Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **466,** 520–539 (2017)
Advance Access publication 2016 November 30

doi:10.1093/mnras/stw3111

# Optical turbulence forecast: ready for an operational application

E. Masciadri,* F. Lascaux, A. Turchi and L. Fini
*INAF Osservatorio Astrofisico di Arcetri, Largo Enrico Fermi 5, I-50125 Florence, Italy*

Accepted 2016 November 28. Received 2016 November 28; in original form 2016 September 7

**ABSTRACT**

One of the main goals of the feasibility study MOSE (MOdelling ESO Sites) is to evaluate the performances of a method conceived to forecast the optical turbulence (OT) above the European Southern Observatory (ESO) sites of the Very Large Telescope (VLT) and the European Extremely Large Telescope (E-ELT) in Chile. The method implied the use of a dedicated code conceived for the OT called ASTRO-MESO-NH. In this paper, we present results we obtained at conclusion of this project concerning the performances of this method in forecasting the most relevant parameters related to the OT ($C_N^2$ , seeing $\varepsilon$, isoplanatic angle $\theta_0$ and wavefront coherence time $\tau_0$). Numerical predictions related to a very rich statistical sample of nights uniformly distributed along a solar year and belonging to different years have been compared to observations, and different statistical operators have been analysed such as the classical bias, root-mean-squared error, $\sigma$ and more sophisticated statistical operators derived by the contingency tables that are able to quantify the score of success of a predictive method such as the percentage of correct detection (PC) and the probability to detect a parameter within a specific range of values (POD). The main conclusions of the study tell us that the ASTRO-MESO-NH model provides performances that are already very good to definitely guarantee a not negligible positive impact on the service mode of top-class telescopes and ELTs. A demonstrator for an automatic and operational version of the ASTRO-MESO-NH model will be soon implemented on the sites of VLT and E-ELT.

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

## 1 INTRODUCTION

The MOSE (MOdelling ESO Sites) project is a feasibility study whose principal goal is to prove the possibility to forecast all the most relevant classical atmospheric parameters for astronomical applications (wind speed intensity and direction, temperature and relative humidity) and the optical turbulence (OT) that means $C_N^2$ profiles with the integrated astroclimatic parameters derived from the $C_N^2$ (i.e. seeing $\varepsilon$, isoplanatic angle $\theta_0$ and wavefront coherence time $\tau_0$) above the two European Southern Observatory (ESO) sites of Cerro Paranal, the site of the Very Large Telescope (VLT) and Cerro Armazones, the site of the European Extremely Large Telescope (E-ELT). The ultimate goal of the project is to investigate the possibility to implement an automatic and operational system for the forecast of these parameters at the VLT and at the E-ELT. In this paper, we will treat the OT that, as will see later on, is the most difficult but also the most challenging parameter to be forecasted among those we treated. In previous papers related to the MOSE project, we treated the abilities of the model in reconstructing wind speed and direction, temperature and relative humidity all along the atmosphere (∼20 km, i.e. vertical stratification of the atmospheric parameters; Masciadri, Lascaux & Fini 2013a) and in reconstructing the same atmospheric parameters close to the surface (Lascaux, Masciadri & Fini 2013, 2015). In all cases, results indicated excellent model performances. In this paper, we will focus our attention on results we obtained on the analysis of the model performances in reconstructing the OT, i.e. the astroclimatic parameters ($C_N^2$ profiles, the seeing $\varepsilon$, the isoplanatic angle $\theta_0$ and the wavefront coherence time $\tau_0$) that are commonly used to optimize the observations supported by the adaptive optics (AO). This paper completes therefore the whole feasibility study carried out for ESO.

The forecast of the OT is crucial for the success of the new-generation telescopes. We refer the reader to the introduction of Masciadri et al. (2013a) for a detailed description of the scientific challenges related to the forecast of the OT. We recall here the key elements. The traditional scheduling of scientific programme using as a criterion only the quality of the scientific programme has, indeed, important drawbacks and limitations. It has been widely accepted by the astronomical community that we had to take into account simultaneously the quality of the scientific programme but also the status of the OT to optimize the use of the telescope; otherwise, we risk that the most challenging scientific programmes

*E-mail: masciadri@arcetri.astro.it

are not carried out and the most important potentials of telescopes and instrumentation are not exploited as they could. The forecast of the OT is therefore extremely important to schedule the scientific programmes, to select the typology of instruments to be used at a specific time of the night and to optimize the AO systems' performances. The AO techniques can be very powerful, at present, in correcting the perturbations induced by the OT on the wavefronts but the AO performances are strongly dependent on the status of the OT and, under particular conditions, they can hardly run or they can even not run at all. More in general the forecast of the OT is crucial for the service mode, i.e. the observation mode of all the top-class facilities of present time and the observing mode of all new-generation facilities. It is the observing strategy that will maximize the possibility to achieve outstanding scientific goals with the ELTs. The service mode implies the knowledge in advance of the status of the atmosphere (atmospheric parameters and OT) and the rank of the scientific programmes. The *forecast* of these parameters plays therefore a crucial role in the context of the high-angular-resolution ground-based astronomy. Besides, we have not to forget that the cost of a night of observations is of the order of a hundred thousand US Dollars, and it is therefore immediate to understand that the forecasts play a crucial role not only in scientific but also in economical terms. The forecast we are dealing about aims to provide information in advance on a time-scale $\Delta T$ that is not inferior to 20 min. This is the typical time required for a beam to be shifted from an instrument to another in a configuration of permanent instruments placed in different focal stations. This is the configuration planned for new-generation telescopes.

