m s^{-1}(%)
PC	75.4	78.7	80.5
EBD	1.1	0.1	0.0
POD_N	81.3	85.8	85.5
POD_E	81.5	89.2	97.5
POD_S	67.0	68.0	71.8
POD_W	80.1	85.1	87.7

67.0 per cent to 81.5 per cent. In Table 11, we show the contingency table statistical indicators calculated with the three different wind speed filters. On raising the filter from 3 m s^{-1} to 10 m s^{-1}, we see that EBD rapidly drops to zero, while PC rises up to 80.5 per cent for wind speeds greater than 10 m s^{-1}. For strong winds, we are able to obtain excellent PODs ranging from 71.8 per cent to 97.5 per cent. If we consider that the wind comes mainly from the south-west (see the wind rose in Fig. 3), we deduce that the model performance is within 71.8 per cent and 87.7 per cent if we consider the threshold at 10 m s^{-1} and within 67 per cent and 80 per cent if we consider the threshold at 3 m s^{-1}.

The model seems to perform slightly worse for winds from the south; however, this is because the wind rose distribution (Fig. 3) shows that a significant portion of the winds come from an angle that is near the intersection of the west and south quadrants. If we rotate the division of quadrants by 45° and we consider south-east = [90°, 180°] and south-west = [180°, 270°], we obtain in these quadrants a POD of 76.6 per cent and 85.5 per cent, respectively, and a similar global PC = 79.2 per cent, with a wind speed threshold of 3 m s^{-1}. POD_{SW}, which corresponds to the sector with the highest wind frequency, is, therefore, an excellent 85.5 per cent.

By observing the cumulative distributions of bias, RMSE and σ computed over each single night, in Appendix A, computed with a wind speed threshold of 3 m s^{-1}, we observe that the median RMSE = 22° and σ = 15° are much lower than those obtained over the whole sample.

By looking at the scatter plots of the subsample of summer and winter (Figs B1 and B2, fourth panel, Appendix B), we observe that the model performance is definitely better in winter than in summer. $RMSE_{rel} = 21\%$ in the summer and RMSE = 13 per cent in the winter. This is because the winter (see Fig. 5) has much stronger winds than the summer (see Fig. 4) and it is easier for the model to reconstruct well the wind direction when the wind speed is strong. This, as observed in Table 11, produces a statistic that is more favourable to model predictions.

6 CONCLUSIONS

In this paper, we presented the results of the validation study on the operational forecast system being developed for LBT, as part of the ALTA project, performed on a large test sample of 144 nights uniformly distributed over the solar years 2014 and 2015. The test was performed on the atmospheric parameters (temperature, RH, wind speed and direction) near ground level using, as a reference term, the telemetry data feed from the weather stations installed on the telescope dome. We used the Meso-Nh model with the grid-nesting technique and horizontal resolutions of the innermost domain of 500 and 100 m, the latter being essential in reconstructing well the wind speed for strong winds. The results obtained in this validation study are extremely good and satisfactory and absolutely comparable to those obtained above Paranal and Armazones. Different statistical operators have been used (bias, RMSE and σ) and more sophisticated operators derived from the computation of the contingency tables, i.e. the percentage of correct detections (PC), the probability of detecting the value of a parameter inside a specific range of values (POD) and the probability of having an extremely bad detection (EBD). The validation of the forecasting system refers to the operational configuration of the model used in the ALTA project. We summarize here the most relevant results:

(i) The model has an excellent degree of reliability in reconstructing the temperature, with bias, RMSE and σ below 1°C. We obtained an excellent PC = 91.7 per cent and all POD_i are between 84 and 99 per cent.

(ii) The model shows good performance in reconstructing the RH. We find a very satisfactory bias = −2.36 per cent, RMSE ≃ 14 per cent and σ = 13.8 per cent. Besides, PC = 71.8 per cent is good too. The most critical POD_i for the RH is POD_3, which is the probability of detecting a RH larger than the second tertile. We obtained POD_3 = 61 per cent. The model correctly discriminates the weak and the strong values of RH. We would like to improve the dispersion of the very high values of RH but the present result is already more than satisfactory.

