MOSE
TECHNICAL REPORT

Doc.No : E-TRE-INA-245-0001
Version : 3.0
Date : 4 July 2013

MOSE: MOdeling Sites ESO

PI: Elena Masciadri
TEAM: Luca Fini, Franck Lascaux, Elena Masciadri

INAF/Osservatorio Astrofisico di Arcetri
L.go Enrico Fermi 5
50125 Florence, Italy

ESO BOARD:
Florian Kerber, Harald Kuntschner, Pierre-Yves Madec, Marc Sarazin

ABSTRACT

The MOSE project (MOdeling ESO Sites) aims at proving the feasibility of the forecast of **(1)** all the classical atmospheric parameters (temperature, wind speed intensity and direction, relative humidity) and **(2)** the optical turbulence OT (C_N^2 profiles) with all the main integrated astro-climatic parameters derived from the C_N^2 i.e. the seeing (ε), the isoplanatic angle (θ_0), the wavefront coherence time (τ_0) above the two ESO sites of Cerro Paranal (site of the Very Large Telescope - VLT) and Cerro Armazones (site selected for the European Extremely Large Telescope - E-ELT).

For what concerns the atmospheric parameters the study aims at studying the ability of the model in reconstructing their vertical stratification along the troposphere and low stratosphere (around 20 km) and their estimation near the ground (the first 30 m). To constrain and validate the model a different set of measurements have been used that are provided by different instruments. At Cerro Paranal: Automatic Weather Station (AWS) [RD1] and radiosoungings [RD2]. At Cerro Armazones: Automatic Weather Station (AWS) and mast [RD3].

For what concerns the optical turbulence and derived integrated astroclimatic parameters measurements from a Generalized Scidar (GS), a Differential Image Motion Monitor (DIMM) and a MASS (Multi Aperture Scintillation Sensor) running simultaneously during the PAR2007 Site testing campaign at Cerro Paranal [RD4] have been used for the model calibration and a preliminary model validation. Independent measurements from a DIMM and MASS taken routinely at Cerro Paranal since 19/08/1998[a] have been used for the final model validation.

In this technical report are summarized the deliverables according to MOSE Project - Phase A (Section 3.2.1 and Table 3-2 [RD5]) and minutes from review meetings (BSCW archive).

---

[a] *http://archive.eso.org/asm/ambient-server*

Contents

Related Documents

RD1 VLT-MAN-ESO-17440-1773: *VLT Astronomical Site Monitor ASM Data User Manual*

RD2 E-TRE-UVA-222-0001: *ADDENDUM: Measuring and Forecasting PWV above La Silla, APEX and Paranal Observatories*

RD3 TMT web-site: http://sitedata.tmt.org/

RD4 W. Dali Ali, A. Ziad, A. Berdja,, J. Maire,, J. Borgnino, M. Sarazin, G. Lombardi, J. Navarrete, H. Vazquez Ramio, M. Reyes, J. M. Delgado, J. J. Fuensalida, A. Tokovinin and E. Bustos: *Multi-instrument measurement campaign at Paranal in 2007. Characterization of the outer scale and the seeing of the surface layer*, A&A, 524, id.A73, 8 pp., 2010

RD5 E-SOW-ESO-245-0933: *Statement of Work for the feasibility study of an operational turbulence forecasts model*

RD6 E-ESO-SPE-066-0283: *Observing Modes in Science Operation*

1 Introduction

The MOSE project (MOdeling ESO Sites) aims at proving the feasibility of the forecast of all the most relevant classical atmospherical parameters for astronomical applications (wind speed intensity and direction, temperature, relative humidity) and the optical turbulence OT (C_N^2 profiles) with the integrated astro-climatic parameters derived from the C_N^2 i.e. the seeing (ε), the isoplanatic angle (θ_0), the wavefront coherence time (τ_0) above the two ESO sites of Cerro Paranal (site of the Very Large Telescope - VLT) and Cerro Armazones (site selected for the European Extremely Large Telescope - E-ELT).
The final outcome of the study is to investigate the opportunity to implement an automatic system for the forecast of these parameters at the VLT Observatory at Cerro Paranal and at the E-ELT Observatory at Cerro Armazones.

1.1 Scientific drivers

The OT forecast is a crucial cornerstone for the feasibility of the ELTs: it is indeed fundamental to support all kind of AO facilities in an astronomical observatory and to perform the flexible-scheduling of scientific programs and instrumentation through the Service Mode.
Most of the scientific programs, associated to the most challenging scientific goals, require, indeed, frequently excellent optical turbulence (OT) conditions to be carried out. The traditional queue system, that is based on the quality of the scientific program but that does not take into account the OT conditions, leads necessarily to a paradox: the higher is the scientific challenge of a scientific program, the lower is the probability to complete the program itself. From this we derive that the Service Mode, that takes into account the status of the OT, is mandatory to optimize the exploitation of the ELTs. However, to implement a Service Mode we need to forecast the OT to know its status at different delayed times ΔT with respect to the time at which the prediction is performed. The optimization of the *use* of a ground-based facility has therefore serious implications on the final scientific impact of the facility itself. We note, moreover, that the cost of a night of observation is of the order of hundreds of K$. The implementation of an OT forecast system leads therefore to a not negligible rationalization of costs versus scientific feedbacks. We remind also that, for evident statistical reasons, the advantage of the Service Mode can be fully achieved only if most of the available observing time is scheduled in this mode. For all these reasons ELTs plan to have, in their baseline configuration, permanent instruments located at different focal stations. It has been estimated that, at the E-ELT, the typical time required to switch the beam from an instrument to another is of the order of 10-20 minutes [RD6]. This is therefore the final minimal time-scale in terms of OT prediction that we can take as a reference. All these premises lead to the main conclusion that the success and the feasibility of the ELTs relies on our ability in forecasting the OT.

A few different methods/techniques have been tested for this purpose: statistical methods, analytical methods, neural network methods, physical methods. Most of these methods are based on strong approximations and/or empirical assumptions. In some cases they can provide typical trend but not really a specific prediction. A detailed description of the methods with their characteristics can be found in the review [1]. The peculiar feature of the numerical approach and more precisely of the mesoscale atmospheric models is the possibility to describe a 3D map of the C_N^2 (i.e. volumetric distribution of the optical turbulence in the troposphere) in a region around a telescope of the order of a few tens of kilometers at a time t_0 placed in the future. **This is the unique method that aims at address directly the solution of the Navier-Stokes equation of the hydrodynamical flow above the Earth. It is therefore intrinsically the most appropriate to detect directly a rapid change of the atmospheric status.**

1.2 Why do we need non-hydrostatic atmospheric meso-scale models ?

The typologies of models that can be used to study thermodynamic evolution of the earth's atmosphere are [2]: the General Circulation Models (GCM), the non-hydrostatic meso-scale models, the models for Large Eddy

Simulations (LES) and the models for Direct Numerical Simulations (DNS). These models differ in their resolution and, as a consequence, also in the extent of the typical domain of the atmosphere that they can reconstruct. We deduce that each of these typologies of model is dedicated to resolve phenomena of different nature and that the numerical approach as well as the physics described inside each of these models is done following a completely different logic.

The OT can not be resolved by the GCMs because the spatio-temporal fluctuation of the OT is much smaller (order of centimeter/meter) than the typical resolution of the GCM ($\sim$16 km). With the DNS the OT can be completely resolved but we can not forecast it because we definitely miss the link with initialization data i.e. with the evolution of the atmospheric flow at large spatial scales. The meso-scale models represent the right trade-off that permits to reconstruct the OT maintaining the link with the external spatio-temporal evolution of the atmospheric flow. The OT is completely parameterized in the meso-scale models. With the LES the OT is partially resolved and partially parameterized. In principle it is possible in perspective to use the LES to improve the resolution of the simulations if initialized with outputs coming from meso-scale models. It has been observed that, when the horizontal resolution increases and reaches values of ΔX smaller than 10 km, the hydrostatic models show a tendency in distorting the spectrum of the gravity waves, particularly in proximity of mountain regions. For this reason, non-hydrostatic models are more suitable to reconstruct the atmospheric flow on such conditions.

1.3 How to produce the OT forecast

The atmospherical meso-scale model is applied over a limited area (orography) of the whole Earth. The extension of the area can change depending on the method used to simulate it (grid-nesting or monomodel). The typical order of magnitude of the innermost domain size in the grid-nesting case is a few tens of kilometers. The horizontal resolution of the atmospheric model for the OT forecast has to be of the order of 500 m or even higher [3]. The meso-scale model is initialized with *analyses* and/or *forecasts* provided by a GCM, in our case from the European Center for Medium-Range Weather Forecasts (ECMWF) GCM. Analyses and forecasts represent therefore the atmospheric status above the whole Earth (and therefore also above the limited area over which the meso-scale model is applied) discretized on a finite number of grid points of the GCM calculated at a precise instant t_0. Fig. 1[1] shows, as an example, how analyses or forecasts appear in a 3D representation as produced by a GCM. In this figure a 3D map of the temperature is displayed through a table of colors. When the meso-scale model is initialized, the initialization data are re-sampled on the grid of the meso-scale model.

Analyses and forecasts are continuously produced by the GCMs at a regular temporal sampling (typically of the order of 6 hours). The GCMs are continuously fed by different kind of measurements: radiosoundings from meteorological stations distributed at different locations on the Earth, satellites measurements, measurements provided by planes and ships... A procedure called "Data Assimilation" permits to control and correct the spatio-temporal evolution of the atmospheric flow above the whole Earth produced by the GCMs comparing these measurements with the outputs of the GCM calculated at the same time. It follows that, in general, the accuracy of the analyses and forecasts provided by the GCMs is higher where the density of measurements is higher.

To explain the principle used to forecast the OT we can use a simple toy model (Fig. 2).

Let's take the time axis and let's imagine to be at 18 hour on the day (J-1). We take then the meteorological forecast, calculated at 6 and 12 hours, that is at 0 and 6 hours of the day J. At these two instants we can know the atmospherical state at the boundaries and on the inner part (cube) of the meso-scale model. Let's start the simulation with the meso-scale model at 00:00 UT of the day J. By the mean of a temporal interpolation we can know the state of the atmospheric flow at the boundaries and in the inner part of the model (cube) at each time step. If we start our simulation at 00:00 hours of the day J, and we link our simulations with the

---

[1] Taken from the review: *La Météorologie*

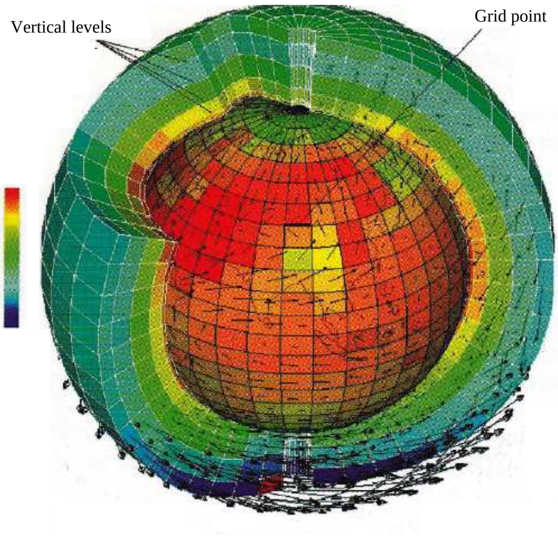


Figure 1: Schematic GCM representation in 3D. The color map refers to the temperature and the arrows to the wind speed and direction.

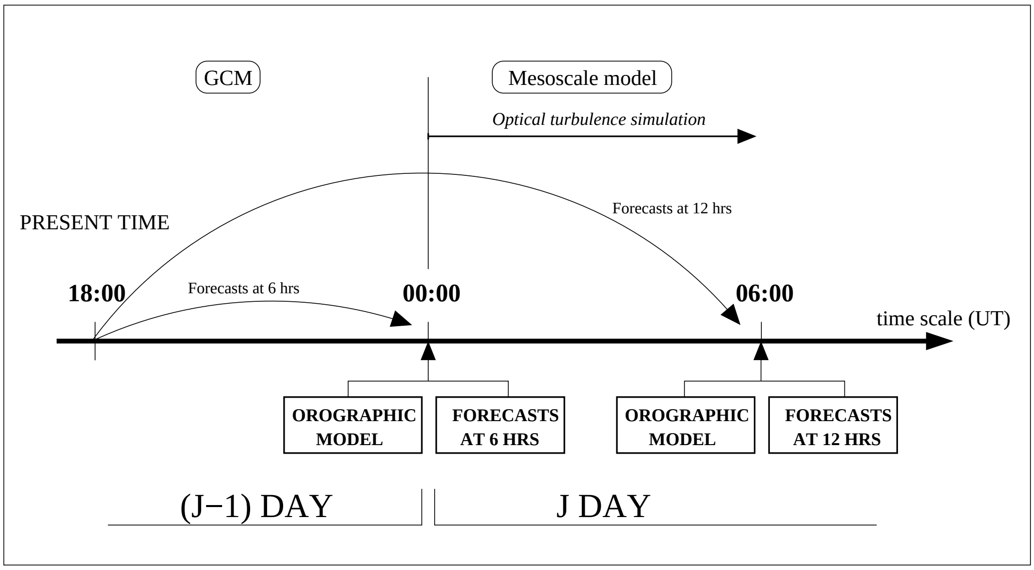


Figure 2: Forecast toy model scheme.

DEM (Digital Elevation Model i.e. the orography map - see Section 3), the temporal evolution of the optical turbulence is a real forecast of the OT because the atmospheric flow in the inner part of the cube (i.e. the region reconstructed by the model) is in constant thermodynamic balance with the external atmospherical flow. This example represents a OT forecast calculated at 6 hours. The procedure is, from a practical point of view, more complex than this toy model because it takes into account two more main elements: **(1)** the simulations done with the meso-scale model are forced every six hours (the synoptic hours) with successive forecasts taken from the ECMWF GCM that are available in the meanwhile; **(2)** the simulations are performed using the grid-nesting technique that permits to increase the resolution on specific portions of the domain using a set of imbricated domains (typically three or four domains).

Playing with different combinations and sequence of initializations products (analyses and/or forecasts) one can conceive OT forecasts at different delayed times. In Section 3 is described the configuration used in Phase A of MOSE project.

1.4 Forecasted parameters

In the context of the MOSE project we focused our attention on the following atmospheric parameters that are strictly related to the optimization of the AO operational activities, all the astroclimatic parameters and the following parameters from which the OT mainly depends on.

- Surface temperature: the temperature near the ground is extremely important to indirectly thermalize the dome and reduce/eliminate the dome seeing, the most important contribution of the seeing experienced on ground-based images. The seeing inside the dome can be very hardly modeled being that the turbulence develops inside a confined environment and the Kolmogorov theory is rarely respected under these conditions. It has been proved [4] that the dome-seeing is mainly determined by $(T_m\text{-}T_i)^{6/5}$ and $(T_i\text{-}T_o)^{6/5}$ where T_m is the primary mirror temperature, T_i is the temperature inside the dome and T_o is the temperature outside the dome, near the ground. Once T_o is forecasted, it is possible to tune T_i and T_m so as to minimize the temperature gradients. The temperature gradient near the ground is moreover extremely critical to well reconstruct the optical turbulence near the ground. It is fundamental for us to know how good or bad is the model performance in reconstructing this parameter to discriminate the potential discrepancies of the model.

- Surface wind speed: the intensity of the wind speed near the ground can strongly affect the AO performances because of vibrations induced by all critical structures such as primary and secondary mirrors. Vibrations produced by wind speed bursts are among the most annoying source of noise for these elements and the accuracy of the AO correction is directly proportional to the noise introduced by such a kind of vibrations. An efficient forecast of the wind speed intensity can usefully optimize the use of the most sophisticated AO techniques and even decide when it is better to avoid AO observations if conditions do not permit to achieve required conditions. Besides, the wind speed shear near the ground is one of the main cause for OT triggering together with the temperature gradient. The interest in studying the ability of the model in reconstructing this parameter is therefore similar to what already said for the surface temperature.

- Surface wind direction: the wind speed direction is one of the atmospheric parameters that are more frequently and easily correlated to the local seeing characteristics. The prediction of the wind direction can therefore give a direct information on the probability to have bad or good seeing. The specifications provided by the ESO Board for this parameter give a Peak-to-Valley (PtV) of 90° meaning that, most of the time, one would like to predict the quadrant from which the wind is blowing.

- Vertical profiles on $\sim$ 20 km: potential temperature, wind speed intensity and direction and relative humidity: Such a stratification can obviously be obtained by GCMs. However, it has been proved ([5],[6],[7]) that GCMs are not reliable in the low part of the atmosphere particularly in proximity of mountain regions because of their too low horizontal resolution and a too smooth orography. It is also worth reminding that a meso-scale model can produce a continuous temporal evolution of the vertical stratification of these parameters with a temporal sampling that can be freely selected while information provided by GCM is sampled only at synoptic hours. *We proved [7] that the wind speed provided by the meso-scale model Meso-Nh is definitely the best solution to calculate the wavefront coherence time at an astronomical Observatory.* Besides, the wind speed intensity shear (together with the potential temperature gradient) are the fundamental cause of the OT triggering. It is therefore important to assure us how the model reconstructs the vertical stratification of these parameters to be able to discriminate the potential causes of model discrepancies or failures.

- C_N^2 profiles: The algorithm used for the C_N^2 parameterization is the one described in [3]. We refer the reader to the same paper for the detailed description of the theory connecting the dynamic turbulence with the optical turbulence and the algorithms used in the Meso-Nh model. In the context of the MOSE project we also test a modified/new algorithm (Section 6) in order to test which of the two ones provides better performances. In a forthcoming paper the algorithm will be presented. The modified algorithm will be indicated hereafter with the '_mod' extension.

- Integrated astro-climatic parameters: in the context of the MOSE project it has been established to investigate the three most important integrated astro-climatic parameters: the **seeing (ε)**, the **isoplanatic angle (θ_0)** and the **wavefront coherence time (τ_0)**. We highlight that the Astro-Meso-Nh package includes also other parameters such as the scintillation rate, the spatial coherence outer scale, ...

2 The Meso-NH model and the Astro-Meso-NH package

2.1 Status of the art

We summarize here the main milestones achieved so far in our researches on the optical turbulence forecasts in terms of model reliability and the novelty represented by MOSE.
In 1999 ([3], [8]) we proved for the first time that a meso-scale model could reconstruct reliable C_N^2 profiles. From a qualitative point of view, it has been proved that the shape of the profile could match the observed one, and from a quantitative point of view the observed and simulated integrated values of the turbulence present in the atmosphere were well correlated. In 2001 [9] a calibration method for the model has been proposed and the whole technique has been validated on a statistic of 10 nights comparing simulations with measurements taken from different instruments (balloons and Generalized SCIDAR) [10]. We proved that the dispersion of the seeing simulated and observed $\Delta\varepsilon_{model,GS} \sim 25\%$ was comparable to the dispersion obtained using different instruments $\Delta\varepsilon_{ballons,GS} \sim 30\%$. In a successive time we applied the same technique in an autonomous way on a whole solar year at San Pedro Martir [11] carrying out a complete statistical analysis of all the astro-climatic parameters permitting us to achieve the first complete and homogeneous analysis of the seasonal variation investigation of the OT ever done so far. In 2011 [12] the model/technique has been validated using a richer statistical sample of GS measurements performed on 43 nights uniformly distributed along one year. Besides, the model has been applied to the internal Antarctic Plateau achieving two major conclusions: (1) we proved that such a model is able to discriminate between the sites on the top of the summit of a plateau characterized by a very thin surface layer (Dome C and Dome A) [13],[14]; (2) thanks to the numerical technique (Meso-NH model) we provided the first estimates of the OT vertical stratification ever done in the free atmosphere at Dome A [15].

The MOSE project aims at overcoming two major limitations that we normally encounter in studies focused on the optical turbulence forecast with atmospheric models: **(1)** the difficulty in having independent samples of measurements for the model calibration and model validation to estimate if and how the correlation between measurements and predictions decreases with the increasing of the number of nights used for the calibration; **(2)** the difficulty in having a large number of simultaneous measurements done with different and independent instruments for the OT estimates (in particular vertical profilers). This project is performed with the non-hydrostatic mesoscale atmospherical model **Meso-NH** [16] joined with the **Astro-Meso-NH package** for the calculation of the optical turbulence [3, 8] to perform the OT forecasts. An extended data-set of observations (meteorological parameters and optical turbulence) have been considered in the project. We took advantage of measurements obtained in the context of the site selection for the TMT (American study) at Cerro Armazones and on measurements taken routinely during the last decade and/or in a dedicated site testing campaign (PAR2007) at Cerro Paranal.

In this project we also tested a different parameterization of the C_N^2 and evaluated its performances in statistical terms (see Section 6).

2.2 Tool of investigation

All the numerical simulations of the nights presented in this study have been performed with the mesoscale numerical weather model **Meso-NH**[2] [16]. The model has been developed by the Centre National des Recherches Météorologiques (CNRM) and Laboratoire d'Aéreologie (LA) de l'Université Paul Sabatier (Toulouse). The Meso-NH model can simulate the temporal evolution of three-dimensional meteorological parameters over a selected finite area of the globe. The system of hydrodynamic equations is based upon an anelastic formulation allowing for an effective filtering of acoustic waves. It uses the Gal-Chen and Sommerville [17] coordinates system on the vertical and the C-grid in the formulation of Arakawa and Messinger [18] for the spatial digitalization. It employs an explicit three-time-level leap-frog temporal scheme with a time filter [19]. It employs a one-dimensional 1.5 turbulence closure scheme [20]. For this study we use a 1D mixing length proposed by Bougeault and Lacarrère [21]. The surface exchanges are computed using ISBA [22] (Interaction Soil Biosphere

---

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

Atmosphere).

The optical turbulence and derived parameters are not an intrinsic part of the Meso-NH model but it has been developed by our team in an independent package. The package (called **Astro-Meso-NH**), includes the algorithms for the C_N^2 parameterization and all the other integrated astro-climatic parameters (seeing, isoplanatic angle, wavefront coherence time, scintillation rate, spatial coherence outer scale,..). The first version of the code has been presented in [3], [8]. Since then the code have been improved during the years in terms of flexibility permitting a set of further outputs and permitting us to carry out different scientific studies mainly addressing the reliability of this technique ([5]-[6], [8]-[12], [23]-[24]). In most recent years, we parallelized the Astro-Meso-NH code for the optical turbulence (the standard version of Meso-NH, without the Astro-Meso-NH package developed by our team, is already efficiently parallelized). To do this we used OPEN-MPI-1.4.3 and then ran the model on local workstations as well as on the HPCF cluster of the ECMWF, in parallel mode so as to gain in computing time. The second solution permitted us to achieve relatively rich statistical estimates of these analyses. Simulations for this contract have been performed on both platforms. Typically we gave preference to the HPCF cluster to carry on systematic simulations and/or expensive simulations in CPU terms.

3 Model configuration and set-up

The grid-nesting technique [25], employed in our study, consists of using different imbricated domains of the Digital Elevation Models (DEM i.e orography) extended on smaller and smaller surfaces, with increasing horizontal resolution but with the same vertical grid. Two different grid-nesting configurations have been employed. The standard configuration includes three domains (Fig. 3 and Table 1) and the innermost resolution is $\Delta X = 500$ m. The second configuration is made of five imbricated domains, the first same three as the previous configuration, and other two centered at both Paranal and Armazones sites, with a horizontal resolution of $\Delta X = 100$ m (Fig. 4 and Table 2). One can notice that using these configurations, we are able to do the forecast at both sites simultaneously. The orographic DEMs we used for this project are the GTOPO[3] with an intrinsic horizontal resolution of 1 km (used for the domains 1 and 2) and the ISTAR[4] with an intrinsic horizontal resolution of 0.5 km (used for the domain 3, 4 and 5). Here, it is important to bear in mind that, even if the ISTAR orographic grid has an intrinsic resolution of 0.5 km, we still gain information when it is interpolated on a Meso-NH $\Delta X = 100$ m grid with respect to a Meso-NH $\Delta X = 500$ m grid. Indeed, on a Meso-NH $\Delta X = 500$ m grid, even if the resolution is equal to the one of the ISTAR orographic grid, the grid points are not co-located and the resulting interpolation leads to an artificial loss of accuracy, which is in the most part reduced if we oversample the Meso-NH grid with $\Delta X = 100$ m (see BSCW archive for more details).
It is important to note also that we used a "two way" grid-nesting that means that the atmospheric flow has a mutual interaction at the interface of each couple of domains: the larger domain (called 'father') and the contiguous smaller domain (called 'son'). This means that there is a retro action on the spatio-temporal evolution of the atmospheric flow in the small domain that does that the atmospheric flow in the smaller domain is constantly in a thermo-dynamic equilibrium with the flow at larger scales and it guarantees the propagation of the gravity waves that can be triggered by the interaction of the atmospheric flow on the Chilean coast and reach the atmosphere above Cerro Paranal or Cerro Armazones in the free atmosphere. Along the z-axis we have 62 levels distributed as follows: a first vertical grid point equal to 5 m, a logarithmic stretching of 20 % up to 3.5 km above the ground, and an almost constant vertical grid size of ∼600 m up to 23.8 km.
All simulations were initialized the day before at 18 UT and forced every 6 hours with the analyses from the ECMWF, and finished at 09 UT of the simulated day (except for the simulations of 2009 which finished at 14 UT of the simulated day). In Phase B it is planned to study the model performances using, as initialization data, forecasts from the ECMWF.
All the vertical profiles (Section 4 for the meteorological profiles and Section 6 for the C_N^2 profiles) extracted from the model computations are available every 2 min. All the computed meteorological parameters near the surface [0-30] m a. g. l. (Section 5) are available at every time step of the innermost domain, which is $\Delta t = 3$ s for the $\Delta X = 500$ m configuration and $\Delta t = 1$ s for the $\Delta X = 100$ m configuration.

From now on and until the end of this report, when the model configuration is not specified, it is implicitly supposed to be the $\Delta X = 500$ m configuration.

---

[3] $http://www1.gsi.go.jp/geowww/globalmap-gsi/gtopo30/gtopo30.html$
[4] Bought by ESO at the ISTAR Company - Nice-Sophia Antipolis, France

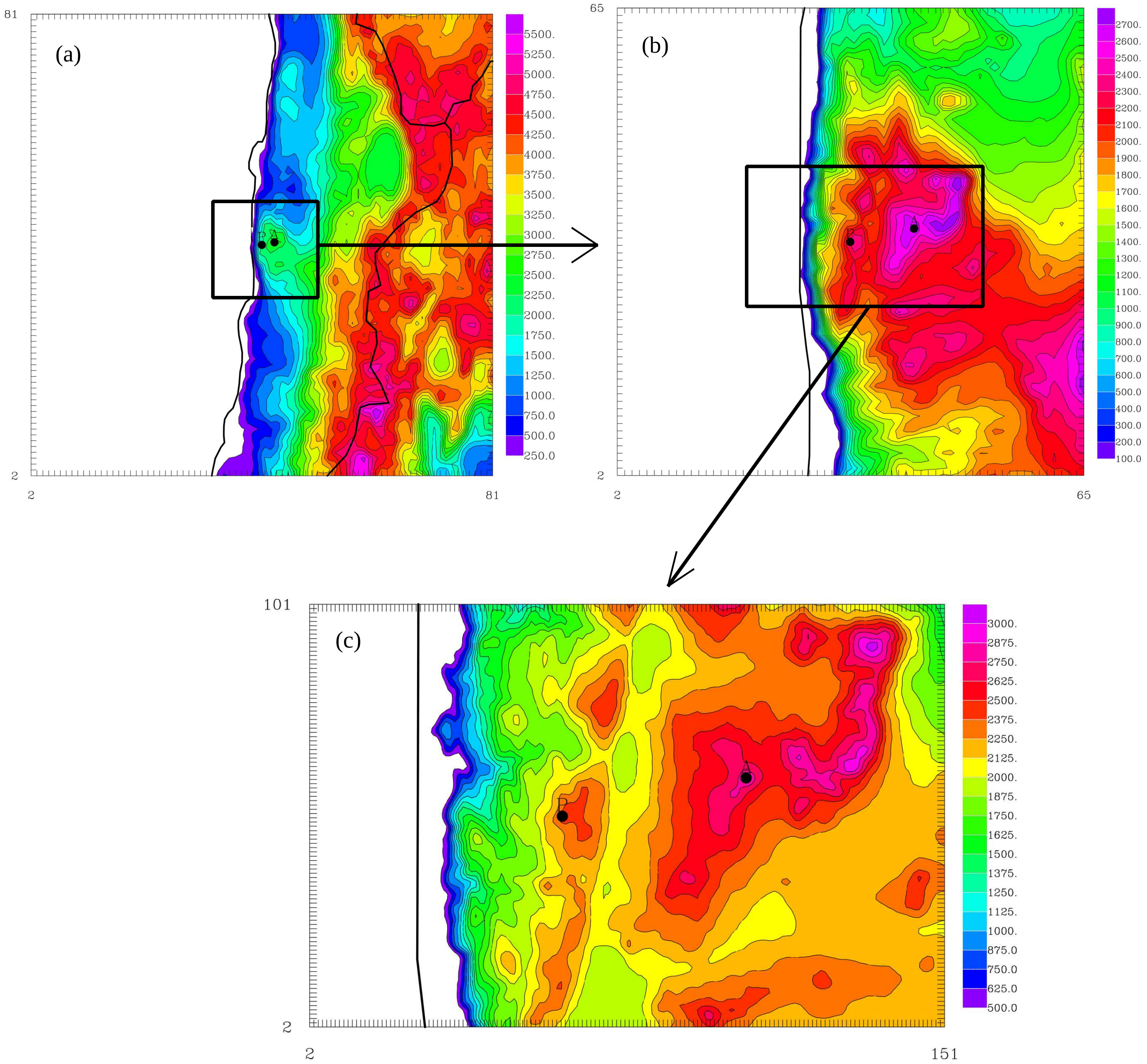


Figure 3: **Orography** (altitude in m) of the region of interest as seen by the Meso-NH model (polar stereographic projection) for all the imbricated domains of the **grid-nested standard configuration**. **(a)** Domain 1 (orographic data from GTOPO), **(b)** Domain 2 (orographic data from GTOPO), **(c)** Domain 3 (orographic data from ISTAR), A dot stands for Cerro Armazones and P dot stands for Cerro Paranal. See Table 1 for the specifications of the domains (number of grid-points, domain extension, horizontal resolution).

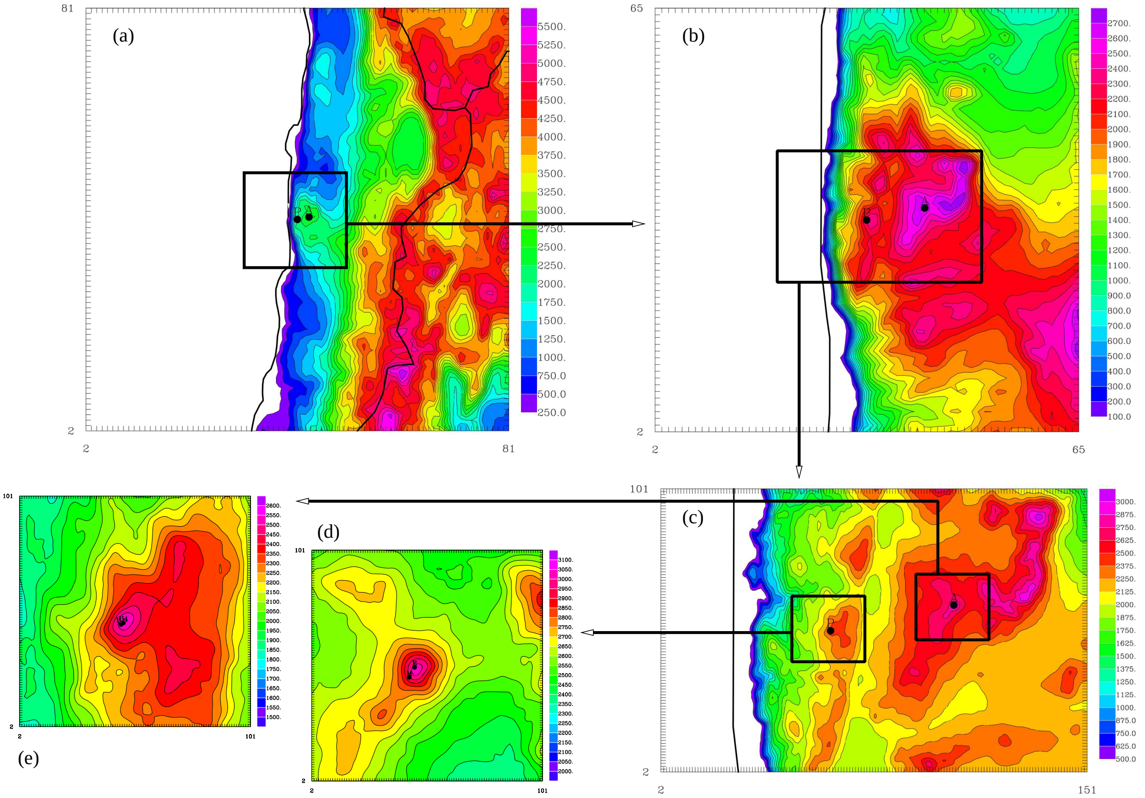


Figure 4: **Orography** (altitude in m) of the region of interest as seen by the Meso-NH model (polar stereographic projection) for all the imbricated domains of the **grid-nested high horizontal resolution configuration**. **(a)** Domain 1 (orographic data from GTOPO), **(b)** Domain 2 (orographic data from GTOPO), **(c)** Domain 3 (orographic data from ISTAR), **(d)** Domain 4 (orographic data from ISTAR), **(e)** Domain 5 (orographic data from ISTAR), A dot stands for Cerro Armazones and P dot stands for Cerro Paranal. See Table 2 for the specifications of the domains (number of grid-points, domain extension, horizontal resolution).

Domain	Grid Points	Domain size (km)	Horizontal resolution
Domain 1	80×80	800×800	$\Delta X = 10$ km
Domain 2	64×64	160×160	$\Delta X = 2.5$ km
Domain 3	150×100	75×50	$\Delta X = 500$ m

Table 1: Meso-NH model **standard grid-nesting configuration**. In the second column the number of horizontal grid-points, in the third column the domain extension and in the fourth column the horizontal resolution ΔX.

Domain	Grid Points	Domain size (km)	Horizontal resolution
Domain 1	80×80	800×800	$\Delta X = 10$ km
Domain 2	64×64	160×160	$\Delta X = 2.5$ km
Domain 3	150×100	75×50	$\Delta X = 500$ m
Domain 4	100×100	10×10	$\Delta X = 100$ m
Domain 5	100×100	10×10	$\Delta X = 100$ m

Table 2: Meso-NH model **high resolution grid-nesting configuration**. In the second column the number of horizontal grid-points, in the third column the domain extension and in the fourth column the horizontal resolution ΔX.

Site	Real altitude (m)	Meso-NH altitude (m) $\Delta X = 1$ km*	Meso-NH altitude (m) $\Delta X = 500$ m	Meso-NH altitude (m) $\Delta X = 100$ m
Paranal	2634	2430 ($\Delta h = 204$ m)	2478 ($\Delta h = 156$ m)	2545 ($\Delta h = 89$ m)
Armazones	3065	2820 ($\Delta h = 245$ m)	2901 ($\Delta h = 164$ m)	3010 ($\Delta h = 55$ m)

* The model was not used in this configuration in this study. It is presented here only as a quantitative example on the degradation of the orography representation with a lower horizontal resolution.

Table 3: Comparisons between sites observed **altitudes** and model interpolated altitudes (in m). Altitudes from the Meso-NH model are taken from the innermost domain with the higher horizontal resolution ΔX.

4 Meteorological parameters: vertical stratification

4.1 General information

- **Site(s)**: Cerro Paranal;
- **Instruments**: balloons radiosondes;
- **Parameters investigated**: wind speed, wind direction, potential temperature, relative humidity;
- **Number of nights**: 23 nights in 2009 (12 during winter, 11 during summer, cf. Table 64 in Annex A for a complete list);
- **Atmospheric region investigated**: 0 - 20 km.

At Cerro Paranal, we had access to 50 radio-soundings (vertical distribution of the meteorological parameters in the 20 km above the ground) launched above this site in the context of an intense site testing campaign for water vapor estimates [26] and covering 23 nights in 2009 (11 nights in summer and 12 nights in winter).

The radiosondes used (Vaisala radiosondes RS92 - SGP) have an accuracy of 0.5 °C for the temperature, an accuracy of 5% for the relative humidity, an accuracy of 0.15 $\mathrm{m \cdot s^{-1}}$ for the wind speed and an accuracy of 2° for the wind direction. They also have a positioning uncertainty on the vertical of 20 m.

The balloon have been launched close to the synoptic hours (00:00 UT, 06:00 UT, and/or 12:00 UT). These 23 nights were chosen to investigate the performances of the model in reconstructing the vertical distribution of the meteorological parameters. To do this, we performed 23 simulations (1 for every night), each starting at 18:00UT the day before and ending at 14:00UT of the simulated day (for a total duration of 20 hours). We then compared observations from the radio-soundings with the simulations performed by the model in the range [3 km - 21 km] a. s. l., because the first 400-500 meters above the ground (h = 2634 m - Paranal summit) constitute, indeed, a sort of 'grey zone' in which it can be meaningless to retrieve any quantitative useful estimates from the radio-soundings because of many different reasons. Among others: (a) the orographic map of the innermost domain has an intrinsic Δh $\simeq$ 156 m with respect to the real summit due to the natural smoothing effect of the model horizontal interpolation of the DEM; (b) the radio-soundings have been launched at around 50 m below the summit. It should be meaningless to compare the observed and simulated values at the summit ground height because in one case we resolve friction of the atmospheric flow near the ground, in the other no; (c) we have to take care about an uncertainty Δh of around 50 m in the identification of the zero point ($\mathrm{h_0}$) probably due to an unlucky procedure performed during the radio-soundings launches on the zero point setting. This uncertainty has basically no more effects above a few hundreds of meters above the ground because the parameters values are affected by phenomena evolving at larger spatial scales. We decided therefore to treat data only above roughly 500 m from the summit. For the model, we take the averaged vertical profiles simulated by the model in around one hour from the time in which the radio-sounding has been launched and one hour later. We considered that the balloon is an in-situ measurement and a balloon needs around 1 hour to cover 20 km from the ground moving up in the atmosphere with a typical vertical velocity of ~6 $\mathrm{m \cdot s^{-1}}$.

To estimate the statistical model reliability in reconstructing the main meteorological parameters we used the averaged values plus two statistical operators: the bias and the root mean square error (RMSE) defined as:

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

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

where X_i are the individual observations, Y_i the individual simulations parameters calculated at the same time and N is the number of times for which a couple (X_i,Y_i) is available with both X_i and Y_i different from zero. At the same time, due to the fact that we are interested in investigating the model ability in forecasting a parameter and not only in characterizing it, it is important to investigate also the correlation observations/simulations calculated night by night and not only in statistical terms. Before computing the statistics on the vertical profiles, the model outputs as well as the balloons observations were re-interpolated on a constant vertical grid with a size equal to the first vertical grid point (5 m) of the model.

4.2 Overall statistical model performances

In this section we consider the overall statistics involving the full sample of 50 radiosoundings. Fig. 5 shows the average vertical profiles, the bias and the RMSE calculated on this sample of 50 radiosoundings, of the wind speed, the potential temperature, the wind direction, and the relative humidity. The bias contains informations on systematic model off-sets. The RMSE contains information on the statistical errors plus the bias i.e. the whole sources of errors. To elaborate theses figures, first data and model profiles have been interpolated on a 5 m vertical grid, then average, bias and RMSE have been computed for each interpolated levels, and finally a vertical moving average with a Δ of 1 km have been applied on the resulting profiles.

Looking at Fig. 5 we can see that the Meso-NH model shows very good performances in reconstructing the wind speed direction on the whole 20 km. Between 5 km a. s. l. and 18 km a. s. l. the model reconstruction is statistically almost perfect. Below 5 km a. s. l. (where the orographic effects are the most evident) the bias is of the order of $\sim$20° with a RMSE that can reach 60°. This means that, night by night, in the vertical slab [3 km - 5 km] a. s. l. = [500 m - 2.5 km] a. g. l., we can have a discrepancy of the order of a few tens of degrees. It is worth highlighting that the accuracy in observing the wind direction can hardly be better than $\sim$20°. The model performances are therefore very satisfactory.

The model shows very good performances in reconstructing the relative humidity. The bias is basically never larger than 10% all along the 20 km. The largest discrepancy (bias of 10%) of simulations with respect to measurements is observed at the jet stream level where the RMSE can reach 15%. Such a satisfactory result has been obtained in spite of the fact we used a cheap scheme (in terms of CPU cost) for the relative humidity. That was possible because of the dryness of the region. Such a solution permits faster simulations. The small bump at the jet-stream level (Fig. 5, bottom right) is highly probably due to the humidity coming from the close ocean.

The model shows also a very good performances in reconstructing the potential temperature. We observe a very small bias of $\sim$ 2°C from the ground up to around 13 km a. s. l. Above 13 km a. s. l., where the potential temperature slope is steeper and steeper, the bias can reach up to 4°C.

The wind speed intensity is very well reconstructed: we have a bias of around 1 $\mathrm{m \cdot s^{-1}}$ in the [5 km - 15 km] a. s. l. range for the wind speed. Above 15 km a. s. l. and in the [3 km - 5 km] a. s. l. range the bias reaches a value of 2 $\mathrm{m \cdot s^{-1}}$. Comparing results (cf. BSCW archive) obtained with the Meso-NH model and the ECMWF analyses coming from the General Circulation Models (GCM) we could conclude that most of the residual biases and RMSEs we described so far are generated by initial conditions and not by the mesoscale model itself.

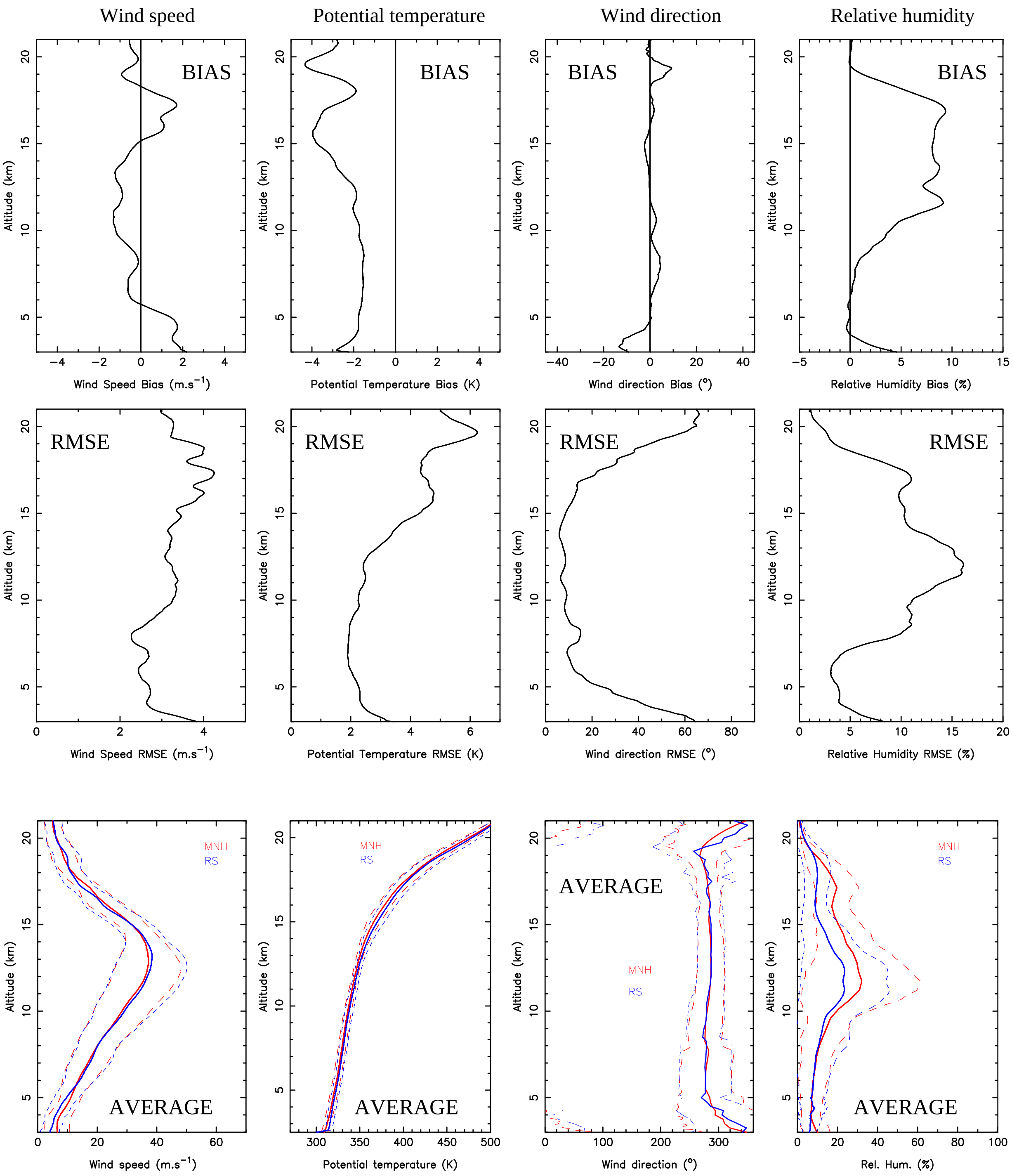


Figure 5: **Bias** and **RMSE** (Meso-NH - Observations) and **average** of **wind speed**, **potential temperature**, **wind direction** and **relative humidity**. Data and model profiles interpolated on a 5 m-vertical grid; then bias and RMSE are computed for each interpolated levels; finally a moving average over 1 km is applied on the resulting profiles. In the bottom figures, dashed lines represent the standard deviation, full thin lines are the average. NB: for the wind direction average profile, not a moving average, only a sub-sample of the points is plotted. The lists of the simulated nights and the hours at which radio-soundings (50 in total) were available is reported in Table 64.

4.3 Individual nights model performances

A comparison (observations/simulations) was performed night by night and at each instant for which a radio-sounding was available (see Annex B). It revealed an excellent agreement in basically all the 50 cases studied. To illustrate this, Fig. 6 shows a comparison of the wind speed observed and simulated in three different instants (00:00, 06:15 and 12:00 UT) of the same night. This case perfectly puts in evidence the excellent performances of the model in adapting itself to the wind speed evolution during the time. In spite of the fact that the observed wind speed strongly modifies its features all along the night at different heights, we note that the model perfectly reconstructs the observed wind speed features in the three different instants. In particular, the lowering of the intensity of the wind speed at the jet-stream level is perfectly reconstructed by the model. A similar behavior is observed in basically all the 50 cases studied.

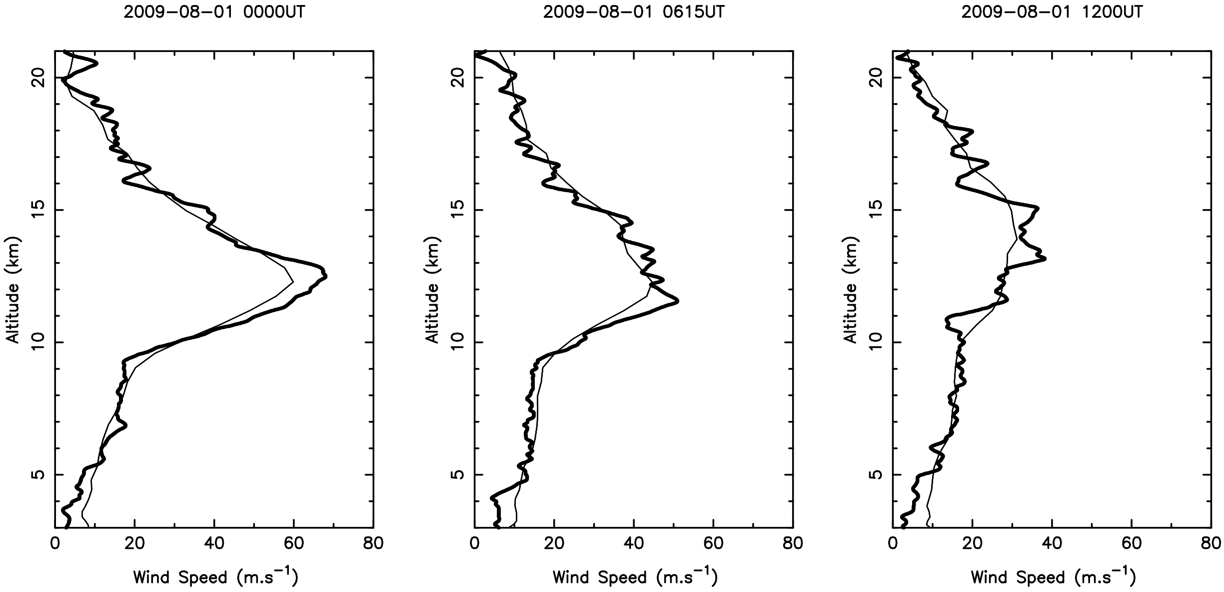


Figure 6: Comparison of the **wind speed** intensity observed (radio-soundings: thick line) and simulated (Meso-NH model: thin line) at three different instants (00:00, 06:15, 12:00 UT) during the same night: 1/8/2009 above **Cerro Paranal**.

This definitely guarantees us the reliability of a tool (the Meso-NH mesoscale model) in reconstructing the temporal evolution of the vertical distribution of the wind speed (V(h,t)) during a whole night. This is a fundamental ingredient (beside to the vertical profiles of the optical turbulence C_N^2(h,t), see Sec. 6) to be used for the calculation of the temporal evolution of the wavefront coherence time τ_0(t).