Our approach implies the use of atmospheric non-hydrostatic mesoscale models, more precisely a model called MESO-NH (Lafore et al. 1998) for the atmospheric parameters joint with a dedicated code developed for the OT (Masciadri, Vernin & Bougeault 1999a). For simplicity, we call this model ASTRO-MESO-NH model. We refer the reader to the introduction of Masciadri et al. (2013a) to know why mesoscale models are necessary instead of other typologies of models [general circulation models (GCM), direct numerical simulations, large eddy simulations (LES)]. Mesoscale models are applied on limited areas of the Earth. There are different typologies of mesoscale models depending on the typical extension of the limited area and the horizontal resolution used. We used here limited areas having a size between 800 and 10 km$^2$ and a subkilometric horizontal resolution in the innermost domain in the neighbouring of the site of interest. The OT is completely parametrized in the mesoscale models. These characteristics guarantee to the ASTRO-MESO-NH model to reconstruct the OT maintaining the link with the spatio-temporal evolution of the atmospheric flow external to the limited areas, i.e. to realize a real 'forecast' of the OT.

The ASTRO-MESO-NH model has been applied in the last decades to many among the most important astronomical sites such as Cerro Paranal in Chile (Masciadri et al. 1999b), San Pedro Mártir in Mexico (Masciadri, Avila & Sanchez 2002, 2004; Masciadri & Egner 2006), Roque de los Muchachos in Canaries Islands (Masciadri, Vernin & Bougeault 2001b), Mt. Graham in Arizona (Hagelin, Masciadri & Lascaux 2011) and Dome C in Antarctica (Lascaux et al. 2009; Lascaux, Masciadri & Hagelin 2010, 2011). For completeness, we recall that other studies concerning the OT forecast on the whole atmosphere have been carried out using other mesoscale models or GCM and similar (or different) approaches in the astronomical context (Cherubini, Businger & Lyman 2011; Ye 2011; Giordano et al. 2013; Liu et al. 2015).

We highlight three important considerations: (1) the paper contains necessarily only a selection of the most relevant results obtained for the OT related to an extended study lasted a few years (and completed recently) that provided a clear indication of the good efficiency of the ASTRO-MESO-NH model for an application to the service mode. These convincing results have induced ESO to propose us to implement a demonstrator for an automatic operational version of the ASTRO-MESO-NH model applied to the sites of the VLT and the E-ELT. This project will start in the next few months. Even if the research on the 'OT forecast' is in a continuum evolution (as well as that of the 'weather forecast') and there is always space for improvements, in this feasibility study we achieved important steps ahead in terms of estimation of the model performances. Due to the fact that we are entering in the new phase of the operational demonstration and more and more observatories are interested in such a kind of application, it is important to provide the state of the art of the performances of our system. (2) Thanks to the development of our most recent algorithm of the $C_N^2$ , we could prove to be able to achieve a vertical resolution of the $C_N^2$ all along the whole atmosphere up to roughly 150 m. This is a crucial achievement that opens interesting new perspectives for the most sophisticated AO systems, i.e. the wide-field adaptive optics (WFAO) such as the ground layer adaptive optics (GLAO; Rigaut 2002), multi-conjugated adaptive optics (MCAO; Beckers 1988; Johnston & Welsh 1994), laser tomography adaptive optics (LTAO; Foy & Labeyrie 1985) and multi-object adaptive optics (MOAO; Assemat et al. 2004). (3) As expected, the model performances in forecasting the OT are not as good as in forecasting the atmospheric parameters (at least so far). This is due to the fact that the spatio-temporal scales on which the OT fluctuates are much smaller than the grid size and also to the fact that the turbulence is a stochastic quantity. This means that it is more difficult to describe numerically the OT. However, in spite of these intrinsic difficulties, we will see that results we achieved are very impressive and, even more important, are objectively already of great support for the service mode.

The plan of the paper is the following: in Section 2, we will describe the observations that we used as a reference to calibrate and to validate the model. In Section 3, we will describe the model configuration used for this study. In Section 4, we will describe the strategy used to calibrate and validate the model and in Section 5 we will show the results obtained. In Section 6, we show the model performances in reconstructing $C_N^2$ profiles with very high vertical resolution. In Section 7, we will present the conclusions and perspectives of this study.

## 2 OBSERVATIONS

Measurements provided by different instruments have been used to carry out this study. Considering the scarcity of OT measurements related to Cerro Armazones, in agreement with ESO the study on the OT has been performed only above Cerro Paranal. An important preliminary analysis having the goal to assure and test the reliability of the measurements has been performed with part of the instruments of the PAR2007 site testing campaign (Dali Ali et al. 2010) useful in our context: (1) a generalized SCIDAR (GS; more precisely the CUTE-SCIDAR III[1]) developed by the Istituto de Astrofisica de Canarias team (Vázquez Ramió et al. 2008) and corrected by Masciadri et al. (2012) to eliminate the error induced by the normalization of the autocorrelation of the scintillation maps by the autocorrelation of the mean pupil [problem identified by Johnston

[1] Hereafter we will call the generalized SCIDAR 'CUTE-SCIDAR III' simply GS.

et al. (2002) and Avila et al. (2009)], (2) a Multi-Aperture Scintillation Sensor (MASS) developed by the Kornilov & Tokovinin team (Kornilov et al. 2003) and (3) a Differential Image Motion Monitor (DIMM), an instrument that since 1988 is running at Cerro Paranal to monitor the seeing, i.e. the integration of the OT developed all along the whole atmosphere (Sarazin & Roddier 1990). We had simultaneous GS and DIMM measurements related to 20 nights. MASS measurements were simultaneous to GS on a subsample of 14 nights. We note that the three instruments are located at basically the same height (GS at 5 m above ground level (a.g.l.), DIMM and MASS at 6 m a.g.l.). Moreover, the VLT is basically a plateau and this guarantees a fair comparison between measurements. The recalibration of the GS measurements was fundamental to assure us a reliable reference. As we will see later on, we decided to use the GS as a reference for the calibration of the ASTRO-MESO-NH model. Besides, a detailed analysis of comparisons of measurements from the GS, MASS and DIMM (Masciadri, Lombardi & Lascaux 2014) permitted us to conclude that the MASS could not be taken as a reference because of three main problems: (1) it underestimates the integrated turbulence ($J$ or seeing) in the free atmosphere with respect to the GS with a relative error of $-32$ per cent in terms of the seeing ($-48$ per cent in terms of $J$); (2) we found important discrepancies between MASS and GS in all the individual layers (reaching relative errors as high as $-65$ per cent in terms of seeing and $-82$ per cent in terms of $J$) with exception of the layers located at 2 and 16 km (layers 3 and 6) in which the relative error remains limited to $+18$ per cent in terms of seeing ($+20$ per cent in terms of $J$). A previous study on a similar topic (Tokovinin et al. 2005, even if it was applied on a poorer statistical sample) found relative errors on individual layers as large as those we found in Masciadri et al. (2014); (3) the particular weighting functions of the MASS, having a triangle shape, do not permit us to identify precisely the height of the boundary layer and in general the height separating a layer from the contiguous one. This represents an important limitation for the calibration of the ASTRO-MESO-NH model.