(iii) The model shows good performance in reconstructing the wind speed with an RMSE ranging from 2.4 to 2.7 m s^{-1}, depending on the horizontal resolution (500 or 100 m). We observed that the resolution of 100 m is necessary only when the wind speed is large (≧8.5 m s^{-1}) otherwise 500 m can provide even better results. We conceived, therefore, a hybrid treatment that considers the model outputs from the run having the highest resolution of 500 m when the average of the wind speed of the night is below 8.5 m s^{-1} and a resolution of 100 m when it is larger than 8.5 m s^{-1}. With such a hybrid configuration, we obtained a very satisfactory PC ≃ 65 per cent, POD_1 = 65.7 per cent, POD_2 = 45 per cent and POD_3 = 84.8 per cent. We highlight that POD_3 is the most important result for observational applications because it refers to the strong wind case that is the critical one for telescope operations. We conclude that model performance is excellent in this respect.

(iv) Also the results obtained for the wind direction are very satisfactory, with an $RMSE_{rel}$ = 16.9 per cent if we consider all data with wind speeds greater than 3 m s^{-1}. Besides, we observed that $RMSE_{rel}$ can reach 8.8 per cent if we consider all data with a wind speed larger than 10 m s^{-1}. This means that the higher the wind speed, the better is the reconstruction of the wind direction by the model. Values of PC and POD_i improve also on passing from a filtering of 3 to 10 m s^{-1} with the best PC = 80.5 per cent. POD_i in all four quadrants is excellent with values always larger than 72 per cent and, when we filter out wind speeds lower than 10 m s^{-1}, model performance achieves percentages larger than 90 per cent (see Table 11). In particular, POD_{SW}, related to the most interesting quadrant from which the wind comes more frequently (south-west), is an excellent 85.5 per cent.

The next step of our investigation for the ALTA Center project will be to use the most recent version of the Astro-Meso-Nh model (Masciadri et al. 2017) to forecast the optical turbulence at LBT. We will take advantage of measurements related to an extended site testing campaign (Masciadri et al. 2010) providing a vertical stratification of the optical turbulence on the whole 20 km above the ground in which a new technique for high-vertical resolution close to the ground (Egner & Masciadri 2007) was employed. A

study (Hagelin, Masciadri & Lascaux 2011) done with a previous version of the Astro-Meso-Nh model has already shown promising perspectives in that context. Astro-Meso-Nh as well as the strategy have evolved since then. The model must, therefore, be recalibrated and revalidated in this perspective.

ACKNOWLEDGEMENTS

The ALTA Center project is funded by the Large Binocular Telescope Corporation. Measurements of surface parameters were provided by the LBT telemetry and annexed instrumentation. The authors thank Christian Veillet, Director of the Large Binocular Telescope, for his continued and valuable support given to this research activity. The authors also thank the LBTO staff for their technical support and collaboration. Part of the numerical simulations were run on the HPCF cluster of the ECMWF using resources from the Project SPITFOT.