Fig. 7 shows the temporal evolution of the wind speed and the potential temperature provided by the Meso-NH model during two different nights. In this example we can appreciate the intrinsic level of the temporal variability of both parameters at different heights above the ground during the night. This is far from being negligible. In other words, the mesoscale predictions provide us a complete information (temporal evolution of the meteorological parameter) with respect to the estimates coming from the General Circulation Models that can provide outputs only at synoptic hours and can not fill the lacking information in between the synoptic hours. This proves to us the invaluable utility of a mesoscale model for the prediction of the astro-climatic parameters. In particular, our results indicate that the wind speed retrieved from the mesoscale model is, at present, certainly the best (and the most practical) solution to calculate the temporal evolution of the wavefront coherence time τ_0(t). We remind that, at present, there is no instrument able to provide autonomously the τ_0(t). The temporal evolutions of the potential temperature and the wind speed contribute both in determining

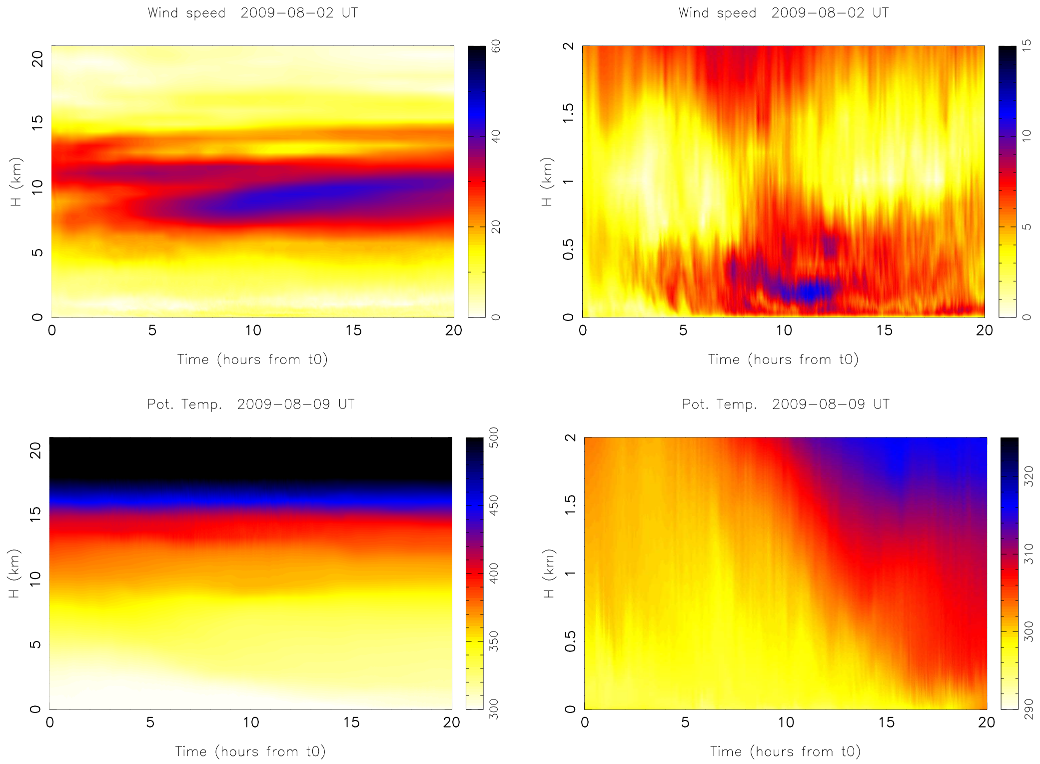


Figure 7: Temporal evolution of the wind speed (top) and potential temperature (bottom) vertical distribution, calculated at the grid point of Paranal and extended along 21 km (left) and 2 km (right) from the ground. The simulation starts at t_0 = 18 UT and lasts 20 hours. The local night (20:00 - 05:00 LT) corresponds to the interval (6 - 15) on the x-axis.

the prediction of the optical turbulence C_N^2(h,t).

A few more words are suitable to comment the wind speed in the vertical slab [3 km - 5 km] a. s. l. that is [0.5 km - 2.5 km] a. g. l. The discrepancy observation/simulation of around 1-2 m·s^{-1} (and a RMSE that can achieve 4 m·s^{-1}) is weak but, differently from the other discrepancies, seems to be the only one derived directly by the mesoscale model and not by the initial conditions (cf. BSCW archive). Even if the discrepancy is weak, it is therefore worth analyzing it more in detail. We counted a roughly comparable number of cases in which the model overestimates and correctly estimates the wind in this vertical slab. The model almost never underestimates the observed values. We did not observe any clear correlation between the discrepancy and the absolute value of the wind speed. We tested the sensitivity to the grid-point selection to check if a not precise selection of the grid-point of the summit could create some anomalous effects on the wind in the low atmosphere: for all the cases in which an overestimate has been observed we calculated the same bias at four different grid-points around the summit but it has not been observed any substantial difference. We note that the radio-sounding is an in-situ measurement and that the balloon moves horizontally along the (x,y) plan during the ascension in the atmosphere. During the ascension time, it therefore senses a volume of atmosphere shifted with respect to the zenithal direction. The radio-sounding lasts for around 4 minutes with a V_z = 6 m·s^{-1} to reach the altitude of 4 km a. s. l. In this temporal interval the balloon can move somewhere (depending on the wind direction) within a circle with a radius of 2.4 km. If we calculate the maximum variation of the wind speed (ΔV) inside such a circle we see that, at 4 km a. s. l., ΔV is of the order of 1.5-2 m·s^{-1}. Table 4 reports these values for a few flights.

The inhomogeneity decreases with the height and disappears in the high part of the atmosphere. In other words, being that the horizontal distribution on the (x,y) plane of the wind speed in the low part of the

Date	Hour (UT)	V_{4km} ($m \cdot s^{-1}$)	ΔV ($m \cdot s^{-1}$)
1/8/2009	00:00	5	1.6
11/11/2009	12:00	10	2
19/11/2009	06:00	5	1.4
19/11/2009	12:00	10	1.6
14/11/2009	12:00	5	2

Table 4: Maximum variability of the wind speed calculated at 4 km a. s. l. inside a circle having a radius proportional to the wind speed observed at 4 km a. s. l. times 4 minutes.

atmosphere is not necessarily homogeneous, this could explain the discrepancy with the simulations in the [3 km - 5 km] a. s. l. range. This argument tells us that the radio-sounding is not an optimal reference for comparisons with simulations to be done in the low part of the atmosphere. A preferable choice should be an instrument based on an optical remote sensing principle. To support this argument we remind that in a recent study [7], indeed, a similar comparison of simulated versus measured wind speed obtained with a remote sensing instrument (a Generalized SCIDAR used for the wind speed measurements) provided a correlation better than 1 $m \cdot s^{-1}$ in the [0.5 km - 1 km] a. g. l. vertical range.

5 Meteorological parameters: surface layer

5.1 General information

- **Site(s)**: Cerro Paranal and Cerro Armazones;
- **Instruments**: Automatic Weather Station (AWS) and masts;
- **Parameters investigated**: wind speed, wind direction, absolute temperature;
- **Number of nights**: 20 nights in summer 2007 (cf. Table 65 in Annex A for a complete list);
- **Atmospheric region investigated**: surface layer ([0, 30 m], cf Table 5).

At Cerro Paranal, observations of meteorological parameters near the surface come from an automated weather station (AWS) and a 30 m high mast including a number of sensors at different heights. Both instruments are part of the VLT Astronomical Site Monitor [27]. Absolute temperature data are available at 2 m and 30 m above the ground (accuracy of 0.1°C). Wind data are available at 10 m and 30 m above the ground (Table 5), with an accuracy of 2% of the measured value for the wind speed and an accuracy of 5.63° for the wind direction. At Cerro Armazones, observations of the meteorological parameters near the ground surface come from the Site Testing Database [28], more precisely from an AWS and a 30 m tower (with temperature sensors and sonic anemometers). Data on temperature and wind speed are available at 2 m, 11 m, 20 m and 28 m above the ground (Table 5). At 2 m (Armazones) temperature measurements from the AWS and the sonic anemometers are both available but we considered only those from the tower (accuracy of 0.1 °C) [29]. Those from the AWS are not reliable because of some drift effects (T. Travouillon, private communication). Wind speed observations are taken from the AWS (at 2 m) and from the sonic anemometers of the tower (at 11 m, 20 m and 28 m). The wind speed measurements from the sonic anemometer (at 11 m, 20 m and 28 m) have an accuracy of 2%-6% (it depends on the angle of the wind vector w.r.t the horizontal). The wind speed accuracy at 2 m from the AWS is 2%. The outputs are sampled with a temporal frequency of 1 minute.

As we will see further on, some biases in the forecasts are not negligible. That's why another statistical parameter was added in the analysis (on the ESO board's request): the bias-corrected RMSE, σ:

$$\sigma = \sqrt{\sum_{i=1}^{N} \frac{[(X_i - Y_i) - (\overline{X_i - Y_i})]^2}{N}} = \sqrt{RMSE^2 - BIAS^2} \tag{3}$$

The bias-corrected RMSE permits us to disentangle the systematic error (bias) from the random error.
To estimate the statistical model reliability in reconstructing the surface meteorological parameters, in addition to the three statistical operators defined in Eq. 1-3, we also use the correlation coefficient cc for the temperature and the wind speed, defined as:

$$cc = \frac{\sum_{i=1}^{N}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{N}(Y_i - \overline{Y})^2}} \tag{4}$$

The case of the wind direction being particular, we use the circular coefficient correlation ccc [30], instead of the correlation coefficient cc (Eq. 4), defined as:

$$ccc = \frac{\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} sin(X_i - X_j) sin(Y_i - Y_j)}{\sqrt{\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} sin^2(X_i - X_j)\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} sin^2(Y_i - Y_j)}} \tag{5}$$

In Eq. 3, 4 and 5, X_i are the individual observations, Y_i the individual simulations parameters calculated at the same time and N is the number of times for which a couple (X_i,Y_i) is available with both X_i and Y_i different from zero.

We performed the analysis on a statistic of 20 nights (Table 65). We selected all the nights of the Paranal site testing campaign of November/December 2007 for which observations (temperature and wind speed) near the surface are available on both sites plus a few others in which measurements above both sites were available. All simulations were initialized the day before at 18 UT and forced every 6 hours with the analyses from the ECMWF. Simulations finished at 09 UT of the simulated day (for a total duration of 15 hours). The statistics is computed only during night local time, from 00 UT to 09 UT[5]. We neglect the first 6 hours related to the day time. In the first 6 hours some spurious effects due to the model spin-up are present (in the first part of the simulation the model adapts itself to the orography). To investigate the model performances during the daytime we suggest to initialize simulations at 06:00 UT. As previously said, in Phase B it is planned to study the model performances using, as initialization data, forecasts from the ECMWF. These are the products to be used in operational conditions.

Parameters	Cerro Paranal	Cerro Armazones
Wind speed	10 m / 30 m	2 m / 11 m / 20 m / 28 m
Wind direction	10 m / 30 m	2 m / 11 m / 20 m / 28 m
Temperature	2 m / 30 m	2 m / 11 m / 20 m / 28 m

Table 5: Altitudes of observation for all meteorological parameters investigated near the surface, at Cerro Paranal and Cerro Armazones.

5.2 Overall statistical model performances

We first proceed in this section to an overall statistical study of the model performances in reconstructing temperature, wind speed and wind direction near the ground at both sites (Cerro Paranal and Cerro Armazones). All the statistical computations are made on the whole sample of data covering 20 nights of observations (Table 65). Fig. 8 displays the scattered plots of the model computed absolute temperature against the observed absolute temperature, for both sites, and at every altitudes for which observations were available (see Table 5). Fig. 9 displays the scattered plots of the model computed wind speed (with the ΔX = 500 m configuration) against the observed wind speed, for both sites, and at every altitudes for which observations were available (see Table 5). Fig. 10 displays the scattered plots of the model computed wind speed (with the ΔX = 100 m configuration) against the observed wind speed, for both sites, and at every altitudes for which observations were available (see Table 5). Finally, Fig. 11 displays the scattered plots of the model computed wind direction against the observed wind direction, for both sites, and at every altitudes for which observations were available (see Table 5). All the bias, RMSE and cc values reported in these figures are also summarized in Tables 6, 7, 8 and 9. In all the aforementioned figures, every point is the result of a 30-minutes average for both observations and model.

The results in terms of absolute temperature are particularly impressive. Considering both sites, and every levels, the bias is very small: it spans from 0.03°C at 28 m at Cerro Armazones (lowest value) to 0.78°C at 2 m at Cerro Armazones (highest value), well below 1°C. Even the RMSE is almost always inferior to 1°C, which is very satisfying (the highest value of 1.10°C is encountered near the ground at 2 m at Cerro Armazones). Both these excellent results are confirmed by the correlation coefficient cc>0.96.

The forecast of the wind direction near the ground is very good. The resulting bias, at every level, is in the interval [0.1° - 9.9°], which is impressive. More over, the RMSE is between 43° and 58°, which also means that the model is capable to predict the good wind direction, or at least its quadrant at +/- 1 standard deviation (corresponding to ∼68% of cases). It has not yet been tested the wind direction with ΔX=100 m but we expect an even better result for the model due to the fact the a higher resolution of the orography should permit to the atmospheric flow to evolve in proximity of the surface following more realistic streamlines.
For the absolute temperature and the direction, given that the bias is very small, the bias-corrected RMSE (σ) is almost equal (or slightly inferior) to the RMSE. As for the wind speed near the surface, we can see (Fig. 9) that

---

[5] here and hereafter we consider the nighttime always as the [00 - 09] UT interval

the model with the standard ΔX = 500 m configuration, underestimates the observed wind velocity, especially at the first level of observations (2 m). The bias goes from 1.07 m$\cdot$s^{-1} at 30 m at Cerro Paranal to 3.58 m$\cdot$s^{-1} at 2 m at Cerro Armazones. The RMSE can reach values as high as 4.74 m$\cdot$s^{-1} (at 11 m at Cerro Armazones). These values are not important in absolute terms but they are not negligible in relative terms. One explanation of this underestimation is the relative smoothing of the orography with ΔX = 500 m that could generate a weaker than observed wind speed over the mountainous peaks (like at Cerro Paranal and Cerro Armazones). One way to overcome this situation is to increase the horizontal resolution. We've done that using ΔX = 100 m, and the results are shown in Fig. 10 and summarized in Table 8. There is an evident improvement of the wind reconstructed over Cerro Paranal and Cerro Armazones by the model of a factor 2. The bias strongly reduces at the higher levels (0.38 m$\cdot$s^{-1} at 28 m at Cerro Armazones and 0.44 m$\cdot$s^{-1} at 30 m at Cerro Paranal). Its highest value is only 1.47 m$\cdot$s^{-1} at 2 m at Cerro Armazones. The RMSE is halved, from a maximum value of 4.74 m$\cdot$s^{-1} (at 11 m at Cerro Armazones) with the ΔX = 500 m configuration to a maximum value of 2.77 m$\cdot$s^{-1} (at 11 m at Cerro Armazones) with the ΔX = 100 m configuration, a much more acceptable value. It is worth noting that in the ΔX = 100 m configuration, the simulated absolute temperature remains very good (cf. BSCW archive).

If we take off the bias from the computations of the RMSE, we obtain lower σ than RMSE that vary from 2.28 m$\cdot$s^{-1} at 2 m at Cerro Paranal to 3.63 m$\cdot$s^{-1} at 28 m at Cerro Armazones with the ΔX = 500 m configuration and from 2.19 m$\cdot$s^{-1} at 2 m at Cerro Armazones to 2.57 m$\cdot$s^{-1} at 28 m at Cerro Armazones with the ΔX = 100 m configuration. Difference between RMSE and σ are obviously more evident for the wind speed (with horizontal resolution of 500 and 100 meters) than for the absolute temperature and wind direction because the biases in the first cases are more important.

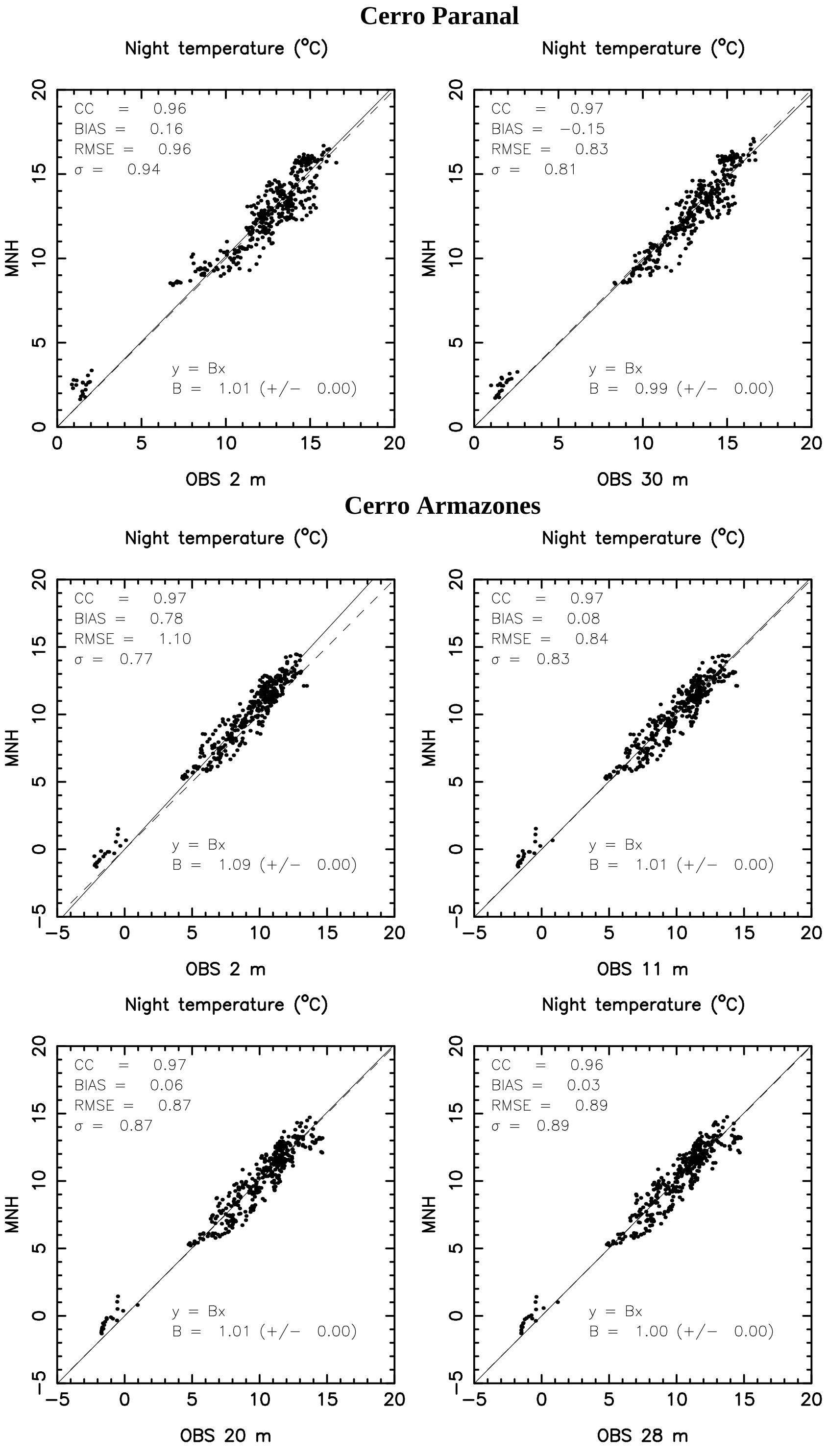


Figure 8: Scattered plot of Meso-NH **temperature** against observations, at 2 m and 30 m at **Cerro Paranal** (top figures); at 2 m, 11 m, 20 m and 28 m at **Cerro Armazones** (central and bottom figures). Every point represents the average over an interval of 30 minutes.

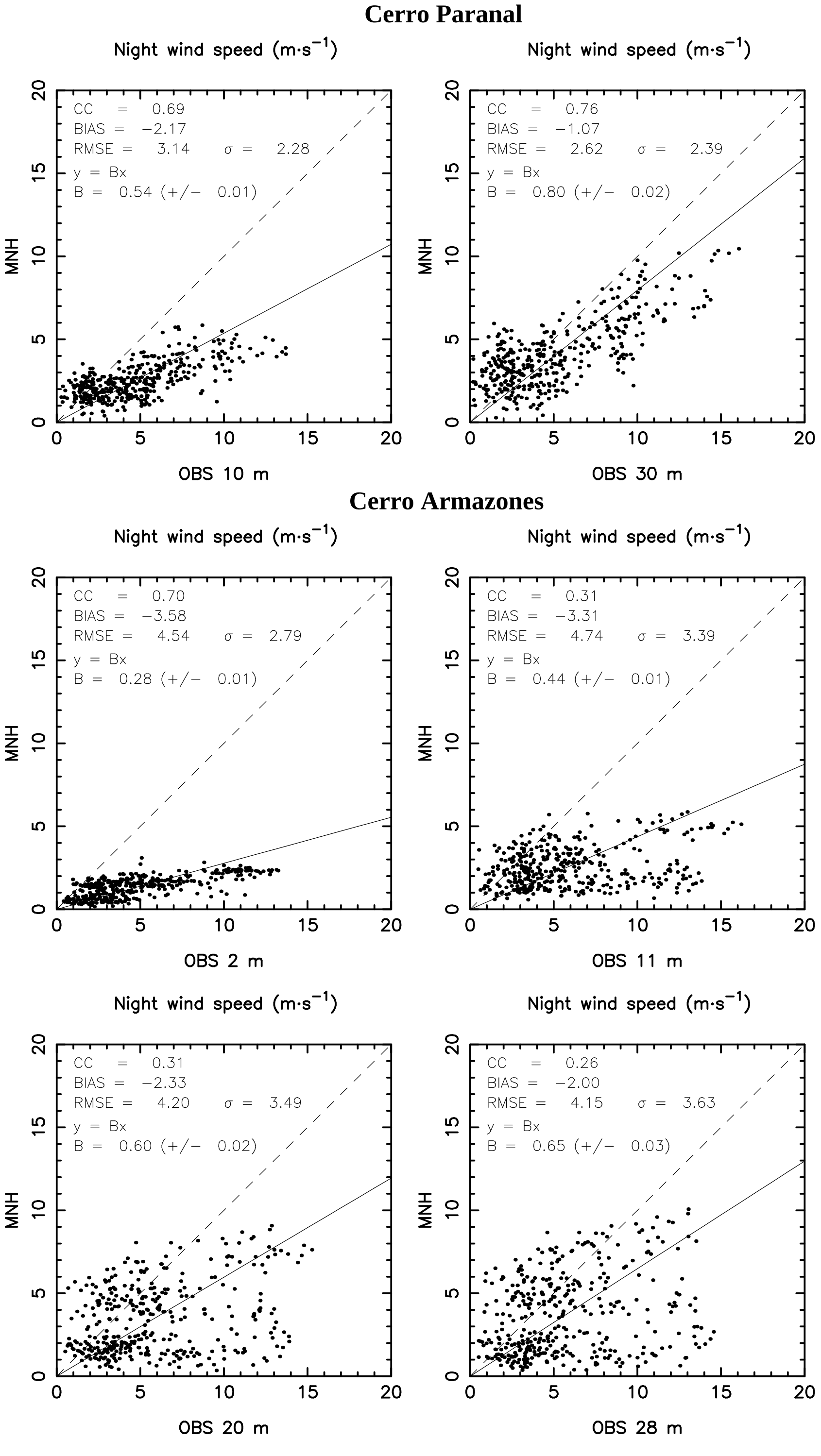


Figure 9: Scattered plot of Meso-NH **wind speed** against observations, at 10 m and 30 m at **Cerro Paranal** (top figures); at 2 m, 11 m, 20 m and 28 m at **Cerro Armazones** (central and bottom figures). Every point represents the average over an interval of 30 minutes.

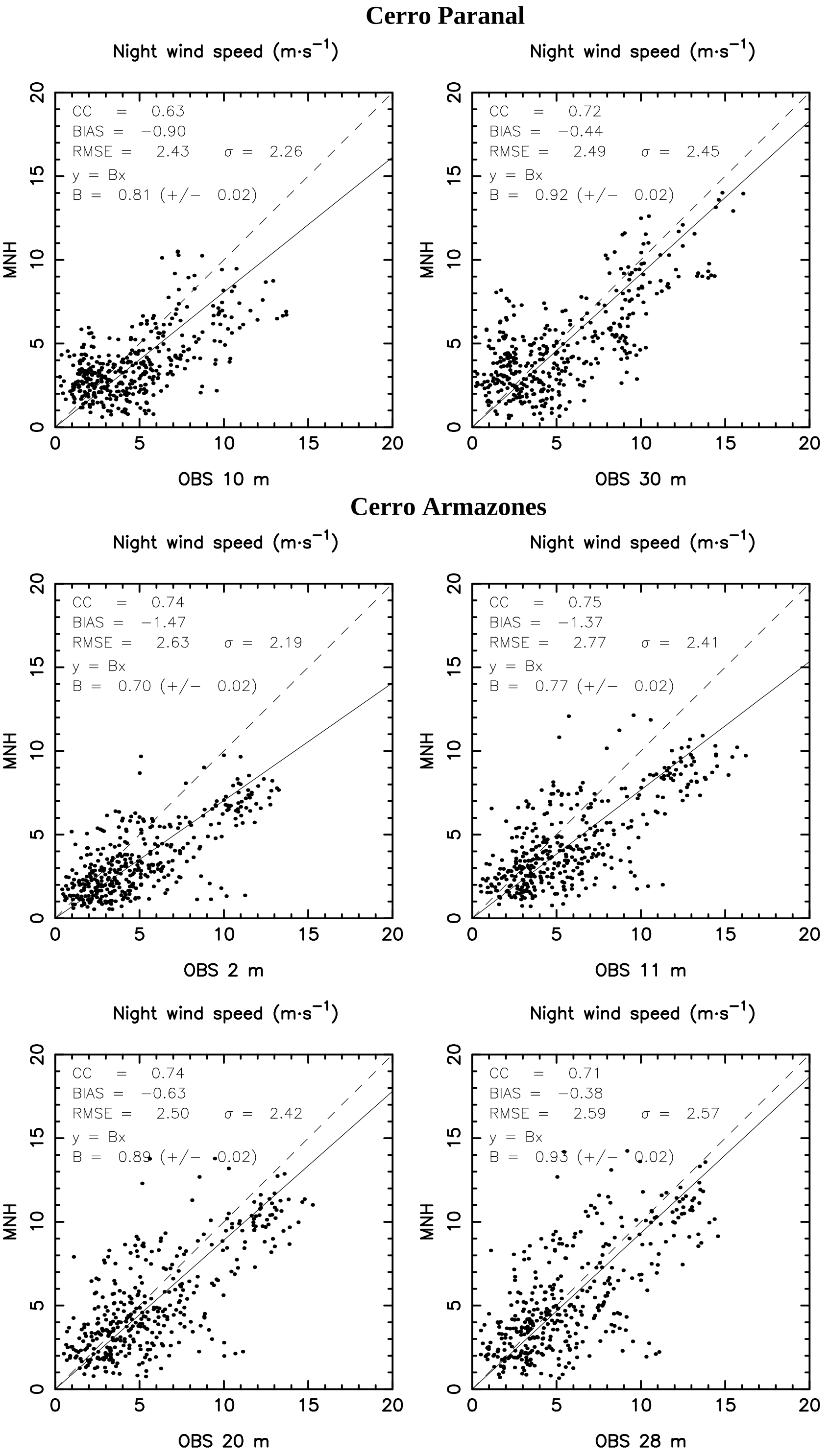


Figure 10: Same as Fig. 9, but with the high horizontal configuration for Meso-NH.

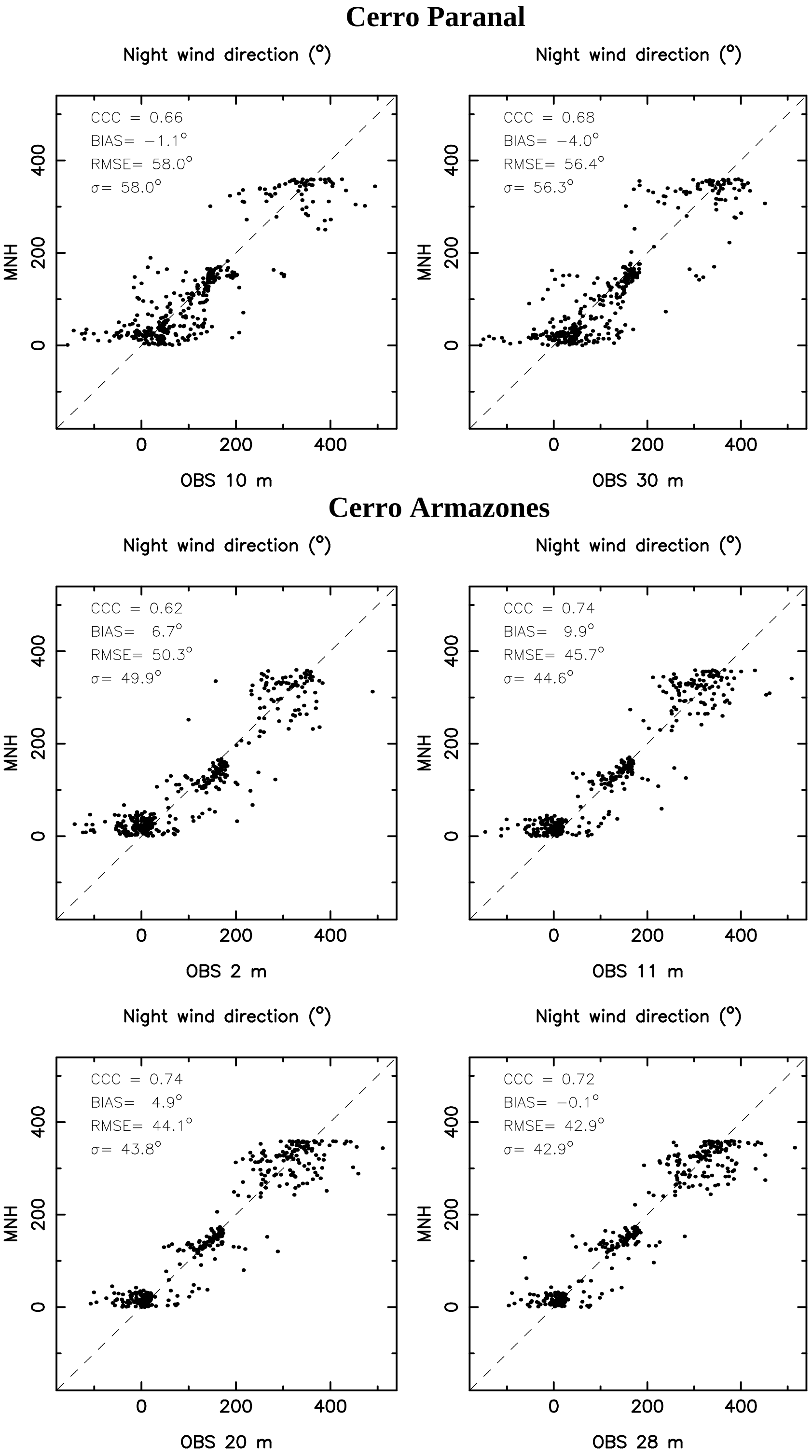


Figure 11: Scattered plot of Meso-NH **wind direction** against observations, at 10 m and 30 m at **Cerro Paranal** (top figures); at 2 m, 11 m, 20 m and 28 m at **Cerro Armazones** (central and bottom figures). Every point represents the average over an interval of 30 minutes. Caution: here is used the circular correlation coefficient [30], adapted for circular data. Notice that no regression line is calculated in the case of wind direction comparisons.

Absolute temperature (°C)

	PARANAL		ARMAZONES			
	2 m	30 m	2 m	11 m	20 m	28 m
BIAS	0.16	-0.15	0.78	0.08	0.06	0.03
RMSE	0.96	0.83	1.10	0.84	0.87	0.89
CC	0.96	0.97	0.97	0.97	0.97	0.96

Table 6: Near surface **temperature correlation coefficient**, **bias** and **RMSE** (Meson-NH minus Observations), using the data with re-sampling average over 30 min.

Wind speed ($m\cdot s^{-1}$) - $\Delta X = 500$ m

	PARANAL		ARMAZONES			
	10 m	30 m	2 m	11 m	20 m	28 m
BIAS	-2.17	-1.07	-3.58	-3.31	-2.33	-2.00
RMSE	3.14	2.62	4.54	4.74	4.20	4.15
CC	0.69	0.76	0.70	0.31	0.31	0.26

Table 7: Near surface **wind speed correlation coefficient**, **bias** and **RMSE** (Meson-NH with the standard configuration - maximum $\mathbf{\Delta X = 500}$ **m** - minus Observations).

Wind speed ($m\cdot s^{-1}$) - $\Delta X = 100$ m

	PARANAL		ARMAZONES			
	10 m	30 m	2 m	11 m	20 m	28 m
BIAS	-0.90	-0.44	-1.47	-1.37	-0.63	-0.38
RMSE	2.43	2.49	2.63	2.77	2.50	2.59
CC	0.63	0.72	0.74	0.75	0.74	0.71

Table 8: Near surface **wind speed correlation coefficient**, **bias** and **RMSE** (Meson-NH with the high horizontal resolution configuration - maximum $\mathbf{\Delta X = 100}$ **m** - minus Observations).

Fig. 12 and Fig. 13 show the temporal evolution of the average, the bias and the RMSE (at different levels and above Cerro Paranal and Cerro Armazones, respectively) of the absolute temperature. Looking at Fig. 12 and Fig. 13 we conclude that, above both Cerro Paranal and Cerro Armazones, we obtain excellent bias and RMSE values: the bias is well below 1°C (at some heights well inferior to 0.5°C) and, even more impressive, the RMSE is basically always inferior to 1°C[6].

Fig. 14 and Fig. 15 show the temporal evolution of the average, the bias and the RMSE (at different levels and above Cerro Paranal and Cerro Armazones, respectively) of the wind speed. We note a general tendency of the model in underestimating the wind speed all along night, particularly at the first level (2 m) even if the model seems to show a better behavior above Cerro Paranal than Cerro Armazones. Even if the bias and the RMSE are still small in absolute terms, considering the typical absolute values of the wind speed at this heights, we can have a non negligible relative error. We have also performed an analog statistical analysis with a different model configuration (five imbricated models with the higher horizontal resolution of 100 m) to check if the model performances in reconstructing the wind speed near the surface in proximity of these peaks can be improved. Fig. 16 and Fig. 17 shows the results in terms of average, bias and RMSE. We found that such a new configuration definitely and substantially improves the model performances in reconstructing the wind speed near the surface of around 50%.

---

[6] We highlight once more that the performances are valid for the night time as described at the beginning of this Chapter. In case one wishes to study the model performances during the day time, we suggest a different instant for the beginning of the simulation and a different sequence of forcing analyses. In other worlds, in Fig.12-Fig.19 we report the temporal evolution of different parameters with respect to the total simulation time but it is meaningless to try to retrieve the model performances during the day time from those figures.

Wind direction (^{o})

	PARANAL		ARMAZONES			
	10 m	30 m	2 m	11 m	20 m	28 m
BIAS	-1.1	-4.0	6.7	9.9	4.9	-0.1
RMSE	58.0	56.4	50.3	45.7	44.1	42.9
CCC	0.66	0.68	0.62	0.74	0.74	0.72

Table 9: Near surface **wind direction correlation coefficient**, **bias** and **RMSE** (Meson-NH with the standard configuration - maximum ΔX = 500 m - minus Observations).

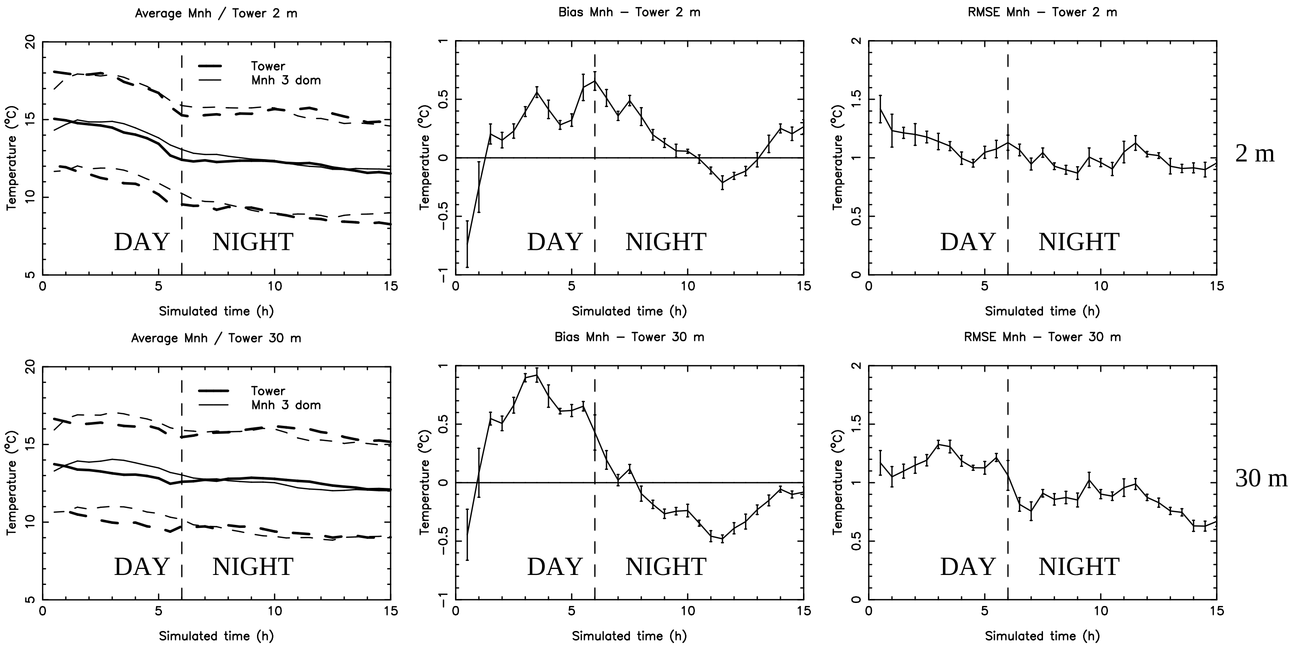


Figure 12: Temporal evolution of the **absolute temperature average** (the bold line is the observation average and the thin line is Meso-NH average), **bias** (Mnh - Observations) and **RMSE** at **Cerro Paranal** (top: at 2 m; bottom: at 30 m). The x-axis represents the time from the beginning of the simulation (00 h is 18 UT of the day before, 06 h is 00 UT, and 15 h is 09 UT). The nights starts at around 06 h (00 UT - 20 LT) and is delimited by the vertical dashed line. As explained in Section 3 the model configuration selected for this study is optimized for the investigation of the night period. Meso-NH is with the ΔX = 500 m configuration. Error bars are $\pm\sigma$ (standard deviation).

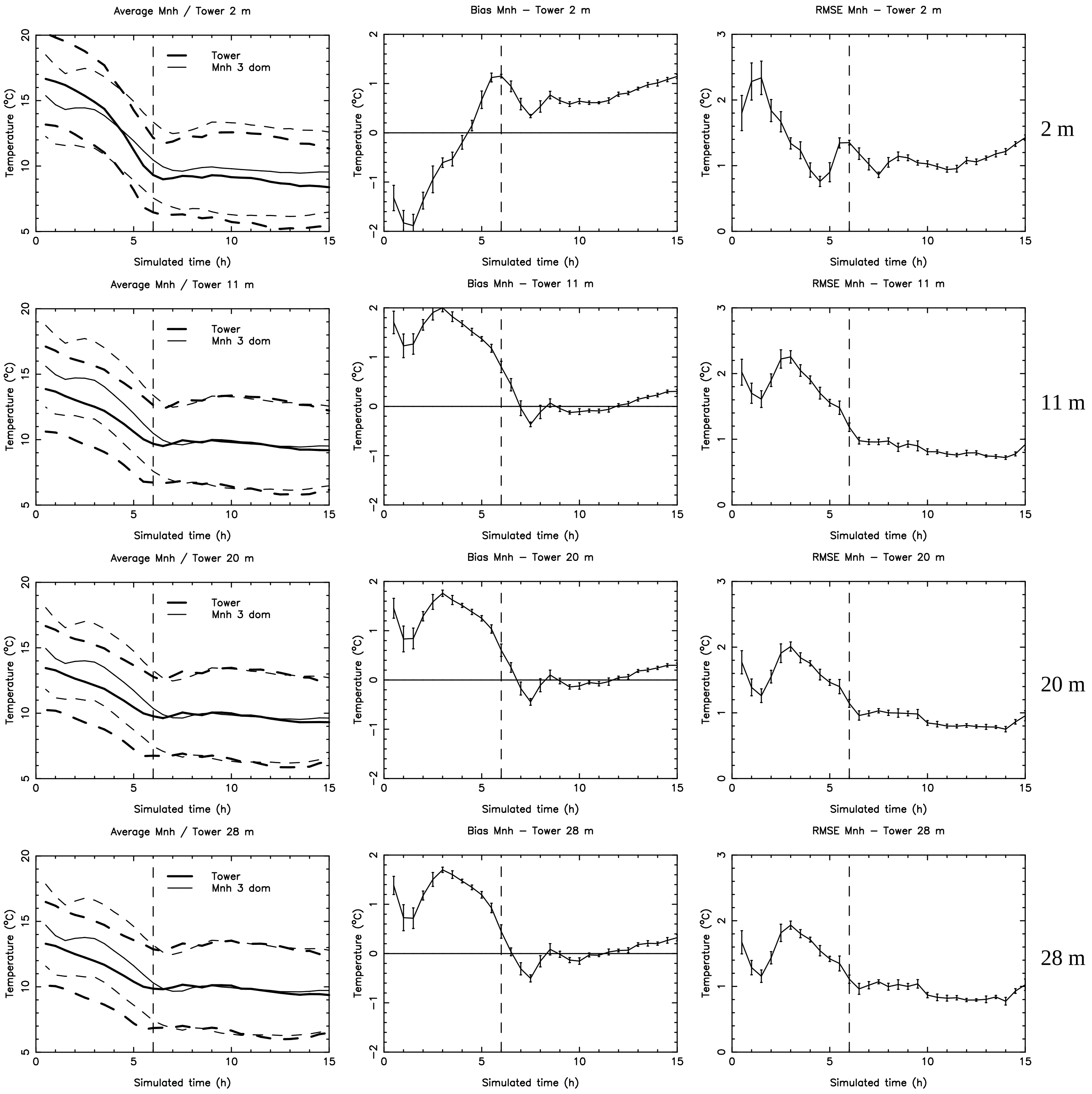


Figure 13: Same as Fig. 12 but for the **absolute temperature** at **Cerro Armazones** (from top to bottom: at 2 m, 11 m, 20 m and 28 m).

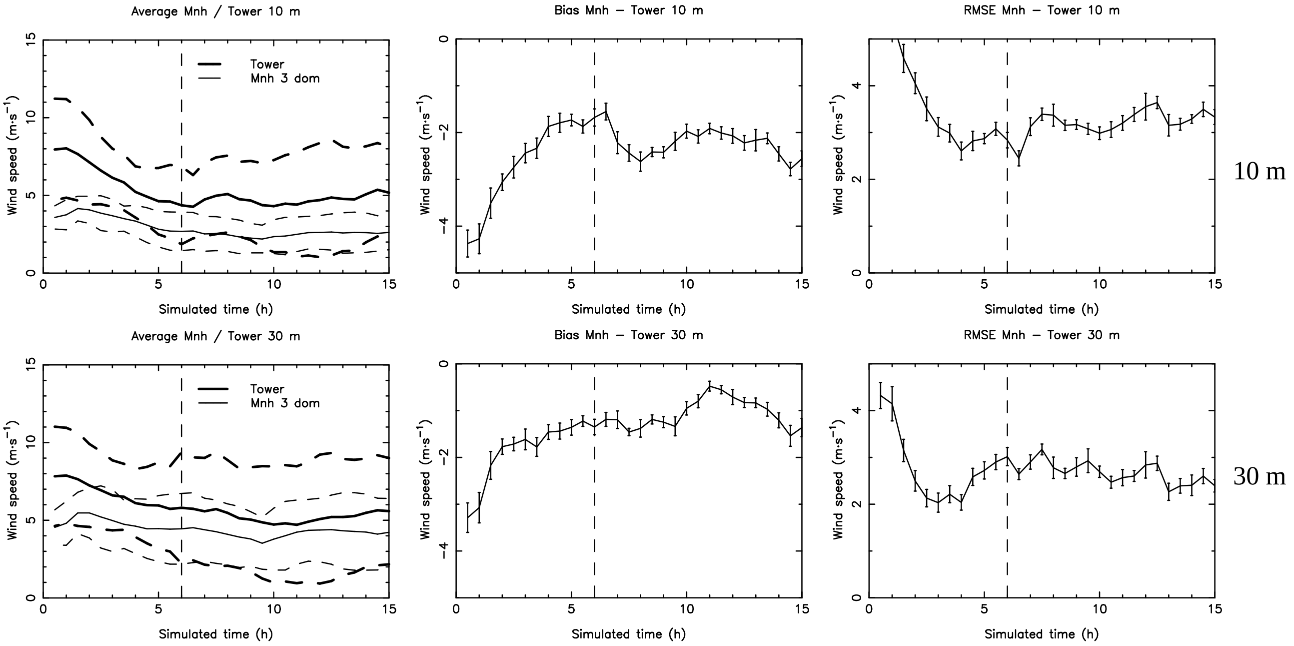


Figure 14: Same as Fig. 12 but for the **wind speed** at **Cerro Paranal** (top: at 10 m; bottom: at 30 m). Meso-NH is used with $\Delta X = 500$ m configuration.

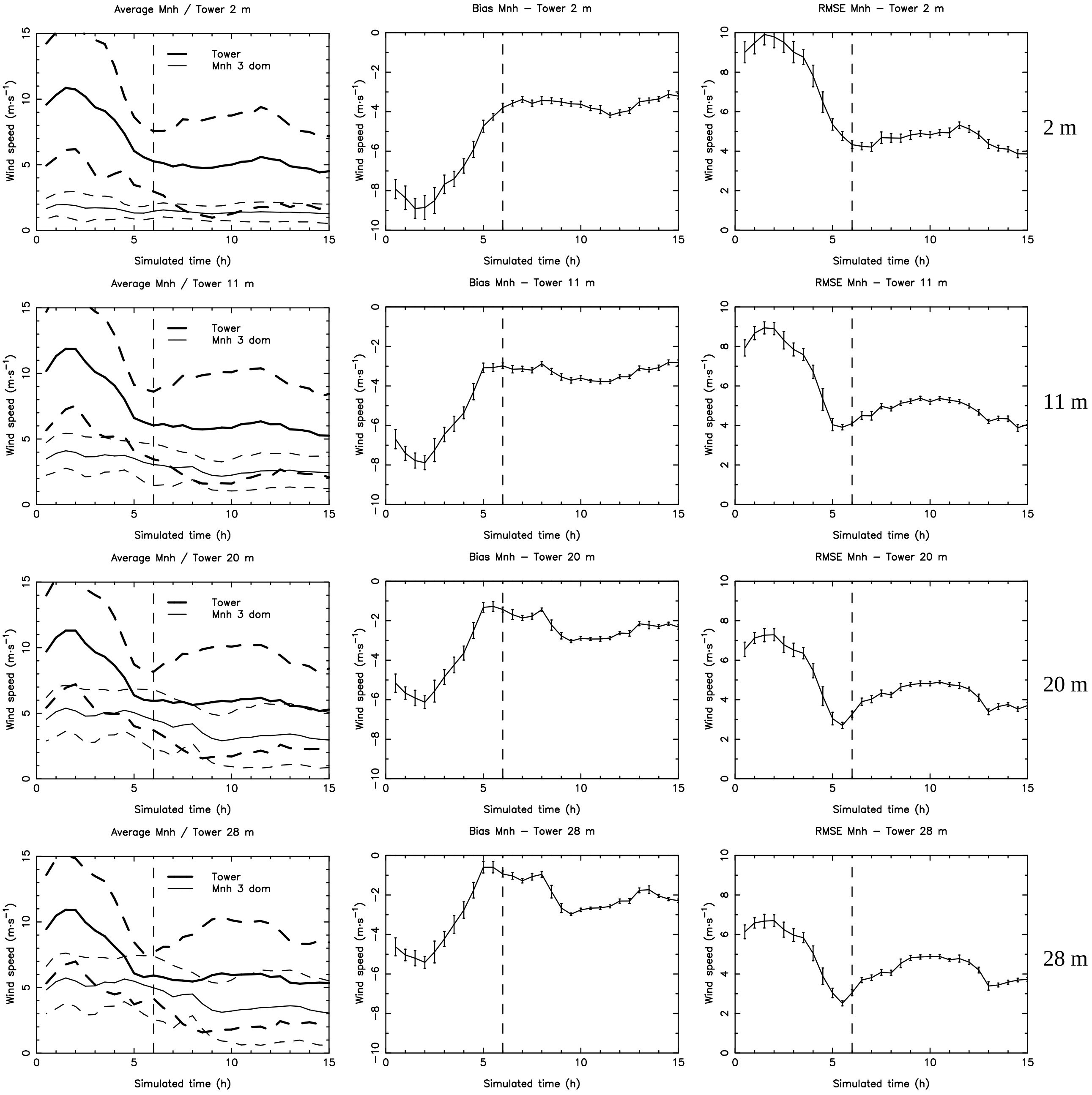


Figure 15: Same as Fig. 12 but for the **wind speed** at **Cerro Armazones** (from top to bottom: at 2 m, 11 m, 20 m and 28 m). Meso-NH is used with $\Delta X = 500$ m configuration.

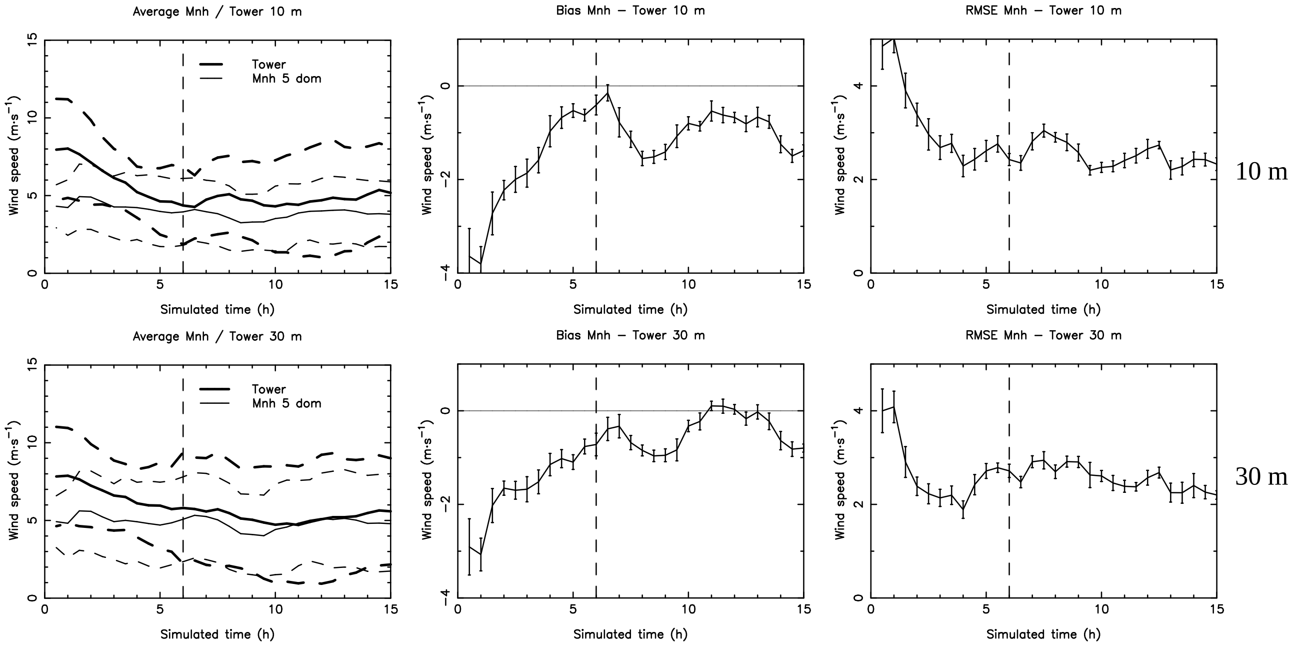


Figure 16: Same as Fig. 14 but Meso-NH is used in the ΔX = 100 m configuration.