Thanks to DIMM measurements (a third independent instrument during the PAR2007 campaign together with GS and MASS), we could prove that the problem causing the discrepancies between GS and MASS came from the MASS. We cite here just the elements useful to justify why MASS measurements could not be used to calibrate the ASTRO-MESO-NH model.[2] An accurate estimate of the $C_N^2$ is indeed important for the model calibration and, because of the reasons we have just discussed, the MASS could not assure that.

Besides, we recall that Masciadri et al. (2014) proved that the isoplanatic angle $\theta_0$ coming from the MASS is reliable (mainly thanks to the good reliability of the OT estimate in layer 6 located at 16 km above the ground) and the $\tau_0$ measurements are reliable too (just a few warnings with respect to this parameter – see the cited paper). This information is useful for the analysis done in this paper.

With these elements in mind, we concluded that we could use MASS measurements of $\theta_0$ and $\tau_0$ to validate the model with respect to these parameters (see Section 4.1). MASS is, indeed, an instrument currently running above Cerro Paranal at VLT Observatory. After the model calibration, we could therefore validate the model with a more extended sample of nights not belonging to the PAR2007 campaign.

[2] We refer the reader to Masciadri et al. (2014) for further details/discussion on the GS/MASS comparison. This is not part of the content of this paper.

**Table 1.** ASTRO-MESO-NH model grid-nesting configuration for the OT simulation. 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 $\Delta X$.

| Domain | Grid points | Domain size (km) | $\Delta X$ (km) |
|---|---|---|---|
| Domain 1 | $80 \times 80$ | $800 \times 800$ | $\Delta X = 10$ |
| Domain 2 | $64 \times 64$ | $160 \times 160$ | $\Delta X = 2.5$ |
| Domain 3 | $150 \times 100$ | $75 \times 50$ | $\Delta X = 0.5$ |

## 3 MODEL CONFIGURATION

As previously said, we used the MESO-NH model[3] interfaced with the ASTRO-MESO-NH code. Both models/codes can simulate the temporal evolution of three-, two- or mono-dimensional parameters over a selected limited area of the globe. In the ASTRO-MESO-NH code, the $C_N^2$ is a 3D parameter, all the integrated astroclimatic parameters ($\varepsilon$, $\theta_0$ and $\tau_0$) are 2D parameters. The two codes are run together. For this reason, we provide the configurations of both models. For what concerns MESO-NH, the system of hydrodynamic equations is based upon an anelastic formulation that permits an effective filtering of acoustic waves. The model uses the Gal-Chen & Somerville (1975) coordinates system on the vertical and the C-grid in the formulation of Arakawa & Messinger (1976) for the spatial digitalization. The model version used for this study employs an explicit three-time-level leap-frog temporal scheme with a time filter (Asselin 1972). The model employs a one-dimensional 1.5 turbulence closure scheme (Cuxart, Bougeault & Redelsperger 2000), and we used a one-dimensional mixing length proposed by Bougeault & Lacarrère (1989). The surface exchanges are computed using the interaction soil biosphere atmosphere module (Noilhan & Planton 1989). The OT and derived integrated parameters are simulated with the ASTRO-MESO-NH code developed by Masciadri et al. (1999a) and since there is continuous development by our group. The geographic coordinates of Cerro Paranal are (24°37′33.″117 S, 70°24′11.″642 W). We used the grid-nesting techniques (Stein et al. 2000) consisting in using different embedded domains of the digital elevation models (i.e. orography) extended on smaller and smaller surfaces, with increasing horizontal resolution but with the same vertical grid. We employed the two-way grid nesting that guarantees a better thermodynamic balance of the atmospheric flow because it takes into account the feedback between each couple of father and son domain. Gravity wave propagation is also better reconstructed than the one-way option. Simulations of the OT consisted of three embedded domains where the horizontal resolution of the innermost domain was $\Delta X = 500$ m (Table 1). Fig. 1 shows the three domains as seen by model. The model is initialized with analyses provided by the General Circulation Model HRES (GCM HRES) of the European Centre for Medium-Range Weather Forecast (ECMWF) having an intrinsic horizontal resolution of 16 km.[4] All simulations we performed with ASTRO-MESO-NH start at 18:00 UT of the day before ($J$-1), and they last in total 15 h. We reject the first 6 h from our analysis because we are interested in data after the sunset, and we therefore take into account ASTRO-MESO-NH outputs starting from 00:00 UT of the day $J$. During the 15 h, the model is forced each six hours (synoptic hours) with the analyses provided by the

[3] http://mesonh.aero.obs-mip.fr/mesonh52 – we used the Masdev4.8 version of the code.

[4] Starting from 2016 March ECMWF products have a horizontal resolution of roughly 9 km.