REFERENCES

ALTA Center Project, 2016, Available at: http://alta.arcetri.astro.it (P.I.: Elena Masciadri – Team members http://alta.arcetri.astro.it/team.php)
Bougeault P., Lacarrère P., 1989, Mon. Weather Rev., 117, 1872
Cuxart J., Bougeault P., Redelsperger J.-L., 2000, Q. J. R. Meteoro. Soc., 126, 1
Egner S., Masciadri E., 2007, PASP, 119, 1441
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
Hagelin S., Masciadri E., Lascaux F., 2011, MNRAS, 412, 2695
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
Lafore J.-P. et al., 1998, Ann. Geophys., 16, 90
Lascaux F., Masciadri E., Hagelin S., Stoesz J., 2009, MNRAS, 398, 1093
Lascaux F., Masciadri E., Fini L., 2013, MNRAS, 436, 3147
Lascaux F., Masciadri E., Fini L., 2015, MNRAS, 449, 1664
Masciadri E., 2003, Rev. Mex. Astron. Astrofis., 39, 249
Masciadri E., Vernin J., Bougeault P., 1999, A&A, 137, 185
Masciadri E., Vernin J., Bougeault P., 2001, A&A, 365, 699
Masciadri E., Stoesz J., Hagelin S., Lascaux F., 2010, MNRAS, 404, 144
Masciadri E., Lascaux F., Fini L., 2013, MNRAS, 436, 1968
Masciadri E., Lascaux F., Turchi A., Fini L., 2017, MNRAS, 00, 00
Noilhan J., Planton S., 1989, Mon. Weather Rev., 117, 536
Stein J., Richard E., Lafore J. P., Pinty J. P., Asencio N., Cosma S., 2000, Meteoro. Atmos. Phys., 72, 203
Turchi A., Masciadri E., Fini L., 2016, Proc. SPIE, 9909, 990938
Veillet C. et al., 2016, Proc. SPIE, 99100S, doi:10:1117/12.2234570

APPENDIX A: CUMULATIVE DISTRIBUTIONS

We report in this section the cumulative distributions of bias, RMSE and σ for all atmospheric parameters at Mount Graham. Sample sizes for each variable are reported in Table 2. Wind direction statistical operators are computed with a wind speed threshold of 3 m s^{-1} (see Section 5.4). Values are computed over the whole two years, 2014 and 2015 (Fig. A1), over the summer seasons of 2014 and 2015 (Fig. A2) and over the winter seasons of the same years (Fig. A3).

APPENDIX B: SEASONAL VALIDATION

In this section, we report the same validation study discussed in Section 5, performed on the subsample of nights from April to September (summer) and in the subsample from October to March (winter). This is done to give a partial validation against the seasonal variation. We refer to the discussion about each parameter in Section 5 for the specific methods used to validate the model. Here we report only the scatter plots and contingency tables obtained for each parameter in the specific seasonal subsample. Reported seasonal variations show a relevant difference regarding wind direction, where the model performs better during winter. This is eventually because during winter the wind speed is significantly higher (see Figs 4 and 5), which in turn reflects on the capability of the model in reconstructing the wind direction, as already discussed in Section 5.4.