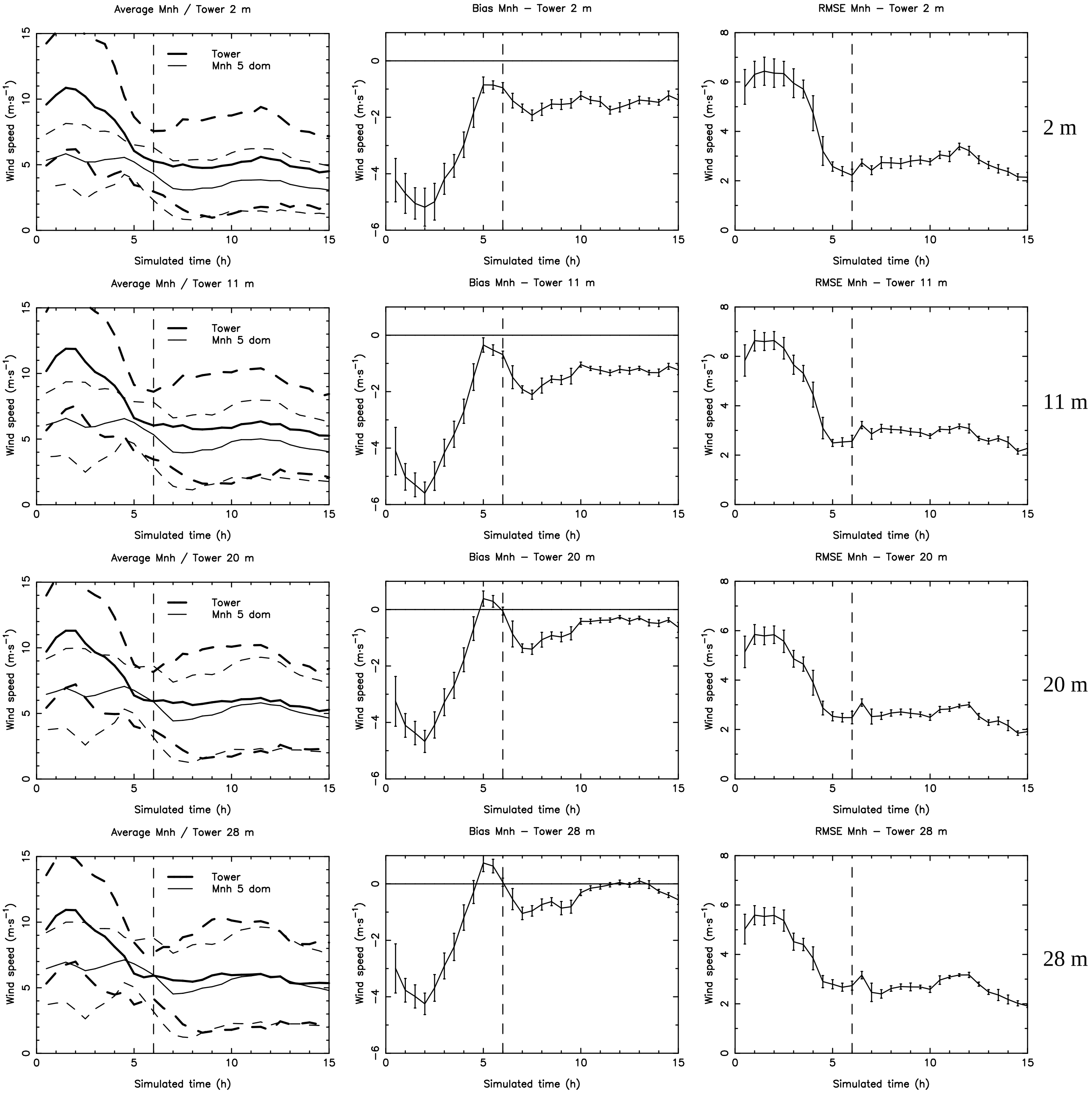


Figure 17: Same as Fig. 15 but Meso-NH is used in the ΔX = 100 m configuration.

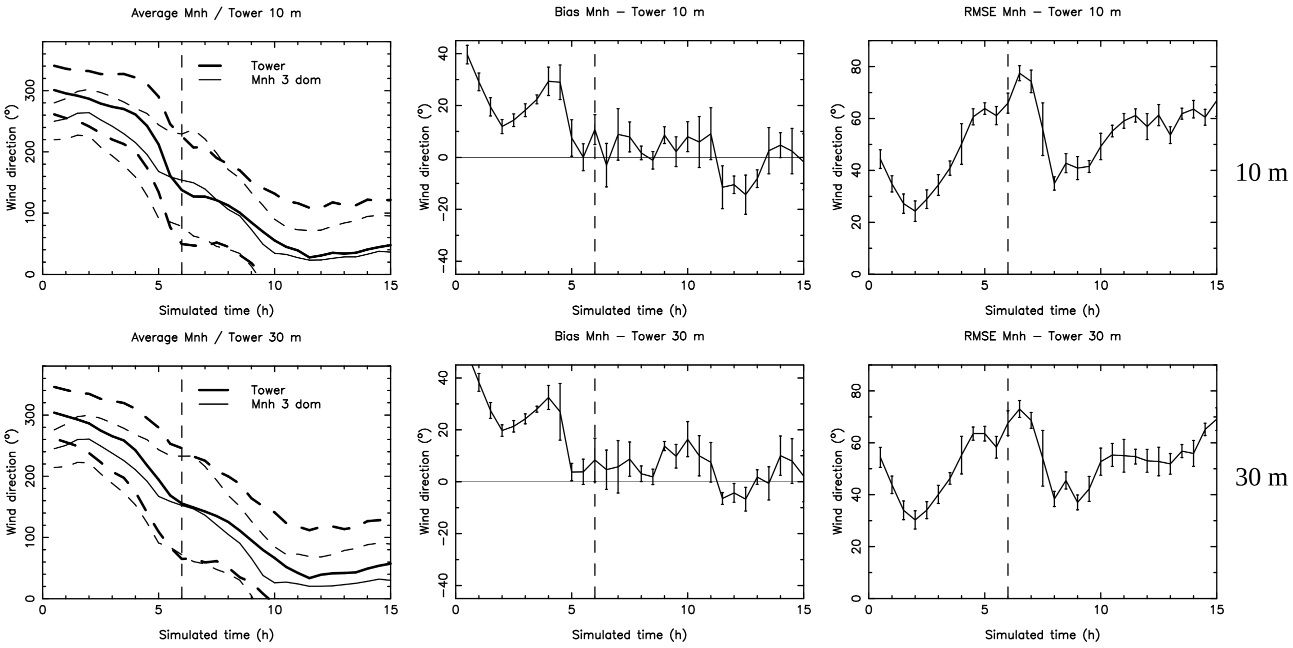


Figure 18: Same as Fig. 12 but for the **wind direction** at **Cerro Paranal** (top: at 10 m; bottom: at 30 m).

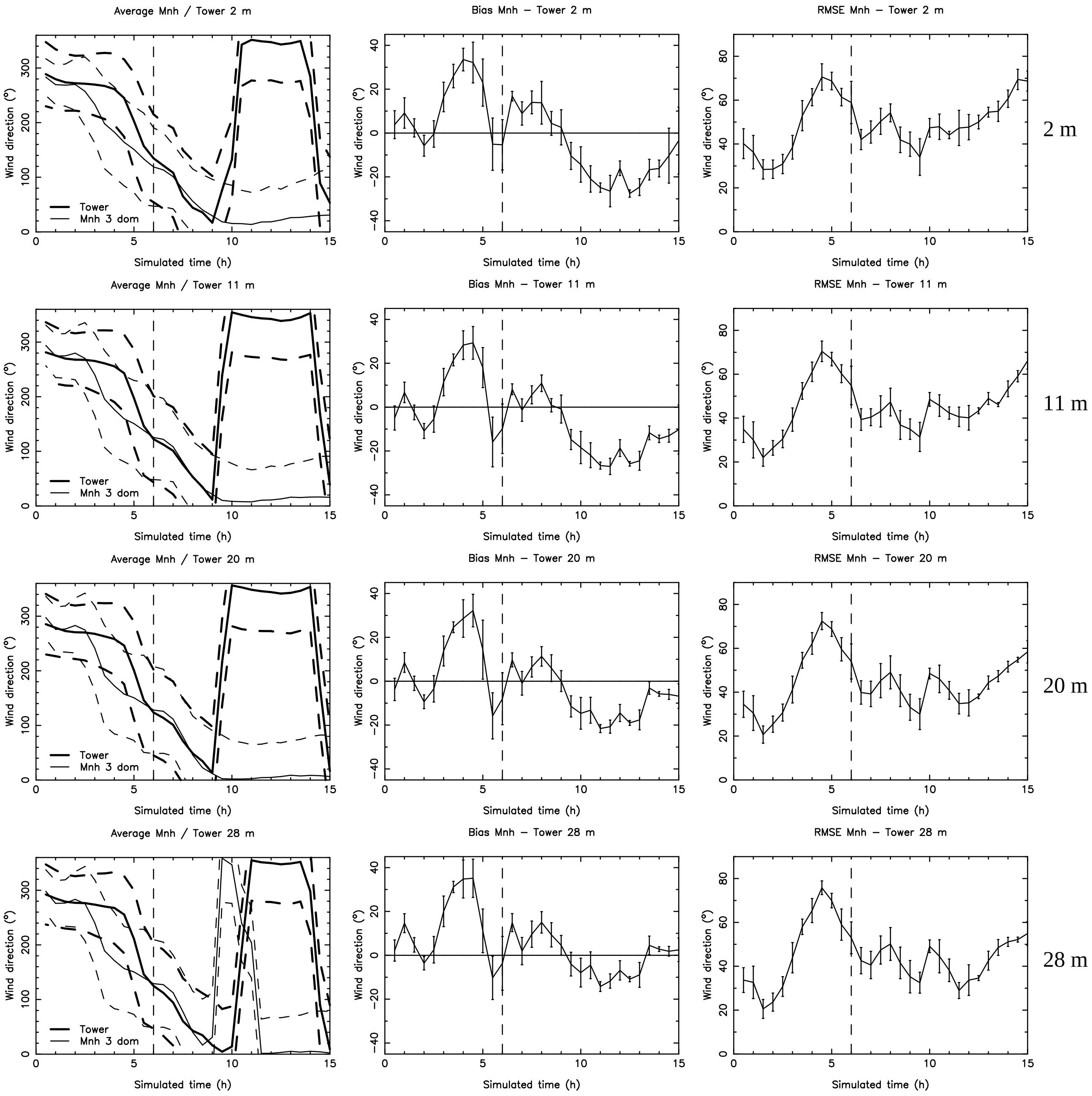


Figure 19: Same as Fig. 12 but for the **wind direction** at **Cerro Armazones** (from top to bottom: at 2 m, 11 m, 20 m and 28 m).

5.3 Individual nights model performances

In this section we present the analysis done on the model performances in reconstructing the atmospheric parameters, night by night. We anticipate that, in general, this is a much more ambitious challenge for evident statistical reasons. We computed for every single night (from the 20 nights sample of Table 65) the bias, RMSE and cc between model and observations, for both sites and at every level near the surface where observations were available, for the absolute temperature and the wind speed and direction. We have then computed the median, first and third quartiles. Results are summarized in Tables 10, 11, 12 and 13, for the temperature, the wind speed and the wind direction, respectively. In Fig. 20 and Fig. 21 are reported the cumulative distributions of bias, RMSE and cc of the 20 nights for the absolute temperature at Cerro Paranal (at 2 m and 30 m) and Cerro Armazones (at 2 m, 11 m, 20 m and 28 m), respectively. In Fig. 22 and Fig. 23 are reported the cumulative distributions of bias, RMSE and cc of the 20 nights for the wind speed at Cerro Paranal (at 10 m and 30 m) and Cerro Armazones (at 2 m, 11 m, 20 m and 28 m), respectively.
In Fig. 24 and Fig.25 are reported the cumulative distributions of bias, RMSE and ccc (circular correlation coefficient) of the 20 nights for the wind direction at Cerro Paranal (at 10 m and 30 m) and Cerro Armazones (at 2 m, 11 m, 20 m and 28 m), respectively.

The very good model performances for the absolute temperature and for the wind direction deduced from the overall statistical analysis are confirmed here. The highest median bias for the temperature is found at Cerro Armazones, at 2 m, and is only 0.64°C. The median RMSE is always good, with values always inferior to 1°C. Even more remarkable is the fact that 75% of the time, the RMSE is inferior to 1.23°C at 2 m at Cerro Armazones (worst value) and inferior to 0.90°C at 30 m at Cerro Paranal (best value).

The wind direction median bias is very close to zero (between -5.45° and 7.28° at all levels), and the median RMSE is in the interval [37.57° - 46.16°]. More over, for 75% of the nights, the RMSE remains alway inferior to 77.45° (which is the highest value, at 2 m at Cerro Paranal), and at some levels can be inferior to 60° for 75% of the nights (at 11 m, 20 m and 28 m at Cerro Armazones). In Section 7 we will discuss the model hit rates (or contingency tables) that is an alternative and complementary method to quantify the model performances. In that context it will be more easy to try to discuss the model performances in predicting the wind direction within the good quadrant.

Concerning the wind speed, the same conclusions than for the overall statistical analysis can be drawn, but the values of the median bias and RMSE are slightly better. The median bias ranges from 0.82 m·s^{-1} at 30 m at Cerro Paranal to 2.40 m·s^{-1} at 2 m at Cerro Armazones. The highest median RMSE is found at Cerro Armazones at 2 m and is equal to 2.71 m·s^{-1}. Performances are significantly improved when the highest horizontal resolution is $\Delta X = 100$ m. The median bias is halved, or better: its best value is found at Cerro Armazones at 28 m and is equal to 0.12 m·s^{-1}, and its worst value remains inferior to 1 m·s^{-1} (0.93 m·s^{-1} at both 2 m and 11 m high at Cerro Armazones). Also, the improvement in the RMSE is not negligible. Even though the RMSE median values are ranging from 1.85 m·s^{-1} at Cerro Paranal at 2 m to 2.18 m·s^{-1} at 20 m at Cerro Armazones, i.e very similar to the $\Delta X = 500$ m configuration, the improvement is visible looking at the 1st and 3rd quartiles. The values of the 3rd quartiles are strongly reduced. Its highest value with the $\Delta X = 500$ m configuration was 6.41 m·s^{-1} (at 11 m at Cerro Armazones), it is now 3.34 m·s^{-1} (at 2 m at Cerro Armazones) with the $\Delta X = 100$ m configuration.

In conclusion it seems that, in spite of the intrinsic higher level of difficulty, the model performances in reconstructing the atmospheric parameters in proximity of the surface night by night is as good as in overall statistical terms and in many cases, even better !

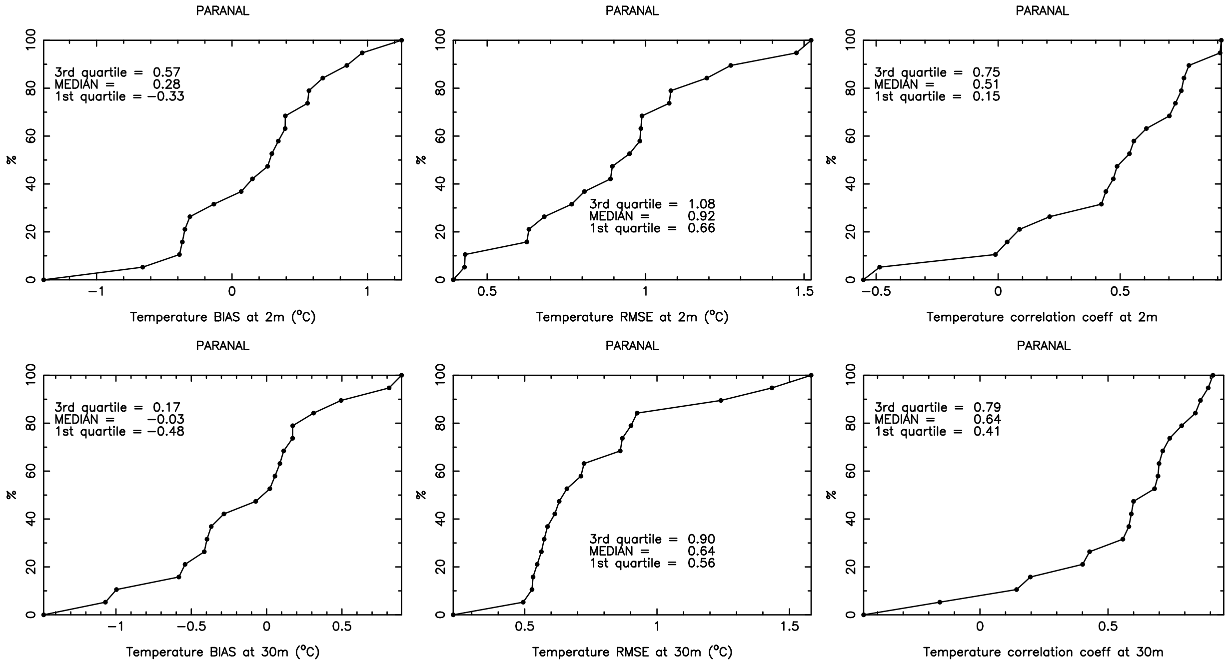


Figure 20: Cumulative distribution of **bias** (Mnh - Observations, on the left) **RMSE** (in the middle) and **correlation coefficient** (on the right) of the **temperature** at **Cerro Paranal** (20 nights sample), at 2 m (top) and 30 m (bottom). See Table 10 for a summarize of the bias and RMSE values. Meso-NH is in the $\Delta X = 500$ m configuration.

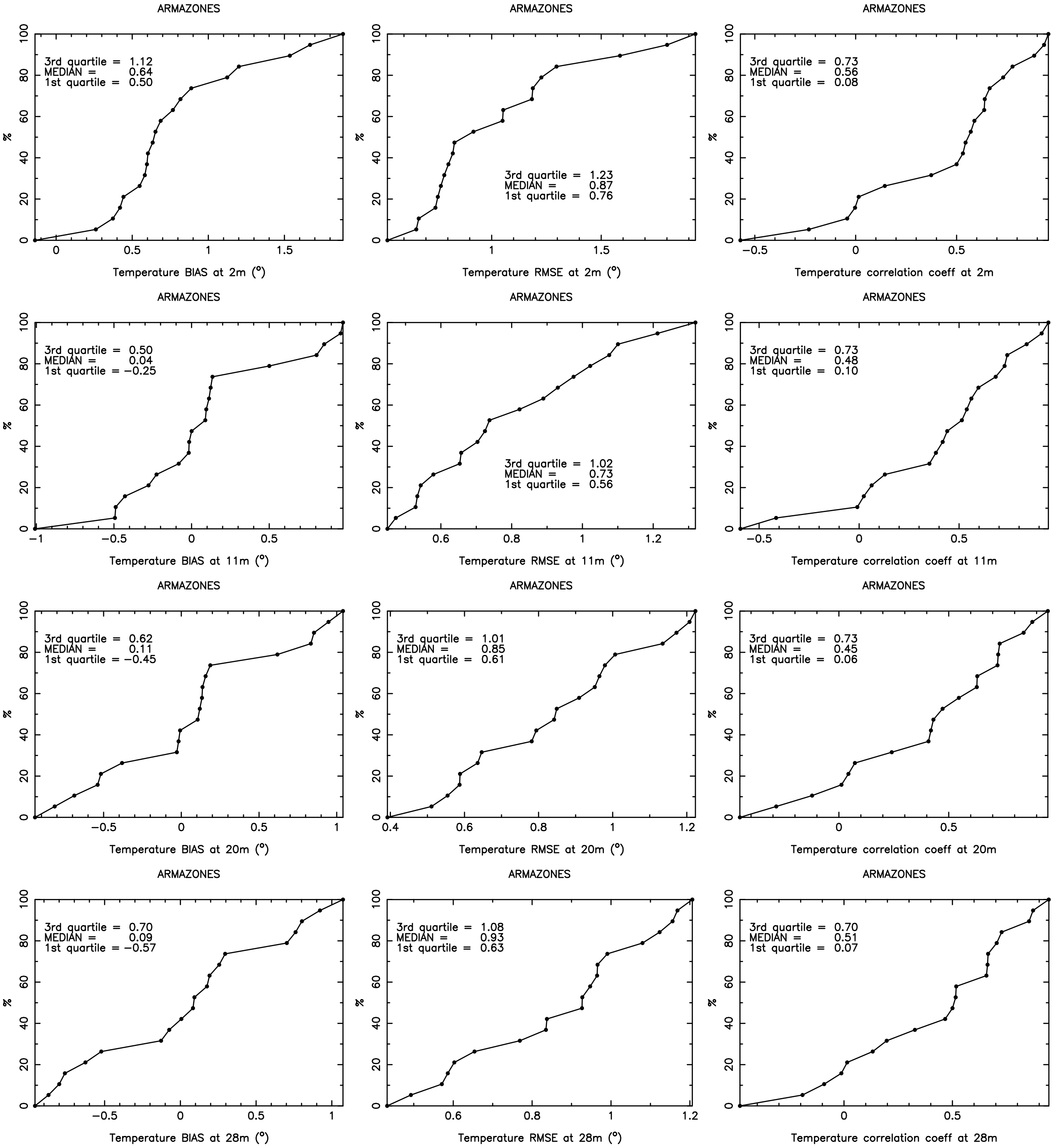


Figure 21: Cumulative distribution of **bias** (Mnh - Observations, on the left) **RMSE** (in the middle) and **correlation coefficient** (on the right) of the **temperature** at **Cerro Armazones** (20 nights sample), at 2 m, 11 m, 20 m and 28 m (from top to bottom, respectively). See Table 10 for a summarize of the bias and RMSE values. Meso-NH is in the $\Delta X = 500$ m configuration.

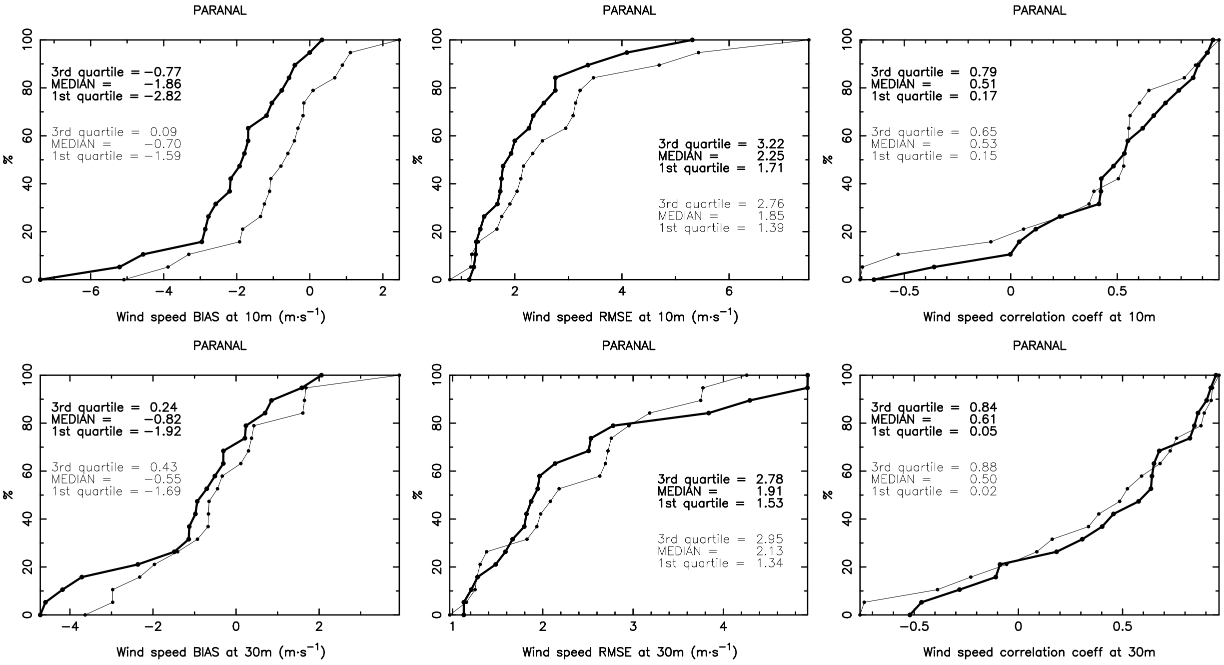


Figure 22: Cumulative distribution of **bias** (Mnh - Observations, on the left) **RMSE** (in the middle) and **correlation coefficient** (on the right) of the **wind speed** at **Cerro Paranal** (20 nights sample), at 10 m (top) and 30 m (bottom). In thick line, Meso-NH is with the standard horizontal resolution ($\Delta X = 500$ m). In thin line, with the high horizontal resolution ($\Delta X = 100$ m). See Tables 11 and 12 for a summarize of the bias and RMSE values.

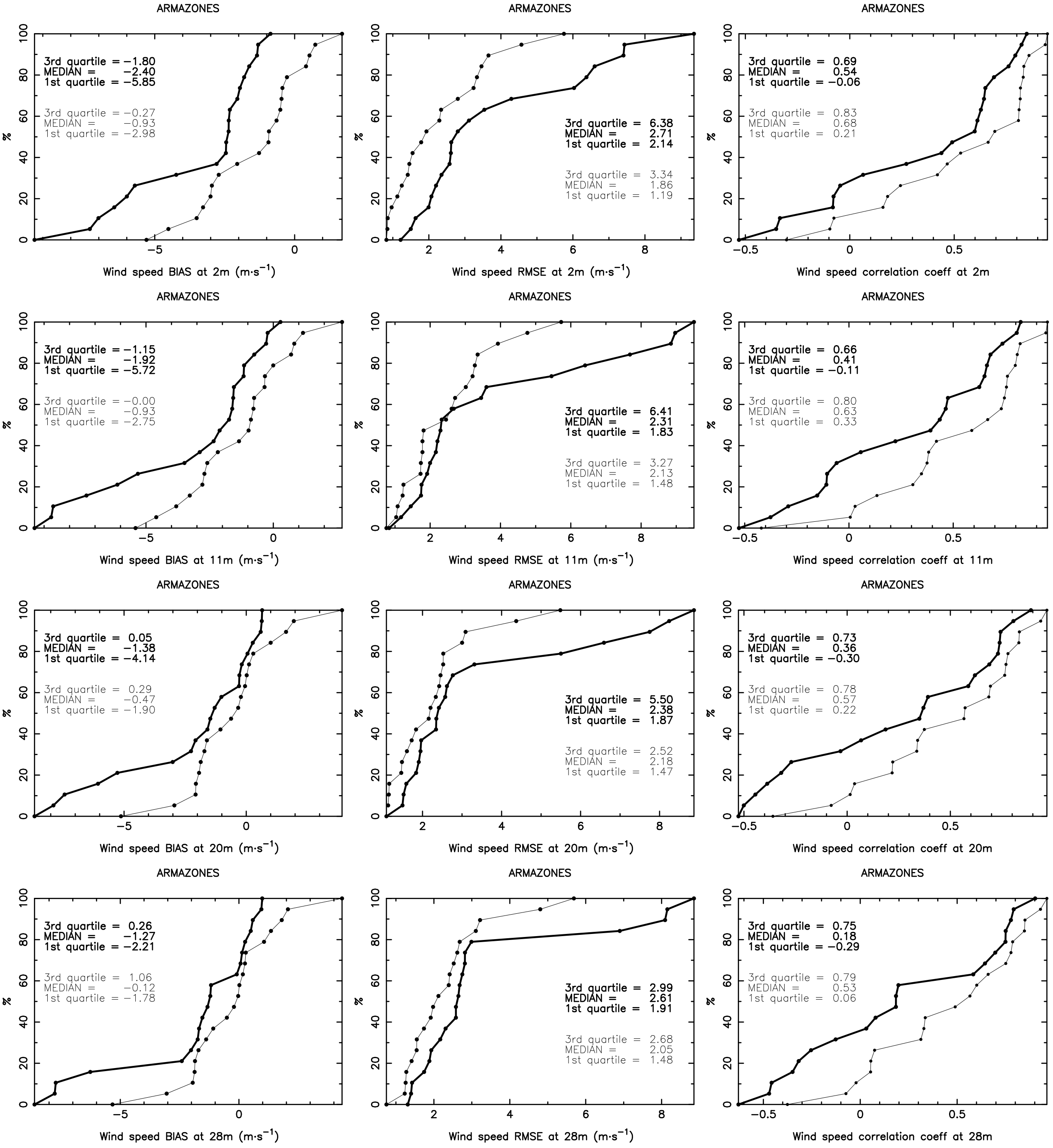


Figure 23: Cumulative distribution of **bias** (Mnh - Observations, on the left) **RMSE** (in the middle) and **correlation coefficient** (on the right) of the **wind speed** at **Cerro Armazones** (20 nights sample), at 2 m, 11 m, 20 m and 28 m (from top to bottom, respectively). See Tables 11 and 12 for a summarize of the bias and RMSE values. In thick line, Meso-NH is with the standard horizontal resolution ($\Delta X = 500$ m). In thin line, with the high horizontal resolution ($\Delta X = 100$ m).

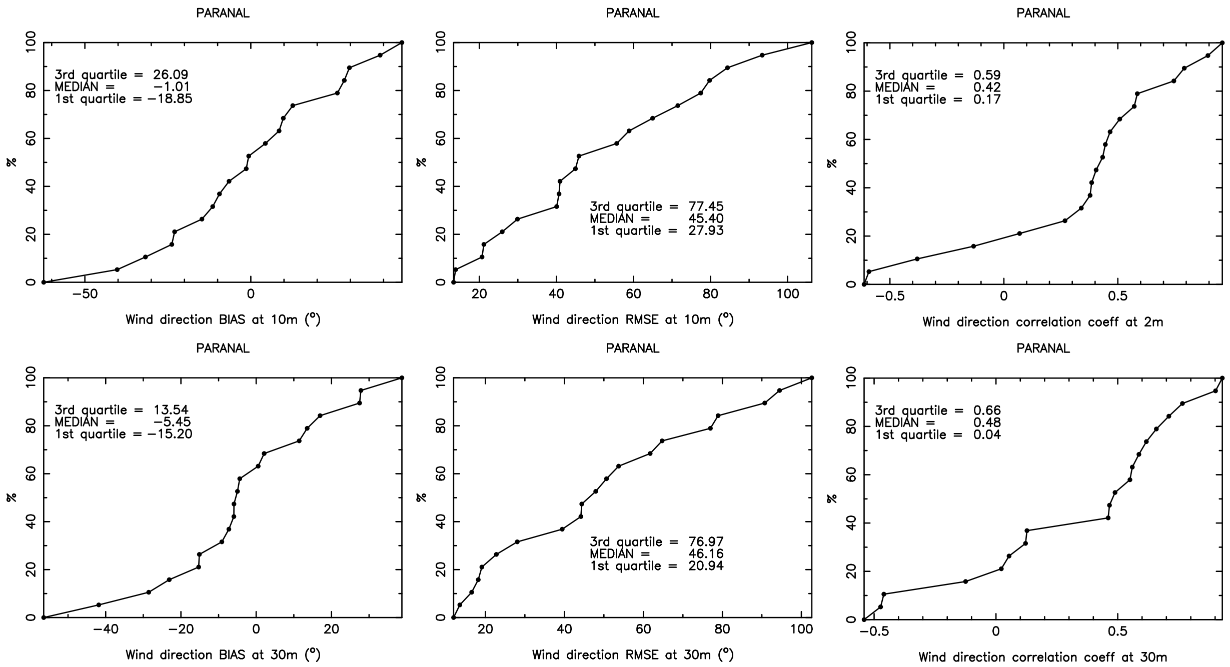


Figure 24: Cumulative distribution of **bias** (Mnh - Observations, on the left) **RMSE** (in the middle) and **circular correlation coefficient** (on the right) of the **wind direction** at **Cerro Paranal** (20 nights sample), at 10 m (top) and 30 m (bottom). See Table 13 for a summarize of the bias and RMSE values. Meso-NH is with the standard horizontal resolution ($\Delta X = 500$ m).

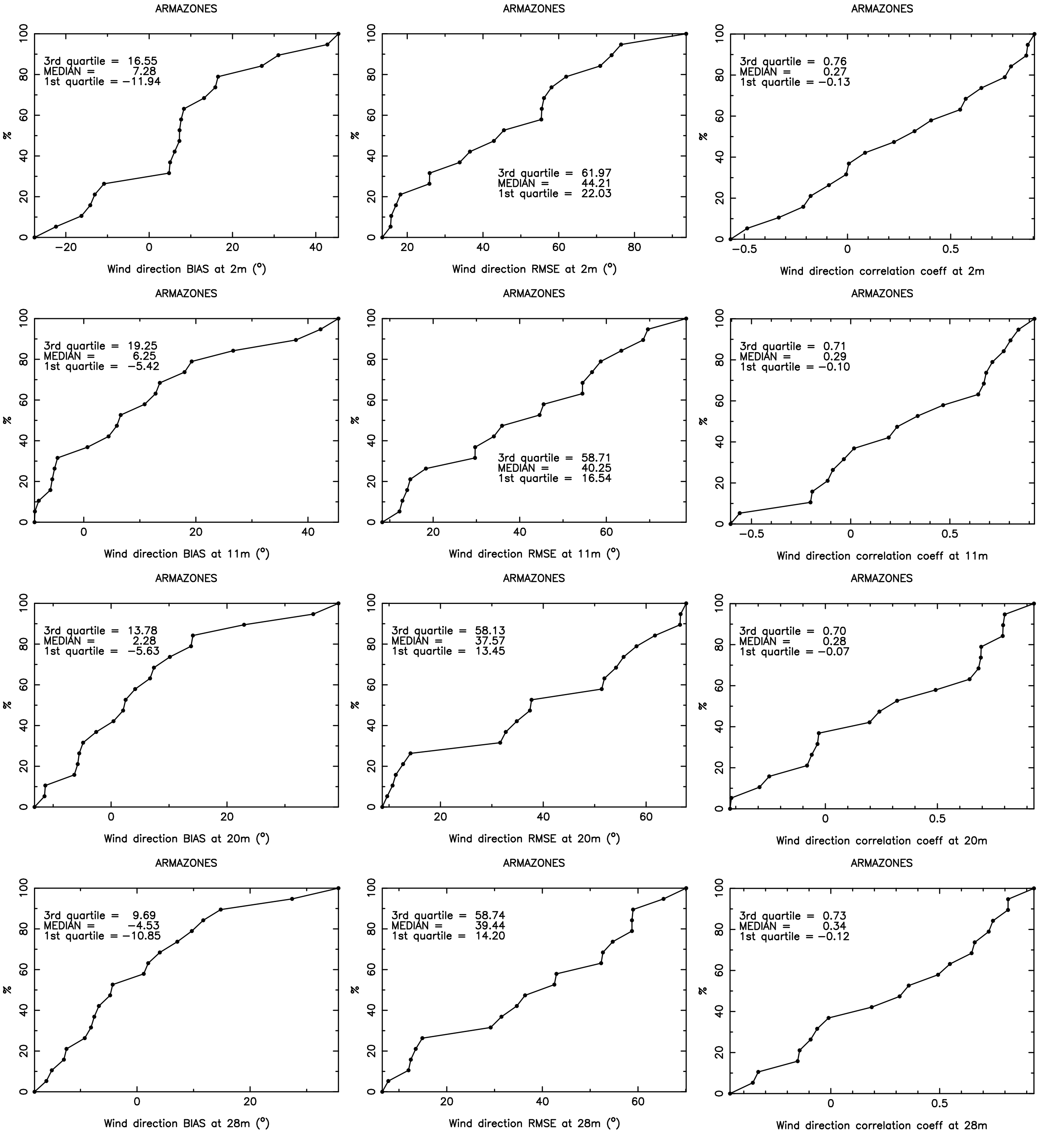


Figure 25: Cumulative distribution of **bias** (Mnh - Observations, on the left) **RMSE** (in the middle) and **circular correlation coefficient** (on the right) of the **wind direction** at **Cerro Armazones** (20 nights sample), at 2 m, 11 m, 20 m and 28 m (from top to bottom, respectively). See Table 13 for a summarize of the bias and RMSE values. Meso-NH is with the standard horizontal resolution ($\Delta X = 500$ m).

Absolute temperature (^{o}C)

	PARANAL		ARMAZONES			
	2 m	30 m	2 m	11 m	20 m	28 m
BIAS	$0.28^{+0.57}_{-0.33}$	$-0.03^{+0.17}_{-0.48}$	$0.64^{+1.12}_{+0.50}$	$0.04^{+0.50}_{-0.25}$	$0.11^{+0.62}_{-0.45}$	$0.09^{+0.70}_{-0.57}$
RMSE	$0.92^{+1.08}_{+0.66}$	$0.64^{+0.90}_{+0.56}$	$0.87^{+1.23}_{+0.76}$	$0.73^{+1.02}_{+0.56}$	$0.85^{+1.01}_{+0.61}$	$0.93^{+1.08}_{+0.63}$
CC	$0.51^{+0.75}_{+0.15}$	$0.64^{+0.79}_{+0.41}$	$0.56^{+0.73}_{+0.08}$	$0.48^{+0.73}_{+0.10}$	$0.45^{+0.73}_{+0.06}$	$0.51^{+0.70}_{+0.07}$

Table 10: Near surface median **correlation coefficient**, **bias** and **RMSE** (Meson-NH minus Observations), of the **temperature** from the single nights (Table 65) values (see Fig. 20 and Fig. 21). In small fonts, the 1st and 3rd quartiles.

Wind speed ($\mathrm{m \cdot s^{-1}}$) - ΔX = 500 m

	PARANAL		ARMAZONES			
	10 m	30 m	2 m	11 m	20 m	28 m
BIAS	$-1.86^{-0.77}_{-2.82}$	$-0.82^{+0.24}_{-1.92}$	$-2.40^{-1.80}_{-5.85}$	$-1.92^{-1.15}_{-5.72}$	$-1.38^{+0.05}_{-4.14}$	$-1.27^{+0.26}_{-2.21}$
RMSE	$2.25^{+3.22}_{+1.71}$	$1.91^{+2.78}_{+1.53}$	$2.71^{+6.38}_{+2.14}$	$2.31^{+6.41}_{+1.83}$	$2.38^{+5.50}_{+1.87}$	$2.61^{+2.99}_{+1.91}$
CC	$0.51^{+0.79}_{+0.17}$	$0.61^{+0.84}_{+0.05}$	$0.54^{+0.69}_{-0.06}$	$0.41^{+0.66}_{-0.11}$	$0.36^{+0.73}_{-0.30}$	$0.18^{+0.75}_{-0.29}$

Table 11: Near surface median **correlation coefficient**, **bias** and **RMSE** (Meson-NH minus Observations) of the **wind speed**, using the Meso-NH ΔX = 500 m configuration from the single nights (Table 65) values (see Fig. 22 and Fig. 23). In small fonts, the 1st and 3rd quartiles.

Wind speed ($\mathrm{m \cdot s^{-1}}$) - ΔX = 100 m

	PARANAL		ARMAZONES			
	10 m	30 m	2 m	11 m	20 m	28 m
BIAS	$-0.70^{0.09}_{-1.59}$	$-0.55^{+0.43}_{-1.69}$	$-0.93^{-0.27}_{-2.98}$	$-0.93^{-0.00}_{-2.75}$	$-0.47^{+0.29}_{-1.90}$	$-0.12^{+1.06}_{-1.78}$
RMSE	$1.85^{+2.76}_{+1.39}$	$2.13^{+2.95}_{+1.34}$	$1.86^{+3.34}_{+1.19}$	$2.13^{+3.27}_{+1.48}$	$2.18^{+2.52}_{+1.47}$	$2.05^{+2.68}_{+1.48}$
CC	$0.53^{+0.65}_{+0.15}$	$0.50^{+0.88}_{+0.02}$	$0.68^{+0.83}_{+0.21}$	$0.63^{+0.80}_{+0.33}$	$0.57^{+0.78}_{+0.22}$	$0.53^{+0.79}_{+0.06}$

Table 12: Near surface median **correlation coefficient**, **bias** and **RMSE** (Meson-NH minus Observations) of the **wind speed**, using the Meso-NH ΔX = 100 m configuration from the single nights (Table 65) values (see Fig. 22 and Fig. 23). In small fonts, the 1st and 3rd quartiles.

Wind direction (^{o})

	PARANAL		ARMAZONES			
	2 m	30 m	2 m	11 m	20 m	28 m
BIAS	$-1.01^{+26.09}_{-18.85}$	$-5.45^{+13.54}_{-15.20}$	$7.28^{+16.55}_{-11.94}$	$6.25^{+19.25}_{-5.42}$	$2.28^{+13.78}_{-5.63}$	$-4.53^{+9.69}_{-10.85}$
RMSE	$45.40^{+77.45}_{+27.93}$	$46.16^{+76.97}_{+20.94}$	$44.21^{+61.97}_{+22.03}$	$40.25^{+58.71}_{+16.54}$	$37.57^{+58.13}_{+13.45}$	$39.44^{+58.74}_{+14.20}$
CCC	$0.42^{+0.59}_{+0.17}$	$0.48^{+0.66}_{+0.04}$	$0.27^{+0.76}_{-0.13}$	$0.29^{0.71}_{-0.10}$	$0.28^{+0.70}_{-0.07}$	$0.34^{+0.73}_{-0.12}$

Table 13: Near surface median **circular correlation coefficient**, **bias** and **RMSE** (Meson-NH minus Observations), of the **wind** direction from the single nights (Table 65) values (see Fig. 24 and Fig. 25). In small fonts, the 1st and 3rd quartiles.

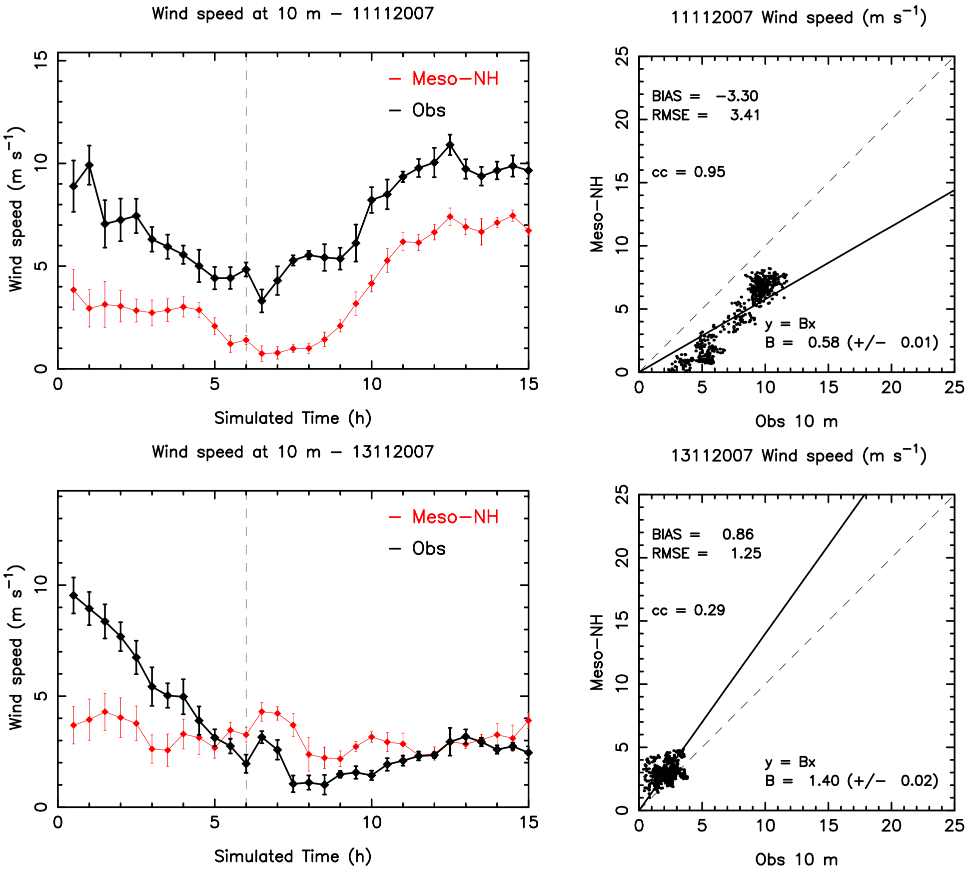


Figure 26: Examples of the **temporal evolution** of the **wind speed** during 2 different nights (on the left) and the corresponding **scattered plots** with the bias, RMSE and **cc** values (on the right).

After a deep analysis we realized that the correlation coefficient cc **(1)** doesn't provide any further information with respect to the bias and the RMSE on the reliability of the model in reconstructing the atmospherical parameters. It doesn't provide any information on the forecast of the trend (as could do a simple temporal evolution average) since it doesn't take into account the time; **(2)** can be misleading in many cases (bad value for visibly good forecasts, and *vice-versa*). This is particularly evident when we consider the night by night case. An example is visible in Fig. 26. We observe that the wind speed matches in a much better reliable way to observations during the night of 13 November than the night of 11 November while the cc for the 11 November night is much better than for the 13 November night (0.95 vs 0.29). For this reason we have decided to discard the analysis of the cc in the rest of the study.

5.4 Correlation between seeing and temperature gradient near the surface

In this section we simply provide a few elements aiming at verifying if it can be useful to investigate the ability of the model in predict the temperature gradient near the surface. Fig. 27 displays the total observed seeing (from both DIMM and GS instruments) against the temperature gradient near the surface (ΔT_{30m} - ΔT_{2m}) at Cerro Paranal. In this plot, only the positive gradients during the local night are considered. This figure clearly shows a trend between bad seeing conditions (with higher temperature gradients) and good seeing conditions (with lower temperature gradients)[7]. As the Meso-NH model succeeds very well in reproducing the temperature near the ground, with a good level of accuracy (the RMSE remains below 1^oC), we can think that such a discrimination could be reproduced by the model also for the temperature gradient. Looking at temporal evolution of the surface temperature (see BSCW archive) it seems that an horizontal resolution of the model of ΔX=100 m provides a better match with observations. This analysis is only preliminary, a more detailed study should be necessary to check which is the required accuracy for the gradient of the temperature and to verify which are the intervals of the gradient of temperature in which the correlation between seeing and the thermal gradient can work in an univocal way. In case of positive answer this should contribute to provide a further indicator for the forecasting of the seeing developed in the surface layer (first 30 meters above the ground).

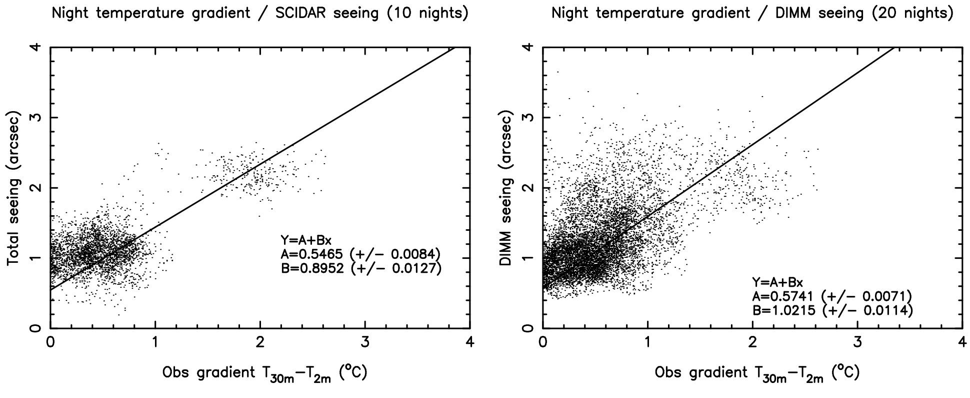


Figure 27: On the left: GS total seeing vs observed temperature gradient between 2 m and 30 m, for a total sample of 10 nights in 2007 (subsample of the 20 nights from Tab. 66 where both temperature observations and GS measurements are available). On the right: DIMM total seeing vs observed temperature gradient between 2 m and 30 m, for a total sample of 20 nights in 2007 (Tab. 66).

---

[7] We precise that DIMM and GS measure turbulence starting from 5 m and 6 m above ground therefore in principle we should compare the seeing with the thermal gradient between 5 and 30 meters.

6 Optical turbulence

6.1 General information

- **Site(s)**: Cerro Paranal;
- **Parameters investigated**: seeing, wavefront coherence time, isoplanatic angle;
- **Number of nights (calibration)**: 20 (CUTE-Scidar) - PAR2007 data;
- **Number of nights (preliminary model validation)**: 20 (CUTE-Scidar, DIMM, MASS) - PAR2007 data;
- **Number of nights (model validation)**: 53 in 2010/2011 and 36 in 2007 (DIMM, MASS). See Tables 67 and 68 for a complete list.

The three samples (20, 36 and 53 nights) are completely independent. The 20 nights used for the model calibration have been selected because we had simultaneous measurements of GS and DIMM (see Table 66). MASS measurements are simultaneous to the GS measurements only on a 14 nights subsample[8].

The selection of the independent sample of nights on which to perform the final model validation after its calibration has been done following the criteria defined on the SOW-MOSE [RD5]. More precisely we intended to cover in a uniform way a whole solar year and to keep at least a minimum of 30 nights. We decided to consider measurements in the period [06/2010 - 05/2011] for which were available DIMM and MASS measurements. For each month we selected the nights corresponding to the 1st, 5th, 15th and 25th of the month where DIMM measurements for both the same night and the night before were available. In case no DIMM measurements were available in these selected nights we looked for the nearest night respecting this criterium. The complete list of 53 nights is reported in Table 67. This list includes 7 nights selected by the ESO-Board.

A second set of simulations (36 nights) have been performed, completely independent from the calibration sample (20 nights) selected on 2007 (Table 68). This further test has been required to investigate some effects observed in results obtained by our analysis on the 53 nights.

6.2 Measurements: CUTE-Scidar, DIMM, MASS

The whole complete GS data-set has been re-calibrated by the MOSE team [31] before to compare OT simulation by Meso-Nh to GS measurement. This was necessary to correct GS data by the bias pointed out by [32], [33].

The general framework of the project was to first calibrate the model and then validate it. The original strategy/plan of the project was to test a new method to calibrate the model using MASS data (instead of GS data) in order to calibrate the model using a sample of nights as rich as we wish and freely selected all along a solar year using MASS data (more easily accessible than the GS measurements in different periods of the year). The method we have normally employed so far has been proposed by Masciadri & Jabouille [9] and validated in many successive studies ([10, 12]). It is now used also by Mauna Kea colleagues. However, during the first phase of the MOSE project we performed a detailed analysis of MASS data compared to the GS and DIMM ones and we put in evidence some substantial discrepancies of the seeing detected by the MASS if compared to that detected by the GS. The discrepancies are present in a more evident way, in some specific layers of the MASS and they are, visibly, a deficiency of the MASS. A detailed analysis of this topic has been treated and summarized in a forthcoming paper. Fig. 29 reports the most important result. It is the scattered plot of the seeing observed by the MASS and by the GS in the six layers defined by the weighting functions of the MASS (Fig. 28). In Fig. 29 the discrepancy of the seeing in most of layers (with exception of layer 3 and 6) is

---

[8] Depending on the software used to reduce the MASS data (Atmos 2.97.3 or Atmos 2.3 and on the parameter studied (seeing, isoplanatic angle or wavefront coherence time) the number of nights with simultaneous measurements with respect to the GS oscillates of some units.