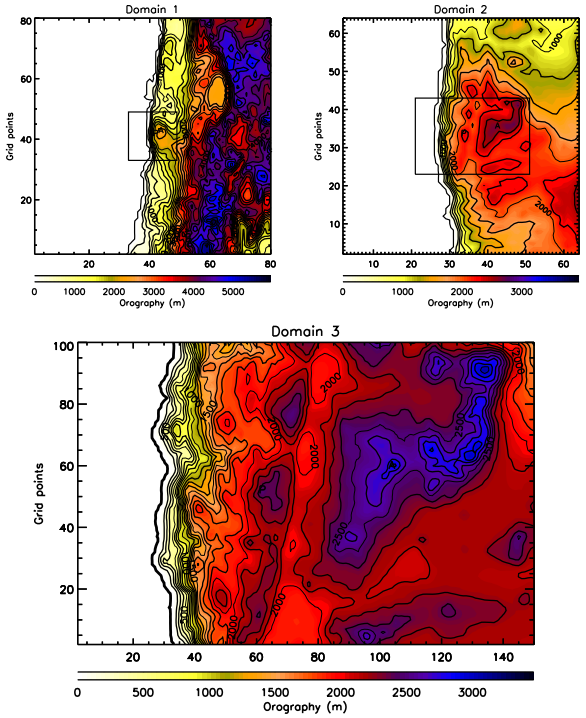


**Figure 1.** Orography (altitude in m) of the region of interest as seen by the MESO-NH model for all the embedded domains of the model grid-nested configuration. (a) Domain 1 (digital elevation model, i.e. orography data from GTOPO), (b) Domain 2 (digital elevation model from GTOPO), (c) Domain 3 (digital elevation model from ISTAR), 'P' stands for Cerro Paranal, 'A' stands for Cerro Armazones. The black square in Domain 1 represents the surface of Domain 2. The black rectangle in Domain 2 represents the surface of Domain 3. See Table 1 for the specifications of the domains (number of grid points, domain extension, horizontal resolution).

GCM related to the corresponding hours. In the grid point of the summit of Cerro Paranal, we print out the $C_N^2$ profiles with a temporal sampling of 2 min from the innermost domain.[5] We selected 2 min because this corresponds more or less to the sampling of typical vertical profilers. There are no major constraints in shortening it but, at present, neither a particular interest in doing that. The vertical grid is the same for all the domains. We have 62 vertical levels with a first grid point of 5 m, a logarithmic stretching of 20 per cent up to 3.5 km above the ground and almost constant vertical grid size of ∼600 m up to 23.8 km. The height of the first grid point is determined by the height above the ground at which the GS, which we used to compare measurements and simulation, is located (5 m). It is important to highlight that this model configuration guarantees the implementation of the model in an operational set-up.

## 4 METHOD OF ANALYSIS

### 4.1 ASTRO-MESO-NH model calibration and validation

We considered in total three completely independent samples of nights (20, 36 and 53 nights). The 20 nights of the PAR2007 campaign (all concentrated in the months of November and December 2007) have been used for the model calibration because we had GS and DIMM simultaneous measurements. Only measurements from the GS have been used in reality to calibrate the model but DIMM measurements have been very useful to assure us the reliability of the GS measurements (see Section 2). To calibrate the ASTRO-MESO-NH model, we need a vertical stratification of the OT, i.e. $C_N^2$ profiles. For the model validation, we could not use GS measurements since no measurements on long sample size are available for Cerro Paranal. We therefore considered for the model validation simulations done in the period [06/2010–05/2011; a solar year] for which DIMM and MASS measurements of the seeing were available. MASS and DIMM are routinely used by ESO to monitor atmosphere at Cerro Paranal. For the validation sample, for each month we selected the nights corresponding to the 1st, 5th, 15th and 25th of the month where DIMM measurements for the same night and the night before were available. In case the constraint was not respected, we looked for the nearest night to the sequence reported above respecting this criterion. A sample of 53 nights has been selected. A second set of 36 nights has been selected in 2007 (but completely independent from the calibration sample of 20 nights of the PAR2007 campaign used for the model calibration) to investigate a specific effect observed in results obtained in our analysis of the 53 nights.[6]

The technique of the ASTRO-MESO-NH model calibration has been proposed by Masciadri & Jabouille (2001) and validated later on by Masciadri et al. (2004). The calibration aims at fixing the value of the minimum turbulent kinetic energy ($E_{\rm min}$) to be given at each fixed vertical slab of the model (or each vertical model level).[7] The $E_{\rm min}$ is the minimum turbulent kinetic energy required by the model to move from its state of equilibrium. Once the $E_{\rm min}$ value is fixed (result of the calibration), all the simulations are run again with this $E_{\rm min}$ (the same for all simulations). $E_{\rm min}$ can be considered a sort of background climatologic noise. In regions in which the dynamic turbulence is well developed, the model rapidly forgets the $E_{\rm min}$ (more frequent in the low part of the atmosphere). However, in stable regions of the atmosphere, Masciadri & Jabouille (2001) proved that $C_N^2 \propto {E_{\rm min}}^{2/3}$; therefore, $E_{\rm min}$ can be retrieved through the minimization of the difference of observed and simulated $C_N^2$ profiles (see equation 8 in Masciadri & Jabouille 2001). In each $C_N^2$ profile, typically we can identify more than one vertical slab in which turbulence is in stable regimes (see Masciadri & Jabouille 2001, fig. 1). Using different values of $E_{\rm min}$ means that we are assuming that the atmosphere has more or less inertia in different regions of the atmosphere, and the thermodynamic instabilities require more or less energy to trigger turbulence in some regions of the atmosphere. Supported by these arguments, the model calibration proposed by Masciadri & Jabouille (2001) has been applied with some different variants, such as the number and the extension of the verticals slabs, in several studies (Masciadri et al. 2004; Masciadri & Egner 2006; Cherubini et al. 2011; Hagelin et al. 2011). In this study, we minimized the number of the vertical slabs so to get the model less dependent on the size of the calibration sample. We divided the 20 km into just three vertical slabs: [0–600 m], [600 m–14 km] and [14 km–20 km]. The first threshold (∼600 m) discriminates regions of the atmosphere in which the model is more or less active and

[5] The $C_N^2$ is calculated obviously by the model each time-step.

[6] The period in which we selected the 53 nights of the final validation was more recent (2010–2011) than that related to the calibration sample (2007). We decided therefore to perform a validation of the model also on a second sample of 36 nights (independent from the calibration sample) but belonging to the same year (2007). The goal of this is exercise is to exclude effects due to the initialization data.