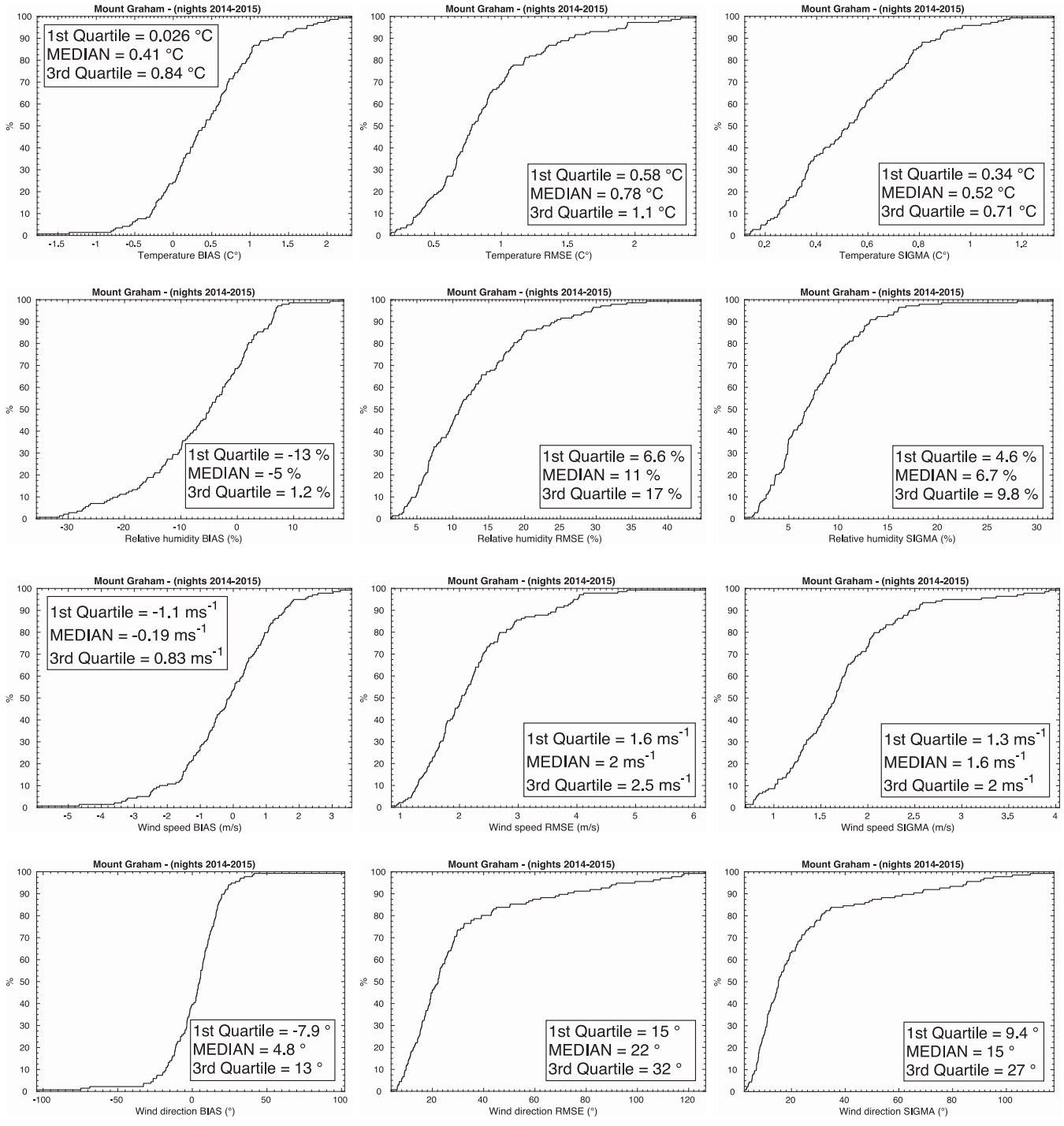

Figure A1. Cumulative distributions of bias, RMSE and σ for the atmospheric parameters. Results relative to the whole years 2014 and 2015.