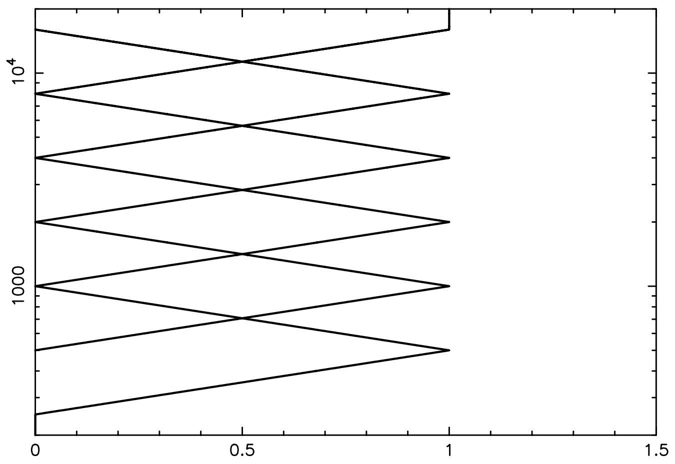


Figure 28: MASS weighting functions.

evident. Fig. 30 shows the scattering plots of the seeing of MASS versus GS in which the contribution from all the 6 layers is considered. Table 14 reports the relative errors calculated for each individual layer. The relative errors in most of the individual layers (with exception of layer 3 and 6) is not negligible and it reaches 50-60% in layers 1, 2 and 5. As a consequence of these results we decided to calibrate the model using the GS data and to use MASS data only in the final phase of model verification (only in some specific contexts in which the MASS limitations were not critical or fundamental). Indeed, passing by a comparison with GS and DIMM data

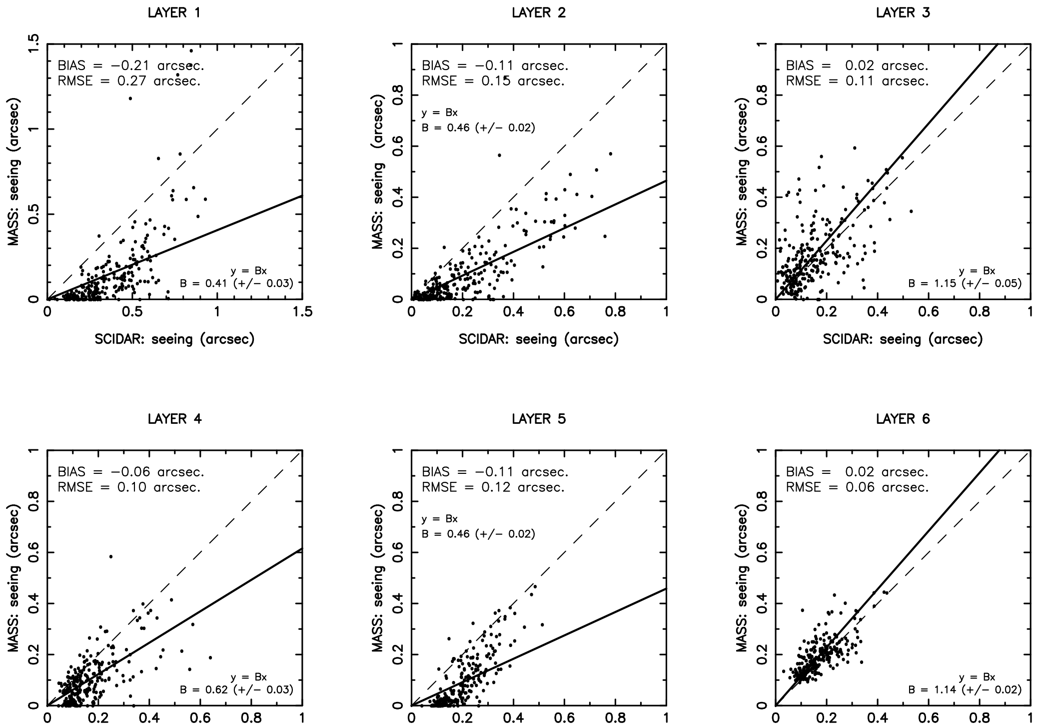


Figure 29: MASS vs GS **scattered plots** of **partial seeings**, layer by layer.

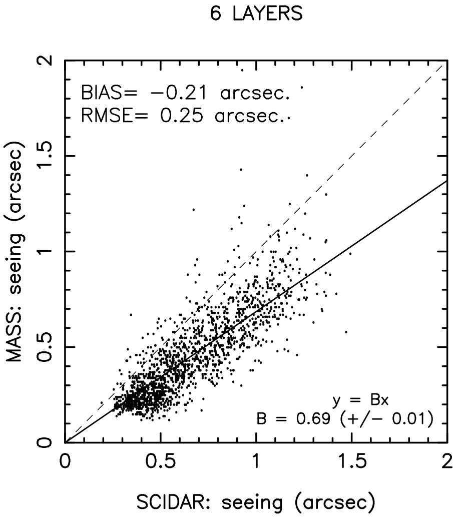


Figure 30: MASS vs GS **scattered plots** of the seeing. For the MASS the contribution from all the 6 layers is considered. The GS is binned like the MASS and we consider the contributions from the corresponding 6 layers.

we concluded that the calculation of the isoplanatic angle by the MASS revealed to be not too much affected by the problem we put in evidence on the seeing in the different layers, probably because θ_0 depends mainly on the turbulence in the high part of the atmosphere (layer 6 is mostly well correlated with GS - see Fig. 29). The comparison of the wavefront coherence time (τ_0) observed by the GS, the DIMM and the MASS (with the version Atmos 2.97.3) also revealed a substantial good agreement between GS and MASS data but in this case it is not so evident to explain why this is possible considering the discrepancy observed for the seeing in the individual layers. We note that the statistical sample of nights on which we draw the conclusion on τ_0 is made of only 14 nights and it should be preferable to confirm this agreement on a richer statistical sample. Kornilov [34] claims that the right τ_0 estimate can be obtained by taking into account a finite exposure time and a weighting function depending on wind speed. However, being that the τ_0 depends on the integral of the C_N^2 and the wind speed on the whole 20 km, and being that the C_N^2 is visibly underestimated in many parts of the free atmosphere, we should expect an overestimate of the MASS with respect to the GS. In spite of that we do not observe such an overestimation. We have an idea of a possible explanation that should be investigated in detail. For the moment we just note that, on a relatively small sample of 14 nights, the τ_0 between GS and MASS seems consistent. It might be possible that, as is the case for θ_0, other elements contribute to get negligible the problem of the seeing provided by the MASS in the estimation of τ_0. In this study we limit ourselves to consider the τ_0 MASS measurements at posteriori just in the final model validation.

	SCIDAR		MASS		
	Average	σ	Average	σ	Relative Error (%)
6 layers	0.66	0.26	0.45	0.23	+32
Layer 1	0.34	0.18	0.12	0.17	+65
Layer 2	0.20	0.17	0.09	0.12	+55
Layer 3	0.15	0.12	0.17	0.13	-13
Layer 4	0.16	0.10	0.10	0.09	+37
Layer 5	0.20	0.10	0.10	0.12	+50
Layer 6	0.17	0.08	0.19	0.07	-18

Table 14: Averages, standard deviation and relative error of seeing from SCIDAR and MASS. Units in arcsec, using a 1 minutes resampling interval.

6.3 Model calibration and preliminary validation

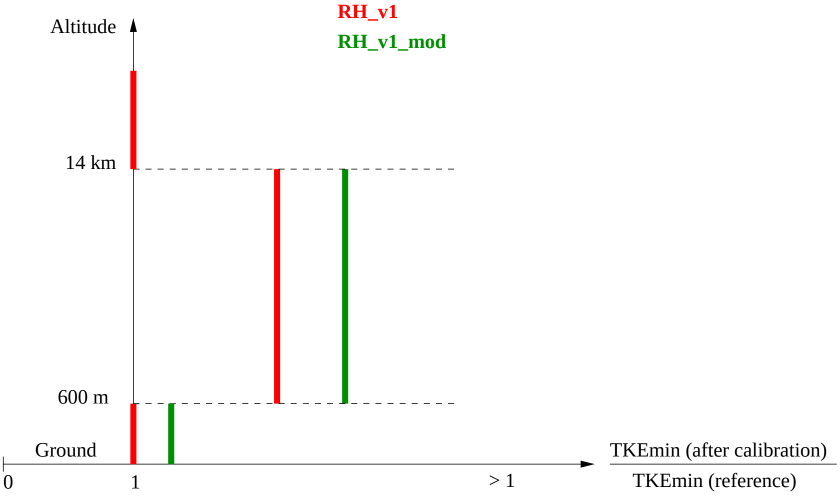


Figure 31: Toy model to illustrate the TKE$_{min}$ calibration.

As a consequence of considerations presented in Section 6.2 we considered as a sample of nights for the model calibration the 20 nights of the PAR2007 site testing campaign [35] in which simultaneous measurements of GS and DIMM are available. The list of nights is reported in Annex A, Table 66. We considered two different algorithms for the C_N^2:

$$C_N^2(z) = 0.58 \cdot \varphi_3(z) \left(\frac{80 \cdot 10^{-6} \cdot P(z)}{T(z)^2} \right)^2 \cdot L(z)^{4/3} \cdot \left(\frac{\partial \theta(z)}{\partial z} \right)^2 \tag{6}$$

that we call model version **v1**.

$$C_N^2(z) = 0.58 \cdot \varphi_3(z) \left(\frac{80 \cdot 10^{-6} \cdot P(z)}{T(z) \cdot \theta(z)} \right)^2 \cdot L(z)^{4/3} \cdot \left(\frac{\partial \theta(z)}{\partial z} \right)^2 \tag{7}$$

that we call model version **v1_mod**. We refer to the paper [3] for the extended descriptions of theory leading to the algorithm of Eq. 6. We briefly remind that P is the atmospheric pressure, T the absolute temperature, L the dynamic outer scale, θ the potential temperature, φ_3 is a dimensionless function depending on the thermal and dynamical stability of the atmosphere, more precisely, the inverse of the Prandtl number. Eq. 7 represents an alternative solution that we are testing. The calibration aims at fixing the value of the minimum turbulent kinetic energy (TKE$_{min}$) to be given at each vertical level [9]. The TKE$_{min}$ is the minimum turbulent kinetic energy required by the model to move from its state of equilibrium. Once the TKE$_{min}$ value is fixed (result of the calibration), all the simulations are re-run with this TKE$_{min}$ (the same for all simulations). From a practical point of view, the model calibration implies two series of runs on the whole calibration sample. In this study we tried to maintain the calibration procedure as light as possible in order to get a more robust model. We divided the 20 km in just three vertical slabs: **[0-600 m], [600 m-14 km] and [14 km-20 km]** (Fig. 31). Inside these three vertical slabs the value of the TKE$_{min}$ is constant. For the v1 version the calibration procedure lead to tune the value of TKE$_{min}$ just inside the interval [600 m-14 km] (see Fig. 31 red line). For the v1_mod version the calibration procedure lead to three different values of the TKE$_{min}$ as indicated in Fig. 31 green line. **The version v1 and v1_mod are called v2 and v2_mod after calibration, respectively**. In Sections 6.3.1 and 6.3.2 we present results obtained after calibration on the calibration sample. In Section 6.3.1 are presented the statistical model performances (statistical results) obtained on the whole calibration sample. In Section 6.3.2 the model performances obtained on individual nights. In Section 6.3.3 a few snapshots on model performances in reconstructing the OT temporal evolution are reported. A complete analysis can be found in the BSCW archive.

The statistical estimators used for this analysis are: bias, RMSE (Eq. 1-2) and regression line passing by the origin. We decided not to include the correlation coefficient (cc) because, after a detailed analysis, we realized that the cc is not an estimator as solid as the bias and the RMSE in our context and it can lead us to misleading conclusions with respect to our scientific goal.

Just as an example Fig. 32 shows the temporal evolution of the total seeing obtained with GS and DIMM for two different nights. On the left side are reported the bias, RMSE and cc in the scattered plot. The cc of the night 17/12/2007 is 0.92 while the cc of the night 22/12/2007 is 0.49. However what is important for us is how much close or far to observations are the values calculated by the model during the night. Looking at Fig. 32 (right side) it is evident that on 22/12/2007 the model has a better performance than on 17/12/2007 in spite of a cc=0.49 (instead of 0.92). Moreover, we note that the cc does not tell us anything about the temporal trend of the estimate. We concluded therefore that the cc does not provide any further useful information with respect to bias, RMSE and regression line and, on the contrary, risks to induce to misleading conclusions.

Fig. 33 shows a set of scattered plots of the total seeing of simultaneous DIMM and GS measurements obtained with different re-sampling with various Δt from 1 minute to 30 minutes. On top left of each figure are reported bias, RMSE and cc. It is visible that starting from Δt = 10 min values of bias and RMSE remain stables. We decided therefore to perform the statistical analysis with $\boldsymbol{\Delta t = 10}$ **min**.

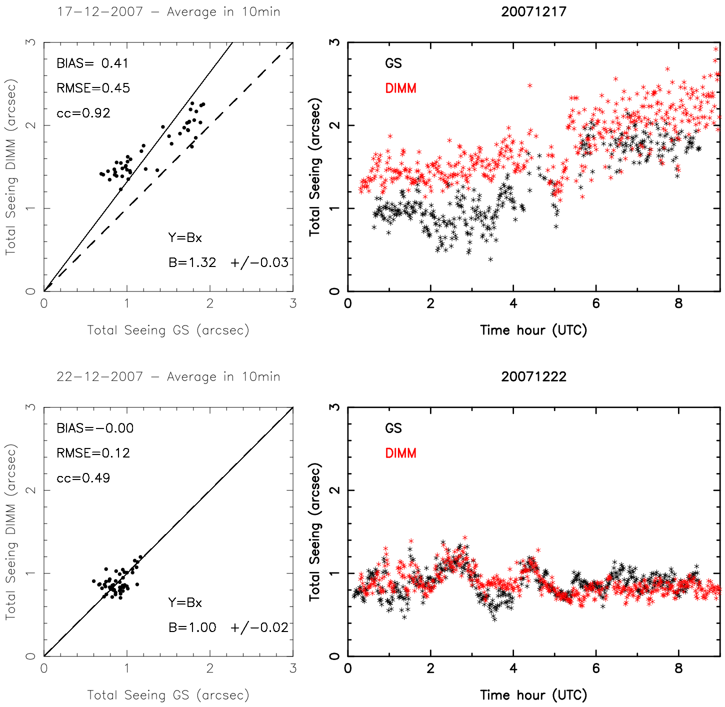


Figure 32: Example of the **temporal evolutions** (on the right) of the **total seeing** for two different nights and the corresponding values of bias, RMSE and cc (on the left). Unit in arcsec.

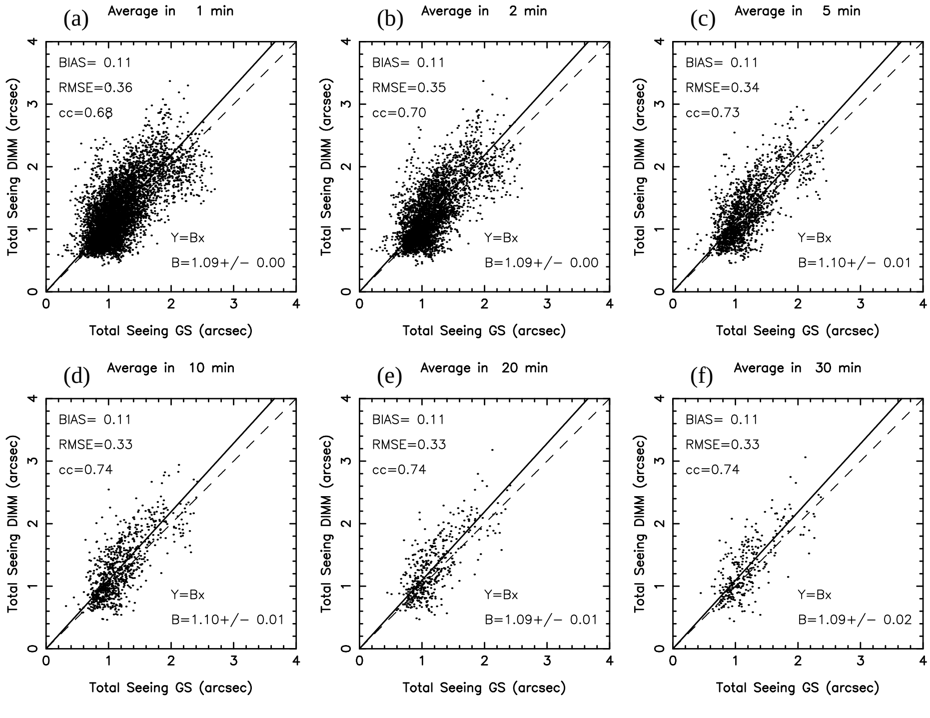


Figure 33: **Scattered plots** of the **total seeing** between DIMM and GS measurements. Each point represents a **(a)** 1-min average, **(b)** 2-min average, **(c)** 5-min average, **(d)** 10-min average, **(e)** 20-min average and **(f)** 30-min average. A sample of 20 nights is analyzed (Table 66). Unit in arcsec.

6.3.1 Overall statistical model performances

C_N^2 PROFILES

The C_N^2 profiles are the prime matter for the calculation of all the astro-climatologic parameters. In this section we discuss how good or bad is the reconstruction of the morphology of the vertical distribution of the turbulence in the whole column of $\sim$ 20 km performed by the model.

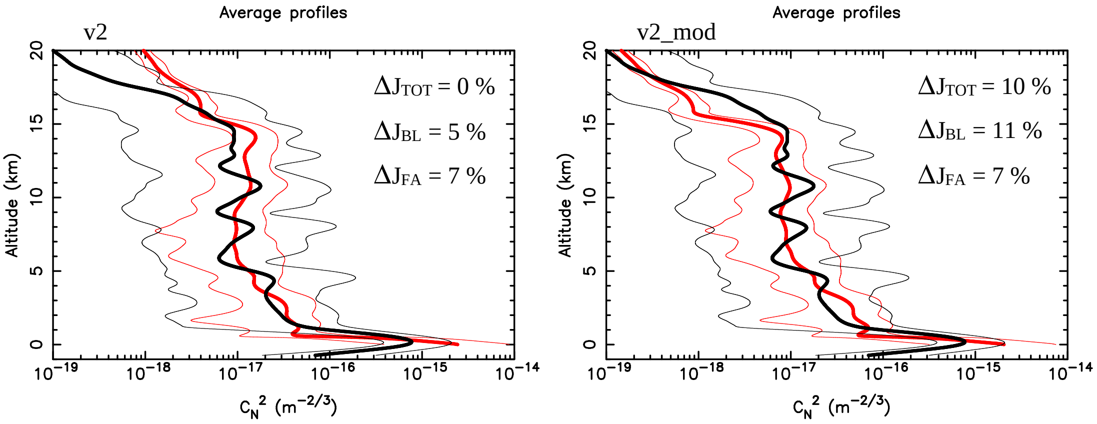


Figure 34: C_N^2 **average vertical profiles** at Cerro Paranal. In red, Meso-NH (version v2 and v2_mod) average vertical profile from the 20-nights sample (Table 66); in black, GS C_N^2 average vertical profiles from the same sample. The thick lines are the average profiles, and the thin lines the minimum and maximum values of the C_N^2 at each level. The latter are an indicator of the night-to-night variability. ΔJ_{TOT}, ΔJ_{BL} and ΔJ_{FA} are the relative error of the integrated turbulence on different vertical slabs in the atmosphere: total atmosphere, boundary layer and free atmosphere respectively.

Fig. 34 shows the vertical stratification of the turbulence (C_N^2 profiles) as obtained with the calibration sample after model calibration (with the v2_mod version). Thick style lines are the average C_N^2 profile as obtained from GS measurements and from the model. Thin style lines are the minimum and the maximum of the C_N^2 on the whole sample of measurements and simulations. Thin lines put in evidence the different level of spatio-temporal variability of the turbulence in the free atmosphere between the simulations and the measurements. The simulations obtained by the model are confined in a thinner range of values with respect to the measurements from the GS. This will be useful to discuss some results later on in the report. The major differences between v2 and v2_mod version are visible above 15 km (where v2_mod seems to have a better decreasing of the turbulence according to measurements with respect to v2) and in the first 600 m where v2 seems to give a better agreement with measurements if we consider the integral of the turbulence inside the vertical slab. From a quantitative point of view, both versions are in any case well correlated to measurements (within 10% in terms of turbulent energy). This is quantified by ΔJ_{TOT}, ΔJ_{BL} and ΔJ_{FA} (Fig.34) that are the relative errors of the integral of the C_N^2 of Fig.34 in the respective vertical slabs. We can note that the average simulated profile is very well correlated to the measured one. This means that the extremely simple solution we selected for the calibration seems to be very efficient to reconstruct a reliable average profile. It has to be considered as a very promising result because it reduces to the minimum terms the invasive action introduced by the calibration, permitting us to preserve a very satisfactory reliability of the model. The v2_mod version shows a better agreement in the high part of the atmosphere with respect to the v2 version. At around 15 km the shape of the simulated C_N^2 shows a decreasing in intensity having a more abrupt change in shape than what seen on observations. This is just due to the calibration procedure. In the context of this study we proved (BSCW archive) that this small difference has basically no impacts on the isoplanatic angle. Also we verified that alternative solutions for the calibration taking into account a TKE_{min} different at each model level provide less performing solutions for the model.

We note that it is meaningless to perform a discussion on the morphology of the vertical distribution of the optical turbulence near the ground. This is because our reference are measurements from the GS that has a vertical resolution:

$$\Delta H(h) = \frac{0.78 \cdot \sqrt{\lambda |h - h_{gs}|}}{\theta} \tag{8}$$

where h_{gs} is the conjugated height under ground typically of the order of 1-2 km, λ is the wavelength (0.5 microns) and θ is the binary star separation. Fig. 35 shows the vertical resolution calculated for h=0 considering all the combinations of θ and h_{gs} values of the PAR2007 site testing campaign. The largest value is around 600 m (therefore at h = 0, ΔH = ± 300 m). This means that near the ground it has no sense to compare the vertical distribution of the OT obtained with the model with that coming from GS measurements. We also note that for this reason i.e. because of its intrinsic vertical resolution (Eq. 8), the GS shows typically a bump at h = 0 with some turbulence reported also under ground. Because of this, the turbulence appearing underground is definitely part of the real turbulence detected at h = 0 by the GS that has to be taken into account when one calculates the seeing. It is traditionally drawn underground just because the zero point is located optically underground at 1-2 km (h_{gs}). From a practical point of view, using a GS, it is impossible to know where turbulence is located within a ΔH given by Eq. 8 centered at h = 0. To validate the model the C_N^2 distribution in the first few hundreds of meters above the ground, a more detailed analysis/verification will be possible using measurements at higher vertical resolution taken as a reference (for example the SLODAR at high resolution). In the literature it has already been put in evidence with measurements that the turbulence stratification decreases in a very sharp way in the first hundreds of meter. Fig. 36 is extracted, as an example, from [36]. It shows the turbulence vertical distribution (J profile) in the boundary layer as measured at Mt.Graham by a HVR-GS having a resolution of 25-30 m in the first kilometer above the ground and at Mauna Kea [37] by LOLAS (vertical resolution of 10-20 m) and a SLODAR (vertical resolution of 40-80 m).

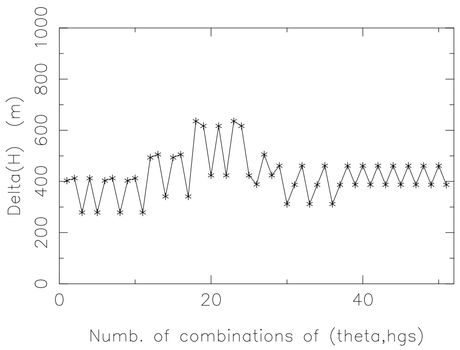


Figure 35: GS vertical resolution calculated at h = 0 m (see Eq. 8) for all the possible combinations of values of θ and h_{gs} from the site testing campaign PAR2007.

In other words, in this study, we took care to verify that the integral of the turbulence developed in the first hundreds of meters above the ground (integral) is correct or not by comparing the simulated and measured relative integrated quantities. For the rest of the study we decided to take as a frontier between the boundary layer and the free atmosphere 600 m above the ground. This means that we are considering a turbulence layer at h = 0 with a thickness of ± 600 m that is typically the height of the boundary layer.

SEEING

Fig. 37 shows the scattered plot of the total seeing observed by DIMM and GS and calculated by the model with the version v2[9]. Table 15 reports the values of bias, RMSE and regression line obtained with version v2

---

[9] This same sequence of panels has been done for the version v2_mod (see BSCW archive). We report just the sequence of

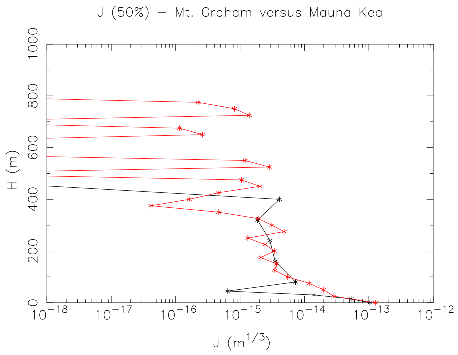


Figure 36: Figure extracted from [36]. Red line: turbulence vertical distribution (J profile) above Mt.Graham. Black line: turbulence vertical distribution (J profile) above Mauna Kea as calculated by [37]. Mauna Kea measurements extend only up to 650 m.

and v2_mod. We observe that the values of bias and RMSE between measurements obtained with different instruments and the model and measurements is of the same order. The bias between model and the GS is even better than that obtained between the GS and the DIMM. Similar results are obtained with the two model versions even if v2 is better for the bias. In bold style (Table 15) is indicated the version of the model that provides the best correlation with measurements.

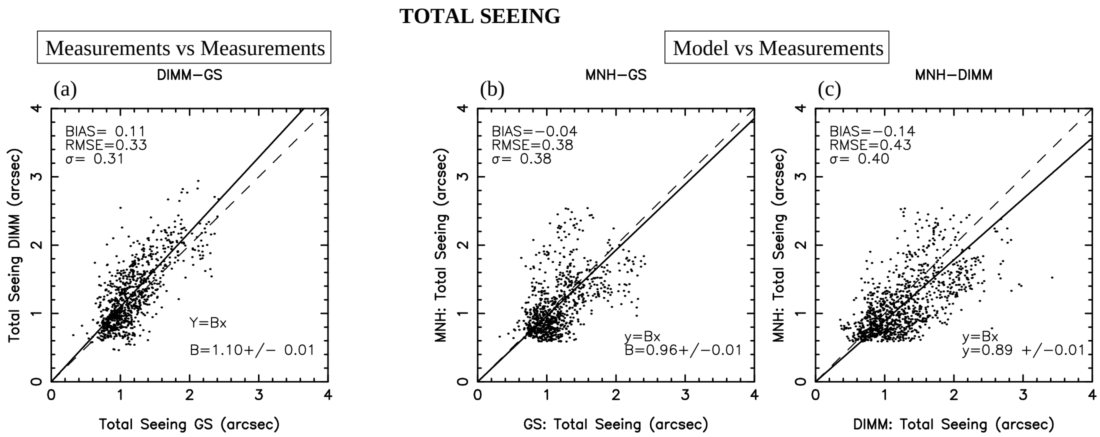


Figure 37: **Scattered plots** of the **total seeing**. **(a)** are DIMM vs GS measurements. **(b) and (c)** are model (version v2) vs measurements plots. Each point represents a 10-min average. A sample of 20 nights is analyzed (Table 66). Unit in arcsec.

Fig. 38 and Table 16 show the scattered plots of the seeing in the boundary layer [5 m-600 m] and the free atmosphere [600 m-20 km] obtained by the GS and the model. We observe that the RMSE is a little bit larger in the boundary layer (0.46 arcsec) than in the total atmosphere (0.38 arcsec). This is not so strange, indeed, the boundary layer is affected in a more evident way by horizontal dis-homogeneities than the total atmosphere [5 m-20 km]. We do not have an equivalent scattering plot of the seeing in the first 600 m obtained by two different instruments and we can not exclude it is of the same order of magnitude than that found between the simulations and observations (0.46 arcsec). In any case, we think that it would be better in the future to repeat the same calculation in a vertical slab [X m-600m] with a X greater than 5 m so that the turbulence can

---

panels related to the most representative version (the complete information can be found in the BSCW archive). We use the same procedure for the other parameters too.

Model Version	Statistical parameter	DIMM - GS	MNH - GS	MNH - DIMM
v2	BIAS	**0.11**	**-0.04**	**-0.14**
	RMSE	**0.33**	**0.38**	**0.43**
	REGRES.	**Y = 1.10·X**	**Y = 0.96·X**	**Y = 0.89·X**
v2_mod	BIAS	0.11	-0.10	-0.20
	RMSE	0.33	0.38	0.44
	REGRES.	Y = 1.10·X	Y = 0.92·X	Y = 0.84·X

Table 15: Overall statistics on the **total seeing**: **bias**, **RMSE** and **regression slopes** between observations (GS, DIMM and MASS) and model forecasts (version v2 and version v2_mod, see text for details). A sample of 20 nights is analyzed (Table 66).

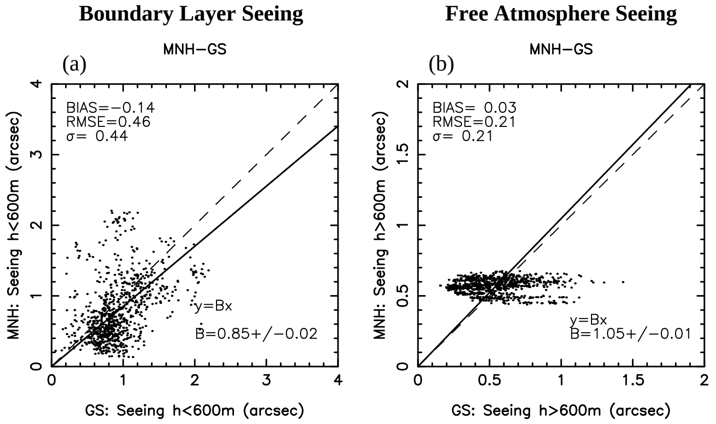


Figure 38: **Scattered plots** of the model (version v2_mod) vs GS **(a) BL seeing** and **(b) FA seeing**. Each point represents a 10-min average. A sample of 20 nights is analyzed (Table 66). Unit in arcsec.

be considered homogenous along the horizontal plane. In other words it should be useful to do measurements at a higher height X (where X respect the condition of homogeneity just described). In this case, we should expect a smaller RMSE in both cases: simulations versus measurements and measurements versus measurements. In this respect it should be important to verify that the height of the telescopes UT is high enough to assure a homogenous turbulence in the horizontal plane. If this is not the case it would be difficult to predict turbulence as well as to measure the turbulence at an horizontal distance different from zero with respect to the UT location without some discrepancy. Here we provide some interesting and positive perspectives. Fig. 39 (J.Navarrete, private communication) shows the scattering plot of all the seeing as measured by the active optics systems of the UT1 and UT2 (distance of ∼ 50 m). Data belong to a whole solar year and they refer

Model Version	Statistical parameter	TOTAL	H < 600 m	H > 600 m
v2	BIAS	**-0.04**	**-0.06**	**0.01**
	RMSE	**0.38**	**0.47**	**0.21**
	REGRES.	**Y = 0.96·X**	**Y = 0.93·X**	**Y = 1.03·X**
v2_mod	BIAS	-0.10	-0.14	0.03
	RMSE	0.38	0.46	0.21
	REGRES.	Y = 0.92·X	Y = 0.85·X	Y = 1.05·X

Table 16: Overall statistics on the **total**, **free atmosphere** and **boundary layer seeing**: **bias**, **RMSE** and **regression slopes** between observations (GS) and model forecasts (version v2 and version v2_mod, see text for details). A sample of 20 nights is analyzed (Table 66).

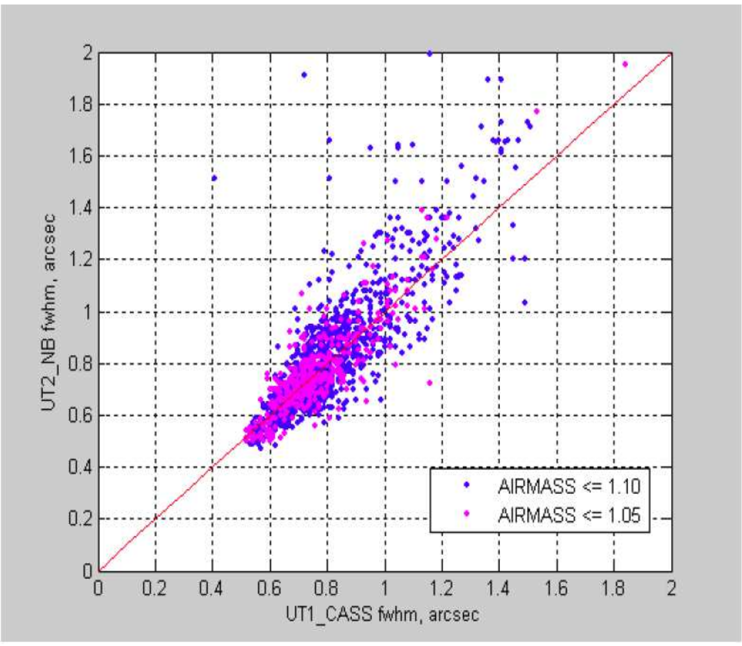


Figure 39: Seeing estimated extracted from the active optics systems of UT1 and UT2. The data sample includes simultaneous measurements from the two UT belonging to one solar year and included inside a cone of semi-angle ∼ 25° with respect to zenith (blue points) and ∼ 20° with respect to zenith (magenta points). (J.Navarrete, private communication)

to simultaneous measurements from both UTs with an airmass < 1.1 (and < 1.06) corresponding to a cone of semi-angle 25° (and 20°) with respect to zenith. Looking at Fig. 39 we deduce that measurements seem well correlated and symmetric with respect to Y=X. This tells us that we should expect a homogeneous turbulence on a spatial scale of ∼ 50 m at ∼ 11 m above the ground. Considering that the dome of the UT is open during the observations and that the primary mirror is located at 11 m we should take this height (11 m) as the height above which we can consider homogeneous the turbulence. It is worth to note, however, that independent site testing studies done at Las Campanas [38], indicated 30 m as a threshold above which to consider homogenous the turbulence even if it is not sure on which extension. The two results are not necessarily in contradiction because this is typically a local effect. In any case, a similar calculation done on UT1 versus UT4 would permit us to verify if such a homogeneity is guaranteed on a longer spatial scale (∼ 100 m). This can not be done now because the active optics systems in UT3 and UT4 requires an adaptation but ESO staff is taking care about this up-date that will be implemented soon. In case we will observe discrepancies this will mean that we go to higher heights above the ground to find a condition of homogeneity on the whole surface covering the four UTs.

Besides that, looking at Fig. 38, we note that in the free atmosphere the range of variability of the model is clearly smaller than that obtained with observations. In other words the model in the high part of the atmosphere succeeds well in reconstructing the average value of the turbulence inside the vertical slab but the temporal variations of the seeing during the night have smaller amplitude (∼ [0.45-0.65] arcsec) than what is observed (∼ [0.2-1.2] arcsec). An example of this effect is shown in Fig. 40-left. The temporal variation of the ε_{FA}, as it is reconstructed by the model, has a smaller amplitude than that observed. We verified that the extension of the variability [min,max] of observations can be reduced if we apply a moving average of 1 hour

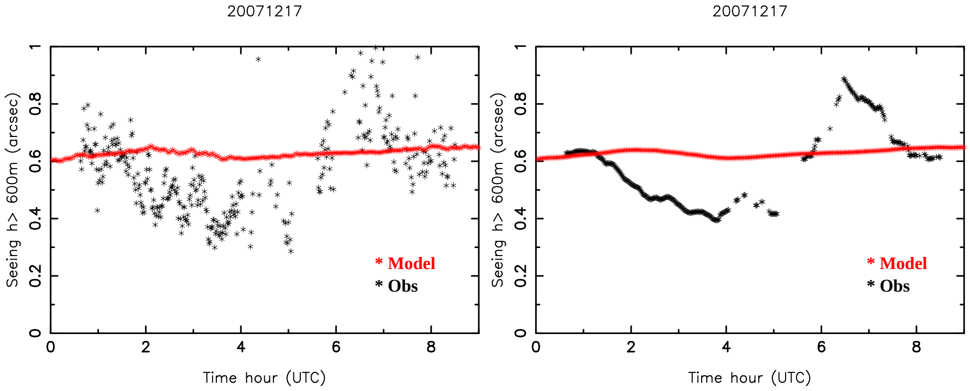


Figure 40: Left: **Temporal evolution** of the free atmosphere seeing for 1 night (17/12/2007), for the model (in red) and for the GS (in black). Unit in arcsec. Model version: v2_mod. Right: A moving average of 1 hour has been applied.

on the temporal sequence of the data-set (measurements and simulations) as shown in Fig. 40-right. However, the difference between the observed amplitude of the min/max peak-to-valley remains larger than that reconstructed by the model. Also we verified that this effect can not be solved with a higher vertical resolution of the meso-scale model. We are at present working on three different strategies related to the parameterization of the C_N^2 and aiming at improving the variability of the C_N^2 in the free atmosphere. As an example, Fig. 41 shows the temporal evolution of the isoplanatic angle simulated during a night using the actual model parameterization (left) and the test for a new parameterization (right); the latter showing a much greater peak-to-valley (PtV) amplitude. In Annex (Fig. 66 and Fig. 67) is reported the temporal evolution of the isoplanatic angle obtained with the model and all the other instruments (when available) using the preliminary new model parameterization. We highlight that these are just tests of sensitivity of the isoplanatic angle to some changes we introduced in the model parameterization and this is not the definitive result. However, such a preliminary result indicate that we should be able to overcome this problem. It is important to remind that, as the reader can observe, even with a larger model PtV amplitude, the uncertainty on measurements is still very important. This remains therefore a fundamental problem to be solved otherwise, even with a larger PtV amplitude it will be very difficult to be able to quantify the model goodness.

We remind that on 25 June 2013 it will be possible to access to initialization data with a better vertical resolution (137 levels instead of 91 levels). We expect therefore a higher sensitivity of the model to the gradients of wind speed and potential temperature and therefore better model performances.

WAVEFRONT COHERENCE TIME

Different techniques/methods are used by the different instruments to retrieve the wavefront coherence time (τ_0). The wavefront coherence time is defined as:

$$\tau_0 = 0.057 \cdot \lambda^{6/5} \cdot \left[\int_0^\infty |V(h)|^{5/3} \cdot C_N^2(h) \cdot dh \right]^{-3/5} \tag{9}$$

For the GS in this study we take the vertical wind speed profiles $V(h)$ from Meso-NH and the $C_N^2(h)$ profiles from the GS instrument[10]. We decided to use the wind speed coming form the model Meso-NH because we already proved that the model can reconstruct a very reliable wind speed on the whole column of 20 km. The

---

[10] In principle it is possible to retrieve V(h) profiles from the GS using the cross-correlation of the scintillation maps but this option is not available with the Cute-SCIDAR.

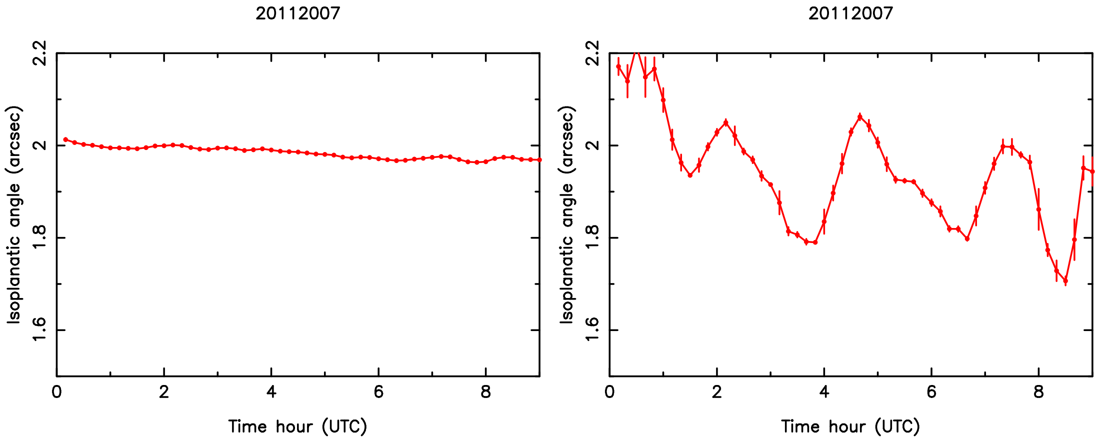


Figure 41: Temporal evolution of the isoplanatic angle simulated during the night 20/11/2007 with the actual model parameterization (left) and with a preliminary new model parameterization (right). Figure on the right shows a much greater PtV amplitude reconstructed by the model.

use of wind speed V(h) extracted by the GCMs as done by [39] is not recommended because they are certainly poorly representative of the wind speed near the ground.
To calculate τ_0 by the DIMM it has been proposed [40] the algorithm:

$$\tau_0 = 0.314 \cdot \frac{r_0}{V} \tag{10}$$

with:

$$V = max(V_{30m}, 0.4 \cdot V_{200mb}) \tag{11}$$

where r_0 is the Fried parameter measured by the DIMM, V_{30m} comes from the AWS and V_{200mb} is the wind speed at the jet-stream level retrieved by the GCMs.
The MASS wavefront coherence time calculated with Atmos 2.97.3 is defined as [34]:

$$\tau_0 = \left[(\tau_{0,GL})^{-5/3} + (\tau_{0,MASS})^{-5/3}\right]^{-3/5} \tag{12}$$

where $\tau_{0,MASS}$ is from the MASS covering the contribution of the free atmosphere and

$$\tau_{0,GL} = 0.314 \cdot \frac{r_{0,GL}}{V_{30m}} \tag{13}$$

is the τ_0 contribution associated to the boundary layer where $r_{0,GL}$ is the Fried parameter associated to the near-ground contribution of r_0 obtained subtracting the free atmosphere turbulent energy (MASS) from the total turbulent energy (DIMM) and V_{30m} comes from the AWS.

$$r_{0,GL} = \left[(r_{0,DIMM})^{-5/3} - (r_{0,MASS})^{-5/3}\right]^{-3/5} \tag{14}$$

The Meso-NH wavefront coherence time is defined in the same way as Eq. 9 but both the wind speed $V(h)$ and $C_N^2(h)$ come from Meso-NH.
We observe that the GS is in principle the unique instrument able to provide the τ_0 in autonomous way[11]. MASS and DIMM require information from other instruments and/or other sources. The τ_0 of the DIMM depends on the wind speed measured by the AWS and the wind speed at 200 mb estimated by the General Circulation Models (GCM). Moreover it assumes a strong approximation expressed by Eq. 11[12]. The τ_0 of MASS depends

---

[11] The GS used for the PAR2007 campaign does not include the measurements of the wind speed [41] but this is not a limitation of the GS from a general point of view, it is a limitation only of that GS.

[12] [11] proved in the past that the relative errors provided by this assumption can provide relative errors as large as 60%.

on the DIMM and the wind speed measured by the AWS. For all of these reasons, differently from the case of the seeing ε, we have to consider the GS as a priority reference when we compare simulations with measurements.

Fig. 42 [(a)-(c)] shows the scattered plot of the τ_0 measured by GS, DIMM and MASS. Fig. 42 [(d)-(f)] shows the scattered plot of the τ_0 simulated by the model (version v2_mod) and measured by GS, DIMM and MASS. Fig. 42 [(g)-(i)] shows the scattered plot of the τ_0 simulated by the model (version v2_mod) but the wind speed in the first 30 m has been modified according to V(h) = V(h)·2. In this way we can investigate the effects on the τ_0 of the model underestimation of the wind speed in the first 30 meters (see Section 5). Table 17 reports the results for both versions (v2 and v2_mod).
For what concerns measurements we note that MASS is, in general, well correlated to GS with a bias of 0.04 ms and RMSE of 1.68 ms. We note, however, a tendency of the MASS in overestimating the τ_0 for large ($>$ 5 ms) values of τ_0 that in any case should be confirmed by a richer statistic. This might be due to the underestimation of the seeing observed for the MASS but, as we said, this is an assumption that we should verified with a richer statistic. When τ_0 is smaller than 5 ms it is highly probably mostly driven by the near-ground component that we suspect mainly given by the DIMM component. Unfortunately the statistic of the calibration sample is done using only 14 nights. It would be important to have a richer sample to better quantify the effect of the MASS seeing underestimation on the bias of the MASS τ_0. This gets a little risky to completely trust on the τ_0 MASS values from a general point of view[13]. The DIMM shows a tendency in underestimating the τ_0 (the bias with respect to the GS is -1.56 ms). For what concerns the model, it shows a very good correlation with the GS (panel d) with a very satisfactory bias of 0.49 ms and an excellent RMSE of 1.48 ms (even smaller than what obtained between MASS and GS: 1.68 ms). Behaviors of the model in panel (e) and (f) are in agreement with the scattering observed between instruments. **If we consider V(h) = V(h)·2, results do not change too much and this means that the tendency in underestimating the wind speed in the first 30 m does not affect in an important way the estimate of the τ_0 that remains, on the contrary, very satisfactory**.

	Statistical parameter	DIMM - GS	MASS - GS	DIMM - MASS
Obs	BIAS	-1.56	0.04	-1.68
	RMSE	2.22	1.68	2.97
	REGRES.	Y = 0.62·X	Y = 1.01·X	Y = 0.63·X
Model Version	**Statistical parameter**	**MNH - GS**	**MNH - DIMM**	**MNH - MASS**
v2	BIAS	-0.78	0.69	-0.67
	RMSE	1.65	1.23	2.25
	REGRES.	Y = 0.81·X	Y = 1.27·X	Y = 0.85·X
v2_mod	BIAS	**-0.49**	0.98	-0.36
	RMSE	**1.48**	1.47	2.12
	REGRES.	**Y = 0.88·X**	Y = 1.38·X	Y = 0.92·X
v2_mod(*)	BIAS	-0.51	0.68	-0.81
	RMSE	1.33	1.24	2.32
	REGRES.	Y = 0.87·X	Y = 1.26·X	Y = 0.82·X

Table 17: Overall statistics on the **wavefront coherence time**: **bias**, **RMSE** and **regression slopes** between observations (GS, DIMM and MASS) and model forecasts (version v2, v2_mod and v2_mod(*), see text for details). A sample of 20 nights is analyzed (Table 66).

ISOPLANATIC ANGLE

For the GS and the Meso-NH model the isoplanatic angle has been calculated using the algorithm:

---

[13] One should expect a tendency in overestimating the τ_0 because of the underestimation of the seeing.

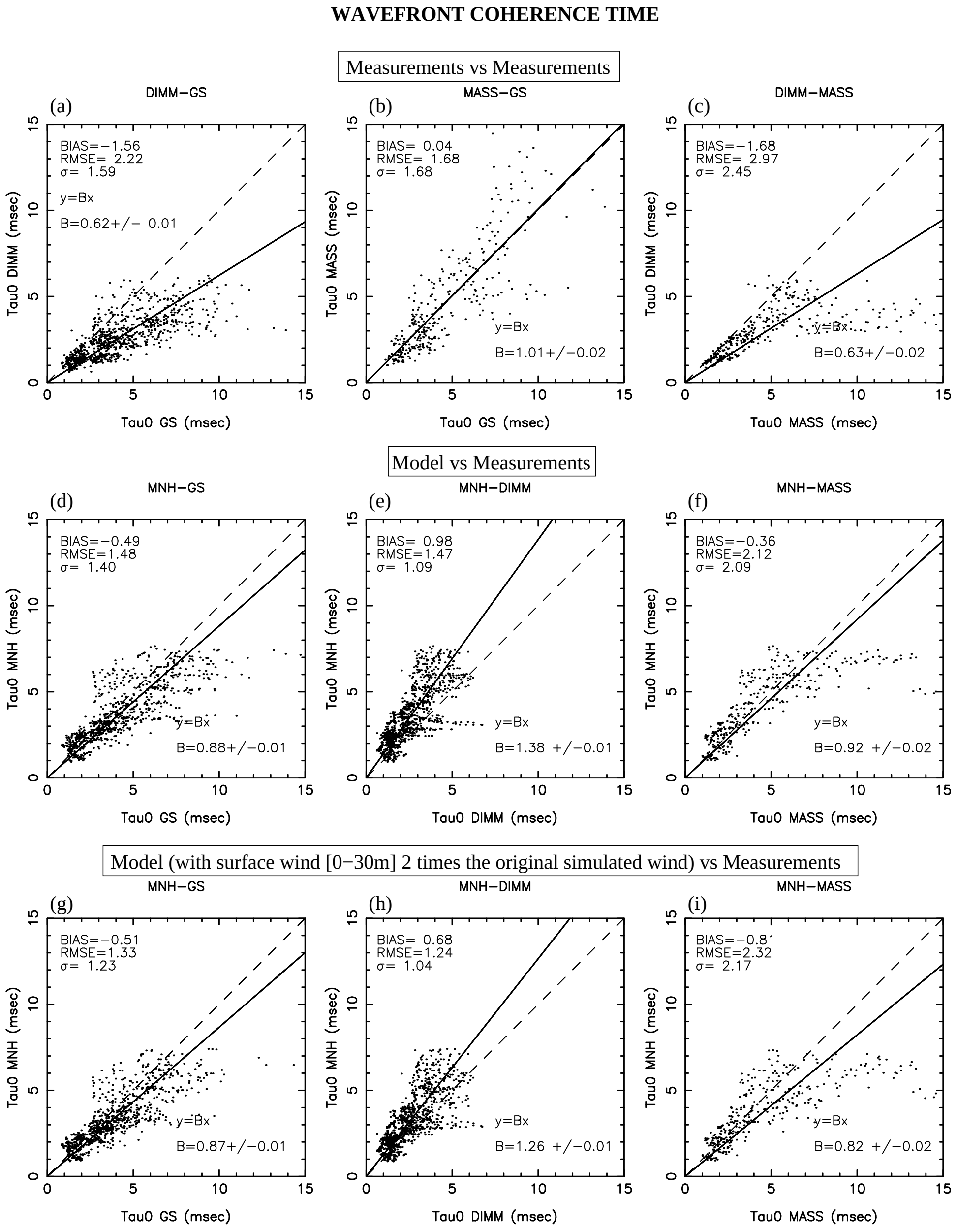


Figure 42: **Scattered plots** of the **wavefront coherence time**. **(a), (b) and (c)** are measurements vs measurements plots. **(d), (e) and (f)** are model (version v2_mod) vs measurements plots. **(g), (h) and (i)** are the same as (d), (e) and (f), respectively, but the model wind up to 30 m a.g.l. has been multiplied by a factor of 2. Each point represents a 10-min average. A sample of 20 nights is analyzed (Table 66), except for when MASS data are involved where a subsample of 14 nights is analyzed. MASS software: **atmos 2.97.3**. Unit in ms.

$$\theta_0 = 0.057 \cdot \lambda^{6/5} \left(\int_0^\infty h^{5/3} C_N^2(h) dh \right)^{-3/5} \tag{15}$$

The DIMM calculates θ_0 using a method proposed by [42] and implemented in the DIMM [40] as well as in the Generalized Seeing Monitor (GSM) [43]. The method consists in calculating θ_0 passing by the calculation of the

scintillation rate imposing a pupil diameter of 10 cm and assuming we are considering turbulence concentrated at 10 km. Again the method is intrinsically less accurate and reliable than a direct integration because it is based on such an approximation. The MASS values are retrieved directly by the data-reduction MASS software (Atmos 2.3 version). We highlight that, in spite of the fact that the MASS is not sensitive to the turbulence in the first $\sim$ 500 m from the ground, this has almost no effect on the calculation of θ_0 because this parameter is weighted by the $h^{5/3}$ function that gets irrelevant whatever the C_N^2 value near the ground is. Also in this case, therefore, the GS has to be considered as a priority reference with respect to DIMM and MASS.
Fig. 43 reports the scattered plots of θ_0 obtained between measurements from different instruments and the model with each instrument. We observe that the DIMM slightly overestimates θ_0 with respect to the GS (a bias of 0.25 arcsec corresponding to around 10% in relative error). The MASS is better correlated to the GS even if we have to remind that the statistic is extended on only 12 nights. The RMSE is more or less of the order of 0.46-0.82 arcsec considering all the couples of instruments. Looking at the model performances we note that the model is well correlated to GS and MASS measurements in terms of bias as well as RMSE. The relative bias and RMSE are absolutely comparable with respect to those obtained between MASS and GS. However, as noted for the free atmosphere seeing, we observe that model predictions appears more confined in a thin range of values [1.7 arcsec - 2.2 arcsec] with respect to measurements [1 arcsec - 4 arcsec] because the model temporal variability of the turbulence in the high part of the atmosphere is underestimated with respect to the observations. Results obtained with DIMM are slightly worse because they reflect the fact that the DIMM slightly overestimates the θ_0. Table 18 summarizes the results obtained for both versions (v2 and v2_mod). As for the case of τ_0, the best results are obtained with the v2_mod version.