[7] We refer the reader to Hagelin et al. (2011) for a discussion on the difference between the two approaches.

weakly/strongly dependent on the calibration procedure. The second threshold (∼14 km) corresponds, in average, to the height at which there is a more rapid change of the structure of the $C_N^2$ at low spatial frequencies. This is also due to the fact that the stratosphere is characterized by different dynamic of the atmosphere than the troposphere. After the calibration, inside these three vertical slabs, the value of the $E_{\rm min}$ assumes three different constant values: $E_{\rm min,1}$, $E_{\rm min,2}$ and $E_{\rm min,3}$.

In conclusion, the strategy of our analysis is therefore summarized in the following sequence of steps: (1) we perform the model calibration using the GS measurements; (2) we fix the free parameters $E_{{\rm min},\,i}$ in the model; (3) we repeat the whole set of simulations on the calibration sample with the same $E_{{\rm min},\,i}$; (4) we perform a preliminary validation, i.e. we quantify the model performances with respect to the sample of night used for the calibration; (5) we perform a model validation, i.e. we quantify the model performances with respect to a totally independent sample of nights. Step (4) revealed to be useful in some circumstances because, as we will see in Section 5, in coincidence of the calibration sample we could access, in some cases, measurements taken simultaneously with different and independent instruments (GS, DIMM and/or MASS). This permitted us to better appreciate the model performances and model uncertainties.

### 4.2 $C_N^2$ algorithm

The algorithm used for the $C_N^2$ is described as follows:

$$C_N^2(z)=0.58\times\varphi_3(z)\left[\frac{80\times10^{-6}\times P(z)}{T(z)\times\theta(z)}\right]^2\times L(z)^{4/3}\times\left[\frac{\partial\theta(z)}{\partial z}\right]^2, \tag{1}$$

where $P$ is the atmospheric pressure, $T$ the temperature, $\theta$ the potential temperature, $\varphi_3$ a dimensionless function depending on the thermal and dynamic stability of the atmosphere proportional to the inverse of the Prandtl number and introduced in MESO-NH by Redelsperger & Sommeria (1981) and $L$ the mixing length [more precisely the BL89 mixing length (from Bougeault & Lacarrère 1989) defined as follows: at any level $z$ in the atmosphere, a parcel of air of given turbulence kinetic energy $e(z)$ can move upwards ($l_{\rm up}$) and downwards ($l_{\rm down}$) before being stopped by buoyancy forces]. These distances are defined by

$$\int_z^{z+l_{\rm up}}\frac{g}{\theta_{v,{\rm ref}}}\left(\theta_v(z)-\theta_v(z')\right){\rm d}z' = e(z) \tag{2}$$

and

$$\int_{z-l_{\rm down}}^{z}\frac{g}{\theta_{v,{\rm ref}}}\left(\theta_v(z')-\theta_v(z)\right){\rm d}z' = e(z), \tag{3}$$

where $L$ is defined as

$$L=\left[\frac{(l_{\rm up})^{-2/3}+(l_{\rm down})^{-2/3}}{2}\right]^{-3/2}, \tag{4}$$

$\theta_v$ is the virtual potential temperature and $\theta_{v,\,{\rm ref}}$ is $\theta_v$ of reference related to the anelastic system. It represents the hydrostatic conditions (where the density depends just on the height $z$). We will call hereafter this algorithm '$C_N^2$ reference algorithm' that is very similar to that proposed by Masciadri et al. (1999a). For completeness, we refer to Appendix A to discuss the few technical differences among the two, which are however irrelevant for the rest of the analysis. To improve the peak-to-valley temporal evolution of some astroclimatic parameters (free-atmosphere seeing, isoplanatic angle), we corrected the $C_N^2$ algorithm of equation (1) by a factor that takes into account the wind shear according to equations (5)–(7):

$$\text{for } z<700\text{ m},\ C_N^2(z)^* = C_N^2(z) \tag{5}$$

$$\text{for } z>700\text{ m},\ C_N^2(z)^* = A(z)C_N^2(z) \tag{6}$$

with

$$A(z)=\begin{cases} A_1(z)=\left[\left(\frac{{\rm d}V_x}{{\rm d}z}\right)^2+\left(\frac{{\rm d}V_y}{{\rm d}z}\right)^2\right]^\beta\Big/\left\langle\left[\left(\frac{{\rm d}V_x}{{\rm d}z}\right)^2+\left(\frac{{\rm d}V_y}{{\rm d}z}\right)^2\right]^\beta\right\rangle \\ \text{or} \\ A_2(z)=\left[\left(\frac{{\rm d}V}{{\rm d}z}\right)^2\right]^\beta\Big/\left\langle\left[\left(\frac{{\rm d}V}{{\rm d}z}\right)^2\right]^\beta\right\rangle \end{cases} \tag{7}$$

with $\beta\in[\frac{1}{2},1,\frac{3}{2}]$. $V_x(z)$ and $V_y(z)$ are the horizontal components of the wind speed. $V(z)$ is the module of the horizontal wind. A forthcoming technical paper focused on this new $C_N^2$ algorithm will treat in detail the motivation for the modification of the $C_N^2$ algorithm, scientific justifications for such a modified algorithm, impact on model performances and perspectives for further improvements. In this paper, we intend to trace the state of the art of the model performances using its best configuration in application to the ground-based astronomy, i.e. the most interesting aspect for astronomers. We limit us to say that the new algorithm improves the peak-to-valley spatio-temporal variability of the model during the night providing a better correlation with measurements.

The threshold of 700 m was selected because in the low part of the atmosphere the problem of the spatio-temporal variability is absent. To test the sensibility of the new algorithm with respect to the selected threshold, we used as a threshold 700 m and 600 m (close to the threshold in which model behaviour changed) and it has been verified that results differed by negligible quantities. This means that we are not very sensitive to the threshold within at least a hundred of metres. On a subsample of nights, we tested $A_1(z)$ and $A_2(z)$ in the new $C_N^2$ algorithm formulation and we found negligible differences on the correction factor. We chose to use $A_2(z)$. Moreover, the best fit was obtained with $\beta=\frac{3}{2}$. The whole analysis presented in this paper is based on this 'new $C_N^2$ algorithm' in which all the $C_N^2$ profiles have been corrected by $A_2(z)$ with $\beta=\frac{3}{2}$. To give an idea of effect of the factor $A_2(z)$, Fig. 2 shows, in the case of one night, the temporal evolution of the $C_N^2$ in its standard form (equation 1) and after the correction (equations 5–7). In the same picture is also shown the temporal evolution of the factor $A_2(z)$.