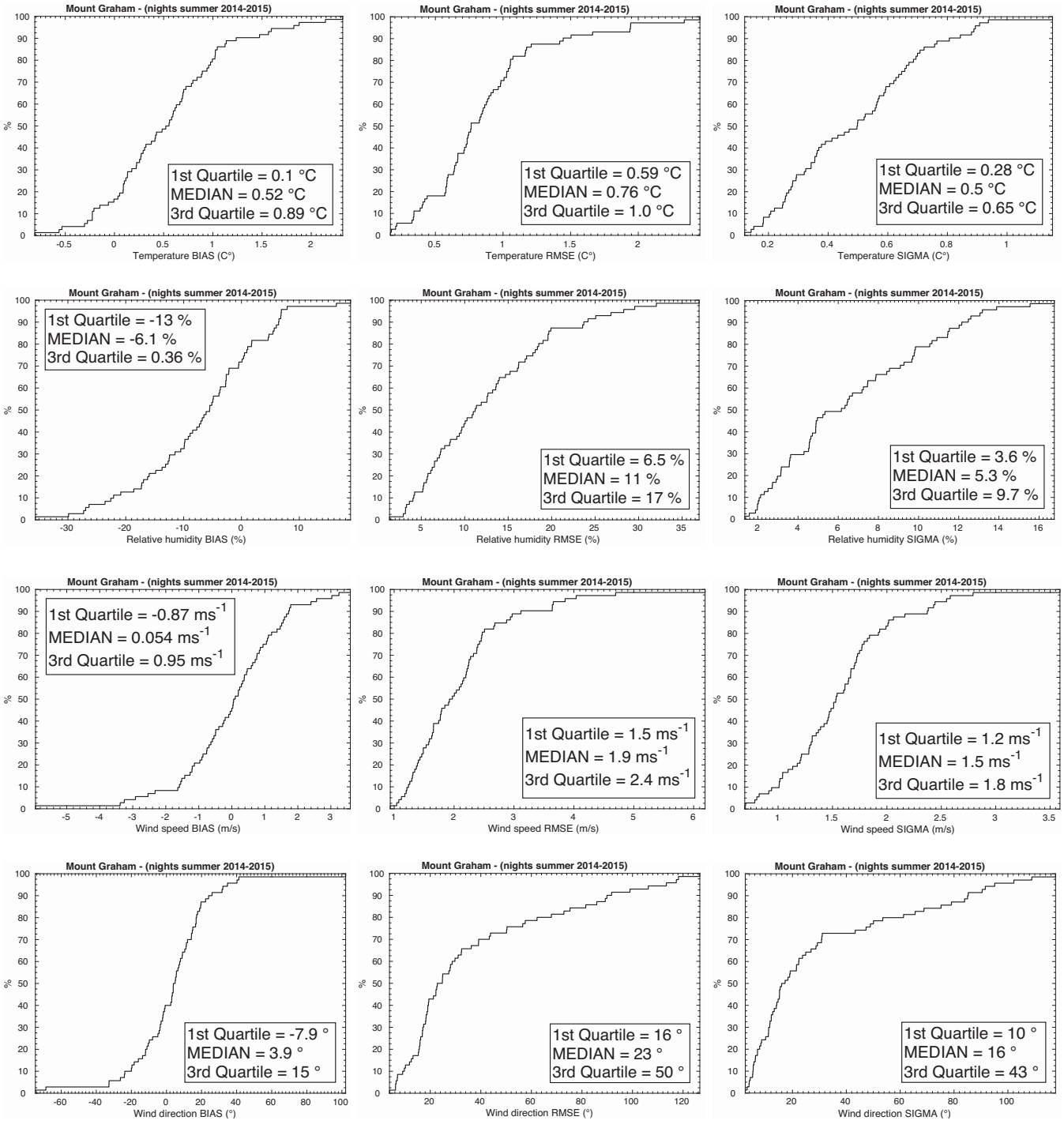

Figure A2. Cumulative distributions of bias, RMSE and σ for the atmospheric parameters. Results relative to the summers (April–September) of 2014 and 2015. Wind direction statistical operators are computed with a wind speed threshold of 3 m s^{-1} (see Section 5.4).