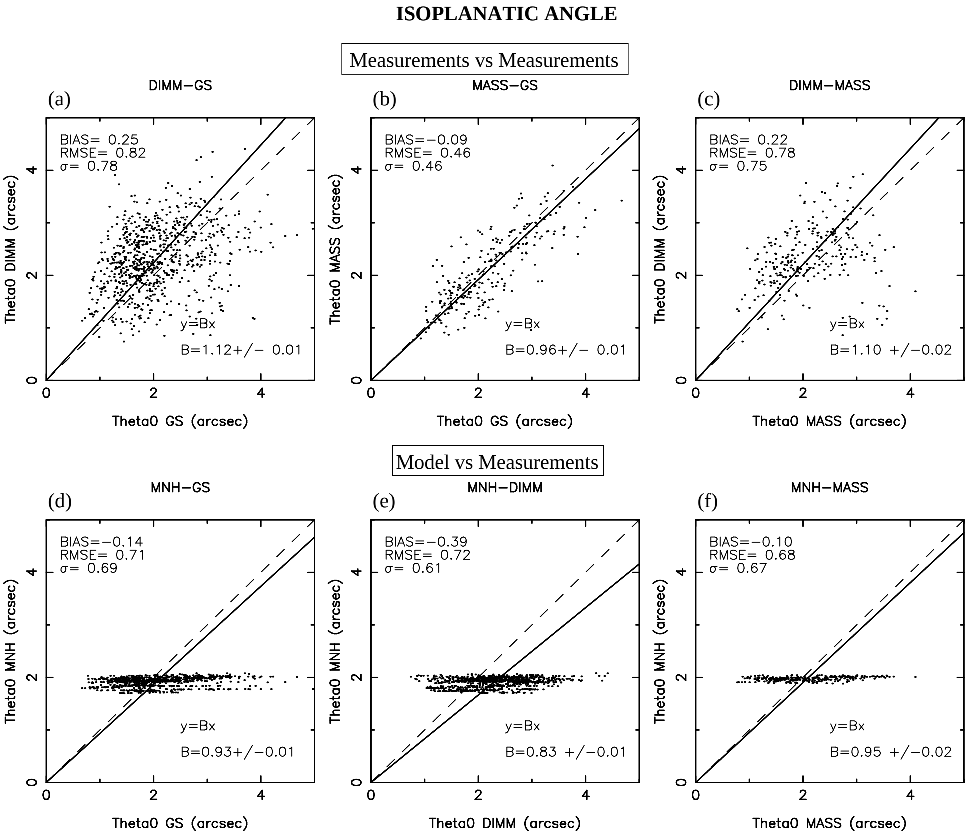


Figure 43: **Scattered plots** of the **isoplanatic angle**. **(a), (b) and (c)** are measurements vs measurements plots. **(d), (e) and (f)** are model (version v2_mod) vs measurements plots. Each point represents a 10-min average. A sample of 20 nights is analyzed (Table 66), except for when MASS data are involved where a subsample of 12 nights is analyzed. MASS software: **Atmos 2.3** version. Unit in arcsec.

	Statistical parameter	DIMM - GS	MASS - GS	DIMM - MASS
Obs	BIAS	0.25	-0.09	0.22
	RMSE	0.82	0.46	0.78
	REGRES.	Y = 1.12·X	Y = 0.96·X	Y = 1.10·X

Model Version	Statistical parameter	MNH - GS	MNH - DIMM	MNH - MASS
v2	BIAS	-0.61	-0.86	-0.61
	RMSE	0.93	1.06	0.91
	REGRES.	Y = 0.70·X	Y = 0.63·X	Y = 0.70·X
v2_mod	BIAS	**-0.14**	-0.39	-0.10
	RMSE	**0.71**	0.72	0.68
	REGRES.	**Y = 0.93·X**	Y = 0.83·X	Y = 0.95·X

Table 18: Overall statistics on the **isoplanatic angle**: **bias**, **RMSE** and **regression slopes** between observations (GS, DIMM and MASS) and model forecasts (version v2, v2_mod, see text for details). A sample of 20 nights is analyzed (Table 66), except for when MASS data are involved where a subsample of 12 nights is analyzed.

6.3.2 Individual nights model performances

SEEING

In this section we report results obtained for the model performances on reconstructing turbulence estimates night by night. Fig. 44 shows the cumulative distribution of the bias and RMSE of the seeing calculated on each night, measured by the instruments and calculated by the model. We note that the model bias with respect to the GS is even better than the one obtained between GS and DIMM. The (model vs. GS) and (GS vs. DIMM) RMSE are well correlated and the order of magnitude of the median bias value is comparable to that obtained for the whole statistical sample (see Fig. 37). The same can be said for the first and third quartiles. For the boundary layer (BL) and free atmosphere (FA) unfortunately we have only estimates between the GS and the model. We have, indeed, no dispersion of measurements obtained with different instruments with which we can discuss our results. A median value of the bias in boundary layer of 0.11 arcsec corresponds to around 20% of the typical seeing value in the boundary layer. The median bias value of 0.03 arcsec in the free atmosphere is definitely a good model performance. More difficult is the judgment of the RMSE values without estimates of dispersions between different instruments. It would be highly suitable to be able to access to the dispersion estimates of this parameter between different instruments in the future. Table 19 and Table 20 summarize values of bias and RMSE obtained by the model with the two versions (v2 and v2_mod) for the total seeing, the seeing in the boundary layer and in the free atmosphere. In agreement with the results obtained in Table 15 and Table 16 for the whole statistical sample, the v2 version appears slightly better than the v2_mod.

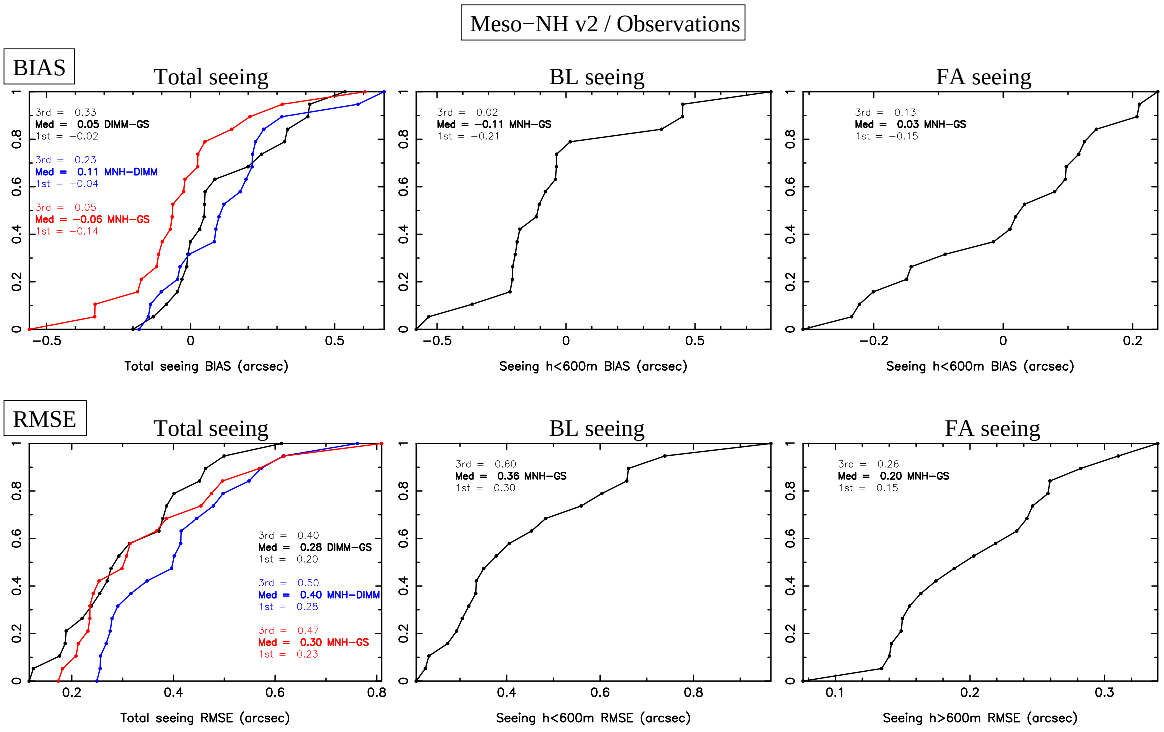


Figure 44: **Cumulative distribution** of the **bias** (top figures) and **RMSE** (bottom figures), between model and observations for the single nights, of the **total seeing**, the **boundary layer seeing** and the **free atmosphere seeing**. Model version is v2. Observations are from DIMM and MASS (only for the total seeing) and GS (total seeing, boundary layer seeing and free atmosphere seeing). Unit in arcsec.

Model Version	Statistical parameter	DIMM - GS	MNH - GS	MNH - DIMM
v2	BIAS	$0.05^{+0.33}_{-0.02}$	$-0.06^{+0.05}_{-0.14}$	$0.11^{+0.23}_{-0.04}$
	RMSE	$0.28^{+0.40}_{+0.20}$	$0.30^{+0.47}_{+0.23}$	$0.40^{+0.50}_{+0.28}$
v2_mod	BIAS	$0.05^{+0.33}_{-0.02}$	$-0.11^{+0.01}_{-0.18}$	$0.15^{+0.32}_{+0.05}$
	RMSE	$0.28^{+0.40}_{+0.20}$	$0.30^{+0.47}_{+0.21}$	$0.41^{+0.52}_{+0.28}$

Table 19: Median **bias** and **RMSE** of the **total seeing** from the 20 single nights from the calibration sample (Table 66). In small fonts, the 1st and 3rd quartiles. Version v2 (GS-MNH) appears slightly better than v2_mod.

Model Version	Statistical parameter	MNH - GS BL (H<600 m)	MNH - GS FA (H>600 m)
v2	BIAS	$-0.11^{+0.02}_{-0.21}$	$0.03^{+0.13}_{-0.15}$
	RMSE	$0.36^{+0.60}_{+0.30}$	$0.20^{+0.26}_{+0.15}$
v2_mod	BIAS	$-0.16^{-0.03}_{-0.30}$	$0.06^{+0.13}_{-0.14}$
	RMSE	$0.38^{+0.61}_{+0.31}$	$0.21^{+0.25}_{+0.15}$

Table 20: Median **bias** and **RMSE** of the **FA seeing** and **BL seeing** from the 20 single nights from the calibration sample (Table 66). In small fonts, the 1st and 3rd quartiles. No major differences between v2 and v2_mod.

WAVEFRONT COHERENCE TIME

For the wavefront coherence time (τ_0) we can compare measurements from three different instruments: GS, DIMM and MASS. In order to get the results more readable Fig. 45 shows the cumulative distribution of the bias and RMSE of the τ_0 obtained night by night and calculated with all combination of couples of instruments. Fig. 46 shows the same cumulative distribution calculated between the model and the three different instruments. Table 21 and Table 22 summarize the results (median values and first and third quartiles) for the bias and RMSE obtained comparing different instruments and between the model and each instrument (version v2 and v2_mod). As discussed previously (Section 6.3.1), when we compare simulated τ_0 with measurements, we have to consider, as a priority reference, the GS that provides the most reliable and direct estimate of this parameter. We briefly remind that, from the analysis done on the overall statistical sample, we had concluded that the DIMM underestimates the τ_0 of a quantity of roughly 1.56 ms while the MASS shows a reasonable good agreement with the GS. However we had highlighted that MASS results should be used with some precaution.

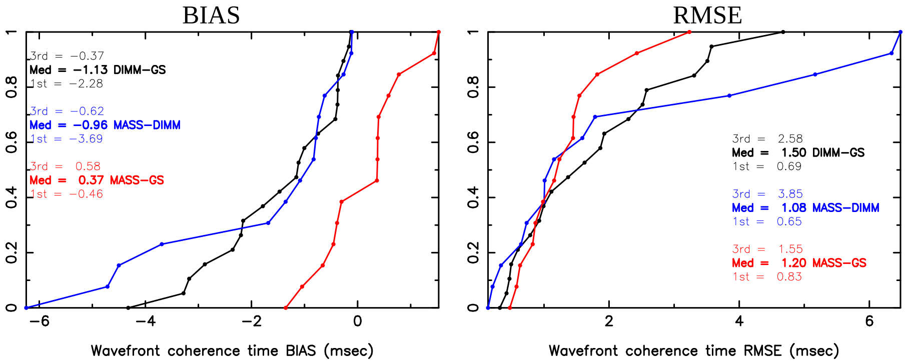


Figure 45: **Cumulative distribution** of the **bias** (left figure) and **RMSE** (right figure), between GS, MASS and DIMM instruments for the single nights, of the **wavefront coherence time**. Unit in ms.

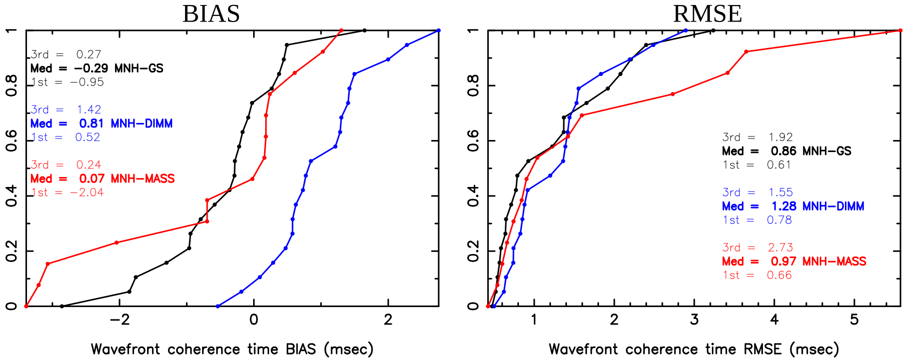


Figure 46: **Cumulative distribution** of the **bias** (left figure) and **RMSE** (right figure), between GS, MASS and DIMM instruments and the Meso-NH model (version v2_mod) for the single nights, of the **wavefront coherence time**. Unit in ms.

Looking at Fig. 45 and Fig. 46 we observe that the median value of the bias MNH-GS (-0.29 ms) and MNH-MASS (0.07 ms) night by night are extremely good and comparable with that obtained with MASS-GS (0.37 ms). The first and third quartiles are also very satisfactory. If we compare the model to the GS this range remains confined within 1 ms: [-0.95 ms - 0.27 ms]. Only the MASS-GS obtains a slightly better result [-0.46 ms - 0.58 ms]. If one look at the median values of the RMSE we observe that the MNH-GS (0.86 ms) and MNH-MASS (0.97 ms) values are better (smaller) than the MASS-GS (1.20 ms). The median value of MNH-DIMM is slightly larger but the first and third quartiles are on the contrary well confined [0.78 ms - 1.55 ms] like the case MNH-GS [0.61 ms - 1.92 ms]. We conclude therefore that the model has very good performances in reconstructing the τ_0 even night by night. If we take as a reference the GS, for both bias and RMSE, the median values of version v2_mod are better than v2.

Statistical parameter	DIMM-GS	MASS-DIMM	MASS-GS
BIAS	$-1.13^{-0.37}_{-2.28}$	$-0.96^{-0.62}_{-3.69}$	$0.37^{+0.58}_{-0.46}$
RMSE	$1.50^{+2.58}_{+0.69}$	$1.08^{+3.85}_{+0.65}$	$1.20^{+1.55}_{+0.83}$

Table 21: Median **bias** and **RMSE** of the **wavefront coherence time** (see Fig. 45). from the 20 single nights from the calibration sample (Table 66). When MASS measurements are involved, the sample reduces to 14 nights only. In small fonts, the 1st and 3rd quartiles, between the three instruments: GS, DIMM and MASS. Unit in ms. Best results obtained by MASS-GS.

Model version	Statistical parameter	MNH-GS	MNH-DIMM	MNH-MASS
v2	BIAS	$-0.46^{-0.10}_{-1.29}$	$0.64^{+1.21}_{+0.27}$	$-0.38^{+0.13}_{-2.64}$
	RMSE	$1.05^{+1.98}_{+0.67}$	$1.03^{+1.36}_{+0.72}$	$0.95^{+3.27}_{+0.60}$
v2_mod	BIAS	$-0.29^{+0.27}_{-0.95}$	$0.81^{+1.42}_{+0.52}$	$0.07^{+0.24}_{-2.04}$
	RMSE	$0.86^{+1.92}_{+0.61}$	$1.28^{+1.55}_{+0.78}$	$0.97^{+2.73}_{+0.66}$

Table 22: Median **bias** and **RMSE** of the **wavefront coherence time** (see Fig. 46). from the 20 single nights from the calibration sample (Table 66). When MASS measurements are involved, the sample reduces to 14 nights only. In small fonts, the 1st and 3rd quartiles, between the Meso-NH model and the three instruments (GS, DIMM and MASS). Unit in ms.

ISOPLANATIC ANGLE

For the isoplanatic angle (θ_0) we can compare measurements from three different instruments: GS, DIMM and MASS. In order to get the results more readable Fig. 47 shows the cumulative distribution of the θ_0 (bias and RMSE) obtained night by night calculated with all combinations of the couples of instruments. Fig. 48 shows the same cumulative distribution calculated between the model and the three different instruments. Table 23 and Table 24 summarize the results (median values and first and third quartiles) for the bias and the RMSE obtained comparing different instruments and between the model and each instrument (version v2 and v2_mod). As discussed previously (Section 6.3.1), when we compare θ_0 simulations with measurements, we have to consider as a priority reference the GS that provides the more reliable and direct estimate of this parameter.

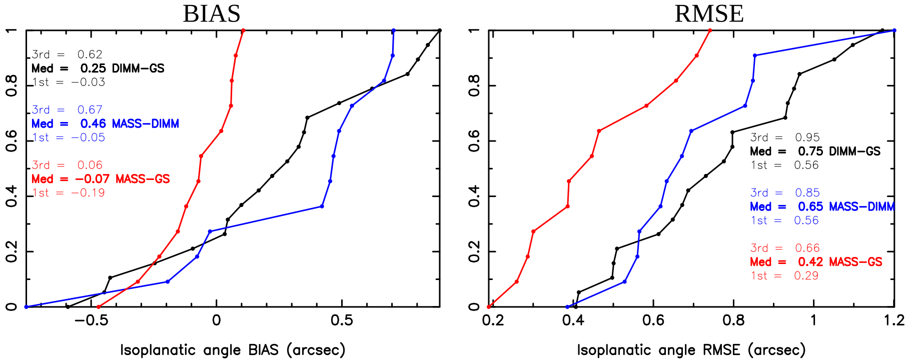


Figure 47: **Cumulative distribution** of the **bias** (left figure) and **RMSE** (right figure), between GS, MASS and DIMM instruments for the 20 single nights (only 12 when the MASS is involved), of the **isoplanatic angle** (see Table 66). Unit in arcsec.

Looking at Fig. 47 we observe that MASS and GS are the best correlated instruments (even if the statistic sample is a little bit poor) with a median bias of -0.07 arcsec and a first-third quartiles range included in the [-0.19 arcsec - 0.06 arcsec] range. The median value of the bias if we consider DIMM-GS and MASS-DIMM is somehow larger (0.25 and 0.46 arcsec). If we look at the RMSE, we observe that the median value MASS-GS is definitely the best one (0.42 arcsec) as well as the extension of the first-third quartiles range [0.29 arcsec - 0.66 arcsec]. If we look at the model performances (Fig. 48) we observe that the median value of the bias MNH-GS (-0.05 arcsec) and MNH-MASS (-0.18 arcsec) are definitely well comparable to those obtained by MASS-GS. The extension first-third quartiles range [-0.43 arcsec - 0.40 arcsec] is a little bit larger than the

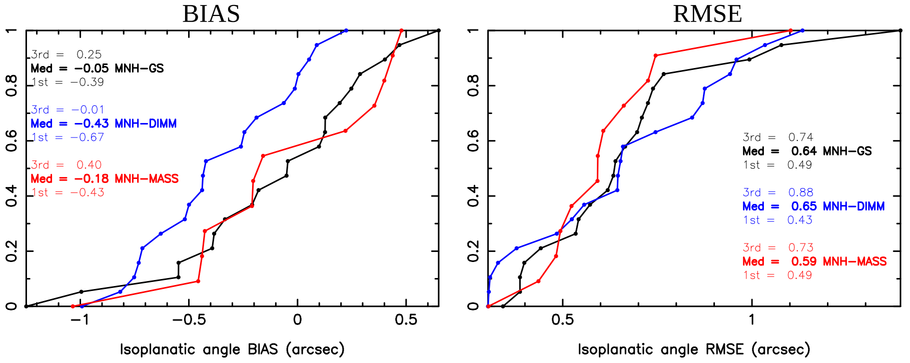


Figure 48: **Cumulative distribution** of the **bias** (left figure) and **RMSE** (right figure), between GS, MASS and DIMM instruments and the Meso-NH model (version v2_mod) for the 20 single nights (only 12 when the MASS is involved), of the **isoplanatic angle** (see Table 66). Unit in arcsec.

MASS-GS [-0.19 arcsec - 0.06 arcsec]. If we look at the RMSE, we observe that the median value MNH-MASS (0.59 arcsec) and MNH-GS (0.64 arcsec) are a little larger than that obtained with MASS-GS (0.42 arcsec) (order of 0.2 arcsec more larger). We conclude therefore that the ability of the model in predicting θ_0 night by night is not worse than what we are able to quantify at present with any couple of instruments with exception of MASS and GS that show a better agreement. Besides, as we already said, the model shows more difficulty in reconstructing the θ_0 temporal variability inside the individual night. Table 23 and Table 24 summarize the results obtained for the median values, first and third quartiles (measurements, model version v2 and v2mod). As for the τ_0 case, version v2_mod appears slightly better than version v2.

Statistical parameter	DIMM-GS	MASS-DIMM	MASS-GS
BIAS	$0.25^{-0.62}_{-0.03}$	$0.46^{+0.67}_{-0.05}$	$-0.07^{+0.06}_{-0.19}$
RMSE	$0.75^{+0.95}_{+0.56}$	$0.65^{+0.85}_{+0.56}$	$0.42^{+0.66}_{+0.29}$

Table 23: Median **bias** and **RMSE** of the **isoplanatic angle** (see Fig. 47), from the 20 single nights from the calibration sample (Table 66), between the three instruments: GS, DIMM and MASS. In small fonts, the 1st and 3rd quartiles. When MASS measurements are involved, the sample reduces to 12 nights only. Unit in arcsec.

Model version	Statistical parameter	MNH-GS	MNH-DIMM	MNH-MASS
v2	BIAS	$-0.54^{-0.25}_{-0.84}$	$-0.92^{-0.52}_{-1.15}$	$-0.66^{-0.12}_{-0.91}$
	RMSE	$0.68^{+1.03}_{+0.46}$	$1.02^{+1.32}_{+0.63}$	$0.74^{+1.01}_{+0.38}$
v2_mod	BIAS	$-0.05^{+0.25}_{-0.39}$	$-0.43^{-0.01}_{-0.67}$	$-0.18^{+0.40}_{-0.43}$
	RMSE	$0.64^{+0.74}_{+0.49}$	$0.65^{+0.88}_{+0.43}$	$0.59^{+0.73}_{+0.49}$

Table 24: Median **bias** and **RMSE** of the **isoplanatic angle** (see Fig. 48), from the 20 single nights from the calibration sample (Table 66), between the Meso-NH model and the three instruments (GS, DIMM and MASS). In small fonts, the 1st and 3rd quartiles. When MASS measurements are involved, the sample reduces to 12 nights only. Unit in arcsec.

6.3.3 OT temporal evolution

In the framework of the MOSE study a detailed analysis of the model performances in reconstructing the temporal evolution of all the astroclimatic parameters have been carried out (see BSCW archive). Fig. 49 shows the most representative of this results: the temporal evolution, during each night of the PAR2007 site testing campaign (20 nights), of the parameter J i.e. the total turbulence energy integrated on the whole atmosphere calculated by the model and measured by the GS and the DIMM. It is possible to appreciate how the model, all along most of the nights, reconstructs in a very satisfactory way the general trend of the total turbulent energy as well as the quantitative estimate with a dispersion that is typically of the same order of magnitude of that obtained by the two instruments.

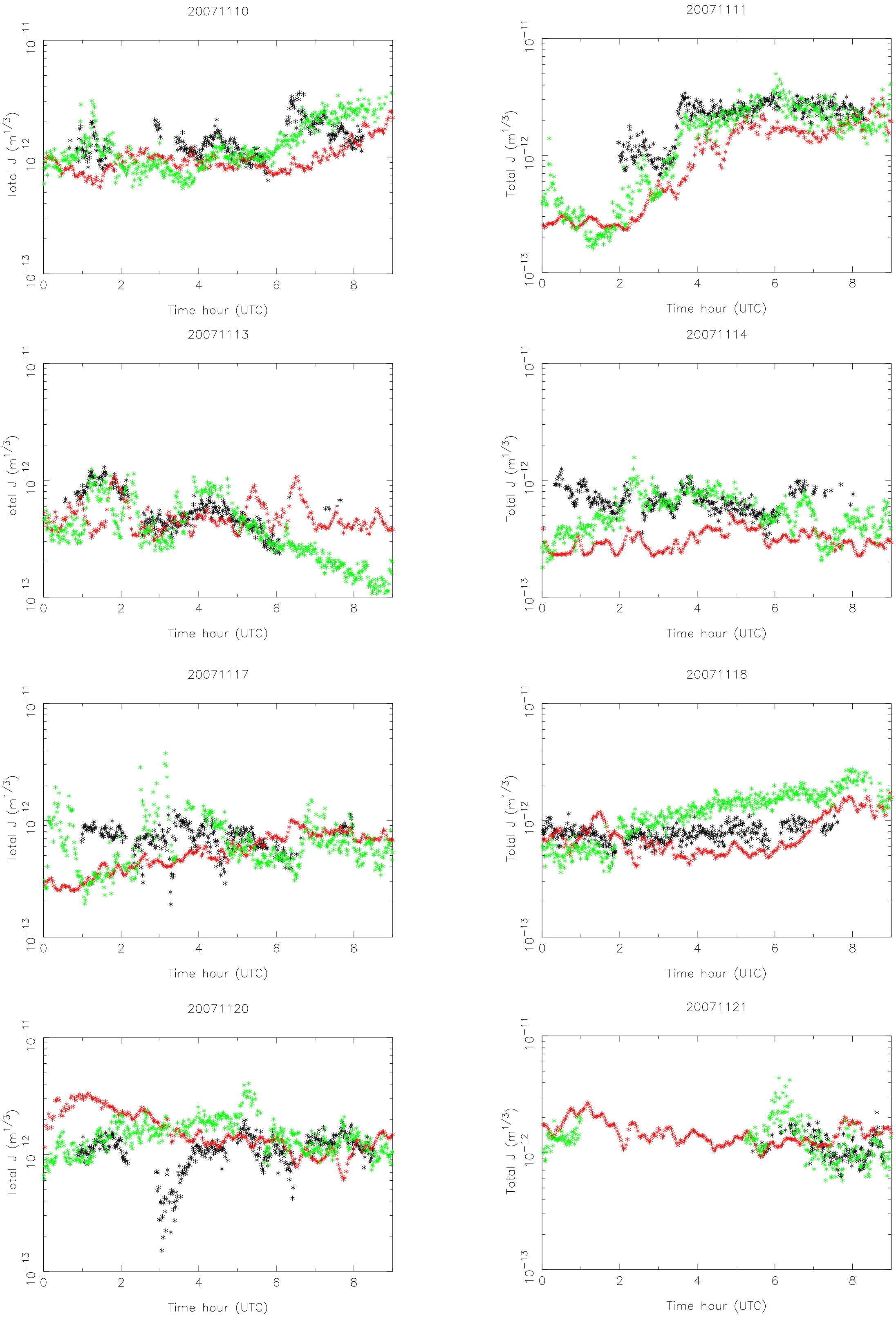


Figure 49: **Temporal evolutions** of J for different simulated nights from the PAR2007 site testing campaign, part 1. In black, the GS; in green, the DIMM; in red, the model (version v2).

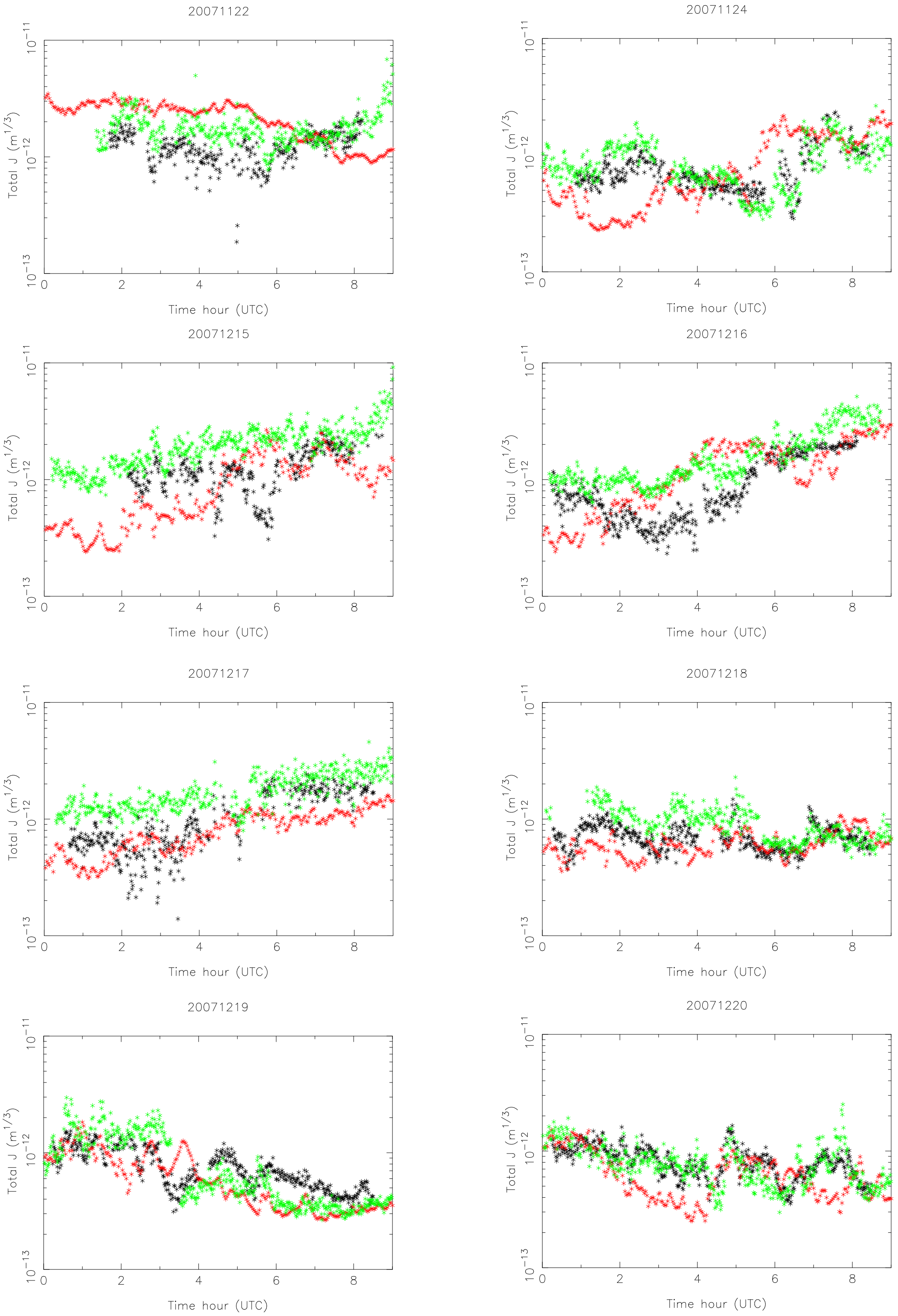


Figure 50: **Temporal evolutions** of J for different simulated nights from the PAR2007 site testing campaign, part 2. In black, the GS; in green, the DIMM; in red, the model (version v2).

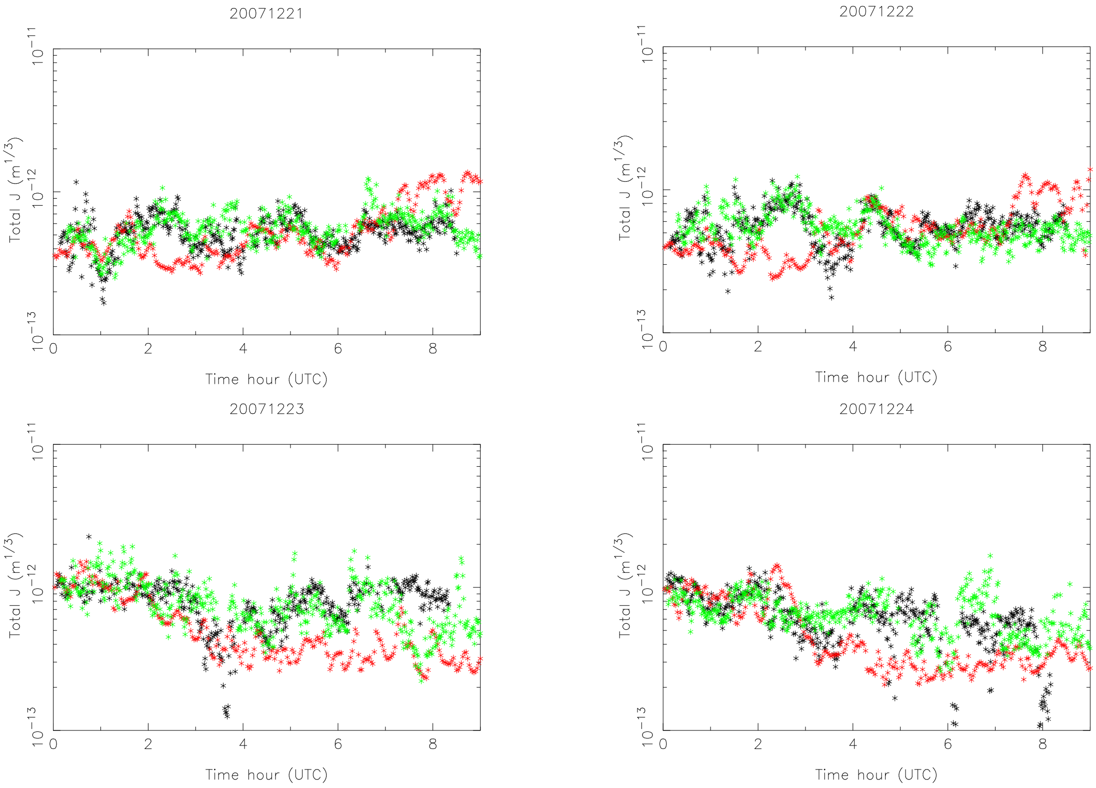


Figure 51: **Temporal evolutions** of J for different simulated nights from the PAR2007 site testing campaign, part 3. In black, the GS; in green, the DIMM; in red, the model (version v2).

6.4 Model validation

As it has been said at the beginning of Section 6.1 the sample selected for the final model validation is composed of 53 nights uniformly distributed in the period [6/2010-5/2011] (Table 67). After a first analysis of the seeing we decided to select a further independent sample of 36 nights in the year 2007 (Table 68) to better discuss results obtained for the seeing.

6.4.1 Overall statistical model performances

C_N^2 PROFILES

Differently from the calibration sample in which we used the GS data as a reference, for the final validation sample we have not a reliable reference providing us a vertical distribution of the turbulence. We can not compare therefore the average C_N^2 profile with some reference. Fig. 52 shows the average C_N^2 profile obtained with the samples of 53 nights and 36 nights and the addition of the two samples (89 nights). They are displayed together with the average C_N^2 profile obtained with the calibration sample just to appreciate how much the profile is modified. Thin lines (red and black) indicate, as in Fig. 34, the minimum and maximum at each level, of the C_N^2 profiles simulated by the model. In this figure it is possible to appreciate that, increasing the number of nights, the range [min,max] increases (the logarithmic scale necessarily gets thinner the differences). The goal of this picture is to put in evidence the vertical spatial variability of the C_N^2 profiles provided by the model. This will reflect on the temporal variability of the integrated astroclimatic parameters.

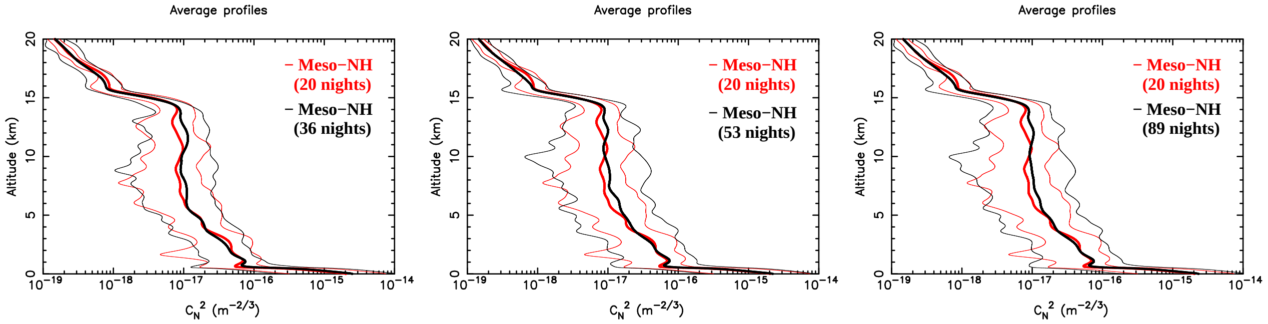


Figure 52: C_N^2 **average vertical profiles** at Cerro Paranal. **Left**: in red, Meso-NH (version v2_mod) average vertical profile from the 20-nights sample (Table 66); in black, Meso-NH average vertical profile from the validation 36-nights sample (Table 68). **Middle**: in red, Meso-NH (version v2_mod) average vertical profile from the 20-nights sample (Table 66); in black, Meso-NH average vertical profile from the validation 53-nights sample (Table 67). **Right**: in red, Meso-NH (version v2_mod) average vertical profile from the 20-nights sample (Table 66); in black, Meso-NH average vertical profile from the validation whole 89-nights sample. In both figures, the thick lines are the average profiles, and the thin lines the minimum and maximum values of the C_N^2 at each level and are an indicator of the night-to-night variability.

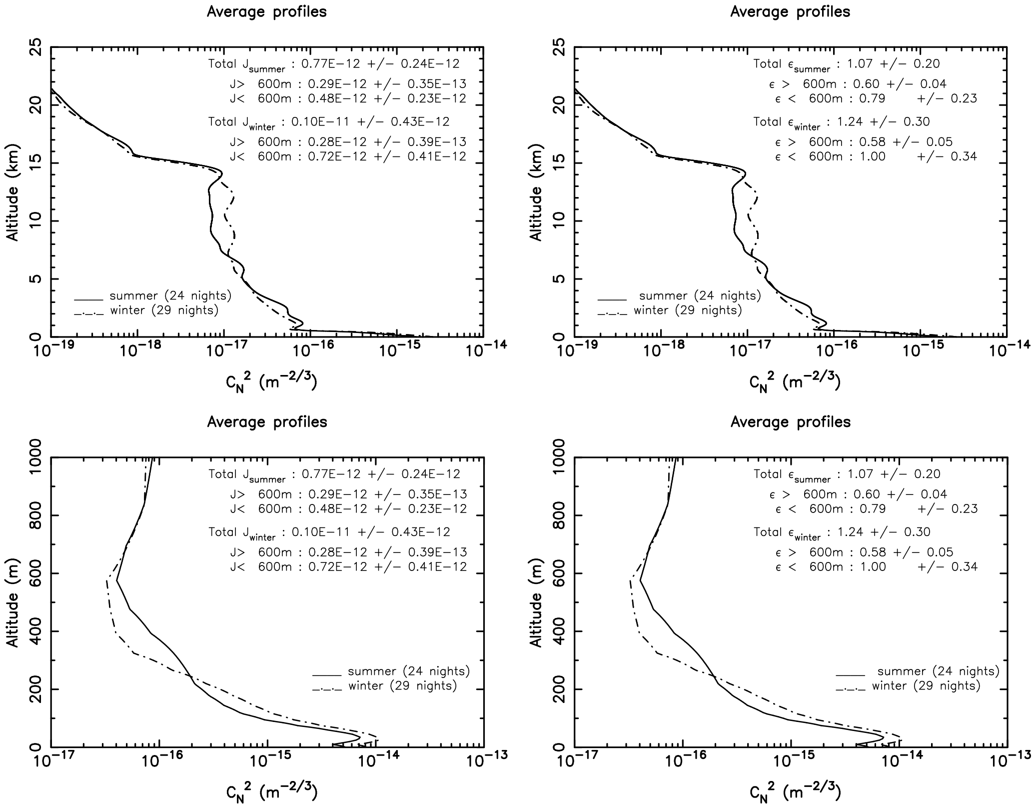


Figure 53: **Average vertical profiles** of the simulated C_N^2. The thin line is the summer vertical profiles (24 nights) and the dashed line is the winter vertical profile (29 nights). By winter we intend the [April 2010 - September 2010] semester, and by summer the [October 2010 - March 2011] semester. Figures in the first and second column are the same but on the first column are reported the value of $J = \int C_N^2(h) \cdot dh$, on the second one the seeing values. Unit of C_N^2 in $m^{-2/3}$.

In Fig. 53 are shown the average vertical profiles of the simulated C_N^2 profiles on the sample of 53 nights, with a distinction between winter [April 2010 - September 2010] and summer [October 2010 - March 2011] seasons. The number of nights is not completely balanced in the two seasons (29 in winter and 24 in summer) because the nights selected by ESO (see Table 67) were almost all in the winter season. In Fig. 53 a seasonal variation is put into evidence. From a quantitative point of view this is evident particularly near the ground where the turbulence in winter is stronger ($\sim$ 0.2 arcsec higher in winter than in summer). Even if in most of the astronomical sites at mid-latitude this is the typical seasonal variation for the seeing, at Chilean regions (tropic), such a seasonal trend is almost not visible [28]. Taking into account the whole set of results that will be discussed in this section, we think that the reason of this is due to fact that the calibration has been performed (necessarily) on a sample of nights concentrated on just a part of the solar year (summer). As it will be argued later we think that this effect should disappear if we consider a sample for the calibration including at least two periods of the solar year in summer and in winter.

SEEING

Fig. 54-top shows the scattered plot of the total seeing as measured by the DIMM and calculated by the model for the whole sample of 53 nights and for the sub-samples related to the winter (29 nights) and summer (24 nights) periods. Unfortunately, for the total seeing we have just one instrument to be taken as a reference and this limits somehow our analysis. We observe a very good bias of 0.08 arcsec and a RMSE that differs of $\sim$ 0.1 arcsec with respect to what we found for the calibration sample (0.53 arcsec in the validation sample vs. 0.38 arcsec (MNH-GS) and 0.43 arcsec (MNH-DIMM) for the calibration sample). Unfortunately we can not estimate if the RMSE for measurements degrades and how much it degrades increasing the number of nights because we have not GS measurements extended on a such large statistical sample. We can not know therefore, what would be the uncertainty obtained with measurements taken in equivalent conditions. For what concerns the RMSE we can not add many other commentary due to this limitation. However, looking at the bias, it is clear that the model seems to behave better during the summer and it overestimates slightly the turbulence in winter time (bias of 0.19 arcsec in winter vs. 0.07 arcsec in summer). This is a difference of more than 0.1 arcsec therefore not negligible. We remind that the calibration sample belongs to the November-December 2007 i.e. the summer period. It seems therefore that the model shows a better behavior in the same season of the calibration sample. The fact that the median bias in winter and summer have opposite signs, tell us that highly probably, if we take a calibration samples at least distributed in two extremes periods of the solar year (summer and winter), we should re-balance the bias and the seasonal variation should disappear. Even better would be to simply performa an independent calibration in summer and in winter time[14].

Being that the period selected for the final validation was more recent (2010-2011) with respect to the calibration sample (2007), we decided to perform a validation of the model also on a second sample (independent from the calibration sample) but belonging to the same year: 2007. This is to be sure to exclude effects due to the initialization data. Fig. 54-bottom shows the same as Fig. 54-top but for the sample of 36 nights of 2007 (Table 68). We observe the same result obtained with the sample of 53 nights that is a substantial good agreement of the model with the measurements in terms of bias and RMSE. If we look at the bias it is evident a better model performance in summer and with the same trend between winter and summer observed in the previous sample of nights in 2010-2011 years. As before it is observed the same seasonal variation trend by the model if compared to measurements. The seasonal variation is also clearly evident if one look at the seeing temporal evolution of measurements and simulations during each night. These results confirm the thesis we have just exposed on the origin of the seasonal effect and on the solution to smooth that. Moreover, the good news is that the model calibration seems to be independent from the initialization data. Another useful information is that the model reproduces the same kind of results in different years. This means that, even if the statistical sample is a few tens of nights this seems representative of a whole year. As we said in the previous sub-section, being that the bias is positive in the winter and negative in the summer period, we conclude that a

---

[14] We highlight that all the other astroclimatic parameters show the same seasonal trends with a better model performances in summer than in winter - see presentation in the Final Review meeting and in a contribution to the AO4ELT 3rd Edit. This confirms the thesis presented here.

little bit richer sample for the calibration preferably with data distributed all along the year, should be enough to re-balance the small off-set still present in data. As already said, the solution we propose is to perform an independent calibration on summer and winter time.

Unfortunately we could not use MASS data for the calibration and GS data were available only on 20 nights. For this reason, for the final model sample, we can not test the free atmosphere and the boundary layer because of the lack of reference. We proved, indeed, that the MASS data are not reliable in many layers. The conclusion is, therefore, that the model performances is associated to a BIAS of 0.08 arcsec and a RMSE of 0.53 arcsec.

Besides we think that, in perspective, it would be useful to have at least a second reference (beside the DIMM) in order to better judge the model performances with respect to the accuracy that the measurements can reach.

Table 25 summarized bias and RMSE values for both samples of 53 and 36 nights. In the BSCW archive are reported also results obtained with version v2 that indicate a larger bias of the model in the same direction shown for v2_mod. Our conclusion is that, for the present calibration, v2_mod is better than v2. We will use in the rest of the report the v2_mod version.

Statistical parameter	2010/1011 **53 nights**	2010/2011 29 winter nights	2010/2011 24 summer nights only
BIAS	0.08	0.19	-0.07
RMSE	0.53	0.53	0.53
Statistical parameter	2007 **36 nights**	2007 18 winter nights	2007 18 summer nights only
BIAS	0.08	0.24	-0.08
RMSE	0.52	0.57	0.46

Table 25: **Bias** and **RMSE** of the **total seeing** between the **model (v2_mod)** and the **DIMM**, for the validation samples (Tables 67,68). See Fig. 54. Unit in arcsec.