## 5 RESULTS: ASTRO-MESO-NH CODE PERFORMANCES

The statistical operators we used in our analysis are the bias, the root-mean-squared error (RMSE) and $\sigma$ that provide fundamental information on the systematic (bias) and statistical uncertainties (RMSE and $\sigma$). Bias and RMSE are defined as

$${\rm Bias}=\sum_{i=1}^{N}\frac{Y_i-X_i}{N} \tag{8}$$

and

$${\rm RMSE}=\sqrt{\sum_{i=1}^{N}\frac{(Y_i-X_i)^2}{N}}, \tag{9}$$

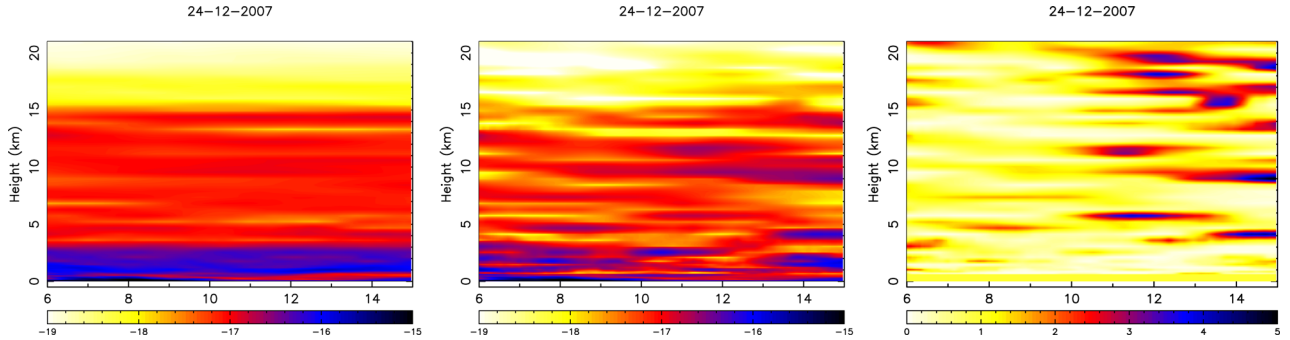


**Figure 2.** Temporal evolution of the $C_N^2$ ($z$) during the night of 24/12/2007 in its standard form (equation 1; left) and with the new algorithm (equations 5–7; centre). On right is shown the temporal evolution of the factor $A_2$. The $x$-axis reports the time in hours from 00:00 UT (six hours after the model initialization at 18:00 UT, i.e. more or less the beginning of the night). The temporal interval shown in the three figures corresponds to the local night. For all figures, a moving average of 1 h has been applied.

where $Y_i$ and $X_i$ are the individual values reconstructed by the model and observed by the instrument, $N$ 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. Starting from the bias and the RMSE, it is possible to retrieve the bias-corrected RMSE, i.e. the $\sigma$:

$$\sigma = \sqrt{\sum_{i=1}^{N} \frac{[(X_i - Y_i) - (\overline{X_i - Y_i})]^2}{N}} = \sqrt{\mathrm{RMSE}^2 - \mathrm{bias}^2} \quad (10)$$

that expresses the intrinsic uncertainty not affected by the bias. We did not use the correlation coefficient because we proved that this operator is not reliable and it can induce misleading conclusions (see Appendix B). Besides, in order to have a more practical estimate of the model score of success, we calculated the 'contingency tables' from which we deduced more sophisticated statistical operators: the per cent of correct detection (PC), the probability to detect a parameter within a specific range of values ($\mathrm{POD}_i$) and the probability of extremely bad detection (EBD). Contingency tables allow for the analysis of the relationship between two or more categorical variables. We refer the reader to Lascaux et al. (2015) for an exhaustive discussion about the role of 'contingency table' of different sizes. We refer the reader to Appendix C for a brief summary of key definitions of contingency tables, PC, POD and EBD to allow the reader to follow the analysis of this paper.

It is also important to mention that we could prove (see Masciadri, Lascaux & Fini 2013b for this issue) that the model calibration for seeing and isoplanatic angle is 'season dependent'. With this we mean that the model should be calibrated with different data related to the two seasons: winter [April–September] and summer [October–March]. However, in our study the unique available GS measurements for the calibration refer to the site testing campaign (PAR2007) performed in November and December, i.e. in summer. Therefore, for the seeing and isoplanatic angle, we report and discuss here only results we obtained for the validation sample in the summer time, i.e. the reliable results. We assume that, using a subsample of SCIDAR measurements taken in winter time, the model will be able to be calibrated for this season.[8] The issue of the dependence of the model calibration from the season is almost not observed on the wavefront coherence time. We think that the reason is due to the fact that $\tau_0$ depends on two parameters: the $C_N^2$ and the wind speed and probably the good model reconstruction of the wind speed can mask and overcome somehow this effect.

Even if the original temporal sampling of measurements and simulation is high (2 min is the temporal sampling of the $C_N^2$ on which depend all the other parameters), all measurements and simulations of $\varepsilon$, $\theta_0$ and $\tau_0$ have been treated with a moving average of 1 h plus a resampling 20 min before to calculate the scattering plots and the contingency tables. The moving average of 1 h has been chosen because we are more interested in the trend of the prediction. It has been observed that the moving average, smoothing out the high frequencies, is more efficient to identify the model and measurement trends than a simple resampling. The high-frequency variability on shorter time-scales is less relevant and useless in our context because the scheduling of scientific programmes cannot be tuned with such high frequency. Astronomers are interested in identifying whether the trend of a parameter is increasing, decreasing or is stationary, in order to be able to take a decision about changing a modality of observation or a scientific programme. The selection of the interval of 20 min for the resampling of measurements is justified by the fact that this is more or less the effective time necessary to switch from one modality of observation to another. The values of 1 h and 20 min have been selected in agreement with ESO staff.[9]

We focused our attention on the analysis of the validation sample that is, obviously, the most important. We reported in this section also some results obtained on the analysis of the calibration but just in some specific contexts when results provided interesting new insights that deserved a discussion. All this to avoid to distract the reader's attention from the main outputs of the study.