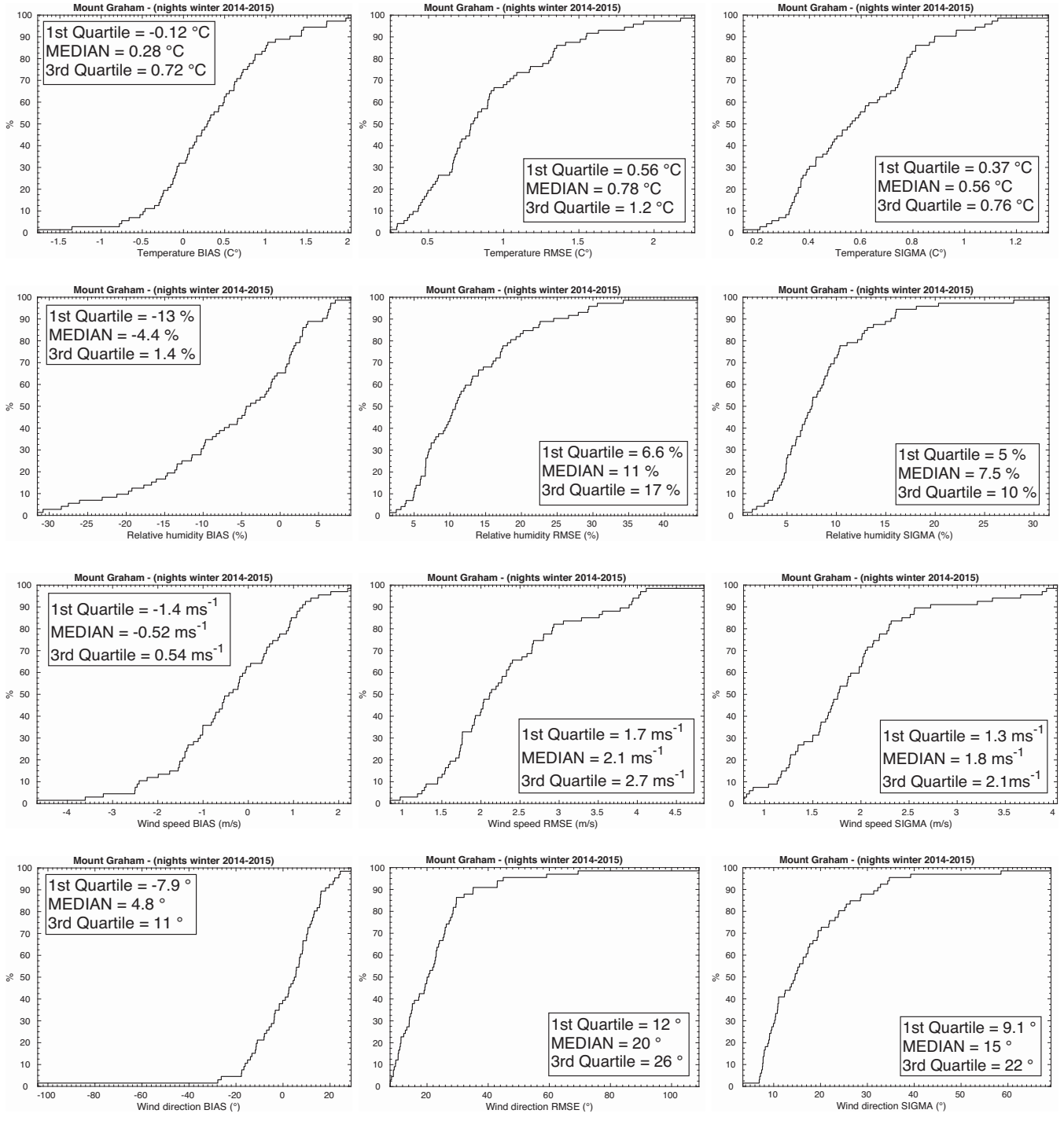

Figure A3. Cumulative distributions of bias, RMSE and σ for the atmospheric parameters. Results relative to the winters (October–March) of 2014 and 2015. Wind direction statistical operators are computed with a wind speed threshold of 3 m s^{-1} (see Section 5.4).

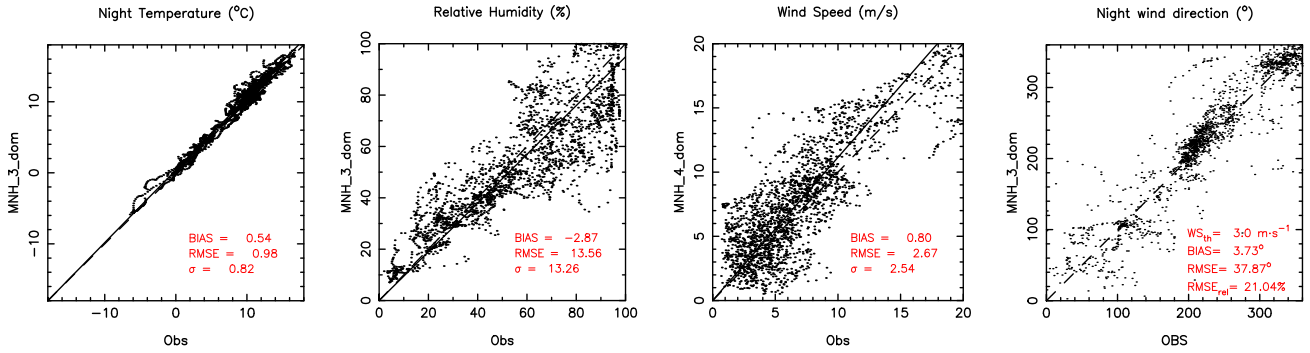

Figure B1. Scatter plots computed for each parameter in the summer subsample (April–September). Wind speed is computed at 100-m horizontal resolution.