WAVEFRONT COHERENCE TIME

Fig. 55-left shows the scattered plot of τ_0 of the DIMM and MASS data-set related to the sample of 53 nights. The bias and RMSE show the same trends obtained in the calibration sample i.e. the DIMM seems to slightly underestimate the τ_0 with respect to the MASS. Fig. 55-center and Fig. 55-right shows the scattered plot of τ_0 obtained between the model and the DIMM and the model and the MASS. The order of magnitude of the RMSE values is similar to that obtained in the calibration sample. The model seems to be better correlated with DIMM measurements than with the MASS ones. From one side this is coherent with what we found for the seeing i.e. the model shows a small tendency in overestimating the turbulence in the whole sample of 53 nights and it is therefore natural that the τ_0 is slightly underestimated particularly with respect to the MASS. On the other side, we remind that the MASS τ_0 values have to be taken with precaution and there is still the possibility that the τ_0 from the MASS is overestimated. **The conclusion of the analysis is that the model performances on the sample of 53 nights remain very satisfactory and coherent with previous results found with the calibration sample of 20 nights. The RMSE is of the order of 1.58 ms or 2.27 ms (depending if we refer to the DIMM or the MASS as a reference, respectively).** We calculated the scattered plots for the different seasons and we found results similar to those obtained for the seeing i.e. in the summer season the behavior of the model improves with respect to the winter (see BSCW archive). Fig. 55-bottom shows the results obtained considering the wind speed in the first 30 meters as V(h) = V(h)·2. In this case we do not observe an improvement of the results because of the same reason: calibration done on a limited sample of nights concentrated in one part (summer) of the solar year. Table 26 summarizes the results obtained with version v2 and v2_mod of the model. **We observe, moreover, that the σ values between the model and observations is definitely comparable (and in some cases even smaller) than those**

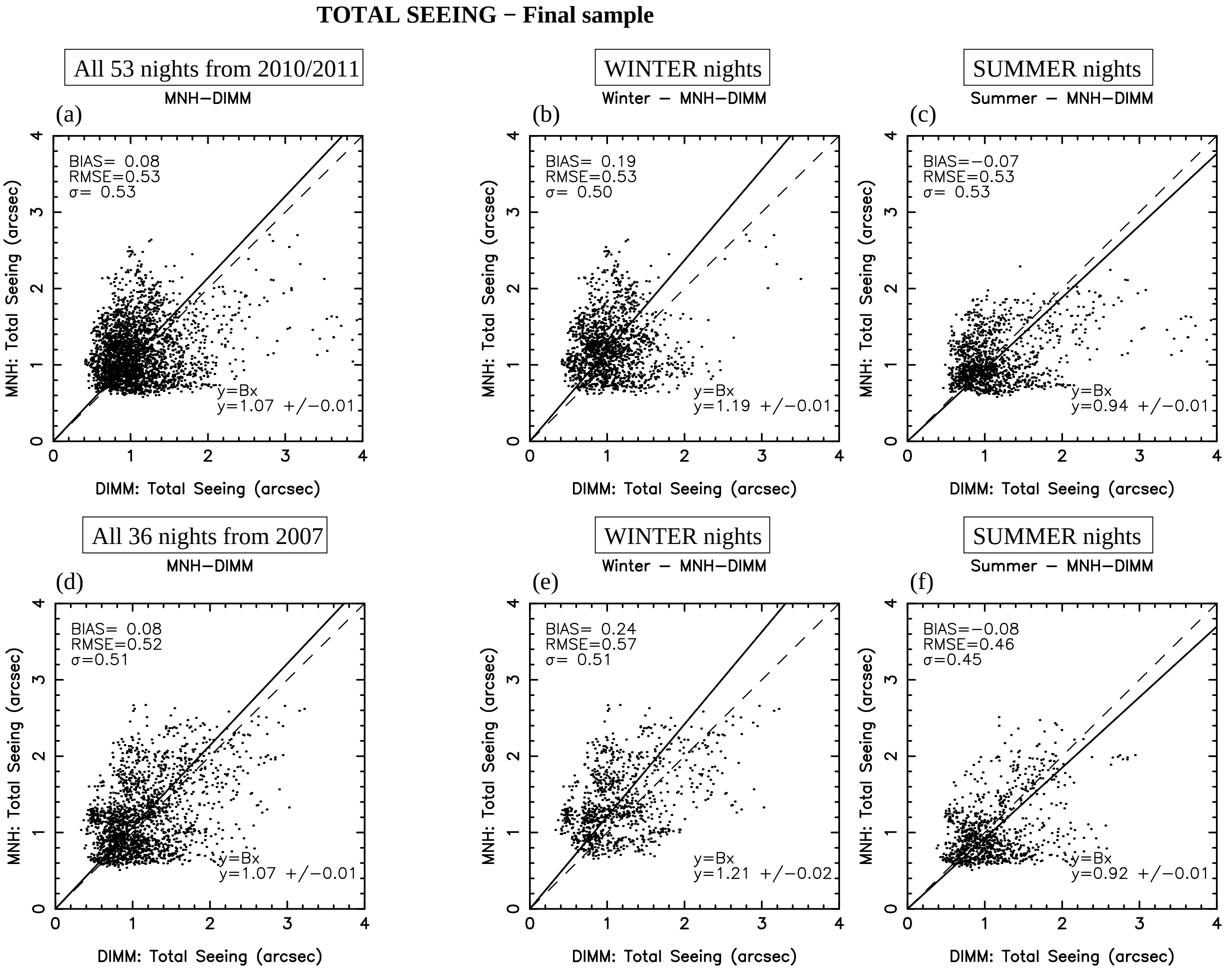


Figure 54: **Scattered plots** of the **total seeing** (Model v2_mod vs. DIMM data). **(a)** whole years 2010/2011 nights sample; **(b)** only winter nights [April 2010 - September 2010] (29 nights); **(c)** only summer nights [October 2010 - March 2011] (24 nights). **(d), (e) and (f)** are similar to (a), (b) and (c), respectively, but considering only nights from the year 2007 sample. Each point represents a 10-min average. 53 nights are analyzed for the years 2010/2011 sample. 36 nights are analyzed for the year 2007 sample. Unit in arcsec.

obtained between observations from different instruments.

Statistical parameter	DIMM - MASS	MNH-DIMM	MNH-MASS	MNH-DIMM $V_{[0-30m]}$ x 2	MNH-MASS $V_{[0-30m]}$ x 2
BIAS	-1.21	-0.30	-1.39	-0.59	-1.66
RMSE	2.18	1.58	2.27	1.69	2.45

Table 26: **Bias** and **RMSE** between the **model (v2_mod)** and the MASS and DIMM instruments of the **wavefront coherence time** for the validation sample of 53 nights (Table 67). See Fig. 55. Unit in ms.

ISOPLANATIC ANGLE

Fig. 56-left shows the scattered plot of measurements from DIMM and MASS. Differently from the calibration sample the bias between the two set of measurements is more important with a tendency of the DIMM in measuring larger θ_0. Fig. 56-center and Fig. 56-right show the scattered plots of model versus the DIMM

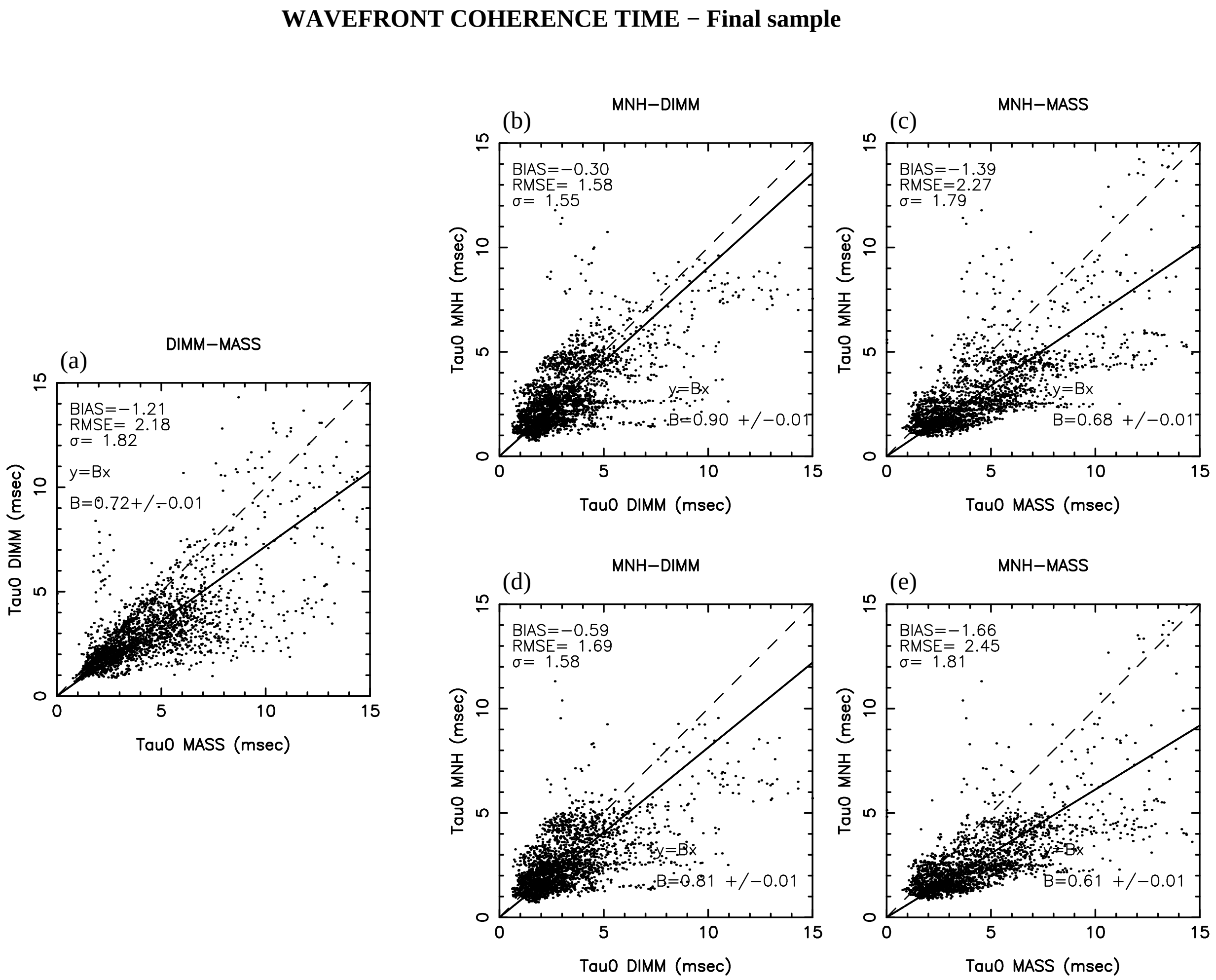


Figure 55: **Scattered plots** of the **wavefront coherence time**. **(a)** DIMM vs. MASS (44 nights); **(b)** Model (v2_mod) vs. DIMM (53 nights); **(c)** Model (v2_mod) vs. MASS (48 nights); **(d) and (e)** as (b) and (c) respectively, but the model wind up to 30 m a.g.l. has been multiplied by a factor of 2. Each point represents a 10-min average. Unit in ms.

and the MASS. The model seems perfectly balanced (bias = -0.06 arcsec) with MASS data while an off-set is evident with respect to DIMM data (bias = -0.83 arcsec). Unfortunately, comparing the results obtained with the calibration and final sample, it is difficult to conclude if it is the MASS that overestimates the θ_0 or the DIMM that underestimates the θ_0. On the other side the model in the high part of the atmosphere (h $>$ 15 km) behaved in a very similar way with respect to the calibration sample (Fig. 52). **All this seems to indicate that it is extremely important to consider rich statistical samples of measurements taken with different instruments running simultaneously to be able to control the status of the turbulence in the atmosphere**. Because the RMSE between the model and the DIMM is visibly larger, we think that, in this case, the RMSE is probably affected by the bias. This argument is in favor of believing that MASS data are more reliable. In favor of this thesis we remind that θ_0 depends mainly on the turbulence in the high part of the atmosphere and looking at the cross-correlation between MASS and GS (Fig. 28) we note that the correlation on the highest layer (layer 6) is definitely satisfactory. The range of variability of θ_0 values is definitely larger [1.5 arcsec - 2.2 arcsec] with respect to that found in the calibration sample. It is however smaller than that observed [0.5 arcsec - 4 arcsec]. We stress out the fact that, at present, our major attention is addressed in trying to improve the model ability in reconstructing the turbulence variability in the free atmosphere. **The model is**

visibly able to reconstruct a bias (-0.06 arcsec) and RMSE (0.65 arcsec) that are comparable to those observed with different instruments. It is however not able to put in evidence, during the same night, changes of θ_0 of the same orders of those observed. This problem is strictly correlated to the one identified for the seeing in the free atmosphere. As previously said we are working on this last months on three different directions to solve this limitation. Table 27 summarizes results obtained with version v2 and v2_mod of the model. **We observe that the σ obtained between the model and observations are of the order of 0.65-0.69 arcsec while the σ between observations from different instruments is of the order of 0.55 arcsec.**

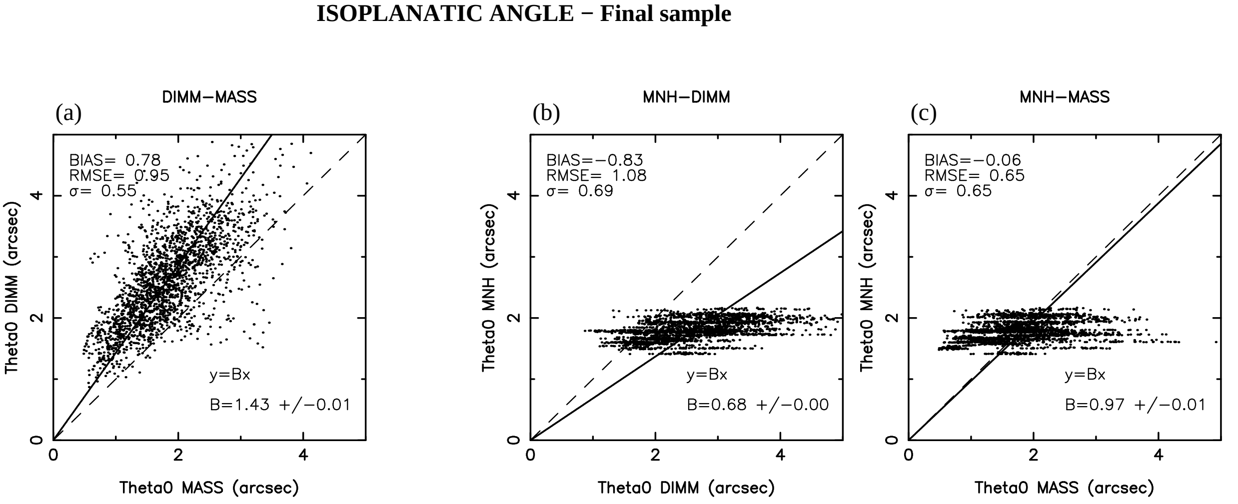


Figure 56: **Scattered plots** of the **isoplanatic angle**. **(a)** DIMM vs. MASS (44 nights); **(b)** Model (v2_mod) vs. DIMM (49 nights); **(c)** Model (v2_mod) vs. MASS (48 nights). Each point represents a 10-min average. Unit in arcsec.

Statistical parameter	DIMM - MASS	MNH-DIMM	MNH-MASS
BIAS	0.78	-0.83	-0.06
RMSE	0.95	1.08	0.65

Table 27: **Bias** and **RMSE** between the **model (v2_mod)** and the MASS and DIMM instruments of the **isoplanatic angle** for a sub-sample of the validation sample of 53 nights (Table 67). See Fig. 56. Unit in arcsec.

6.4.2 Individual nights model performances

Fig. 57-left (top and bottom) shows the cumulative distribution of the bias and RMSE of the total seeing measured by the DIMM and calculated by the model on individual nights (sample of 53 nights). The median value of the bias is excellent (-0.06 arcsec) with first and third quartiles in the range [-0.34 arcsec - 0.16 arcsec]. The statistical sample is well centered on zero. **The median value of the RMSE is 0.42 arcsec with first and third quartiles of 0.33 arcsec and 0.53 arcsec, respectively**.

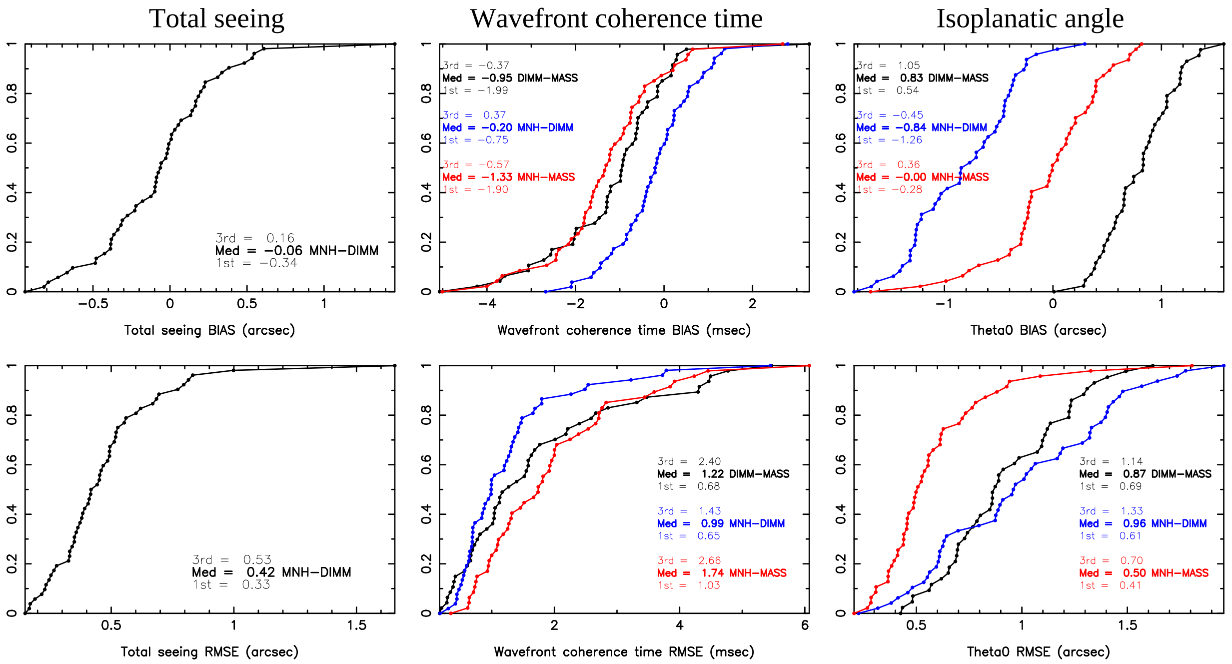


Figure 57: **Cumulative distribution** of the **total seeing bias** and **RMSE** (left), of the **wavefront coherence time bias** and **RMSE** (center) and of the **isoplanatic angle bias** and **RMSE** (right). The sample is the 53 nights validation sample (Table 67). For the seeing, only Model-DIMM comparison. For the wavefront coherence time and the isoplanatic angle, in blue comparisons Model-DIMM, in red comparisons Model-MASS and in black comparisons DIMM-MASS. Model version is v2_mod.

Fig. 57-center shows the cumulative distribution of the bias and RMSE of the wavefront coherence time calculated by the model and measured by DIMM and MASS night by night. The median value of the RMSE between DIMM and MASS is 1.22 ms with a first-third quartile range of [0.68 ms - 2.40 ms]. The median value of the cases model-MASS and model-DIMM are 1.74 ms and 0.99 ms while the first-third quartiles ranges are respectively [1.03 ms - 2.66 ms] and [0.65 ms - 1.43 ms]. As observed in the analysis done for the overall statistical sample, the model is slightly better correlated to the DIMM than the MASS. Taking as a reference the two instruments we conclude therefore that, in 75% of the cases, the RMSE is inferior to 1.43 ms (DIMM) and to 2.66 ms (MASS). As previously mentioned MASS values of τ_0 have to be taken with precaution. On the other side, on the calibration sample, the DIMM seems to slightly underestimate the τ_0 with respect to the GS and the MASS (see calibration sample). At present it is therefore difficult to decide which of the two instruments we have to trust more. It would be useful to have a serie of measurements (GS, MASS and DIMM) in winter (when the wind speed at the jet-stream level is stronger) to cross-correlate them and to check the existence of potential biases. This might provide useful informations on the opportunity of the constant of 0.4 in Eq.11. The calibration sample belongs, indeed, to the summer period. Fig. 57-right shows the cumulative distribution of the bias and RMSE of the isoplanatic angle calculated by the model and measured by DIMM and MASS. As already observed in the analysis of the overall statistical sample, the best median value is the one obtained between the model and the MASS. The median value of the RMSE between MASS and DIMM is

0.87 arcsec while the RMSE between model and DIMM and model and MASS is respectively 0.96 arcsec and 0.50 arcsec. This tells us that the dispersion of the model with respect to the measurements is even better than the dispersion between measurements provided by different instruments. Also we note that the typical dispersion observed between measurements (0.87 arcsec) is not negligible and equivalent to ∼ 35 % of the typical θ_0 (∼ 2.5 arcsec). Table 28 summarizes the results obtained with model version v2 and v2_mod.

Astroclimatic parameter	Statistical parameter	DIMM - MASS	MNH - DIMM	MNH - MASS
ε_{tot}	BIAS	-	$-0.06^{+0.16}_{-0.34}$	-
	RMSE	-	$0.42^{+0.53}_{+0.33}$	-
τ_0	BIAS	$-0.95^{-0.37}_{-1.99}$	$-0.20^{+0.37}_{-0.75}$	$-1.33^{-0.57}_{-1.90}$
	RMSE	$1.22^{+2.40}_{+0.68}$	$0.99^{+1.43}_{+0.65}$	$1.74^{+2.66}_{+1.03}$
θ_0	BIAS	$0.83^{+1.05}_{+0.54}$	$-0.84^{-0.45}_{-1.26}$	$0.00^{+0.36}_{-0.28}$
	RMSE	$0.87^{+1.14}_{+0.69}$	$0.96^{+1.33}_{+0.61}$	$0.50^{+0.70}_{+0.41}$

Table 28: Median **bias** and **RMSE** of the **total seeing**, the **wavefront coherence time** and the **isoplanatic angle**, when available from the 53 single nights from the calibration sample (Table 67), when measurements are available. In small fonts, the 1st and 3rd quartiles. See Fig. 57.

7 Model reliability, hit rates and model score of success

7.1 Model reliability and accuracy

One of the most important output we expect from this study is the estimate of the accuracy with which the model can reconstruct the value of the different astroclimatic parameters: ε, ε_{BL}, ε_{FA}, τ_0 and θ_0. Statistically speaking, the accuracy is the difference between an estimate and the 'true value'. The 'true value' is in general assumed to be equal to a measured value. In the field of the forecasts, the accuracy is basically the difference between the forecasts and the observations. In the case of the turbulence (highly non linear and stochastical parameter) it can be extremely dangerous to refer to measurements provided by just one instrument, in other words it is extremely difficult to establish and define a 'true value'. It is more suitable to replace the concept of 'true value' with the dispersion of measurements obtained by different instruments: Δ_{obs} that is basically the reference that represents our specifications (Δ_{obs}=$\Delta_{reference}$). To quantify the accuracy of the model we compare the dispersion between the model estimates and observations Δ_{mod} with the dispersion between measurements done with different instruments Δ_{obs}. In this context we are using the term *dispersion* in a very general sense to indicate the *error* between forecasts and measurements. If we want to give a quantitative value to this term we have to replace it with a statistical operator. Which statistical operator ? In this respect there is a not a univocal habit. **In this study we used so far to quantify the accuracy, the bias (or mean error) (Eq.1) and the root mean squared errors (RMSE - Eq.2)**. Both provides useful informations. The bias range is from minus infinity to plus infinity and a perfect score is zero. It tells us if a system present some systematic trends (in case the bias is different from zero). The RMSE is the squared difference between the forecasts and the observations and provides us an information on the typical dispersion between forecasts and observations including both systematic and statistical errors. The RMSE range is between zero and infinity with a perfect score of zero. In all the statistical analysis (scattering plots) we also calculated the bias-corrected RMSE that we called σ (Eq.3) that permits us to disentangle the systematic errors from the random errors. Of course the bias and the RMSE are not exhaustive in the sense that they do not include all the informations we could require. The bias, for example, is not exhaustive because it is possible to reach a perfect score for a dataset with large errors, if there are compensating errors of a reverse sign. It might be useful therefore, in the future, to calculate the bias of the absolute errors (Eq.16) or even better the median of the absolute errors (Eq.17):

$$BIAS_{abs} = \sum_{i=1}^{N} \frac{|Y_i - X_i|}{N} \tag{16}$$

$$median(|Y_1 - X_1|, |Y_2 - X_2|, |Y_3 - X_3|, ...) \tag{17}$$

The median is less sensitive to large errors in case of a poor statistic. Also the bias and the RMSE do not represent the unique way to represent the systematic and statistical errors. For example, the RMSE, being a second order moment, is much more sensitive to large forecasts errors than the bias or the median of the absolute difference of forecasts and observations. In case of a limited statistic it might be altered by a few large errors. In other words, the RMSE has a tendency in overestimating the error. The same can be said for σ. The concept of accuracy in application to the forecasts verification is therefore alternatively associated in the literature to the RMSE or the bias depending on the application on which one is interested on[44]. Another useful element that tells us that bias and RMSE are useful but not exhaustive in shown in Fig.43. In that figure we can observe that the RMSE of the model with respect to the instruments is comparable to the RMSE of the GS vs. the DIMM and the DIMM vs. the MASS. However it is evident that the scattering plot of Fig.43-(d)-(e)-(f) (model vs. measurements from the three instruments) is mainly flat. The reason of the flatness of the cloud of points is due to the problem of the model in reconstructing the PtV turbulence fluctuations in this free atmosphere and therefore in the isoplanatic angle. This effect is not represented by the RMSE but this does not mean that the RMSE is useless. However the similar RMSE between model and measurements and between DIMM vs. MASS and DIMM vs. GS tells us that, even if the measurements have not the problem of the weak PtV just described for the model, the discrepancy between the measurements of these instruments is in any case comparable to that of the model compared to the measurements. In other words, the RMSE is a reliable parameter that gives us some useful information but not all the informations. It is necessary to calculate also other statistical operators.

Table 29 summarizes the Δ_{obs}, Table 30 the Δ_{mod}. The intervals/ranges [...] are calculated taking into account all the Δ_{obs} obtained considering the calibration and final samples[15]. Table 29 reports just one value for the ε_{TOT} and not a range of values because we have been able to treat just two instruments, the GS and the DIMM, during the 20 nights of the PAR2007 campaign.
Comparing the two tables the most relevant result is that, in most of cases (where the comparison is possible), the Δ_{mod} is definitely comparable to the corresponding Δ_{obs}. We precise that, in these two tables, we included all the results found in our study, however a similar qualitative conclusion is obtained even if we consider only results related to the final sample of 53 nights (Table 32). In some cases (see details in Section 6) we proved that results can be easily improved: for example, the Δ_{mod} of the total seeing ε_{TOT} can be reduced with a calibration done using nights belonging to summer and winter. A calibration performed on a couple of periods (summer and winter) should produce an improvement of model performances. In other cases, we argued that a result from the instrument X has higher level of reliability than instrument Y. For example, the estimate of θ_0 on the final validation sample of 53 nights by the DIMM seems less probable than the MASS one. An important result is however that the presence of several instruments running simultaneously is definitely fundamental, not only to have a more realistic idea of the model performances, but also to have a more realistic idea of the accuracy of measurements and on the uncertainties that still exist.

Table 32 is the same as Table 30 but only results obtained from the statistical sample of 53 nights (independent from the calibration sample of 20 nights) have been used. It is possible to note that results are very similar and the conclusions just described for Table 30 are also valid for Table 32. We highlight that this is the first model validation ever achieved with a sample of nights completely independent from the calibration sample.

In the field of the forecasts of atmospheric parameters an alternative approach is frequently used to judge the quality of the forecasts products [45]. It consists in the calculation of the **category tables (or the hit-rates tables)** from which it is possible to calculate a set of several statistical operators indicating the quality of the prediction (false alarm, miss rate,...).

	OBS - OBS		
	$\Delta_{25,reference}$	$\Delta_{50,reference}$	$\Delta_{75,reference}$
τ_0(*) (ms)	[0.65-0.83]	[1.08-1.50]	[1.55-3.85]
θ_0(*) (arcsec)	[0.29-0.69]	[0.42-0.87]	[0.66-1.14]
ε_{tot}(**) (arcsec)	0.20	0.28	0.40

(*) DIMM-GS, MASS-DIMM, MASS-GS (20 nights)
DIMM-MASS (53 nights)
(**) DIMM-GS (20 nights)

Table 29: Table summarizing the RMSE value ranges for all the astroclimatic parameters, between DIMM, MASS and GS instruments.

---

[15] This is because we did not want to skip-off the most representative instrument: the GS.

	MNH - OBS		
	$\Delta_{25,model}$	$\Delta_{50,model}$	$\Delta_{75,model}$
τ_0(*) (ms)	[0.61-1.03]	[0.86-1.74]	[1.43-2.73]
θ_0(*) (arcsec)	[0.41-0.61]	[0.50-0.96]	[0.70-1.33]
ε_{tot}(**) (arcsec)	[0.21-0.33]	[0.30-0.42]	[0.47-0.53]
ε_{BL}(***) (arcsec)	0.31	0.38	0.61
ε_{FA}(***) (arcsec)	0.15	0.21	0.25

(*) MNH-GS, MNH-DIMM, MNH-MASS (20 nights)
MNH-DIMM, MNH-MASS (53 nights)
(**) MNH-GS, MNH-DIMM (20 nights)
MNH-DIMM (53 nights)
(***) MNH-GS (20 nights)

Table 30: Table summarizing the RMSE value ranges for all the astroclimatic parameters, between Meso-NH and observations. Values obtained by both calibration sample of final validation sample of 53 nights are considered in the statistic.

	OBS - OBS		
	$\Delta_{25,reference}$	$\Delta_{50,reference}$	$\Delta_{75,reference}$
τ_0(*) (ms)	0.68	1.22	2.40
θ_0(*) (arcsec)	0.69	0.87	1.14

(*) DIMM-MASS (53 nights)

Table 31: Table summarizing the RMSE value ranges for all the astroclimatic parameters, between DIMM, MASS instruments.

	MNH - OBS		
	$\Delta_{25,model}$	$\Delta_{50,model}$	$\Delta_{75,model}$
τ_0(*) (ms)	[0.65-1.03]	[0.99-1.74]	[1.43-2.66]
θ_0(*) (arcsec)	[0.41-0.61]	[0.50-0.96]	[0.70-1.33]
ε_{tot}(**) (arcsec)	0.33	0.42	0.53

(*) MNH-DIMM, MNH-MASS (53 nights)

Table 32: Table summarizing the RMSE value ranges for all the astroclimatic parameters, between Meso-NH and observations. Only values obtained with the statistical sample of 53 nights (independent form the calibration sample) are used.

7.2 Hit rates

7.2.1 Optical Turbulence

A complementary tool that provides the percentage of success of a prediction is the hit-rate table. We consider a reference: for example the observations of the parameter X. We calculate the tertiles: 33.3% and 66.6% of the distribution of X. We consider then the model predictions of the same parameter X and we count the number of times in which the model matches the same range of observed values normalized by the total number of counts. Both observations and models are resampled on a 10 minutes time scales. We obtain therefore a (3x3) matrix in which the diagonal (matrix elements: [1,1], [2,2] and [3,3]) corresponds to the cases in which the model provides a good prediction because the prediction falls in the same ranges of observations. Matrix elements [1,3] and [3,1] correspond to the percentage of cases in which the model provides a bad prediction because the prediction is as far as possible from the observed value. Matrix elements [1,2], [2,1], [2,3] and [3,2] represent the cases in which the model provides a prediction of medium quality. Tables 33,37,41 summarize the "hit rates" of the model for the wavefront coherence time prediction, the isoplanatic angle and the total seeing, respectively. The percentiles G, N and B represent the percentage of time with good, medium and bad (respectively) prediction from the model with respect to the instrument taken as a reference, or the percentage of time for which two instruments match is good, medium or bad (respectively). In some texts [45] the parameter G is even defined as *"accuracy"* of the model. The definitions of G, N and B are:

$$G = \sum_{i=1}^{3} [i,i] \tag{18}$$

$$N = [1,2] + [2,1] + [2,3] + [3,2] \tag{19}$$

$$B = [1,3] + [3,1] \tag{20}$$

The perfect case has [1,1]=[2,2]=[3,3]=33,33 and the other matrix elements equal to zero. By definition G+N+B = 100%, however the values reported in Tables 33,37,41 are rounded values and the total sums may slightly differ from 100%. Also the addition of all the elements of each column corresponds to 33,33%. **It is worth highlighting that the percentage of each matrix element does not represent the percentage of success of the model but simply the percentage of time in which the model hits the same values observed $\sim$ 1/3 of time**. In many cases it can be much more useful to use as threshold limits some more significative values for the specific application instead of the first and third tertiles that simply divide in three equal part the distribution of the observed values [45]. A few examples will be discussed later.

WAVEFRONT COHERENCE TIME

Looking at Table 33 we note that the good hit-rate of the model for the **wavefront coherence time** τ_0 (for the validation sample of 53 nights) is almost the same (56.2%) than that obtained by different instruments (55.0% and 56.2%) if we consider the DIMM as a reference and it is slightly inferior (46.0%) if we consider the MASS as a reference. The bad hit-rate of the model is very low ($\sim$ 5-7%). The medium hit-rate of the model is between 39-47% depending on the reference selected. The presence of a bias between the couple of data-set is visible on the matrix elements: [1,2], [2,1], [2,3], [3,2]. When there is a bias [1,2] and [2,3] are larger/smaller than [2,1] and [3,2]. The case of MNH vs. DIMM (bottom-right) is the most balanced case among all and it corresponds also to the best agreement of the model to observations. In conclusion the hit-rate of the model is very close to the hit-rate of instruments for this parameter with a general small bad hit-rate. Hit-rate success of the model with respect to DIMM for the sample of 89 nights (Table 34) is very similar to that obtained with 53 nights and it provides basically the same values of G, N and B: G=56.1%; N=39.0%; B=4.9%. Looking at the best results of the model (bottom-right) we observe that, when the model predicts extremes values of τ_0 ($<$ 2 ms and $>$ 3.34 ms) these are in general good predictions. It might be interesting to investigate if the same relationship between τ_0 retrieved from MASS and DIMM observed on the 53 nights sample is also observed on the sample of 36 nights of 2007. Unfortunately there are no MASS measurements during the months [1/2007-10/2007].

From the point of view of the application to the turbulence prediction for the scheduling, it is indeed more interesting to select some more significative thresholds than the simple repartition in three equal parts of the observed data. In other words, it is much more interesting for us to know if the model is able to discriminate between τ_0 larger or smaller than a certain value X that can represent some specific critical value for the observational point of view (independently from the fact that this is observed frequently or not) instead than to know if the model is able to reconstruct the distribution in three equal parts of the observed values. In this respect we noted that the performances (of both the model and the instruments) change depending on the threshold values (of course this would not happen in a perfect model or with perfect instruments). Table 34 shows, for example, results obtained considering as thresholds the values of **3 msec and 6 msec**. **We selected 6 msec because this is the typical median value of a good site and 3 msec is clearly an indictor of a small τ_0. In this way we are checking if the model is able to disentangle between the categories of the good and bad τ_0.** Under these conditions the good hit-rates of the model (Table 35) are 54% and 71% (taking as a reference the MASS or the DIMM) while the good hit-rate for the instruments is 61%. The good model performances of 71% remain invariant if we consider the 89 nights sample (Table 36). In other words, we obtained better model performances and better agreement between the two instruments. The dependency of the hit-rates on the threshold values is not a surprise. It simply means that model and/or instruments are more efficient in discriminating some values than others. Some dichotomies are more easily reconstructed by the model than others. In other words, the goodness of the model is not an absolute concept but it depends on the question we are asking for. **In the context of the verification of the quality of a model the crucial issue is therefore (a) to create a set of categories and verify if the model provides good performances under these conditions or alternatively (b) to identify the categories in which the model provides good performances and evaluate if these informations can be useful or not for the operational scheduling[16]. The option (b) is more pragmatic and highly suggested. This means that, with the same data-set of predictions one could in principle build several different category tables suitable to answer to more than one question. In some cases the model can be not efficient enough, in other yes. A more complete analysis of the category tables might include in the future the calculation of the miss rates, the false alarms, etc.**

---

[16] With the term 'category' we mean the identification of threshold values.

		MASS τ_0 (ms) - 48 nights		
		<2.70	[2.70,4.80]	>4.80
DIMM τ_0	<2.70	30.8%	18.3%	5.3%
	[2.70,4.80]	1.6%	14.4%	18.1%
	>4.80	0.9%	0.6%	9.8%

G=55.0% N=38.7% B=6.3%

		DIMM τ_0 (ms) - 48 nights		
		<2.06	[2.06,3.20]	>3.20
MASS τ_0 (ms)	<2.06	14.6%	1.9%	0.6%
	[2.06,3.20]	14.3%	10.7%	1.7%
	>3.20	4.4%	20.8%	31.0%

G=56.2% N=38.7% B=5.0%

		MASS τ_0 (ms) - 48 nights		
		<2.70	[2.70,4.81]	>4.81
MNH τ_0 (ms)	<2.70	31.2%	24.8%	7.3%
	[2.70,4.81]	2.0%	6.7%	17.9%
	>4.81	0.1%	1.8%	8.1%

G=46.0% N=46.6% B=7.4%

		DIMM τ_0 (ms) - 53 nights		
		<2.08	[2.08,3.34]	>3.34
MNH τ_0 (ms)	<2.08	23.4%	13.2%	3.6%
	[2.08,3.34]	8.7%	13.4%	10.3%
	>3.34	1.3%	6.7%	19.4%

G=56.2% N=38.9% B=4.9%

Table 33: Hit rates table for the **wavefront coherence time** using the tertiles of the distribution of observations as a threshold (method a). G is the percentage of good prediction, N is the percentage of medium prediction and B is the percentage of bad predictions. G = [1,1] + [2,2] + [3,3]; N = [1,2] + [2,1] + [2,3] + [3,2] and B = [1,3] + [3,1], where [i,j] is the value at the ith row and the jth column . These values are obtained using the final validation sample of **53 nights**, with the exclusion of the nights for which an instrument gave no measurements (Table 67 - 48 nights only when the MASS is involved in the comparison).

		DIMM τ_0 (ms) - 89 nights		
		< 1.85	[1.85,3.07]	> 3.07
MNH τ_0 (ms)	<1.85	21.9%	10.3%	2.53%
	[1.85,3.07]	9.0%	14.5%	11.2%
	>3.07	2.4%	8.5%	19.6%

G=56.1% N=39.0% B=4.9%

Table 34: Same as Table 33 bottom-right but the sample is **89 nights** (method a).

		MASS τ_0 (ms) - 48 nights		
		<3	[3,6]	>6
DIMM τ_0	<3	36.8%	20.8%	4.1%
	[3,6]	2.3%	19.2%	9.7%
	>6	0.5%	0.5%	5.9%

G=61.9% N=33.4% B=4.7%

		DIMM τ_0 (ms) - 48 nights		
		<3	[3,6]	>6
MASS τ_0 (ms)	<3	36.8%	2.3%	0.6%
	[3,6]	20.9%	19.2%	0.5%
	>6	4.1%	9.7%	5.9%

G=61.9% N=33.4% B=4.7%

		MASS τ_0 (ms) - 48 nights		
		<3	[3,6]	>6
MNH τ_0 (ms)	<3	38.0%	27.1%	3.4%
	[3,6]	1.6%	12.1%	13.0%
	>6	0.0%	1.4%	3.5%

G=53.5% N=43.1% B=3.4%

		DIMM τ_0 (ms) - 53 nights		
		<3	[3,6]	>6
MNH τ_0 (ms)	<3	51.35%	14.8%	1.6%
	[3,6]	8.2%	16.4%	2.9%
	>6	0.2%	1.3%	3.2%

G=71.0% N=27.2% B=1.8%

Table 35: As Table33 but with thresholds: **3 msec and 6 msec** (method b).

		DIMM τ_0 (ms) - 89 nights		
		< 3	[3,6]	> 6
MNH τ_0 (ms)	<3	53.9%	12.6%	1.5%
	[3,6]	10.9%	15.2%	2.2%
	>6	0.3%	0.8%	2.6%

G=71.2% N=26.4% B=1.8%

Table 36: Same as Table 34 but with thresholds: **3 msec and 6 msec** (method b).

ISOPLANATIC ANGLE

Looking at Table 37 related to the **isoplanatic angle** θ_0 and considering as threshold values the tertiles, we note, as a first, the evident bias between the two instruments. The ranges defining the tertiles are obviously different for the two instruments because of the presence of a bias also seen in the corresponding scattering plot (Fig.56). In both cases G is of the order of 40%, N of the order of 45% and B of the order of 14-15%. This certainly makes the comparison between simulations and observations more complicated. Looking at the model performances we discuss separately the case in which we assume, as a reference, the MASS and the DIMM. In both cases we consider the sample of 53 nights. In the first case (bottom-left), the good hit-rate of the model is G=34.7%, a little bit smaller than that obtained by different instruments (∼ 40%). The bad hit-rate of the model is very low (3%). It is however evident that the model predictions are concentrated in a thin range of values with respect to observations and this is due to the limitation of the model in reconstructing the correct amplitude of variability of the turbulence in the high part of the atmosphere. In other words the model reconstructs well the mean values but it does not achieve to reach the extreme values. If we look at the model vs. the DIMM (bottom-right) we found an unsatisfactory behavior of the model: whatever is the observation, the model provides values always in the same thin range of values. The model succeeds when the observations produce small values of θ_0 (< 2.25 arcsec) just because the DIMM observed values are visibly biased with respect to the MASS. In this case the bad hits from the model are as frequent as the good ones (B=33%). The comparison MNH versus DIMM is not really representative because, independently from the model performances, we think that the isoplanatic angle provided by the DIMM is very weakly reliable. MASS and GS measurements are more correlated also in terms of temporal evolution. It is known that the DIMM provides an approximated expression for θ_0. If we look at the model versus the MASS (bottom left) we observe that the model provides good estimates when the θ_0 corresponds to the average estimate but it fails in reconstructing the extremes values of θ_0. This matches with the discussion we did in Section 6.4-Fig.56. If we consider the sample of 89 nights (53 nights + 36 nights in Table 38), the model performances improve slightly but maintaining clearly the same limitation. For what concerns the isoplanatic angle we refer the reader to the discussion done in relation to Fig.41, Fig.66 and Fig.67.

As seen in the case of τ_0 if we calculate the hit-rate tables using other thresholds values, for example **1.5 arcsec and 3 arcsec**, results are much more performant for both instruments and model (Table 39 and Table 40). The good hit-rate percentage for the instruments is G=46.3% while the model has a G=61% in both cases (with respect to MASS and DIMM). If we consider the 85 nights (Table 40) the value of G is equal to 64%. The values of B is of the order of maximum 2% in both cases (instruments and model). Of course the limitation of the model in reconstructing the extreme values of θ_0 during the night remains. This is visible in the fact that are less frequent for the model the θ_0 larger than 3 arcsec and smaller than 1.5 arcsec with respect to what is observed. However the statistical performances of the model in terms of number of times in which the model hits the observed values is even better than that observed between the two instruments. Once again this shows that the model validity is really dependent on the question we are interesting on.

		MASS θ_0 (") - 44 nights		
		<1.50	[1.50,2.04]	>2.04
DIMM θ_0 (")	<1.50	4.73%	0.1%	0.0%
	[1.50,2.04]	14.3%	4.9%	2.05%
	>2.04	14.2%	28.7%	31.3%

G=40.5% N=45.2% B=14.2%

		DIMM θ_0 (") - 44 nights		
		<2.20	[2.20,2.91]	>2.91
MASS θ_0 (")	<2.20	31.4%	28.8%	14.1%
	[2.20,2.91]	1.5%	3.8%	14.5%
	>2.91	0.4%	0.7%	4.6%

G=39.9% N=45.5% B=14.6%

		MASS θ_0 (") - 48 nights		
		<1.51	[1.51,2.05]	>2.05
MNH θ_0 (")	<1.51	2.5%	1.9%	1.2%
	[1.51,2.05]	29.6%	29.1%	29.0%
	>2.05	1.1%	2.3%	3.1%

G=34.7% N=63.0% B=2.3%

		DIMM θ_0 (") - 49 nights		
		<2.25	[2.25,2.96]	>2.96
MNH θ_0 (")	<2.25	33.3%	33.3%	33.3%
	[2.25,2.96]	0.0%	0.0%	0.0%
	>2.96	0.0%	0.0%	0.0%

G=33.3% N=33.3% B=33.3%

Table 37: Hit rates table for the **isoplanatic angle** (method a). G is the percentage of good prediction, N is the percentage of medium prediction and B is the percentage of bad predictions. G = [1,1] + [2,2] + [3,3], N = [1,2] + [2,1] + [2,3] + [3,2] and B = [1,3] + [3,1], where [i,j] is the value at the ith row and the jth column. These values are obtained using the final validation sample of **53 nights**, with the exclusion of the nights for which an instrument gave no measurements (Table 67 - 48 nights only for the MNH/MASS comparison, 49 nights for the MNH/DIMM comparison and 44 nights for the MASS/DIMM comparison).

		DIMM θ_0 (") - 85 nights		
		< 2.06	[2.06,2.79]	> 2.8
MNH θ_0 (")	<2.06	33.0%	32.0%	29.4%
	[2.06,2.79]	0.4%	1.3%	4.0%
	>2.79	0.0%	0.0%	0.0%

G=34.3% N=36.3% B=29.4%

Table 38: Same as Table 37 bottom-right but the sample is **85 nights** (method a).

		MASS θ_0 (") - 44 nights		
		<1.5	[1.5,3]	>3
DIMM θ_0 (")	<1.5	4.7%	0.1%	0.0%
	[1.5,3]	26.4%	38.0%	1.2%
	>3	2.0%	23.9%	3.6%

G=46.3% N=51.7% B=2.0%

		DIMM θ_0 (") - 44 nights		
		<1.5	[1.5,3]	>3
MASS θ_0 (")	<1.5	4.7%	26.0%	2.0%
	[1.5,3]	0.1%	38.0%	23.9%
	>3	0.0%	1.2%	3.6%

G=46.3% N=51.7% B=2.0%

		MASS θ_0 (") - 48 nights		
		<1.5	[1.5,3]	>3
MNH θ_0 (")	<1.5	1.5%	1.6%	0.0%
	[1.5,3]	31.2%	59.8%	5.8%
	>3	0.0%	0.0%	0.0%

G=61.3% N=38.6% B=0.0%

		DIMM θ_0 (") - 49 nights		
		<1.5	[1.5,3]	>3
MNH θ_0 (")	<1.5	0.0%	3.7%	0.1%
	[1.5,3]	4.3%	60.9%	31.4%
	>3	0.0%	0.0%	0.0%

G=61.0% N=39.9% B=0.1%

Table 39: As Table 37 but with thresholds: **1.5 arcsec and 3 arcsec** (method b).

		DIMM θ_0 (”) - 85 nights		
		< 1.5	[1.5,3]	> 3
MNH θ_0 (”)	<1.5	0.1%	2.3%	0.0%
	[1.5,3]	8.0%	64.3%	25.2%
	>3	0.0%	0.0%	0.0%

G=64.4% N=35.5% B=0.0%

Table 40: Same as Table 38 but with thresholds: **1.5 arcsec and 3 arcsec** (method b).

TOTAL SEEING

Looking at Table 41-bottom and considering as threshold values the tertiles we note that the good hit-rate of the model for the **total seeing** ε (for the validation sample of 53 nights) is of the order of 37%, the medium hit-rate equal to 42% and the bad hit-rate equal to 21%. These values remains almost the same if the sample of individual nights increases up to 89 nights. If we compare the hit-rate of the model vs. measurements with measurements vs. measurements on the calibration sample of 20 nights (Table 41-top-right) results are better and comparable to what obtained between measurements from different instruments (Table 41-top-left). On this smaller sample the good hit-rate of success of the model is almost twice ($\sim$ 50%) and only slightly inferior with respect to the good hit-rate of success of measurements (60%). The bad hit-rate in both cases is very small (5% and 7%) while the medium hit-rate is 42% and 34% for the model and the measurements respectively.
It seems that, going from the calibration sample to the validation sample of 89 nights, the ability of the model in detecting the good seeing decreases of a factor 1.3 (if we look at the G value) and 2.6 (if we look at the B value). However, if we consider as threshold values **1 arcsec and 2 arcsec** the model performances and instruments are much better (Table 42). These two values are taken according to the following logic: typically an AO system works with its best performances for seeing values smaller than 1 arcsec. Beyond 2 arcsec an AO system definitely does not work at all. The intermedium range of [1,2] arcsec can still be interesting because an AO system can still run even with more modest performances, for example by reducing the number of modes[17]. We note that the value of 1 arcsec corresponds also to the median value of the observed DIMM measurements related to the sample of 89 nights considered in this study[18]. The value of G for the GS and DIMM case (20 nights) is 66% while G=59.8% for the GS and model case (20 nights) and G=56.9% for the DIMM and model case (20 nights). On the validation sample of 53 nights (DIMM vs. model) we have a G=48% while for the 89 nights we have G=49%. Also the values of B are much better (i.e. smaller) than for the tertiles case for both instrument and model. B is maximum equal to 1.6%. On the validation sample the model performances decreases with respect to the calibration sample of a factor 1.2 instead of 1.3 (for the tertiles cases) for G. B increases a little in percentage but it remains within 1.6% and it is therefore negligible and definitely better than the tertile case (B=18.2%). We finally verified that, if we integrate the turbulence simulated by the model starting from 20 m a.g.l. (Table 43) we obtain little improvements in G and B values[19]. The good hit-rate G=50% for model and DIMM (53 nights) but this is a small effect compared to the difference obtained if we consider the tertiles or some fixed threshold values.

Of course alternative categorizations might be envisaged with a different sub-division or with a more refined sub-division. We verified that, using [1,1.5] arcsec instead of [1,2] arcsec, the model performances and the instruments matching both decrease. This is probably due to the fact that the extension of the sub-division is smaller than the previous one and the scattering plots (Fig.37 and Fig.54) have a tendency in increasing the discrepancy for large seeing values. It is therefore possible that also increasing the number of the sub-divisions would lead to an equivalent decreasing of the model performances and instruments matching. In other words, it is preferable to maintain a relatively broad division, at least in a first baseline operational approach. It would be useful, in perspective, to improve the seeing model performances and the seeing matching between different instruments also in the range [1,1.5].

As discussed previously, we think that part of the decreasing model performances on the validation sample is due to the fact the calibration has been done on a number of nights concentrated on summer time (see discussion done in Section 6.4). This means that the factor 3 (for the tertiles case) and 1.2 (for the fixed thresholds case) could be reduced simply with a calibration done in summer and in winter (thesis proved in Section 6.4). This can not be considered an intrinsic limitation of the model but preferably a limitation of the contextual conditions. Moreover it is worth remembering that, unfortunately, we have no similar estimates for different instruments calculated on the same sample of nights. We do not know, at present, how the hit-rate between GS

---

[17] In reality a maximum value of 1.5 should be more appropriated for an AO system but the model performances and the matching of different instruments is worse under this condition.

[18] Monthly averages of seeing values in the period 2010-2011 (*http://www.eso.org/gen-fac/pubs/astclim/paranal/seeing/singstory.html*) indicate a mean of the order of 1-1.1 arcsec at Cerro Paranal. This value is in agreement with our estimate.

[19] This test is done to verify the role played by the inferior limit of the turbulence integration and check if any statistical improvements might be obtained on the global model performances.

and DIMM changes with increasing the number of nights. It might be that the value of G in the measurements case decreases too. The relevant comment in this respect is that for the seeing and the wavefront coherence time, that are mainly affected by what happens in the first meters above the ground, it would be better to perform in the future some tests of model performances integrating the turbulence starting from a higher height above the ground so to assure us to consider a turbulence more homogenous along the horizontal direction. This condition should guarantee also a better correlation between measurements and a more suitable goal for the model too. The ideal solution should be to place a GS at the focus of the UT. We conclude therefore that, for the total seeing hit-rate of the model we could improve G and reduce B simply by doing the calibration on a richer sample of simultaneous GS and DIMM measurements. However, at the same time, it should be fundamental to re-conceive the concept of total seeing integrating from a higher height above the ground.

As a general commentary we intend in the future to try to transform the percentage of N values in G values. We also think that it might be useful to conceive hits-rate table assuming a simpler division in bad and good hits-rates (2x2 tables) basically fixing, for each parameter, a threshold X so that we can disentangle values $<$ or $>$ than X. From the point of view of the scheduling it is important to define some typical thresholds values that are the most significative from an operational point of view. For example: it is much more important that the model can distinguish the very good and the very bad seeing conditions than to be able to distinguish between two contiguous values of medium seeing i.e. 1 and 1.2 arcsec.