### 5.1 Seeing – $\varepsilon_0$

We analyse in this section 3 × 3 contingency tables, with two different sets of thresholds. In the first case (case 1), we use, as thresholds of the total seeing, the first and second tertiles of the cumulative distribution calculated on the sample of nights considered in the validation sample. In the second case (case 2), we use as thresholds of the total seeing 1 arcsec and 1.4 arcsec. A seeing weaker than 1 arcsec and larger than 1.4 arcsec should still provide very useful information to help for the decision to be taken *in loco* and it can be

[8] A site testing campaign covering a whole solar year will be done with a Stereo-SCIDAR in the near future at Cerro Paranal (Osborn et al. 2016). We will be able therefore to perform a dedicated calibration for the winter time for the time in which we will implement the 'demonstrator' of the operational system for the OT forecast at the VLT.

[9] We remind to the reader the importance to have for both, measurements and simulations, an original high temporal sampling. That is very different from having an output each hour that definitely does not permit us to quantify a trend for a stochastic parameter such as the OT.

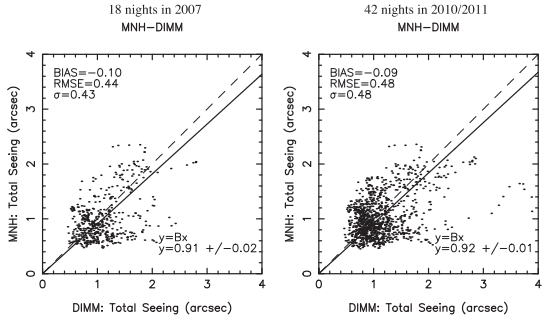


**Figure 3.** Validation samples: scattered plot of the total seeing between MESO-NH outputs and DIMM measurements, for the summer periods of 2007 (case B) on the left and 2010/2011 (case A) on the right.

considered a challenging achievement.[10] Only the DIMM is used as a reference when we treated the large validation sample for the seeing since it is the only instrument providing seeing values on the whole atmosphere. GS measurements are available only for the calibration sample (20 nights related to the PAR2007 campaign). We treated two samples: (A) the whole sample of 53 plus 36 nights, i.e. 89 nights in 2010, 2011 and 2007 and (B) the subsample of 36 nights in 2007, the season for which the calibration of the model was done. Sample B is a subsample of sample A.

### 5.1.1 *Validation sample*

As explained before, we focus our attention on the summer time (42 nights in the sample A and 18 in the sample B). Tables D1 and D2 are the contingency tables for both cases (1 and 2), related to the 42 nights in summer time of sample A. Tables D3 and D4 (in Appendix D) are the contingency tables for both cases (1 and 2), related to the 18 nights in summer time of sample B. Fig. 3 displays the scattered plot of the total seeing between MESO-NH outputs and DIMM measurements, for the 18 nights of summer time of sample B (left), and that of the 42 nights of summer time of sample A (right). We can observe an almost negligible bias and a $\sigma$ of 0.43 arcsec and 0.48 arcsec. Looking at the contingency table just cited, it is possible to see that the values of $POD_i$ are in general smaller than those we obtained for the atmospheric parameters (see Lascaux et al. 2015); however, the external $POD_i$ ($POD_1$ and $POD_3$) are well above the value of 33 per cent corresponding to the random case. From one side, this tells us that, as expected, the prediction of astroclimatic parameters is more difficult. However, from an astronomical point of view, it is much more critical to know in advance the probability to detect the smallest seeing values, i.e. the most interesting and critical range is $POD_1$. $POD_1$ is the probability that the model predicts seeing weaker than the first threshold (case 1) or the probability that the model detects a seeing smaller than 1 arcsec (case 2). In Tables D1–D4, we observe that, in all cases, $POD_1$ has very high percentages. Looking at the largest sample (sample A), the probability to predict a seeing weaker than 1 arcsec is 62.0 per cent (Table D2), while the probability to predict a seeing weaker than the first tertile is 48.1 per cent (Table D1). If we consider just the 2007 data (sample B), results are even better: we observe that the probability to predict a seeing weaker than 1 arcsec is 72.5 per cent (Table D4) while the probability to predict a seeing weaker than the first tertile is 53.1 per cent (Table D3). These values are well above the random cases of 33 per cent. We can say therefore that, in spite of the fact that there is obviously space for improvements, model performances in detecting the weakest values are therefore already very good, particularly in the case in which we relax the threshold from the first tertile to 1 arcsec. We analysed separately the year 2007 because it gave some better results with respect to the total sample (2007/2010/2011). We think this might be due to the fact that the calibration sample belongs to the 2007 (also if the calibration sample is totally independent from the validation sample). This seems to indicate that there is space in the future to further improve the results with more sophisticated calibrations. The fact that $POD_2$ and $POD_3$ provide worse performances is justified by the fact that, in the scattering plot (Fig. 3), the points show a tendency in increasing the dispersion for large values of the seeing. On the other side, this is not necessarily a problem of the model (or only of the model) as we will see in Section 5.1.2.

### 5.1.2 *Calibration sample*

We report here the results for the calibration sample (20-night sample). The interest of this analysis is that, for this sample, we can compare the performance of the model versus an instrument taken as a reference with the dispersion between two different instruments. This exercise provides us a very important reference on the real uncertainty with which we are at present able to estimate the seeing with measurements. This is therefore the ultimate limit over which a model cannot go.