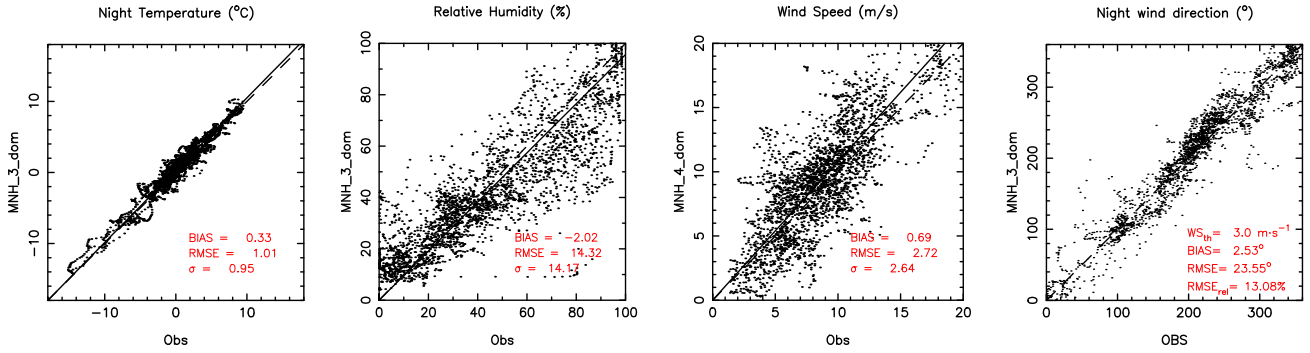

Figure B2. Scatter plots computed for each parameter in the winter subsample (October–March). Wind speed is computed at 100-m horizontal resolution.

Table B1. 3 × 3 contingency table for the absolute temperature during the night on the summer subsample. We use the Meso-Nh ΔX = 500 m configuration.

Temperature (°C) Summer		Observations $T < 6.9$	$6.9 < T < 10.7$	$T > 10.7$
Model	$T < 6.9$	771	10	0
	$6.9 < T < 10.7$	10	538	63
	$T > 10.7$	0	219	704

Sample size: 2315; PC: 87.0 per cent; EBD: 0.0 per cent; POD_1: 98.7 per cent; POD_2: 70.1 per cent; POD_3: 91.8 per cent.

Table B2. 3 × 3 contingency table for the relative humidity during the night on the summer subsample. We use the Meso-Nh ΔX = 500 m configuration.

Relative humidity (%) Summer		Observations RH < 37.7	37.7 < RH < 67.4	RH > 67.4
Model	RH < 37.7	562	78	4
	37.7 < RH < 67.4	94	546	321
	RH > 67.4	0	126	548

Sample size: 2279; PC: 72.7 per cent; EBD: 0.2 per cent; POD_1: 85.7 per cent; POD_2: 72.8 per cent; POD_3: 62.8 per cent.

Table B3. 3 × 3 contingency table for the wind speed during the night on the summer subsample. We use the Meso-Nh ΔX = 500 m configuration.

Wind speed (m s^{-1}) Summer – 500-m resolution		Observations WS < 4.2	4.2 < WS < 7.4	WS > 7.4
Model	WS < 4.2	380	184	12
	4.2 < WS <7.4	282	386	182
	WS > 7.4	22	185	682

Sample size: 2315; PC: 62.5 per cent; EBD: 1.5 per cent; POD_1: 55.6 per cent; POD_2: 51.1 per cent; POD_3: 77.9 per cent.

Table B4. 3 × 3 contingency table for the wind speed during the night on the summer subsample. We use the Meso-Nh ΔX = 100 m configuration.