		GS ε (”) - 20 nights		
		<0.97	[0.97,1.23]	>1.23
DIMM ε (”)	<0.97	20.1%	10.8%	0.8%
	[0.97,1.23]	8.5%	8.8%	4.2%
	>1.23	4.7%	13.8%	28.2%

G=57.1% N=37.3% B=5.5%

		GS ε (”) - 20 nights		
		<0.96	[0.96,1.23]	>1.23
MNH ε (”)	<0.96	24.4%	22.7%	3.9%
	[0.96,1.23]	6.0%	6.3%	9.0%
	>1.23	3.0%	4.3%	20.3%

G=51.0% N=42.0% B=6.9%

		DIMM ε (”) - 20 nights		
		< 1.5	[1.5,3]	> 3
MNH ε (”)	<0.98	26.9%	18.5%	5.8%
	[0.98,1.41]	5.8%	11.5%	13.5%
	>1.41	0.7%	3.3%	14.0%

G=52.3% N=41.2% B=6.5%

		DIMM ε (”) - 53 nights		
		<0.86	[0.86,1.14]	>1.14
MNH ε (”)	<0.86	7.9%	7.1%	7.4%
	[0.86,1.14]	12.0%	11.1%	8.1%
	>1.13	13.4%	14.9%	17.7%

G=37.0% N=42.2% B=20.8%

		DIMM ε (”) - 89 nights		
		<0.86	[0.86,1.17]	>1.17
MNH ε (”)	<0.86	9.5%	8.1%	7.0%
	[0.86,1.17]	12.7%	11.1%	7.6%
	>1.17	11.1%	14.1%	18.6%

G=39.2% N=42.6% B=18.2%

Table 41: Hit rates table for the **total seeing** (method a). G is the percentage of good prediction, N is the percentage of medium prediction and B is the percentage of bad predictions. G = [1,1] + [2,2] + [3,3], N = [1,2] + [2,1] + [2,3] + [3,2] and B = [1,3] + [3,1], where [i,j] is the value at the ith row and the jth column. On the top-left: DIMM vs. GS for the calibration sample of 20 nights; top-right: MNH vs. DIMM for the calibration sample; bottom-left: MNH vs. DIMM for the final validation sample of 53 nights; bottom-right: MNH vs. DIMM for the final validation sample of 89 nights (53 nights of 2010-2011 years + 36 nights of the 2007 year).

		GS ε (”) - 20 nights		
		<1	[1,2]	>2
DIMM ε (”)	<1	23.4%	10.8%	0.%
	[1,2]	14.4%	40.7%	2.0%
	>2	0.35%	6.32%	2.0%

G=66.1% N=33.5% B=0.3%

		GS ε (”) - 20 nights		
		<1	[1,2]	>2
MNH ε (”)	<1	29.7%	24.6%	0.5%
	[1,2]	8.7%	30.1%	3.5%
	>2	0.0%	3.2%	0.0%

G=59.8% N=39.7% B=0.5%

		DIMM ε (”) - 20 nights		
		< 1	[1,2]	> 2
MNH ε (”)	<1	28.2 %	24.7%	0.2%
	[1,2]	6.5%	28.0%	8.9%
	>2	0.0%	2.9%	0.6%

G=56.9% N=42.9% B=0.2%

		DIMM ε (”) - 53 nights		
		<1	[1,2]	>2
MNH ε (”)	<1	23.0%	17.5%	0.5%
	[1,2]	29.0%	25.2%	2.8%
	>2	0.9%	2.8%	1.1%

G=48.0% N=50.8% B=1.3%

		DIMM ε (”) - 89 nights		
		<1	[1,2]	>2
MNH ε (”)	<1	22.8%	17.1%	0.6%
	[1,2]	26.5%	25.1%	2.9%
	>2	0.8%	2.4%	0.9%

G=49.0% N=49.4% B=1.6%

Table 42: As Table 41 but with thresholds: **1 arcsec and 2 arcsec** (method b).

		GS ε (”) - 20 nights		
		<1	[1,2]	>2
DIMM ε (”)	<1	23.4%	10.8%	0.%
	[1,2]	14.4%	40.7%	2.0%
	>2	0.35%	6.32%	2.0%

G=66.1% N=33.5% B=0.3%

		GS ε (”) - 20 nights		
		<1	[1,2]	>2
MNH ε (”)	<1	31.7%	26.2%	0.5%
	[1,2]	6.6%	28.6%	3.5%
	>2	0.0%	2.9%	0.0%

G=60.4% N=39.1% B=0.5%

		DIMM ε (”) - 20 nights		
		< 1	[1,2]	> 2
MNH ε (”)	<1	28.2 %	24.7%	0.2%
	[1,2]	6.5%	28.0%	8.9%
	>2	0.0%	2.9%	0.6%

G=56.9% N=42.9% B=0.2%

		DIMM ε (”) - 53 nights		
		<1	[1,2]	>2
MNH ε (”)	<1	28.1%	21.3%	0.6%
	[1,2]	23.1%	21.6%	2.8%
	>2	0.7%	1.4%	0.3%

G=50.1% N=48.6% B=1.3%

		DIMM ε (”) - 89 nights		
		<1	[1,2]	>2
MNH ε (”)	<1	27.2%	21.0%	0.7%
	[1,2]	22.2%	21.7%	3.0%
	>2	0.8%	2.4%	1.0%

G=49.0% N=49.4% B=1.5%

Table 43: As Table 41 but with thresholds: **1 arcsec and 2 arcsec** (method b) and integrating the turbulence of the model from 20 m a.g.l.

7.2.2 Atmospherical surface parameters

The analysis performed in Section 7.2.1 on the astroclimatic parameters is done here on the meteorological surface parameters at Cerro Paranal and Cerro Armazones. For the wind speed and the temperature, the same 3x3 hit rates tables of the previous section are presented and analyzed. For the temperature we used the same hit-rates table used for the astroclimatic parameters. However, since we have identified a statistical negative bias in the forecast of the surface wind speed (the wind speed provided by the model is statistically underestimated), it was decided to evaluate the predictive power of the model using a forecasted wind speed corrected by the multiplicative bias B_{mult} computed as:

$$B_{mult}(k) = \frac{\overline{WS_{mod}}(k)}{\overline{WS_{obs}}(k)} \tag{21}$$

with $\overline{WS_{mod}}(k)$ the average of the forecasted wind speed over the whole sample of the 20 nights (Table 65), and $\overline{WS_{obs}}(k)$ the average of the observed wind speed from the same sample, at the level k (10 m and 30 m for Cerro Paranal and 2 m, 11 m, 20 m and 28 m for Cerro Armazones). Table 44 reports all the $B_{mult}(k)$ for each level k. The choice of the multiplicative bias is justified by the fact that we want the highest values of the wind speed to be more corrected than the lowest values. The underestimation observed in the model is indeed more evident when the observed wind speed is important and almost absent when the wind speed is weak. A multiplicative bias is a sort of weighted correction factor. On the contrary, an additive correction factor would add the same correction to every wind speed, whatever its intensity, thing that would not answer to our necessity.
After correction, the wind speed $WS_{mod}(k)$ at each the level k used to calculate the hit-rates (Table 47 and Table 48) is:

$$WS_{mod,corr}(k) = WS_{mod,ori}(k)/B_{mult}(k) \tag{22}$$

with $WS_{mod,ori}$ and $WS_{mod,corr}$ the original and corrected wind speed, respectively.

Concerning the wind direction hit rates tables, we decided to categorize the sample in four interval of 90° each, in order to take into account the circularity of the wind direction values (between 0° and 360°). More over, this permits us to evaluate the ability of the model in determining from which quadrant the wind is coming from. One should note that the computation of the G, N and B percentages are different for this case. We now define:

	10 m	30 m
C. Paranal ΔX = 500 m	0.54	0.80
C. Paranal ΔX = 100 m	0.81	0.92

	2 m	11 m	20 m	28 m
C. Armazones ΔX = 500 m	0.28	0.44	0.60	0.64
C. Armazones ΔX = 100 m	0.70	0.77	0.89	0.93

Table 44: Wind speed multiplicative biases $B_{mult}(k)$ (from Eq. 21).

$$G = \sum_{i=1}^{4}[i,i] \tag{23}$$

$$B = [1,3] + [2,4] + [3,1] + [4,2] \tag{24}$$

$$N = [1,2] + [1,4] + [2,1] + [2,3] + [3,2] + [3,4] + [4,1] + [4,3] \tag{25}$$

Concerning the absolute temperature (Table 45 and Table 46), the accuracy of the model (characterized by G) is outstanding. The worst G value is obtained at 2 m at Cerro Armazones (51.3%), but all the others values (at both sites and all levels) oscillate between 74.7% and 79%. More over, and almost as important, the bad predictions (B values) are almost inexistent (the highest

B is 3.9% at 2 m at Cerro Armazones). Concerning the temperature at 2 m above Cerro Armazones we highlight that we used measurements from the tower sensor because the one from the AWS was not reliable (T.Travouillon, private communication).

		OBS at 2 m T (°C)		
		<11.98	[11.98,13.78]	>13.78
M-NH T (°C)	<11.98	29.7%	6.9%	0.0%
	[11.98,13.78]	3.6%	20.6%	8.9%
	>13.78	0.0%	5.8%	24.4%

G=74.7% N=25.2% B=0.0%

CERRO PARANAL

		OBS at 30 m T (°C)		
		<12.33	[12.33,14.06]	>14.06
M-NH T (°C)	<12.33	31.9%	7.5%	0.0%
	[12.33,14.06]	1.4%	21.9%	9.7%
	>14.06	0.0%	3.9%	23.6%

G=77.4% N=22.5% B=0.0%

Table 45: Hit rates table for the **absolute temperature** during the night at **Cerro Paranal**, obtained with the sample of 20 nights (Table 65).

		OBS at 2 m T (°C)		
		<7.01	[7.01,9.00]	>9.00
M-NH T (°C)	<7.01	16.1%	0.0%	0.0%
	[7.01,9.00]	13.3%	1.9%	0.0%
	>9.00	3.9%	31.4%	33.3%

G=51.3% N=44.7% B=3.9%

CERRO ARMAZONES

		OBS at 11 m T (°C)		
		<8.54	[8.54,10.63]	>10.63
M-NH T (°C)	<8.54	27.8%	0.8%	0.0%
	[8.54,10.63]	5.3%	18.1%	0.3%
	>10.63	0.3%	14.4%	33.1%

G=79.0% N=20.8% B=0.3%

		OBS at 20 m T (°C)		
		<9.22	[9.22,11.41]	>11.41
M-NH T (°C)	<9.22	28.9%	4.4%	0.0%
	[9.22,11.41]	4.4%	21.4%	4.7%
	>11.41	0.0%	7.5%	28.6%

G=78.9% N=21.0% B=0.0%

		OBS at 28 m T (°C)		
		<9.29	[9.29,11.41]	>11.41
M-NH T (°C)	<9.29	28.1%	5.3%	0.0%
	[9.29,11.41]	5.3%	18.9%	3.6%
	>11.41	0.0%	9.2%	29.7%

G=76.7% N=23.4% B=0.0%

Table 46: Hit rates table for the **absolute temperature** during the night at **Cerro Armazones**, obtained with the sample of 20 nights (Table 65).

		OBS at 10 m WS (m/s)		
		<2.85	[2.85,5.42]	>5.42
MNH WS (m/s)	<2.85	10.6%	8.1%	1.1%
	[2.85,5.42]	21.4%	21.1%	8.6%
	>5.42	1.4%	4.2%	23.6%

G=55.3% N=42.3% B=2.5%

ΔX = 500 m
CERRO PARANAL

		OBS at 30 m WS (m/s)		
		<2.83	[2.83,6.09]	>6.09
MNH WS (m/s)	<2.83	9.4%	9.7%	0.0%
	[2.83,6.09]	21.4%	20.0%	8.6%
	>6.09	2.5%	3.6%	24.7%

G=54.1% N=43.3% B=2.5%

Table 47: Hit rates table for the **wind speed** (corrected from the multiplicative bias) during the night at **Cerro Paranal**, obtained with the sample of 20 nights (Table 65). **ΔX = 500 m**.

		OBS at 2 m WS (m/s)		
		<2.88	[2.88,5.49]	>5.49
MNH WS (m/s)	<2.88	16.1%	10.0%	0.3%
	[2.88,5.49]	11.4%	11.4%	7.8%
	>5.49	5.8%	11.9%	25.3%

G=52.8% N=41.1% B=6.1%

ΔX = 500 m
CERRO ARMAZONES

		OBS at 11 m WS (m/s)		
		<3.54	[3.54,6.48]	>6.48
MNH WS (m/s)	<3.54	10.8%	7.5%	6.4%
	[3.54,6.48]	13.3%	11.9%	13.3%
	>6.48	9.2%	13.9%	13.6%

G=36.3% N=48.0% B=15.6%

		OBS at 20 m WS (m/s)		
		<3.46	[3.46,6.44]	>6.44
MNH WS (m/s)	<3.46	18.6%	10.0%	10.8%
	[3.46,6.44]	5.6%	6.4%	6.9%
	>6.44	9.2%	16.9%	15.6%

G=40.6% N=39.4% B=20%

		OBS at 28 m WS (m/s)		
		<3.44	[3.44,6.46]	>6.46
MNH WS (m/s)	<3.44	19.2%	11.4%	10.6%
	[3.44,6.46]	4.7%	5.6%	5.8%
	>6.46	9.4%	16.4%	16.9%

G=41.7% N=38.3% B=20.0%

Table 48: Hit rates table for the **wind speed** (corrected from the multiplicative bias) during the night at **Cerro Armazones**, obtained with the sample of 20 nights (Table 65). **ΔX = 500 m**.

		OBS at 10 m WS (m/s)		
		<2.85	[2.85,5.42]	>5.42
MNH WS (m/s)	<2.85	9.4%	11.9%	1.4%
	[2.85,5.42]	21.1%	17.5%	11.1%
	>5.42	2.8%	3.9%	20.8%

G=47.7% N=48.0% B=4.2%

ΔX = 100 m **CERRO PARANAL**		**OBS at 30 m WS (m/s)**		
		<2.83	[2.83,6.09]	>6.09
MNH WS (m/s)	<2.83	10.0%	11.9%	0.8%
	[2.83,6.09]	19.7%	18.1%	10.6%
	>6.09	3.6%	3.3%	21.9%

G=50.0% N=45.5% B=4.4%

Table 49: Hit rates table for the **wind speed** (corrected from the multiplicative bias) during the night at **Cerro Paranal**, obtained with the sample of 20 nights (Table 65). $\mathbf{\Delta X = 100\ m}$.

		OBS at 2 m WS (m/s)		
		<2.88	[2.88,5.49]	>5.49
MNH WS (m/s)	<2.88	18.3%	10.6%	2.2%
	[2.88,5.49]	11.9%	16.4%	7.8%
	>5.49	3.1%	6.4%	23.3%

G=58.0% N=36.7% B=5.3%

ΔX = 100 m **CERRO ARMAZONES**		**OBS at 11 m WS (m/s)**		
		<3.54	[3.54,6.48]	>6.48
MNH WS (m/s)	<3.54	20.3%	9.2%	2.5%
	[3.54,6.48]	10.3%	18.1%	6.9%
	>6.48	2.8%	6.1%	23.9%

G=62.2% N=32.5% B=5.3%

		OBS at 20 m WS (m/s)		
		<2.88	[2.88,5.49]	>5.49
MNH WS (m/s)	<3.46	19.4%	9.2%	2.5%
	[3.46,6.44]	11.4%	17.2%	6.9%
	>6.44	2.5%	6.9%	23.9%

G=60.5% N=34.4% B=5.0%

		OBS at 28 m WS (m/s)		
		<3.54	[3.54,6.48]	>6.48
MNH WS (m/s)	<3.44	18.3%	10.6%	2.5%
	[3.44,6.46]	12.5%	15.6%	7.8%
	>6.46	2.5%	7.2%	23.1%

G=57.0% N=38.1% B= 5.0%

Table 50: Hit rates table for the **wind speed** (corrected from the multiplicative bias) during the night at **Cerro Armazones**, obtained with the sample of 20 nights (Table 65). $\mathbf{\Delta X = 100\ m}$.

CERRO PARANAL		**OBS at 10 m WD (°)**			
		[0,90]	[90,180]	[180,270]	[270,360]
MNH WD (°)	[0,90]	31.4%	6.9%	3.1%	9.4%
	[90,180]	2.5%	20.3%	5.3%	1.9%
	[180,270]	1.1%	0.0%	0.0%	0.0%
	[270,360]	4.8%	1.1%	3.1%	9.2%

G=60.9% N=31.9% B=7.2%

Table 51: Hit rates table for the **wind direction** during the night at **Cerro Paranal**, at 10 m, obtained with the sample of 20 nights (Table 65).

CERRO PARANAL		**OBS at 30 m WD (°)**			
		[0,90]	[90,180]	[180,270]	[270,360]
MNH WD (°)	[0,90]	33.3%	7.5%	2.2%	8.3%
	[90,180]	2.5%	23.6%	0.3%	1.7%
	[180,270]	0.3%	0.3%	0.3%	0.0%
	[270,360]	6.1%	0.8%	3.1%	9.7%

G=66.9% N=28.1% B=5.0%

Table 52: Hit rates table for the **wind direction** during the night at **Cerro Paranal**, at 30 m, obtained with the sample of 20 nights (Table 65).

CERRO ARMAZONES		**OBS at 2 m WD (°)**			
		[0,90]	[90,180]	[180,270]	[270,360]
MNH WD (°)	[0,90]	27.8%	1.7%	2.8%	20.3%
	[90,180]	1.7%	20.3%	1.7%	0.3%
	[180,270]	0.6%	0.3%	2.2%	1.9%
	[270,360]	1.9%	0.3%	4.4%	11.9%

G=62.2% N=33.9% B=3.9%

Table 53: Hit rates table for the **wind direction** during the night at **Cerro Armazones**, at 2 m, obtained with the sample of 20 nights (Table 65).

CERRO ARMAZONES		**OBS at 11 m WD (°)**			
		[0,90]	[90,180]	[180,270]	[270,360]
MNH WD (°)	[0,90]	21.4%	1.7%	1.4%	21.9%
	[90,180]	1.7%	20.8%	1.1%	0.3%
	[180,270]	0.3%	0.0%	1.9%	1.4%
	[270,360]	3.3%	0.8%	6.7%	15.3%

G=59.4% N=37.8% B=2.8%

Table 54: Hit rates table for the **wind direction** during the night at **Cerro Armazones**, at 11 m, obtained with the sample of 20 nights (Table 65).

CERRO ARMAZONES		**OBS at 20 m WD (°)**			
		[0,90]	[90,180]	[180,270]	[270,360]
MNH WD (°)	[0,90]	21.9%	1.4%	0.8%	15.8%
	[90,180]	1.7%	20.3%	1.1%	0.3%
	[180,270]	0.3%	0.3%	2.2%	1.9%
	[270,360]	6.7%	0.6%	6.9%	17.8%

G=62.2% N=35.8% B=1.9%

Table 55: Hit rates table for the **wind direction** during the night at **Cerro Armazones**, at 20 m, obtained with the sample of 20 nights (Table 65).

CERRO ARMAZONES		**OBS at 28 m WD (°)**			
		[0,90]	[90,180]	[180,270]	[270,360]
MNH WD (°)	[0,90]	24.7%	1.4%	0.3%	11.4%
	[90,180]	1.4%	18.9%	2.2%	0.6%
	[180,270]	0.3%	0.3%	2.5%	2.2%
	[270,360]	8.3%	0.6%	5.8%	19.2%

G=65.3% N=33.1% B=1.7%

Table 56: Hit rates table for the **wind direction** during the night at **Cerro Armazones**, at 28 m, obtained with the sample of 20 nights (Table 65).

Concerning the wind direction (Table 51-Table 56), the hit rates tables put into evidence the very good performances of the model. At every level, and for both sites, the good predictions occur between 59.4% and 66.9% of the times, and the bad predictions are in all cases contained within a few percents (always inferior to 7.2%).

Looking at the bias-corrected wind speed (Table 47-Table 50), the model performances are also very satisfactory at Paranal with a ΔX = 500 m configuration and at Armazones with the ΔX = 100 m configuration. At Cerro Paranal, it is worth noting that the ΔX = 500 m configuration has a similar or even better accuracy than the ΔX = 100 m configuration (G values of 55.3% and 54.1% against 47.7% and 50%, at 10 m and 30 m, respectively) after the correction with the multiplicative term (Table 47 and Table 49). This is particularly interesting because it would permit to reach the suitable performances using the model in a less expensive configuration[20]. On the contrary, at Cerro Armazones, the ΔX = 100 m configuration performs better. The ΔX = 100 m configuration permits, in any case, to obtain very satisfactory performances above both sites.. G oscillates between 57% and 62.2% whereas it oscillates between 36.3% and 52.8% with the ΔX = 500 m configuration (Table 48 and Table 50). These differences in terms of performances between the two configuration may be explained by the limited sample of 20 nights. To definitely confirm the very promising performances the next step should be to extend the same calculation on a richer statistical sample of the nights. **Also it is worth noting that the bad predictions (B) occur only in a limited percent of the times with the ΔX = 100 m configuration, where B is always inferior to 5.3%. Using the ΔX = 500 m configuration at Armazones B can reach values also of the order of 15-20%**.

At this point we can say that looking at the 3x3 and 4x4 category tables, the models appears to be very reliable for what concerns the prediction of the absolute temperature and the wind direction. However the overall accuracy of the wind speed prediction, though acceptable, is slightly smaller than for the wind direction and the temperature. The 3x3 category table for the wind speed, as is, gives valuable insight on how the model is accurate, but doesn't allow to answer more specific (in some cases even more important) questions. For example, one may want to know how good is the model in predicting strong winds. To answer such a specific question, we can use a 2x2 contingency table, where the limit (threshold) on the wind speed is defined by the user. **For the purpose of our demonstration, let's use V_{lim} =6 m·s^{-1} as a limit for the "strong" winds**. This limit could be redefined later in agreement with ESO, for example the wind intensity that does not permit to an AO system to be run or the wind intensity that does not permit to the secondary mirror to handle vibrations harmlessly. **Is the model able to well detect wind stronger than V_{lim}?** Tables 57-59 answer such a question. In the case of the 2x2 contingency table, G and B are defined as in the case of the 3x3 tables. Let's define in the 2x2 case the **probability of detection POD** as:

$$G_{2\times 2} = \sum_{i=1}^{2} [i, i] \tag{26}$$

and

$$POD = \frac{[1, 1]}{[1, 1] + [2, 1]} \tag{27}$$

The probability of detection can be seen as the percentage of times in which the model succeeded in forecasting the observed strong winds. **Looking at Tables 57-59 it is evident that the model is capable of detecting wind speed stronger than V_{lim} =6 m·s^{-1} with a good accuracy (G), with the ΔX = 500 m configuration at Paranal and ΔX = 100 m configuration at Armazones. At 10 m at G = 86.7%, and at 30 m, G=84.7%. The probability of detection are high and equal to 69.7% (at 10 m) and 74.2% (at 30 m). At Cerro Armazones we need to use ΔX = 100 m configuration to obtain equivalent satisfactory results: G is between 77.5% and 81.7%, and the probability of detection is of the order of 69% at all levels**.

---

[20] The term "expensive configuration" is used to refer to a configuration that requires more CPU time

		OBS at 10 m WS	
		$> V_{lim}$	$< V_{lim}$
MNH	$> V_{lim}$	19.2%	5.0%
	$< V_{lim}$	8.3%	67.5%

G=86.7% B=13.3%
POD = 69.7%

ΔX = 500 m
CERRO PARANAL

		OBS at 30 m WS	
		$> V_{lim}$	$< V_{lim}$
MNH	$> V_{lim}$	24.7%	6.7%
	$< V_{lim}$	8.6%	60.0%

G=84.7% B=15.3%
POD = 74.2%

Table 57: Hit rates table for the **wind speed** (corrected from the multiplicative bias) during the night at **Cerro Paranal**, obtained with the sample of 20 nights (Table 65). The Meso-NH configuration is the ΔX = 500 m configuration. Here $V_{lim} = 6m \cdot s^{-1}$.

		OBS at 2 m WS	
		$> V_{lim}$	$< V_{lim}$
MNH	$> V_{lim}$	20.8%	9.2%
	$< V_{lim}$	9.2%	60.8%

G=81.7% B=18.3%
POD = 69.4%

ΔX = 100 m
CERRO ARMAZONES

		OBS at 11 m WS	
		$> V_{lim}$	$< V_{lim}$
MNH	$> V_{lim}$	25.6%	10.8%
	$< V_{lim}$	11.7%	51.9%

G=77.5% B=22.5%
POD = 68.7%

Table 58: Hit rates table for the **wind speed** (corrected from the multiplicative bias) during the night at **Cerro Armazones** at 2 m and at 11 m, obtained with the sample of 20 nights (Table 65). The Meso-NH configuration is the ΔX = 100 m configuration.

		OBS at 20 m WS	
		$> V_{lim}$	$< V_{lim}$
MNH	$> V_{lim}$	25.6%	10.8%
	$< V_{lim}$	11.4%	52.2%

G=77.8% B=22.2%
POD = 69.2%

ΔX = 100 m
CERRO ARMAZONES

		OBS at 28 m WS	
		$> V_{lim}$	$< V_{lim}$
MNH	$> V_{lim}$	25.0%	11.1%
	$< V_{lim}$	11.4%	52.5%

G=77.5% B=22.5%
POD = 68.7%

Table 59: Hit rates table for the **wind speed** (corrected from the multiplicative bias) during the night at **Cerro Armazones** at 20 m and at 28 m, obtained with the sample of 20 nights (Table 65). The Meso-NH configuration is the ΔX = 100 m configuration.

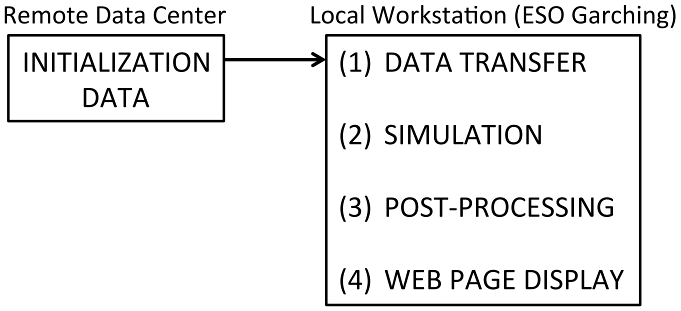
8 Operational forecast system overview

In the Introduction a simple toy model describing how an OT forecast works has been presented (Section 1 Fig. 2). Besides, a general descriptions of the ingredients necessary to perform the OT forecast and a complete sequence of processes necessary to perform the OT forecast has been included. From an operational point of view, the mandatories steps to perform a real-time OT forecasts at an ESO site are summarized in the scheme of Fig. 58.

A soon as the GCM (in our case, the ECMWF) makes available the meteorological files coming from its real-time global forecasts (T+12h, T+18h, T+24h, T+36h, T+48h, etc. forecasts), these files are sent in a remote data center (belonging to the ECMWF or an intermediate meteo service from which the initialization data are bought). The contractor provides, at a fixed and precise daily frequency, the data on this remote data center.

Automatic daily procedures (using for example bash scripts and crontab) will then copy these files in the local workstation (or cluster, or super-computer...), for example at ESO-Garching, where the Meso-NH model is installed and ready to perform the OT forecast (**step (1)** of Fig. 2) Immediately after the end of the transfers, the preparation of the simulation can begin (it consists in creating the initialization and coupling files for the given night). Once the files created, the Meso-NH simulation itself (the OT forecast) can start (this is **step (2)** of Fig. 2). At the end of the simulation, the Meso-NH outputs are post-processed (diagnostics computed using Fortran and idl, **step (3)** of Fig. 2) in order to generate exploitable figures including predicted parameters to be displayed on a dedicated web-page freely accessible by astronomers, instruments scientists, etc....(**step (4)** of Fig. 2).

All these steps are part of an automated process that we have to implement through bash scripts and crontab on a Linux machine.



(1) Copy from the remote data center of the forecast files from the GCM used for initialization and coupling;
(2) Preparation of the initialization and coupling Meso-NH files (using the forecast files from step (1)), and Meso-NH simulation;
(3) Post-processing of Meso-NH outputs and preparation of exploitable figures;
(4) Display on a web-page of the post-processed figures of step (3).

Figure 58: Operational forecast system overview.

9 Conclusions and perspectives

We summarize here the most important conclusions we achieved in this feasibility study.

ATMOSPHERIC PARAMETERS: VERTICAL STRATIFICATION

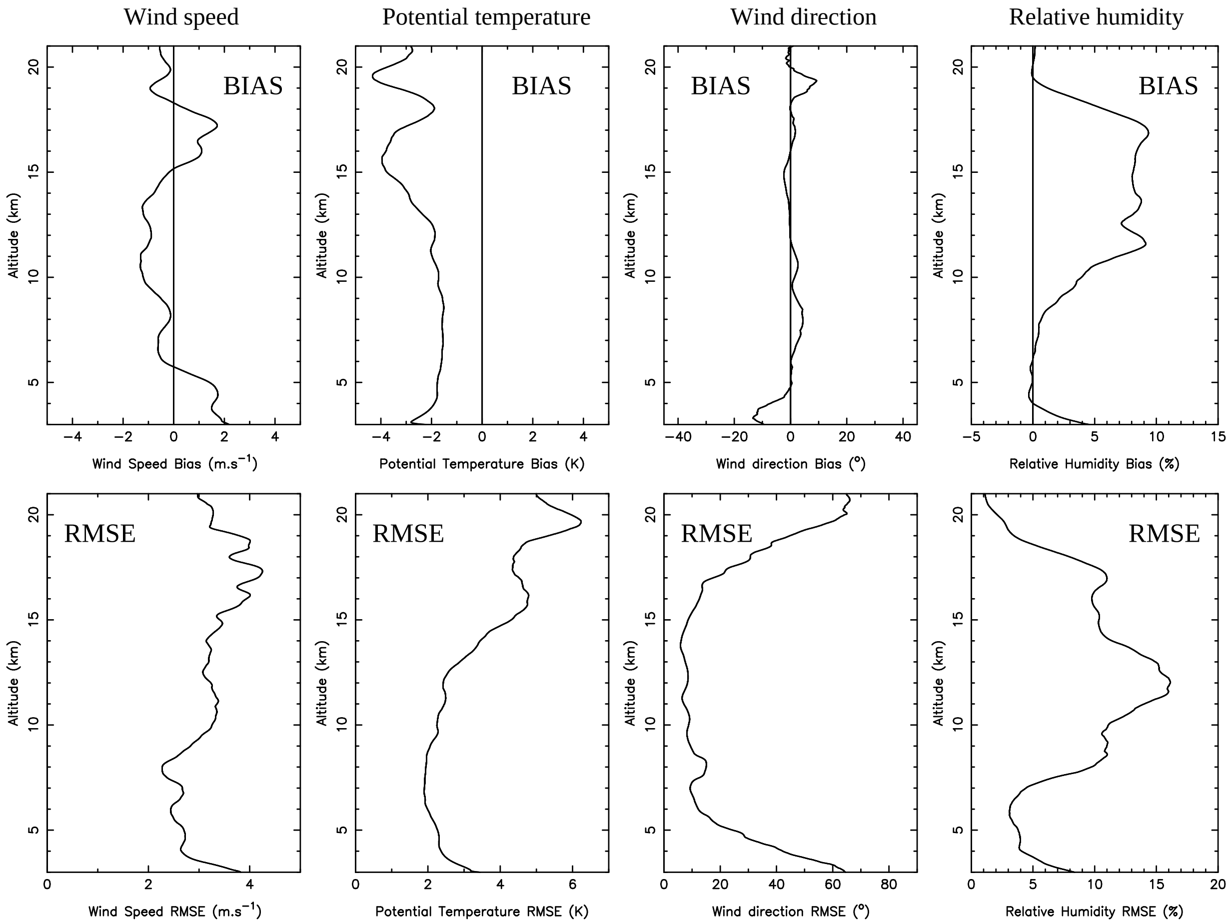


Figure 59: **Bias** and **RMSE** (Meso-NH - Observations) of **wind speed**, **potential temperature**, **wind direction** and **relative humidity**. Data and model profiles interpolated on a 5 m-vertical grid; then bias and RMSE are computed for each interpolated levels; finally a moving average over 1 km is applied on the resulting profiles.

- For the vertical stratification of the atmospheric parameters (potential temperature, wind speed and direction and relative humidity) over a column of 20 km we proved that the Meso-NH model, used in the grid-nesting configuration with three domains having a highest horizontal resolution of 500 m, provides excellent results. Fig. 59 indicates that the **RMSE** is within [2-4] m/s for the wind speed; [10-20] (°) in the 5-15 km a .g. l. and [10-50] (°) below 5 km a. g. l. and above 15 km for the wind direction; [2-4] K in the [0-15] km a. g. l. range and [4-6] K for h $>$ 15 km for the potential temperature. The relative humidity shows a RMSE within 10 % in the [0-10] km a. g. l. range and 15% at the jet stream level. We emphasize here that a very simple scheme for the relative humidity has been used because our major interest was focused on the parameters from which the turbulence depends on. This means that results on the relative humidity do not have to be considered as the best obtainable by the model Meso-Nh in absolute terms.

The **bias** is within [1-2] m/s for the wind speed on the 20 km range; within a few degrees in the [4-19] km a. g. l. range and within 10 (°) below 4 km and above 19 km; within 2 (K) below 12 k a. g. l. and within [2-4] (K) in the [12-20] km a. g. l. range. For the relative humidity the bias is almost zero below 10 km but it is of the order of 8 % in the region [10-17] km a. g. l. range. We proved also that, in the high part of the atmosphere, the meso-scale model Meso-NH behave in a very similar way to the GCMs, as we expect[21].

- We highlight that, measurements of atmospherical parameters have a high level of reliability, precision and accuracy. The goodness of the model performances is therefore strongly supported by the comparison with these measurements.

- It can be observed that, in proximity of 3 km a.s.l., the RMSE is in general, a little bit larger. As we discussed extensively in Sec. 4, this has not to be considered necessarily as an uncertainty of the model. We proved indeed that, this discrepancy can be explained by the fact that our reference (balloon) is drifted by the wind speed horizontally and therefore the radio-soundings are not the suitable reference to be used at this heights above the ground where the wind speed is not necessarily homogenous along the horizontal direction. Remote sensing reference should be preferable to be compared to model in the thin vertical slab [3-5] km a.s.l.

- The great advantage in using the Meso-NH model in an operational mode is that this permits us to have the temporal evolution of the vertical stratification of all the classical atmospheric parameters (wind speed, potential temperature, etc...) during the whole night with the desired temporal sampling. GCMs provides outputs only at synoptic hours and they are not reliable in the low part of the atmosphere for our applications.

- We proved (as indicated also in our previous studies on other sites - [7], [12]) that, to calculate the wavefront coherence time in operational conditions, the wind speed intensity as calculated by the Meso-NH model is definitely the best solution (and the most practical solution).

- Potential temperature and wind speed difference calculated on each individual night of the total sample of 23 nights has been provided (Annex B reports the wind speed case) and model performances revealed to be excellent also in this case.

ATMOSPHERIC PARAMETERS: FIRST 30 METERS FROM THE GROUND

- Table 60 and Table 61 summarizes results obtained for all the atmospherical parameters in proximity of the surface (first 30 m). They report the median value, the first and the third quartiles of the cumulative distribution of the bias and the RMSE calculated on a sample of 20 nights for the three parameters (wind speed, wind direction and absolute temperature). Both statistical indicators for the wind speed has been also calculated using a different model configuration having the horizontal resolution of the innermost domain of $\Delta X = 100$ m. Estimates are provided at different heights from the ground corresponding to location of observations for temperature and wind speed. **For the bias we have excellent results for the temperature and wind direction if we consider the model performances night by night: the worst median value for the absolute temperature bias is 0.64 K; for the wind direction is 7.28°. For the RMSE also results are extremely satisfactory for the same two parameters: the worst median value for the absolute temperature is 0.93 K; for the wind direction is 46.16°. These values are definitely very promising considering that the absolute values for the temperature (for this sample of nights) are: [11-13] °C at 2 m and 30 m at Paranal; [8-10] °C at 2 m, 11 m, 28 m and 30 m at Armazones. Bias and RMSE (always for model performances night by night) are less performant for the wind speed when the horizontal resolution is 500 m. The worst median value of the bias for the wind speed is -2.40 m/s and 2.71 m/s for the RMSE. The wind speed has, in average, an absolute value of 5 ms^{-1} at 10 m (Paranal) and at 2 m (Armazones); 5-6 ms^{-1} at the higher heights at Paranal**

---

[21] We highlight that these results can not be compared to those obtained in [26]. We discovered that the RMSE calculated by the authors do not correspond to the analytical definition of the RMSE. To perform a clean comparison we should conceive a benchmark test. A more extended discussion can be found in the BSCW archive.

while 6-7 ms^{-1} at the higher heights at Armazones. This means that the relative errors of the bias and the RMSE can be important and reach order of 50% in some cases. However, when the horizontal resolution is 100 m the wind speed is much better reconstructed by the model. The bias is only 0.93 m/s and the RMSE is 2.18 m/s that, if consider the absolute values, provide a relative error within 20%. We also proved that the better model behavior when the resolution is 100 m is due to the more realistic reconstruction of the interaction between the atmospheric flow and a summit (better resolved by the more detailed orography) that should generate an appropriated acceleration of the flow near the surface.

- Model performances providing a bias and a RMSE smaller than 1 K let us think that the Meso-NH model could definitely contribute to solve the problem of the thermalization of the dome and the elimination of the dome seeing contribution (the most critical contribution in the total turbulence budget).

- A preliminary analysis indicates that such a good performance of the model in reconstructing the temperature near the ground should reflect an equivalent good performance in reconstructing the temperature gradient near the surface. This aspect requires a dedicated analysis (may be in Phase B, TBD with ESO Board). Being that this is strictly correlated to the seeing developed in the first tens of meters above the ground, we conclude that the utilization of the Meso-NH model might provide an extremely further useful seeing indicator to be applied to the AOF (and others) facilities running at the VLT Observatory. Of course this should be properly tested and evaluated in the future.

- The hit-rate tables analysis told us that the absolute temperature and the wind direction can be reconstructed by the model in a very efficient way. For the temperature the good hit-rate G is of the order of 75-77% at Paranal and 77-79% at Armazones with exception of the temperature at 2 m at Armazones where G=51%. For the wind direction G is of the order of 61-67% at Paranal and 59-65% at Armazones. For both parameters (temperature and wind direction) the bad hit-rate is very small (order of a few percents). The maximum B for the temperature is 4% and the maximum B for the wind direction is 7%.

- The hit-rate tables analysis of the wind speed has been calculated using a corrector factor that depends on the bias of the model with respect to the observation (Eq.21). Such a multiplicative factor permits to obtain much better model performances above the two sites. **Using the highest horizontal resolution of 500 m the good hit-rate G is of the order of 55% at Paranal. Using the highest horizontal resolution of 100 m the good hit-rate G is of the order of 57-62% at Armazones. Using the respective configurations B is as low as 2.5% at Paranal and around 5% at Armazones.** At Paranal equivalent performances have been reached by the model (with the corrector factor) for an horizontal resolution of 500 m and 100 m. This permits to use a less expensive configuration, at least above this site (see discussion in Section 7).

- We proved that the model has a probability of detection (POD) of the wind speed greater than 6 ms^{-1} that is always greater than 68% above the two sites. At Paranal POD=69.7% (at 10 m) and 74.2% (at 30 m) with $\Delta X = 500$ m; at Armazones POD is around 69% at all levels with $\Delta X = 100$ m. This is very promising in order to know in advance the conditions characterized by strong wind speed.

- The hit-rate tables provide for all the atmospheric parameters a good hit-rate G greater than 55% and negligible bad hit-rates B at Paranal. The hit-rate tables provide a good hit-rate G greater than 50% and negligible bad hit-rates B at Armazones for temperature and wind direction. A good hit-rate G greater than 57% and negligible bad hit-rate are also found for the wind speed if we use a horizontal resolution of 100 m. .

	PARANAL		ARMAZONES			
	2 m	30 m	2 m	11 m	20 m	28 m
Temperature (^{o}C)	$0.28^{+0.57}_{-0.33}$	$-0.03^{+0.17}_{-0.48}$	$0.64^{+1.12}_{+0.50}$	$0.04^{+0.50}_{-0.25}$	$0.11^{+0.62}_{-0.45}$	$0.09^{+0.70}_{-0.57}$
Wind speed (m·s^{-1}) (ΔX = 500 m)	$-1.86^{-0.77}_{-2.82}$	$-0.82^{+0.24}_{-1.92}$	$-2.40^{-1.80}_{-5.85}$	$-1.92^{-1.15}_{-5.72}$	$-1.38^{+0.05}_{-4.14}$	$-1.27^{+0.26}_{-2.21}$
Wind speed (m·s^{-1}) (ΔX = 100 m)	$-0.70^{+0.09}_{-1.59}$	$-0.55^{+0.43}_{-1.69}$	$-0.93^{-0.27}_{-2.98}$	$-0.93^{+0.00}_{-2.75}$	$-0.47^{+0.29}_{-1.90}$	$-0.12^{+1.06}_{-1.78}$
Wind direction (^{o})	$1.01^{+22.96}_{-19.36}$	$5.45^{+15.29}_{-12.48}$	$-7.28^{+13.05}_{-16.22}$	$-6.25^{+5.62}_{-18.62}$	$-2.28^{+5.75}_{-11.96}$	$4.53^{+12.47}_{-8.41}$

Table 60: Table summarizing the BIAS values for all the surface meteorological parameters forecasted by the model.

	PARANAL		ARMAZONES			
	2 m	30 m	2 m	11 m	20 m	28 m
Temperature (^{o}C)	$0.92^{+1.08}_{+0.66}$	$0.64^{+0.90}_{+0.56}$	$0.87^{+1.23}_{+0.76}$	$0.73^{+1.02}_{+0.56}$	$0.85^{+1.01}_{+0.61}$	$0.93^{+1.08}_{+0.63}$
Wind speed (m·s^{-1}) (ΔX = 500 m)	$2.25^{+3.22}_{+1.71}$	$1.91^{+2.78}_{+1.53}$	$2.71^{+6.38}_{+2.14}$	$2.31^{+6.41}_{+1.83}$	$2.38^{+5.50}_{+1.87}$	$2.61^{+2.99}_{+1.91}$
Wind speed (m·s^{-1}) (ΔX = 100 m)	$1.85^{+2.76}_{+1.39}$	$2.13^{+2.95}_{+1.34}$	$1.86^{+3.34}_{+1.19}$	$2.13^{+3.27}_{+1.48}$	$2.18^{+2.52}_{+1.47}$	$2.05^{+2.68}_{+1.48}$
Wind direction (^{o})	$45.40^{+77.45}_{+27.93}$	$46.16^{+76.97}_{+20.94}$	$44.21^{+61.97}_{+22.03}$	$40.25^{+58.71}_{+16.54}$	$37.57^{+58.13}_{+13.45}$	$39.44^{+58.74}_{+14.20}$

Table 61: Table summarizing the RMSE values for all the surface meteorological parameters forecasted by the model.

OPTICAL TURBULENCE

- We corrected the whole data-set of GS measurements of the PAR2007 site testing campaign from an incorrect normalization of the autocorrelation of the scintillation maps by the average of the intensity of the scintillation maps. **We published these results in a paper: Masciadri, E. et al., 2012, MNRAS [31]**.

- Table 62 and Table 63 summarize the dispersion of measurements Δ_{obs} and simulations provided by the model Δ_{mod} (including results from the calibration and the validation sample). These are obtained starting from the cumulative distribution of the RMSE calculated in each night and retrieving the median values and the first and third quartiles. From these tables we retrieve that the model provides dispersion mostly comparable to that provided by measurements. In Section 7-Table 32 we proved that these same results are valid also if we consider only results obtained from the validation sample of 53 nights. **For the total seeing we have a Δ_{obs} = 0.28 arcsec vs. a Δ_{mod} = [0.30-0.42] arcsec. For the τ_0 we have a Δ_{obs} = [1.08-1.50] msec vs. a Δ_{mod} = [0.86-1.75] msec. For the θ_0 we have a Δ_{obs} = [0.42-0.87] arcsec vs. a Δ_{mod} = [0.50-0.96] arcsec**. Unfortunately we could not provide the dispersion for the boundary (ε_{BL}) and free atmosphere (ε_{FA}) seeing calculated on an independent sample because we proved that the MASS estimates are not sufficiently reliable for our calculations.

- If we limit our analysis to the calibration sample (20 nights) (Fig.44 and Table 20) and to the comparison model vs. GS measurements, we conclude that **the Δ_{mod} for ε_{BL} is of the order of 0.36 arcsec and for ε_{FA} is 0.20 arcsec**. Obviously this can be considered only as a preliminary result. We should be able to have access to reliable observations of the free atmosphere seeing on the range [600 m - 20 km] on a completely independent sample of measurements to obtain a model performance as done for the other astroclimatic parameters. For what concerns ε_{BL}, it will be possible to perform this calculation in the future with alternative instruments (for example SLODAR at high vertical resolution that is routinely running at the VLT Observatory). In both cases (free atmosphere and boundary layer) the best solution should be to have more than one instrument measuring the same parameters and being totally independent to provide homogeneous estimates.

- We observed that the bias and the RMSE are just a part of the statistical operators useful to describe the model goodness and the quality of a prediction. The RMSE, being a second order moment, is mainly weighted by the large errors and it therefore shows a tendency in overestimating the errors. Moreover, it has been observed that it can depend on the absolute values of the specific parameter. For this reason a complementary statistical operator has been calculated: the hit-rate tables recommended by the ECMWF protocols [44],[45].

- Besides to the sample of 53 nights, we considered a further sample of independent nights (36 nights) from the calibration sample uniformly distributed all along the same solar year of the calibration sample (2007) and we verified that the model presents the same behavior and the same statistical estimators for the total seeing. This tells us that the calibration is independent on the initialization data period and makes our conclusion more solid.

- A detailed investigation done comparing observed and simulated C_N^2 profiles, night by night, indicates that the model reconstructs, in a very reliable way, the average vertical distribution of the turbulence on the 20 km during a night by detecting the major turbulence layers. The amplitude of the variability (min,max) is satisfactory in the boundary layer while in the free atmosphere it is confined in a thinner range of variability with respect to measurements. This is responsible for a smoother temporal variability of the isoplanatic angle and the free atmosphere seeing provided by the model.

- The temporal evolution of the seeing (ε_{TOT}), the wavefront coherence time (τ_0) and the seeing in the boundary layer (ε_{BL}) is definitely excellent showing a good model ability in reconstructing the temporal evolution of these parameters all along a night. The model well reconstructs the average values during the night of the seeing in the free atmosphere (ε_{FA}) and the isoplanatic angle (θ_0) but it shows some limitations in reconstructing a temporal variability (peak-to-valley) with a similar amplitude than that observed. Even if a definite result is not yet achieved, in this feasibility study we could identify the origin of the problem. We are at present following three different approaches to overcome this problem. Preliminary results shows us that there are good premises to improve the model PtV fluctuations in the free atmosphere and to solve the problem (see details in Section 6.3.1 and Fig.66 and Fig. 67.)

- Results presented in this study have been obtained using an extremely simplified model calibration that makes the whole technique more solid. We obtained evidence of the fact that a sample of measurements extended to the winter period to be used for the model calibration could visibly improve the results obtained by the model. The model indeed shows some better results in summer that coincides with the period to which the calibration sample belongs.

- We proved that, taking into account the present results, the Meso-NH model joint with the Astro-Meso-NH package, can already provide a great number of fundamental information that could support the management of facilities such as the AOF.

- We calculated the hit-rate tables 3x3 considering two different approaches: **(a)** the first one uses as threshold values the tertiles of the distribution of the observed values; **(b)** the second one uses as threshold values some pre-defined fixed values chosen, for each parameter, in a way to be useful for an hypothetic application to the scientific program scheduling. We obtain much better results using the method (b). **For the wavefront coherence time (τ_0) (thresholds: 3 msec and 6 msec) the hits-rate of the model vs. the observations for the calibration sample (20 nights) gives us a G = 62% vs. a G=66% obtained between GS and DIMM. For the validation sample (53 and 89 nights) we obtain a G=53% (with respect to MASS) and a G=70% (with respect to DIMM). In all cases the bad hit-rates B is smaller than 5%. For the isoplanatic angle θ_0 (thresholds: 1.5 arcsec and 3 arcsec) and the validation sample (53 and 89 nights) the good hit-rate between the observations is G=46.3% while G= 61-64% for the model vs. the MASS and the DIMM. In all cases the values of the bad hit-rate B is smaller than 2%. For the total seeing the good hits-rate of the observations on the calibration sample (20 nights) is G=66% (GS vs. DIMM) while G=60.4% for model vs. GS and G=56.9% for model vs. DIMM. On the validation sample (53 and 89 nights) the good hit-rate of the model vs. DIMM is G=50% (53 nights) and G=49% (89 nights). In all cases the bad hit-rate is smaller than 1.5%. We conclude therefore that the model goodness on the calibration sample decreases of just a factor 1.2 passing from the calibration sample to the validation sample**. If we consider as a thresholds the tertiles of the observed values (method (a)) the model performances on the validation sample decreases of a factor 1.3 with respect to the calibration sample but the value of B is much larger (order of 18.2%). Also the G value is smaller of around 10% with respect to the method (b).In this kind of exercise the key point is to identify the threshold values that might provide useful information for the programs scheduling and verify the model performances correspondent to that couple of thresholds. In other words, we could figure out to use more than one hit-rate table for each astroclimatic parameter. A set of other statistical operators can be calculated in the future (miss rate, false alarm, probability of detection, etc...).

- We conclude that it should be suitable to access a richer statistical sample of simultaneous measurements of the total seeing provided by different instruments to proceed and progress on this analysis and establish how the G values for measurements decreases with the increasing the number of nights.

- Results reported in Table 62 on Δ_{obs} tell us that the specification of accuracy for a few astroclimatic parameters supported AO facilities as indicated in document E-TRE-ESO-528-0795 are strongly underestimated. An accuracy of 0.01 arcsec for the seeing and 0.1 ms for τ_0 are certainly underestimated and are not realistic. Our results indicate much larger values.

	OBS - OBS		
	$\Delta_{25,reference}$	$\Delta_{50,reference}$	$\Delta_{75,reference}$
τ_0(*) (ms)	[0.65-0.83]	[1.08-1.50]	[1.55-3.85]
θ_0(*) (arcsec)	[0.29-0.69]	[0.42-0.87]	[0.66-1.14]
ε_{tot}(**) (arcsec)	0.20	0.28	0.40

(*) DIMM-GS, MASS-DIMM, MASS-GS (20 nights)
DIMM-MASS (53 nights)
(**) DIMM-GS (20 nights)

Table 62: Table summarizing the RMSE value ranges for all the astroclimatic parameters, between DIMM, MASS and GS instruments.