Wind speed (m s^{-1}) Summer – 100-m resolution		Observations WS < 4.2	4.2 < WS < 7.4	WS > 7.4
Model	WS < 4.2	327	191	22
	4.2 < WS < 7.4	299	324	57
	WS > 7.4	58	240	797

Sample size: 2315; PC: 62.5 per cent; EBD: 3.5 per cent; POD_1: 47.8 per cent; POD_2: 42.9 per cent; POD_3: 91.0 per cent.

Table B5. 4 × 4 contingency table for the wind direction during the night on the summer subsample. We use the Meso-Nh ΔX = 500 m configuration. Wind speed threshold is 3 m s^{-1}.

Wind direction WS > 3 m s^{-1} – Summer		Observations North	East	South	West
Model	North	328	37	20	95
	East	44	201	46	5
	South	16	24	516	31
	West	33	1	181	301

Sample size: 1879; PC: 71.6 per cent; EBD: 2.2 per cent; POD_N: 77.9 per cent; POD_E: 76.4 per cent; POD_S: 67.6 per cent; POD_W: 69.7 per cent.

Table B6. 3 × 3 contingency table for the absolute temperature during the night on the winter subsample. We use the Meso-Nh ΔX = 500 m configuration.

Temperature (°C) Winter		Observations $T < -1.2$	$-1.2 < T < 2.4$	$T > 2.4$
Model	$T < -1.2$	629	85	0
	$-1.2 < T < 2.4$	128	1044	88
	$T > 2.4$	0	135	777

Sample size: 2886; PC: 84.9 per cent; EBD: 0 per cent; POD_1: 83.1 per cent; POD_2: 82.6 per cent; POD_3: 89.9 per cent.

Table B7. 3 × 3 contingency table for the relative humidity during the night on the winter subsample. We use the Meso-Nh ΔX = 500 m configuration.

Relative humidity (%) Winter		Observations RH < 24	24 < RH < 67.2	RH > 67.2
Model	RH < 24	624	94	13
	24 < RH < 67.2	165	1103	324
	RH > 67.2	0	70	493

Sample size: 2886; PC: 76.9 per cent; EBD: 0.5 per cent; POD_1: 79.1 per cent; POD_2: 87.1 per cent; POD_3: 59.4 per cent.

Table B8. 3 × 3 contingency table for the wind speed during the night on the winter subsample. We use the Meso-Nh ΔX = 500 m configuration.

Wind speed (m s^{-1}) Winter – 500-m resolution		Observations WS < 5.6	5.6 < WS < 9.4	WS > 9.4
Model	WS <5.6	378	253	26
	5.6 < WS < 9.4	170	622	340
	WS > 9.4	25	179	688

Sample size: 2681; PC: 63.0 per cent; EBD: 1.9 per cent; POD_1: 66.0 per cent; POD_2: 59.0 per cent; POD_3: 65.3 per cent.

Table B9. 3 × 3 contingency table for the wind speed during the night on the winter subsample. We use the Meso-Nh ΔX = 100 m configuration.

Wind speed (m s^{-1}) Winter – 100-m resolution		Observations WS < 5.6	5.6 < WS < 9.4	WS > 9.4
Model	WS < 5.6	338	201	11
	5.6 < WS < 9.4	171	508	183
	WS > 9.4	64	345	860

Sample size: 2681; PC: 63.6 per cent; EBD: 2.8 per cent; POD_1: 59.0 per cent; POD_2: 48.2 per cent; POD_3: 81.6 per cent.

Table B10. 4 × 4 contingency table for the wind direction during the night on the winter subsample. We use the Meso-Nh ΔX = 500 m configuration. Wind speed threshold is 3 m s^{-1}.

Wind direction WS > 3 m s^{-1} – Winter		Observations North	East	South	West
Model	North	409	16	0	50
	East	19	316	79	0
	South	5	39	661	43
	West	52	0	253	602

Sample size: 2544; PC: 78.1 per cent; EBD: 0.2 per cent; POD_N: 84.3 per cent; POD_E: 85.2 per cent; POD_S: 66.6 per cent; POD_W: 86.6 per cent.

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