	MNH - OBS		
	$\Delta_{25,model}$	$\Delta_{50,model}$	$\Delta_{75,model}$
τ_0(*) (ms)	[0.61-1.03]	[0.86-1.74]	[1.43-2.73]
θ_0(*) (arcsec)	[0.41-0.69]	[0.50-0.96]	[0.70-1.33]
ε_{tot}(**) (arcsec)	[0.21-0.33]	[0.30-0.42]	[0.47-0.53]
ε_{BL}(***) (arcsec)	0.31	0.38	0.61
ε_{FA}(***) (arcsec)	0.15	0.21	0.25

(*) MNH-GS, MNH-DIMM, MNH-MASS (20 nights)
MNH-DIMM, MNH-MASS (53 nights)
(**) MNH-GS, MNH-DIMM (20 nights)
MNH-DIMM (53 nights)
(***) MNH-GS (20 nights)

Table 63: Table summarizing the RMSE value ranges for all the astroclimatic parameters, between Meso-NH and observations.

Among the **perspectives** we list the following items:

- The good results obtained with all the atmospherical parameters (temperature, wind speed and direction) on a sample of 20 nights motivated us in extending this analysis to richer statistical sample to confirm the model performances. At the status of art, the model seems to be already in the conditions to provide useful informations and answers for the support of scheduling in terms of all the atmospherical parameters.

- **The most urgent goal now is for us to overcome the problem of the limited temporal variability of the turbulence (peak-to-valley amplitude) in the free atmosphere during the night (typical stable regime)**. We are at present working on this topic following three different approaches. Preliminary results indicate that there are good premises to achieve this goal.

- This study permitted us to collect a huge amount of data-set from the mesoscale model that so far have been analyzed in statistical terms. The next step will be to look at the rich data-set with a different perspective i.e. looking at the individual nights to retrieve more detailed informations on model performances. More precisely, it is important to check if it is possible to identify some features of the model outputs of the different atmospheric parameters associated to a model score of success for the OT more or less good. This should help us in identifying the conditions associated to a low-performance model predictions for the OT. Also it should be important to quantify the frequency of the bad initialization data (i.e. cases in which the initialization data are bad) and their impact on the final score of success.

- It has been agreed with the ESO Board that, for future developments, it would be probably better to filter the data on through a low-pass frequency filter of 1 hour so to eliminate the random fluctuations at high frequency and to analyze the statistic results on time longer time scales.

- Starting from 25/6/2013 new initialization data with 137 levels (instead of 91) covering the same vertical range of atmosphere (model top 0.01 hPa) will be available from the ECMWF through the integrated forecasting system (IFS) cycle 38r2. We expect therefore an improvement in the quality of the initialization data as well as a more suitable data to resolve thinner turbulence layers in the free atmosphere. If the vertical resolution of the model at low spatial frequency extended on the whole Earth (from which we select our initialization data) increases, we can figure out that the initialization data, and in particular the vertical stratification, will be better described. We know that the optical turbulence is mainly dependent on gradient of temperature and wind speed. Of course the vertical resolution is not the only factor from which the quality of the initialization depend on. For example the quality of the initialization data depends also on the density and frequency of observations that constitute the data assimilation procedure (see Introduction, Section 1.3).

- It should be crucial to set-up a working group made by MOSE team and ESO members to define the set of parameters specifically addressed to describe in exhaustive way what an operator would need to check in real time to perform the scheduling of observations. This is a delicate but fundamental step and we expect a multi-iteration progression in this direction.

- To validate how the model reconstructs the vertical stratification of the turbulence in the first hundreds of meters it should be necessary the use of at least two instruments providing the turbulence vertical stratification at high vertical resolution in the boundary layer. The SLODAR at high vertical resolution and LuSci could be used for this goal provided the two instruments are validated in an independent way. In case those measurements reveal useless, a dedicated campaign would be desired employing alternative instrumentation (TBD with ESO).

- It should be important to quantify the height above the ground at which the turbulence can be considered homogeneous along the horizontal direction on a spatial scale of at least 100 m. This step is fundamental to decide where to place the DIMM, if one DIMM is enough or not and what is the height starting from which it is useful to integrate the turbulence and compare measurements vs. model predictions for the operational scheduling application. If the DIMM is not representative of the turbulence arriving at the focus of the UT it become useless.

- So far we performed comparisons of model versus measurements provided by different instruments all conceived for OT measurements in an astronomical context. We know, however, that the final goal should be to produce prediction as close as possible to the turbulence perturbations at the telescope focus. The best solution should be therefore to consider, among the constraints we are using as a reference, also measurements of the OT that can be retrieved by the AO running routinely during the astronomical observations. It is known that AO measurements contain this information and some studies on this topic exist in the literature ([46]-[49]). However, no systematic studies on the relationship existing between this information and that provided by standard instruments for the OT measurements (DIMM, MASS, GS, etc...) exist at present. We think that, from one side, it would be useful to include among our list of constraints (measurements) with which we control the behavior of the model also those measurements obtained with the AO. On the other side a dedicated study should be done, in parallel, to quantify the reliability of this measurement in order to be able to treat it as a reference. TBD with ESO how to proceed in this direction.

- Results we obtained so far clearly indicate that a more intense investigation on the absolute estimate of the astroclimatic parameters provided by different instruments should be done and it is important to plan a road-map at long time scale within a wide community: AO, OT modeling with atmospheric models and instruments for OT measurements. We recently promoted with the French colleagues from Observatoire de Paris and ONERA, a working group including several European colleagues working in the field of the optical turbulence in astronomical context aiming at defining a set of site testing/experiments suitable to answer (among others) to these requirements [50]. Our intention is to promote a site testing campaign in which different vertical profilers plus DIMM run simultaneously. Position of each instrument and modality of running of each instruments is important so that we can have an answer to the uncertainties still existent and the doubts we still have. Details can be found in [50].

- A useful analysis to be done to control and manage the model behavior is the comparison of the simulated and measured ground fluxes. Such an exercise was originally planned for MOSE-Phase A but we could not carried out it within the contract time scale because instrumentation was not yet installed at the sites. ESO is taking care about that in the last months. We think therefore that the original plan will be able to be carried out in the future.

- As described in MOSE-Phase B plan (RD5), once evaluated the model performance obtained using analyses from the GCMs as initialization data (MOSE-Phase A), it would be useful to estimate the model performances using, as initialization data, meteorological forecasts from the GCM so to quantify the dependency of the model performances from these data. This would permit us to perform an OT forecast at an established delay ΔT.

- The output of this analysis should provide us the elements to set-up the most suitable combination of initialization data and the suitable architecture (platform) to perform operational OT prediction at the VLT Observatory and, in perspective, for the E-ELT. The complete automated process is described in Section 8.

ANNEX A

Simulated Nights - radiosoundings observations				
2009-07-29 12:00 (UT)	2009-07-30 12:15 (UT)	2009-07-31 00:30 12:00 (UT)	2009-08-01 00:00 06:15 12:00 (UT)	2009-08-02 00:00 06:15 (UT)
2009-08-04 00:00 (UT)	2009-08-05 01:00 06:00 (UT)	2009-08-06 01:00 06:15 (UT)	2009-08-07 00:00 06:20 12:00 (UT)	2009-08-08 00:00 06:00 (UT)
2009-08-09 00:00 (UT)	2009-08-10 01:10 12:00 (UT)	2009-11-09 12:00 (UT)	2009-11-10 12:00 (UT)	2009-11-11 00:00 06:00 12:00 (UT)
2009-11-12 00:00 06:00 12:15 (UT)	2009-11-13 00:00 06:00 12:00 (UT)	2009-11-14 00:00 06:00 12:00 (UT)	2009-11-15 00:00 06:00 12:20 (UT)	2009-11-16 00:00 06:15 (UT)
2009-11-17 00:00 06:00 12:00 (UT)	2009-11-18 00:00 06:00 12:00 (UT)	2009-11-19 00:00 06:00 12:00 (UT)		

Table 64: List of the 23 simulated nights with available radiosoundings observations in 2009. Below each date are reported the hours at which a balloon was launched.

Simulated Nights - meteorological surface observations				
2007-11-02	2007-11-03	2007-11-04	2007-11-05	2007-11-06
2007-11-07	2007-11-11	2007-11-13	2007-11-14	2007-11-17
2007-11-18	2007-11-20	2007-11-27	2007-11-28	2007-11-29
2007-11-30	2007-12-20	2007-12-22	2007-12-23	2007-12-24

Table 65: List of the 20 simulated nights with available meteorological observations near the surface. These 20 nights were used for the validation of the model meteorological performances.

Simulated Nights - model calibration with GS observations				
2007-11-10	2007-11-11	2007-11-13	2007-11-14	2007-11-17
2007-11-18	2007-11-20	2007-11-21	2007-11-22	2007-11-24
2007-12-15	2007-12-16	2007-12-17	2007-12-18	2007-12-19
2007-12-20	2007-12-21	2007-12-22	2007-12-23	2007-12-24

Table 66: List of the 20 simulated nights with available GS data. These 20 nights were used for the validation of the model OT performances.

Simulated Nights (2010/2011) - OT final validation

DATE	MASS		DIMM	
	τ_0	θ_0	τ_0	θ_0
2010-06-05	x	x	x	
2010-06-08*	x	x	x	
2010-06-17	x	x	x	
2010-06-26	x	x	x	x
2010-07-01	x	x	x	x
2010-07-05			x	x
2010-07-07*	x	x	x	x
2010-07-12*	x	x	x	x
2010-07-15	x	x	x	x
2010-07-25	x	x	x	x
2010-08-01	x	x	x	x
2010-08-05	x	x	x	x
2010-08-15	x	x	x	x
2010-08-20*	x	x	x	x
2010-08-25	x	x	x	x
2010-08-31	x	x	x	x
2010-09-05	x	x	x	x
2010-09-13*	x	x	x	x
2010-09-15	x	x	x	x
2010-09-25	x	x	x	x
2010-10-01	x	x	x	x
2010-10-05	x	x	x	x
2010-10-17	x	x	x	x
2010-10-25	x	x	x	x
2010-11-01	x	x	x	x
2010-11-05	x	x	x	
2010-11-15	x	x	x	x
2010-11-25	x	x	x	x
2010-12-02	x	x	x	x
2010-12-05	x	x	x	x
2010-12-13	x	x	x	x
2010-12-25	x	x	x	x
2011-01-02	x	x	x	x
2011-01-05	x	x	x	x
2011-01-15	x	x	x	x
2011-01-28	x	x	x	x
2011-02-01	x	x	x	x
2011-02-05	x	x	x	x
2011-02-15	x	x	x	x
2011-02-24			x	x
2011-03-03			x	x
2011-03-05			x	x
2011-03-15	x	x	x	x
2011-03-25	x	x	x	x
2011-04-01	x	x	x	x
2011-04-05			x	x
2011-04-15	x	x	x	x
2011-04-25	x	x	x	x
2011-05-01	x	x	x	x
2011-05-05	x	x	x	x
2011-05-11*	x	x	x	x
2011-05-15	x	x	x	x
2011-05-23*	x	x	x	x

*Night chosen by ESO

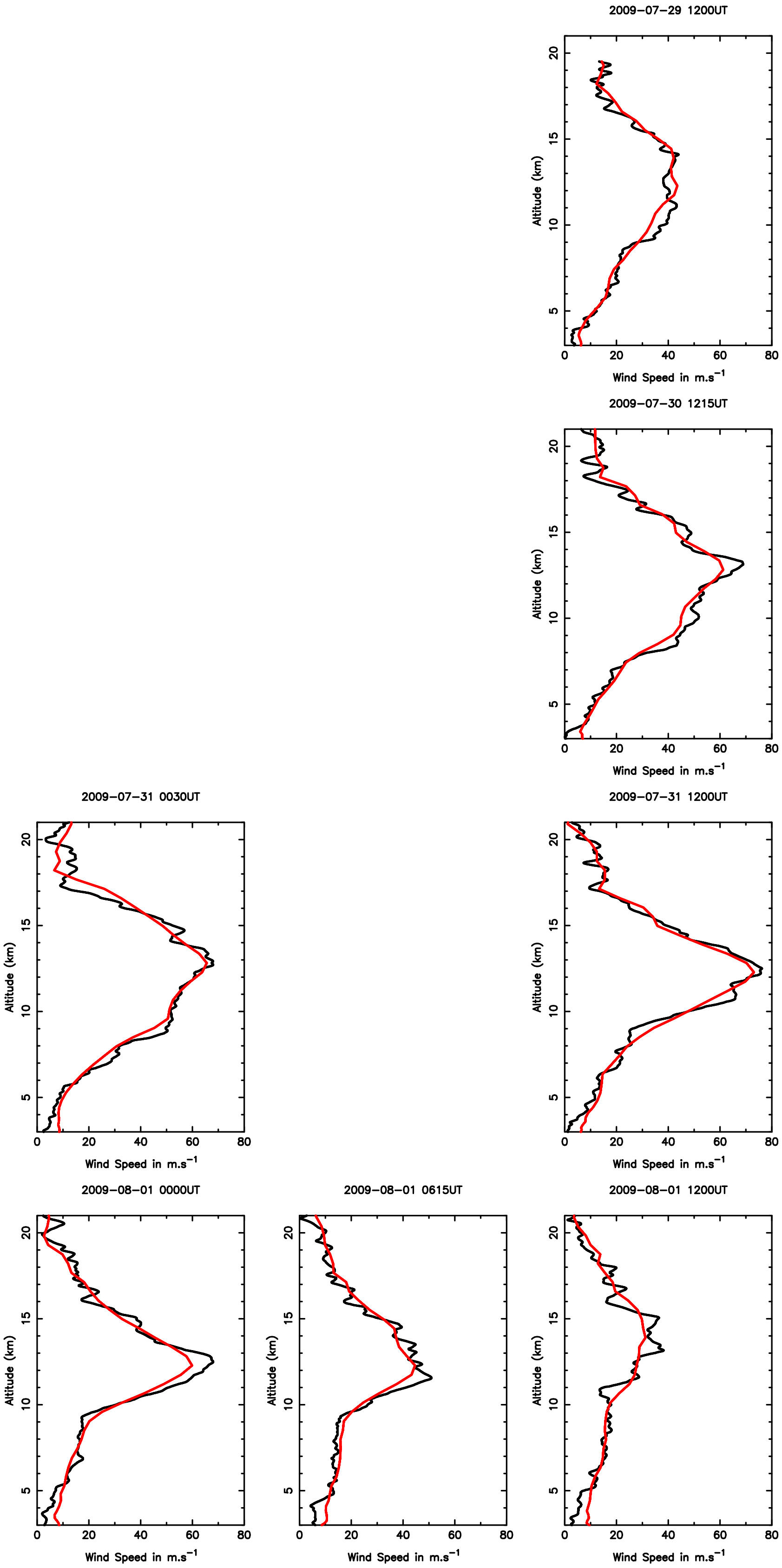
Table 67: List of the 53 simulated nights in 2010/2011 with available meteorological observations near the surface. These 53 nights (all different from the 20 nights of table 65) were used for the validation of the model performances after calibration, on an independent sample.

Simulated Nights (2007) - OT final validation				
05-01-2007	18-01-2005	25-01-2007	05-02-2007	19-02-2007
25-02-2007	05-03-2007	15-03-2007	25-03-2007	05-04-2007
15-04-2007	27-04-2007	07-05-2007	15-05-2007	27-05-2007
07-06-2007	17-06-2007	27-06-2007	05-07-2007	15-07-2007
25-07-2007	05-08-2007	15-08-2007	25-08-2007	05-09-2007
15-09-2007	24-09-2007	05-10-2007	15-10-2007	25-10-2007
05-11-2007	15-11-2007	25-11-2007	05-12-2007	14-12-2007
25-12-2007				

Table 68: List of the 36 simulated nights in 2007 with available meteorological observations near the surface. These 36 nights (all different from the 20 nights of table 65) were used for the validation of the model performances after calibration, on an independent sample.

ANNEX B

In Fig.60-Fig.65 are reported the 50 different vertical profiles of the wind speed taken from radiosondes measurements (in black) and the corresponding forecasted profiles from Meso-NH (in red). Each row is a different night of the total sample of 23 nights (Table 64) are represented.


2009-07-29 1200UT

2009-07-30 1215UT

2009-07-31 0030UT

2009-07-31 1200UT

2009-08-01 0000UT

2009-08-01 0615UT

2009-08-01 1200UT

Figure 60: Wind speed vertical profiles observed from radiosondes measurements (in black) and simulated by the mode (in red). Each row is a different night, and all the nights from the Table 64 are represented. Unit in m·s^{-1}.

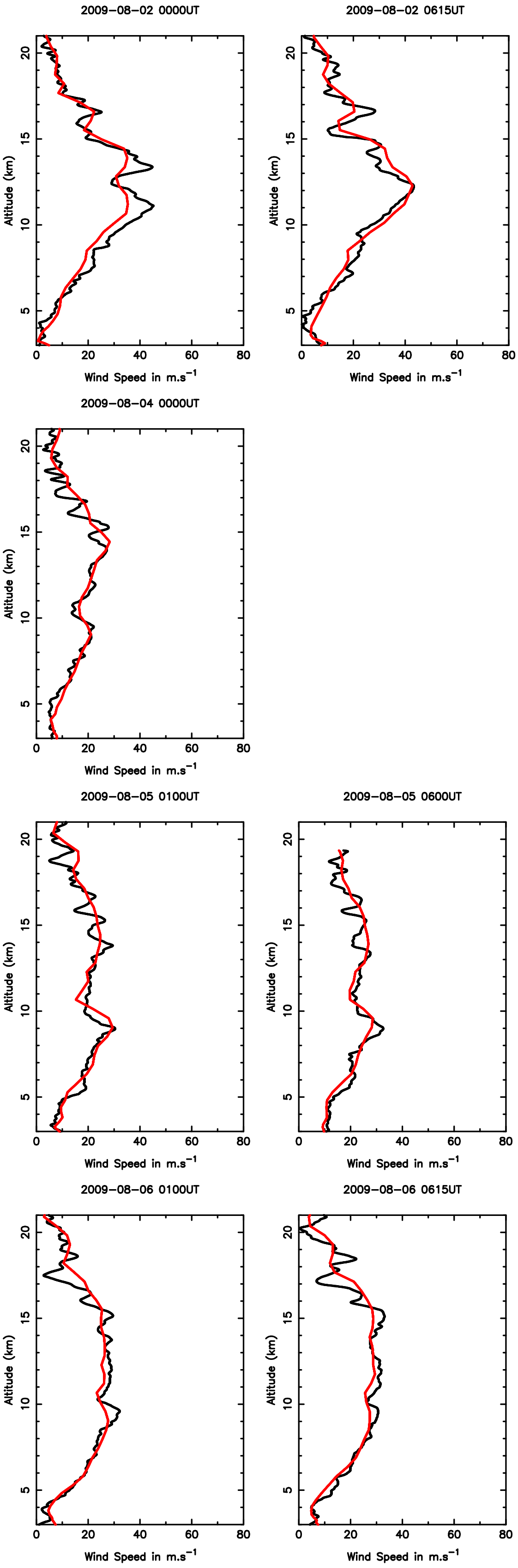


Figure 61: Following of Fig. 60.

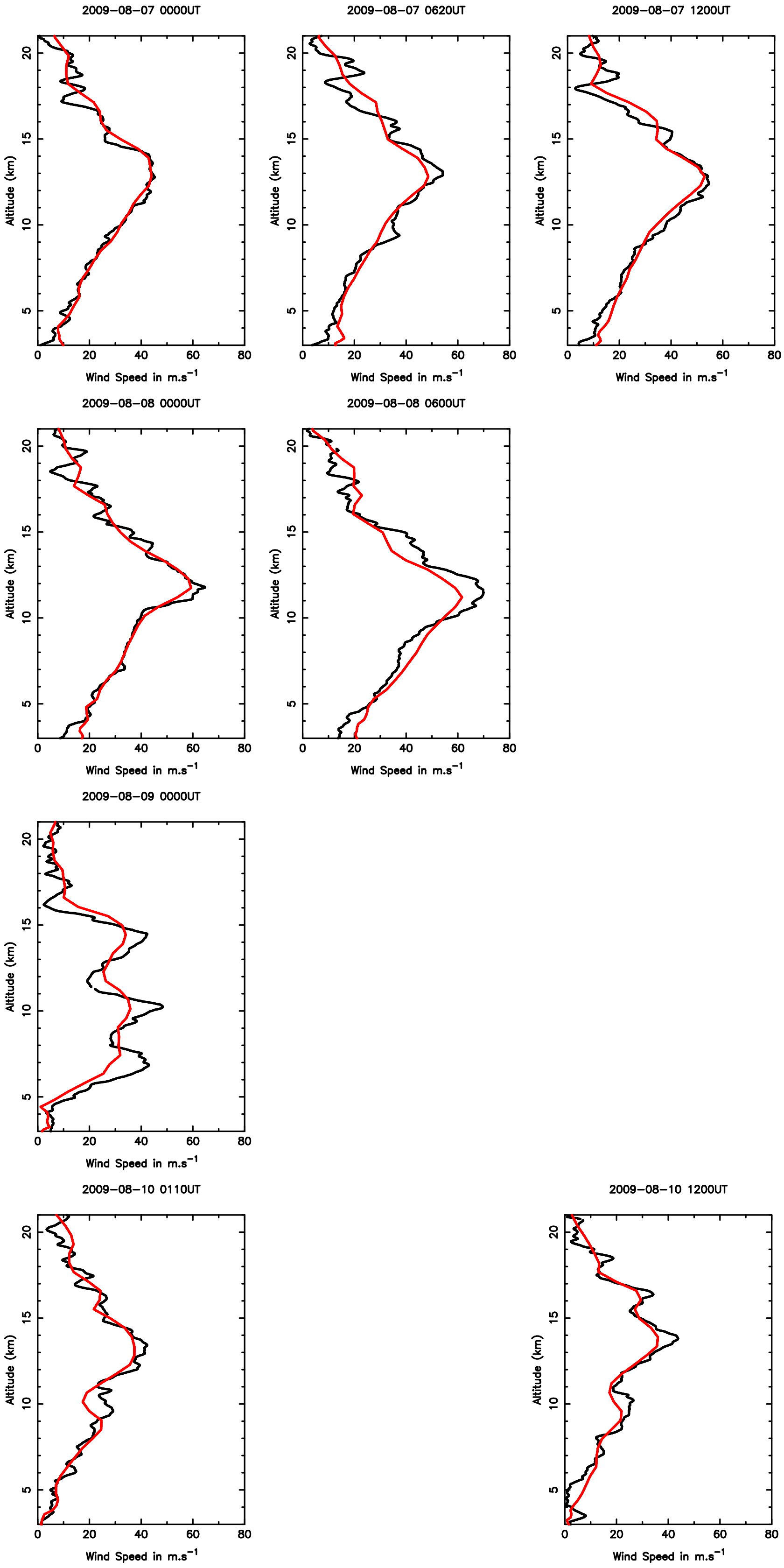


Figure 62: Following of Fig. 60.

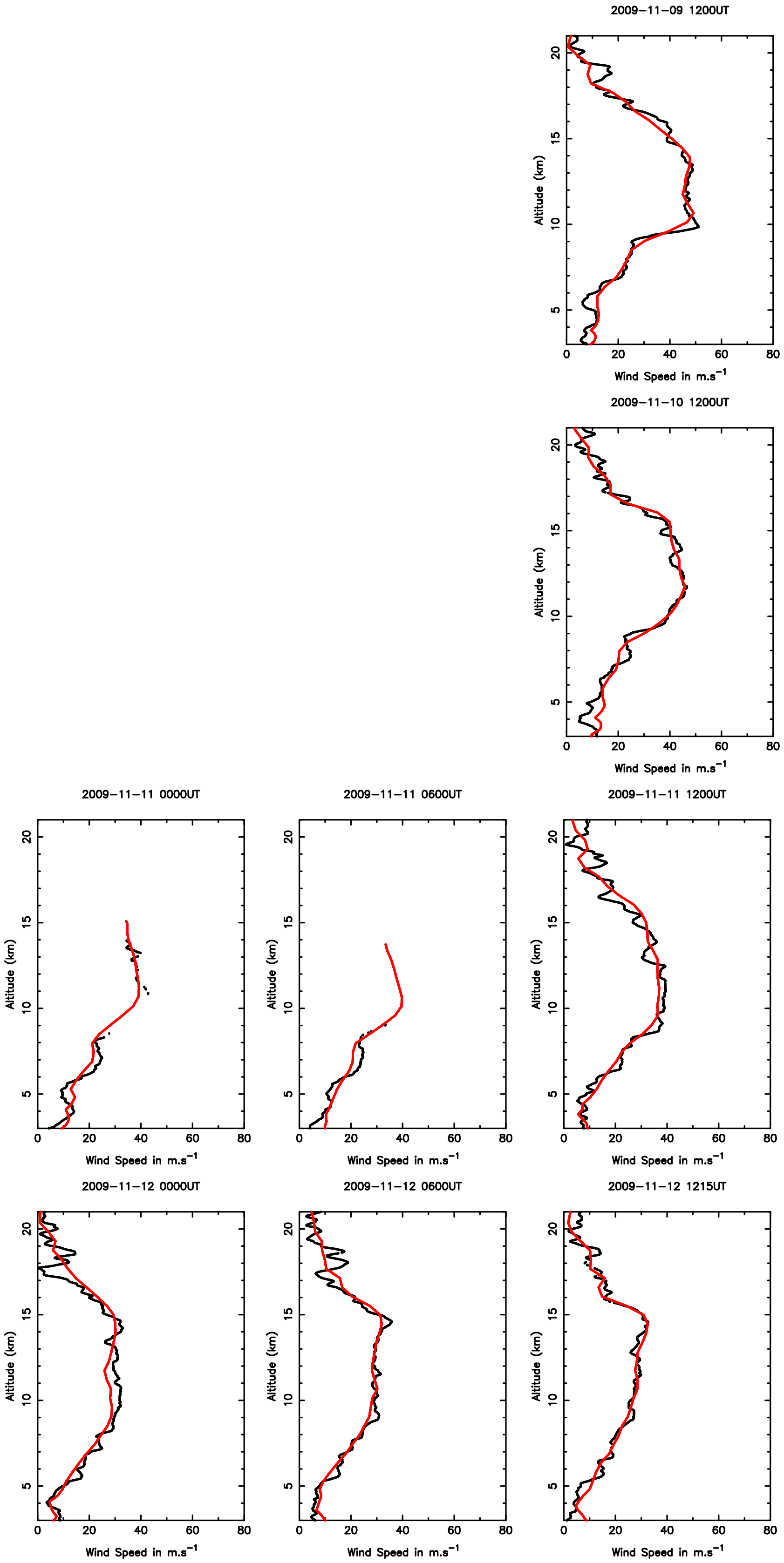


Figure 63: Following of Fig. 60.

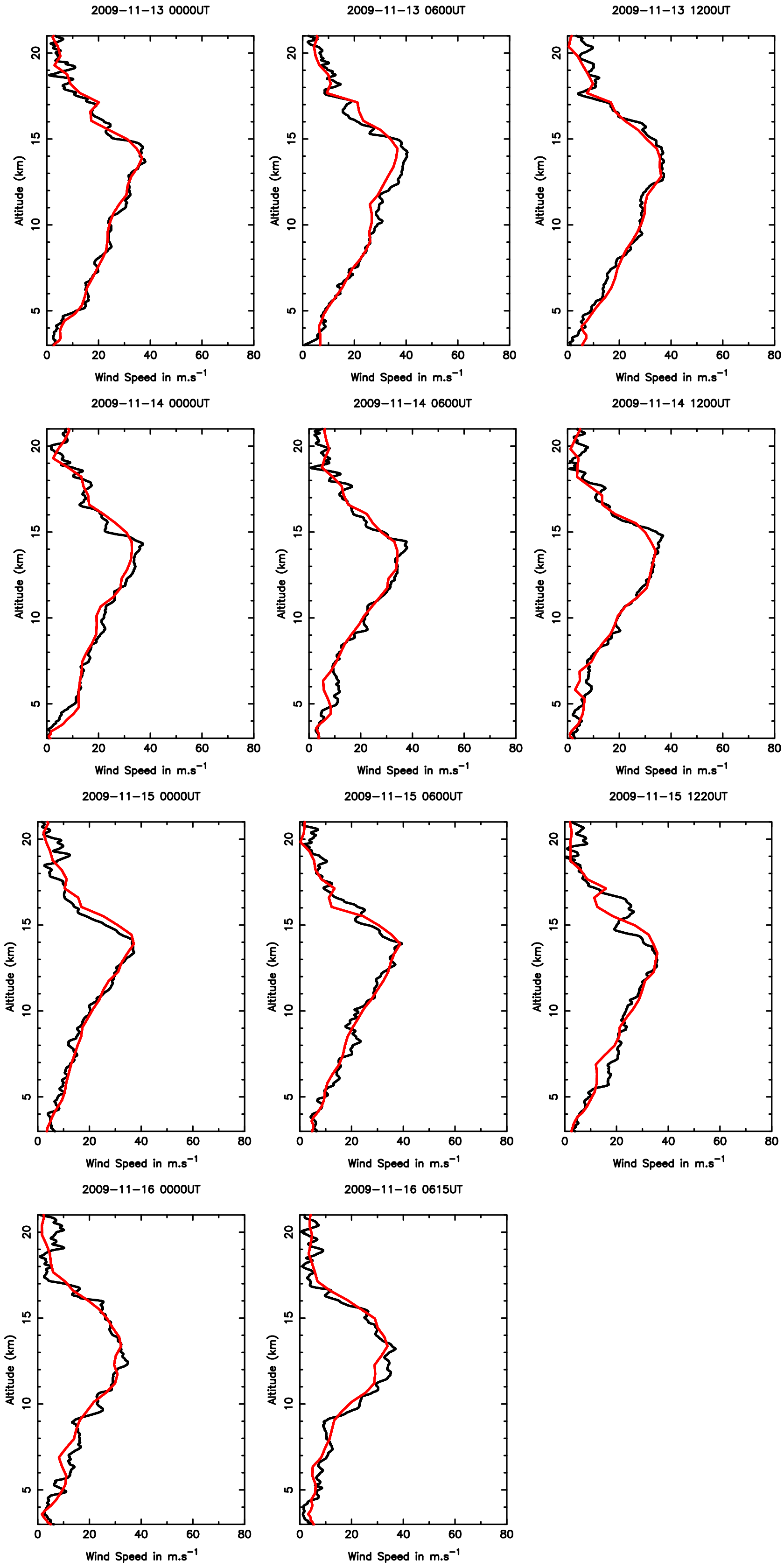


Figure 64: Following of Fig. 60.

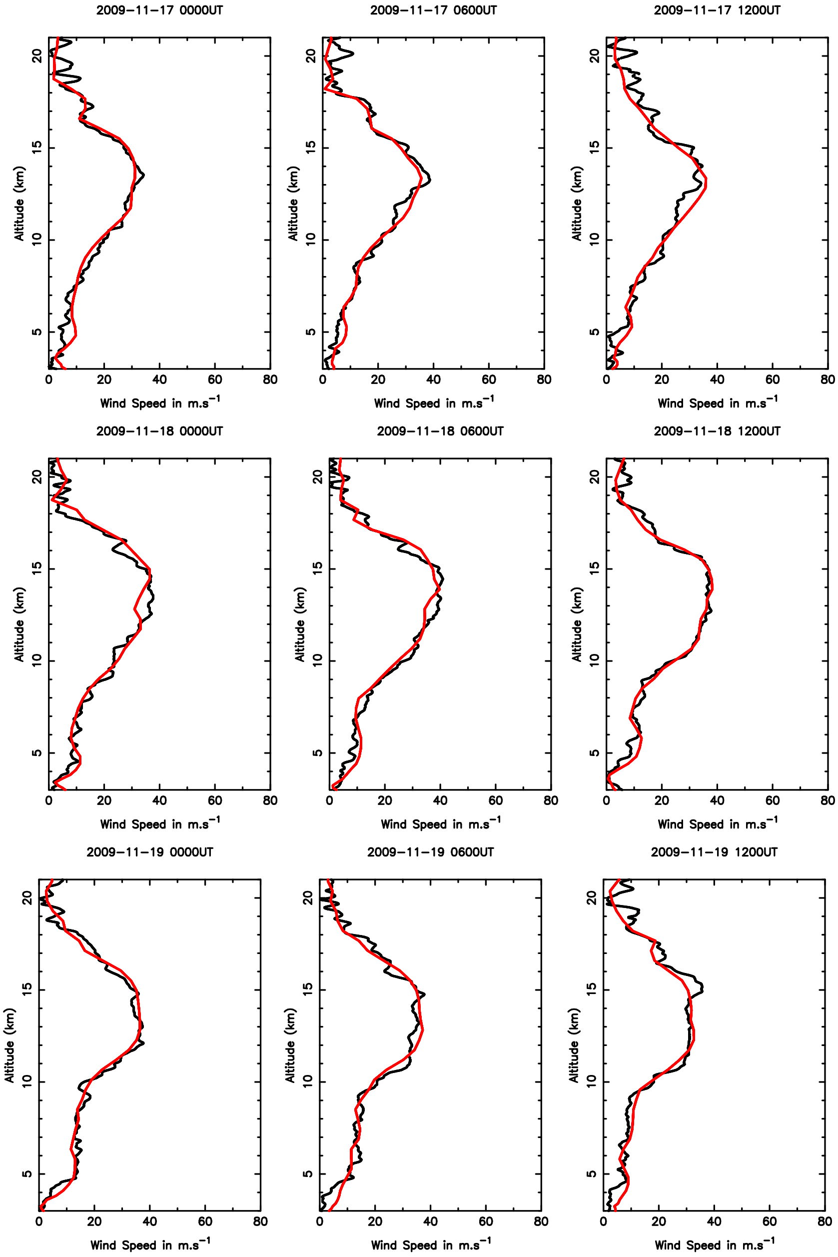


Figure 65: Following of Fig. 60.

ANNEX C

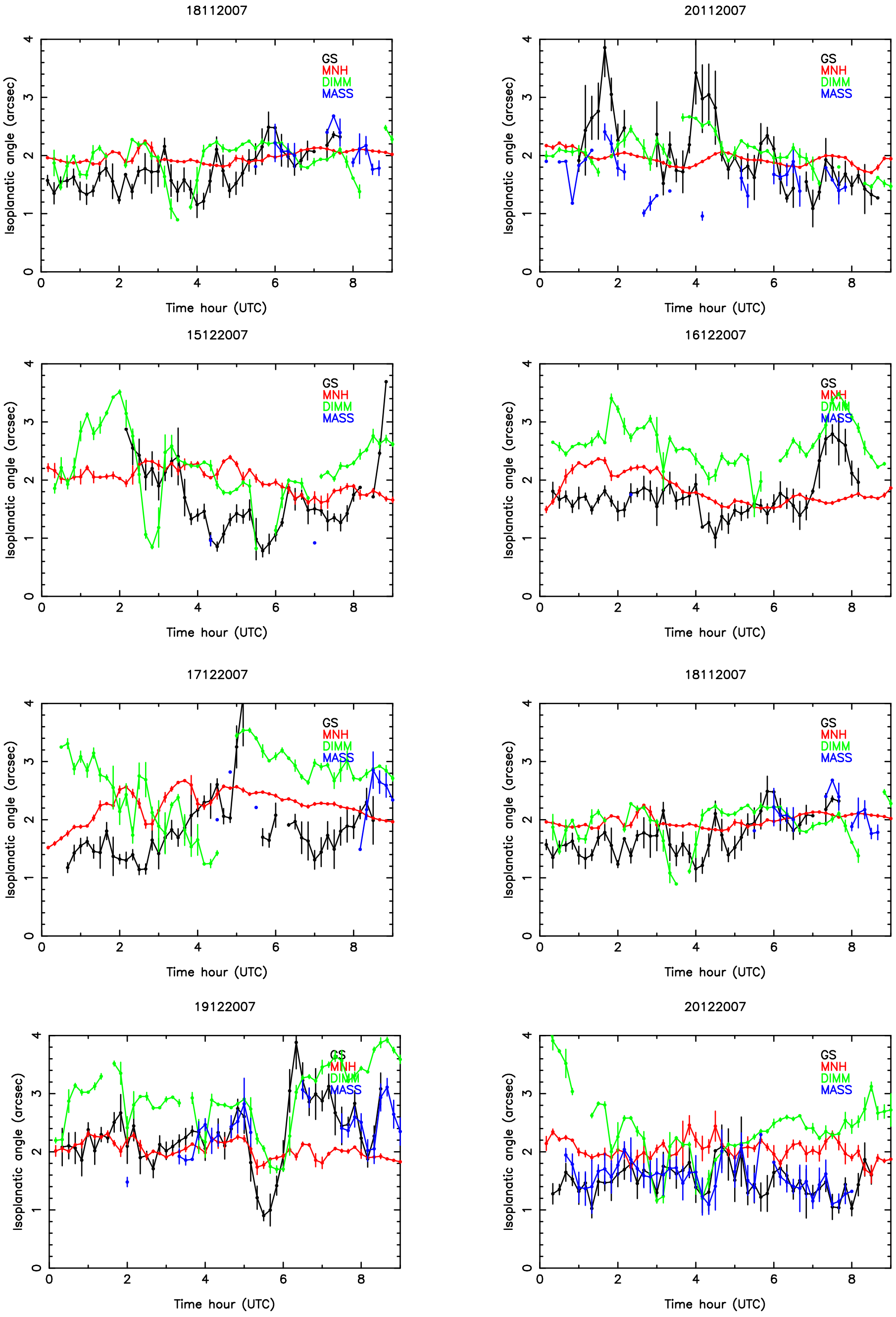


Figure 66: Temporal evolution of the isoplanatic angle for each single night of the calibration sample, from observations and model using a new model parameterization. Study on-going. Unit in arcsec.

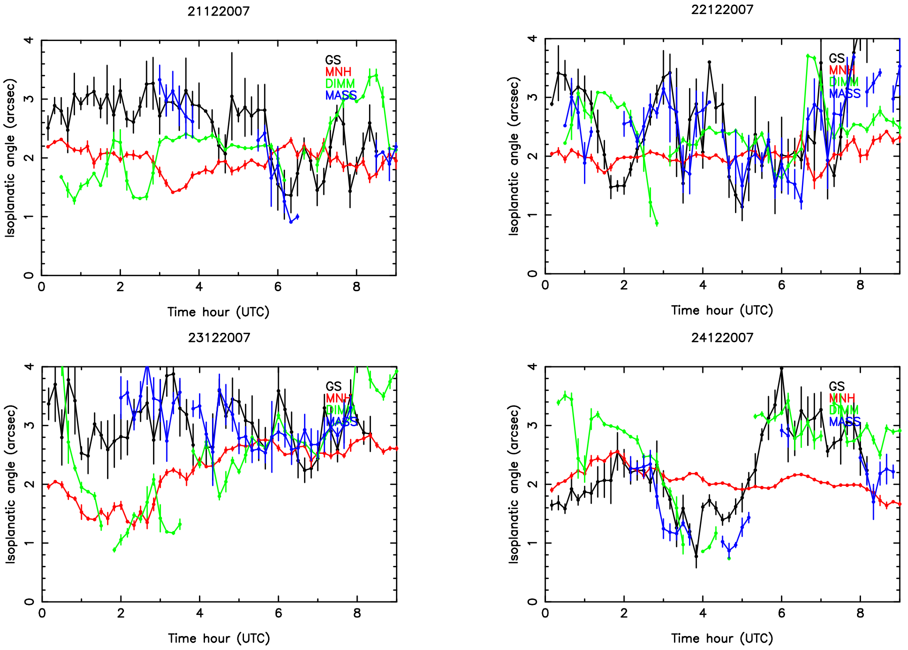


Figure 67: Temporal evolution of the isoplanatic angle for each single night of the calibration sample, from observations and model using a new model parameterization. Study on-going. Unit in arcsec.

References

[1] E. Masciadri: *Optical Turbulence Modeling and Forecasts - Toward a new era for the ground-based astronomy*, Book Title: "Seeing Clearly: An Introduction to Atmospheric Optical Turbulence, its Measurement and Mitigation", Ed. S. Businger & T. Cherubini, Virtual Book Worm Publishing, pp. 131-164, 2011a.

[2] E. Masciadri, F. Lascaux and S. Hagelin: *Optical turbulence forecast with non-hydrostatical mesoscale models*, Second International Conference on Adaptive Optics for Extremely Large Telescopes, in Victoria, Canada. Online at http://ao4elt2.lesia.obspm.fr, id.79, 2011b.

[3] E. Masciadri, J. Vernin and P. Bougeault: *3D mapping of optical turbulence using an atmospheric numerical model. I. A useful tool for the ground-based astronomy*, A&ASS, 137, 185, 1999a

[4] R. Racine, D. Salmon, D. Cowley and J. Sovka: *Mirror, dome, and natural seeing at CFHT*, PASP, 1991, 103, 1020, 1991

[5] E. Masciadri and T. Garfias: *Wavefront coherence time season variability and forecasting at San Pedro Martir site*, A&A, 366, 708, 2001a

[6] E. Masciadri: *Near ground wind simulations by a meso-scale atmospherical model for the ELTs site selection*, RMxAA, 39, 249, 2003

[7] S. Hagelin, E. Masciadri and F. Lascaux: *Wind speed vertical distribution at Mt Graham*, MNRAS, 407, 2230, 2010

[8] E. Masciadri, J. Vernin and P. Bougeault, *3D mapping of optical turbulence using an atmospheric numerical model. II. First results at Cerro Paranal*, A&ASS, 137, 203, 1999b

[9] E. Masciadri and P. Jabouille, *Improvement in the Optical Turbulence Parametrization for 3D simulations in a region around a telescope*, A&A, 376, 727, 2001b

[10] E. Masciadri, R. Avila and L.J. Sanchez: *Statistic reliability of the Meso-Nh atmospherical model for 3D C_N^2 simulations*, RMxAA, 40, 3, 2004

[11] E. Masciadri and S. Egner: *First seasonal study of optical turbulence with an atmospheric model*, PASP, 118, 1604, 2006

[12] S. Hagelin, E. Masciadri and F. Lascaux: *Optical turbulence simulations at Mt Graham using the Meso-NH model*, MNRAS, 412, 2695, 2011

[13] F. Lascaux, E. Masciadri, J. Stoesz and S. Hagelin: *Mesoscale optical turbulence simulations at Dome C*, MNRAS, 398, 1093, 2009

[14] F. Lascaux, E. Masciadri and S. Hagelin: *Mesoscale optical turbulence simulations at Dome C: refinements*, MNRAS, 403, 1714, 2010

[15] F. Lascaux, E. Masciadri and S. Hagelin: *Mesoscale optical turbulence simulations above Dome C, Dome A and South Pole*, MNRAS, 411, 693, 2011

[16] J.-P. Lafore, J. Stein, N. Asencio, P. Bougeault, V. Ducrocq, J. Duron, C. Fischer, P. Hereil, P. Mascart, V. Masson, J.-P. Pinty, J.-L. Redelsperger, E. Richard and J. Vilà-Guerau de Arellano: *The Meso-NH Atmospheric Simulation System. Part I: adiabatic formulation and control simulations*, Annales Geophysicae, 16, 90, 1998

[17] T. Gal-Chen and C.J. Sommerville, "On the use of a coordinate transformation for the solution of the Navier-Stokes equations", *J. Comput. Phys.*, 17, pp. 209-228, 1975.

[18] A. Arakawa and F. Messinger, "Numerical methods used in atmospheric models", *GARP Tech. Rep.*, 17, WMO/ICSU, Geneva, Switzerland, 1976.

[19] R. Asselin, "Frequency filter for time integration", *Mon. Weather. Rev.*, 100, pp. 487-490, 1972.

[20] J. Cuxart, P. Bougeault and J.-L. Redelsperger, "A turbulence scheme allowing for mesoscale and large-eddy simulations", *Q. J. R. Meteorol. Soc.*, 126, pp. 1-30, 2000.

[21] P. Bougeault and P. Lacarrère, "Parameterization of orographic induced turbulence in a mesobeta scale model", *Mon. Weather. Rev.*, 117, pp. 1972-1890, 1989.

[22] J. Noilhan and S. Planton, "A simple paramterization of land surface processes for meteorological models", *Mon. Weather. Rev.*, 117, pp. 536-549, 1989.

[23] E. Masciadri, J. Vernin and P. Bougeault, *3D numerical simulation of optical turbulence at Roque de Los Muchachos Observatory using the atmospherical model Meso-Nh*, A&A, 365, 699, 2001c

[24] E. Masciadri, R. Avila and L.J. Sanchez: *First evidence of the finite horizontal extent of the optical turbulence layers. Implications for the new adaptive optics techniques*, A&A , 382, 387, 2002

[25] J. Stein, E. Richard, J.-P. Lafore, J.-P. Pinty., N. Asencio and S. Cosma, "High-Resolution Non-Hydrostatic Simulations of Flash-Flood Episodes with Grid-Nesting and Ice-Phase Parameterization", *Meteorol. Atmos. Phys.*, 72, pp. 203-221, 2000.

[26] A. Chácon, O. Cuevas, D. Pozo, J. Marìn, A. Oyanadel, C. Dougnac, L. Cortes, L. Illanes, M. Caneo, M. Curé, M. Sarazin, F. Kerber, A. Smette, D. Rabanus, R. Querel and G. Tompkins, *Measuring and forecasting of PWV above La Silla, APEX and Paranal Observatories*, RMxAC, 41, 20-23, 2011

[27] S. Sandrock and R. Amestica, *VLT Astronomical Site Monitor - ASM Data User Manual*, Doc no.: VLT-MAN-ESO-17440-1773, 1999

[28] M. Schoeck, S. Els, R. Riddle, W. Skidmore, T. Travouillon, R. Blum, E. Bustos, G. Chanan, S. G. Djorgovski, P. Gillett, B. Gregory, J. Nelson, A. Otárola, A. Seguel, J. Vasquez, J. Walker, A. Walker, and D. Wang, ,*Thirty Meter Telescope Site Testing I: Overview*, PASP, 121, 384-395, 2009

[29] W. Skidmore, T. Travouillon and R. Riddle, *Report of the calibration of the T2-Armazones, 30-m tower air temperature sensors and sonic anemometers, the cross comparison of weather stations and sonic anemometers and turbulence measurements of sonic anemometers and finewire thermocouples*, Internal TMT Report, 2007

[30] N. I. Fisher and A. J. Lee, *A Correlation Coefficient for Circular Data*, Biometrika, 70, 327-332, 1983

[31] E. Masciadri, F. Lascaux, J. J. Fuensalida, G. Lombardi and H. Vasquez-Ramio, *Recalibrated generalized SCIDAR measurements at Cerro Paranal (the site of the Very Large Telescope)*, MNRAS, 420, 2399-24168, 2012

[32] R. A. Johnston, C. Dainty, N. J. Wooder, R. G. Lane, *Generalized scintillation detection and ranging results obtained by use of a modified inversion technique*, App. Optics, 41, 6788-6772, 2002

[33] R. Avila and S. Cuevas, *On the normalization of scintillation autocovariance for generalized SCIDAR*, Optics Express, 17, 3, p. 10926, 2009

[34] V. Kornilov, *Stellar scintillation in the short exposure regime and atmospheric coherence time evaluation*, A&A, 530, id.A56, 7 pp., 2011

[35] W. Dali Ali, A. Ziad, A. Berdja, J. Maire, J. Borgnino, M. Sarazin, G. Lombardi, J. Navarrete, H. Vazquez Ramio, M. Reyes, J. M. Delgado, J. J. Fuensalida, A. Tokovinin and E. Bustos, *Multi-instrument measurement campaign at Paranal in 2007 - Characterization of the outer scale and the seeing of the surface layer*, A&A, 524, id.A73, 8 pp., 2010

[36] E. Masciadri, J. Stoesz, S. Hagelin and F. Lascaux, *Optical turbulence vertical distribution with standard and high resolution at Mt Graham*, MNRAS, 404, 144-158, 2010

[37] M. Chun, R. Wilson, R. Avila, T. Butterley, J.-L. Aviles, D. Wier and D. Benigni, *Mauna Kea ground-layer characterization campaign*, MNRA, 394, 1121-1130, 2009

[38] J.E. Thomas-Osip, G. Prieto, A. Berdja, K.W. Cook, S. Villanueva, Jr., D.L. DePoy, J.L. Marshall, J.P. Rheault, R.D. Allen, D.W. Carona, CHaracterizing Optical turbulence at the GMT site with MooSci and MASS-DIMM, PASP, 124, 84, 2012

[39] H. Vazquez-Ramio, J. Vernin, C. Munoz-Tunon, M. Sarazin, A. M. Varela, H. Trinquet, J. M. Delgado, J. J. Fuensalida, M. Reyes, A. Benhida, Z. Benkhaldoun, D. G. Lambas, Y. Hach, M. Lazrek, G. Lombardi, J. Navarrete, P. Recabarren, V. Renzi, M.Sabil and R. Vrech, *European Extremely Large Telescope Site Characterization. II. High Angular Resolution Parameters*, 124, 918, 868-884, 2012

[40] M. Sarazin & A. Tokovinin, *The Statistics of Isoplanatic Angle and Adaptive Optics Time Constant derived from DIMM Data*, ESO Conference and Workshop Proceedings, 58, 321-328, 2002

[41] Avila, R., Vernin, J., Cuevas, S., *Turbulences profiles with Generalized Scidar at San Pedro Mártir Observatory and isoplanatic study*, PASP, 110, 1106, 1998

[42] J. Krause-Polstorff, A. Edmund, and L. W. Donald, *Instrument comparison: corrected stellar scintillometer versus isoplanometer*, Appl. Opt. 32, 40514057, 1993

[43] A. Ziad, R. Conan, A. Tokovini, F. Martin and J. Borgnino, *From the Grating Scale Monitor to the Generalized Seeing Monitor*, App. Opt., 39, 5415-5425, 2000

[44] Nurmi, P., *Recommendations on the verification of local weather forecasts (at ECMWF Member States) - Consultancy Report (ECMWF Operations Department)*, 2003

[45] Thornes, J.E., Stephenson. D.B., *How to judge the quality and value of weather forecast products*, Meteorol. Appl., 8, 307, 2001

[46] T. Fusco, G. Rousset, D. Rabaud, E. Gendron D. Mouillet, F. Lacombe, G. Zins, P.-Y. Madec, A.-M. Lagrange, J. Charton, D. Rouan, N. Hubin and N. Ageorges, *NAOS on-line characterization of turbulence parameters and adaptive optics performance*, J, Opt. A., 6, 585-596, 2004

[47] J. Kolb, N. Muller, E. Aller-Carpentier, P. Andrade and J. Girard, *What can be retrieved from Adaptive Optics real-time data?*, Proc. of SPIE: Adaptive Optics Systems III, 8447, 84475U, 2012

[48] J. Kolb, *Turbulence characterization at the Nasmyth focal plane of the VLT Melipal*, Proc. of SPIE: Adaptive Optics Systems.7015, 70154O, 2008

[49] P. Martinez, J. Kolb, M. Sarazin, J. Navarrete, *Active optics Shack-Hartmann sensor: using spot sizes to measure the seeing at the focal plane of a telescope*, MNRAS, 421, 3019-3026, 2012

[50] E.Masciadri, G.Rousset, T.Fusco, P.Bonifacio, J.Fuensalida, C.Robert, M.Sarazin, R.Wilson, A.Ziad, A roadmap for a new era turbulence studies program applied to the ground-based astronomy supported by AO, AO4ELT 3rd Edit., Florence 27-31 May, 2013