In Fig. 4 are shown the scattering plots of the total seeing between GS and DIMM (left) and between the MESO-NH and GS (centre) and MESO-NH and DIMM (right). We observe that the $\sigma$ value between the two instruments is 0.30 arcsec while those of the model with respect to the instruments is 0.34 and 0.36 arcsec, therefore substantially comparable. Tables D5 and D6 report the contingency tables of the two instruments (GS and DIMM) in which we take, as a reference, one or the other instrument. Table D7 reports the contingency table between MESO-NH and the GS for the total seeing. Table D8 reports the contingency table between MESO-NH and the DIMM for the total seeing. If we look at $POD_1$, i.e. the probability of the model to forecast a seeing weaker than the first tertile, this is 82.1 per cent (Table D7) or 88.1 per cent (Table D8) depending on whether we consider the GS or the DIMM as a reference. These results are even better than those obtained comparing the two different instruments. Indeed, the $POD_1$, i.e. the probability to measure a seeing weaker than the first tertile, is 64.3 per cent (Table D5) or 71.4 per cent (Table D6) depending on the reference (GS or the DIMM). This tells us that the model fails on a percentage that is not larger than the intrinsic uncertainty with which we can estimate the turbulence.

Besides that, to have information on the trends, we calculated the temporal evolutions. Figs E1–E3 show the temporal evolution of the turbulence integrated on the whole atmosphere (see equation 11) as estimated by the ASTRO-MESO-NH model and as measured by the GS and the DIMM in all the 20 nights of the PAR2007 campaign,

$$J = \int_0^\infty C_N^2(h)\mathrm{d}h. \tag{11}$$

It is possible to appreciate how the model, during all the nights, reconstructs in a very satisfactory way not only the quantitative estimate of $J$ but also the general trend (the morphology of the trends is well reconstructed). It is also interesting to observe that it is extremely useful to run simultaneous different instruments

[10] The values of 1 and 1.4 arcsec have been decided after consultation with the ESO MOSE Board and colleagues from the AO group of the Arcetri Astrophysical Observatory.

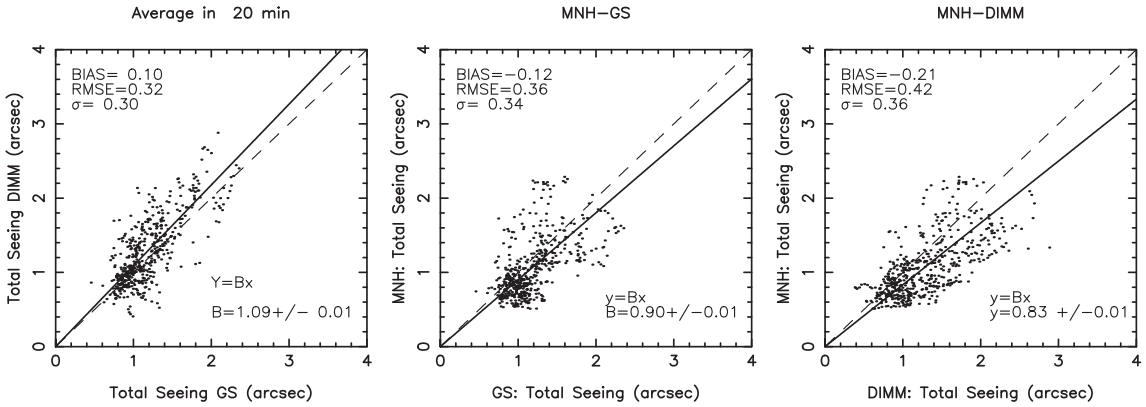


**Figure 4.** Scattered plot of the total seeing between MESO-NH outputs and DIMM and GS measurements (centre and right) and between the GS and the DIMM (left) for the calibration sample of 20 nights.

because in some cases, the model appears more in agreement with one instrument than the other. Therefore, the fact to have more than one instrument permits us to better appreciate the real model performances. In other words, the intrinsic accuracy with which we can at present quantify the OT with an instrument is lower than that achievable for a classical atmospheric parameter (such as temperature or wind); therefore, the use of multi-references is extremely important in the analysis.

### 5.2 Wavefront coherence time – $\tau_0$

In the case of $\tau_0$, we could compare the ASTRO-MESO-NH estimates with measurements coming from the DIMM and the MASS. When the model is compared to the DIMM, we treated the whole 89 nights (2010, 2011 and 2007). When it is compared to MASS, the analysis is limited to 48 nights because unfortunately there were no MASS measurements on 2007. For $\tau_0$, we could perform an analysis of model performances on summer as well as on winter time because the model calibration did not show particular limitations for this parameter.

#### *5.2.1 Validation sample*

All the scattered plots for the wavefront coherence time $\tau_0$ are summarized in Fig. 5. As can be seen in Fig. 5(a), there is a not negligible bias between the DIMM and MASS measurements. This indicates that one or both instruments still have some problems to be arranged for $\tau_0$. We do not want to enter in a deep discussion on this issue because it is visibly not the main topic of the paper. What is important is that, assuming that the problem on the specific instrument is identified, this bias between the model and the instrument can be corrected by a multiplicative coefficient. The bias appears indeed visibly multiplicative. Even in case we can establish that the MASS is the correct one and the DIMM is wrong, a multiplicative coefficient which corrects the regression line according to Figs 5(f) and (g) can be implemented. The panels to be retained are therefore panels (b) and (c) if DIMM is considered as reliable and panels (f) and (g) if MASS is considered reliable. For what concerns the scattering plots, we note that the values of $\sigma$ obtained between DIMM and MASS and between MESO-NH and MASS or MESO-NH and DIMM are comparable. In some cases, the $\sigma$ of MESO-NH with respect to the instrument (for example the $\sigma$ of MESO-NH and DIMM in summer and winter is respectively 1.66 ms and 1.20 ms) is even smaller than that obtained considering the two instruments ($\sigma = 1.73$ ms).

From a different perspective, Tables D9 and D10 report the contingency tables between MASS and DIMM for $\tau_0$ on the whole sample of 48 nights of the 53 simulated nights of the period 2010–2011 in which we had simultaneous measurements from DIMM and MASS. In the first table, the DIMM is taken as a reference, and in the second table the MASS is taken as a reference. The very different values of $POD_1$ and $POD_3$ in both tables just reflect the problem we discussed for the scattering plot. MASS and DIMM have a visible not negligible bias that, however, in phase of analysis of model performances, can be corrected with a multiplicative coefficient [see panels (f) and (g) of Fig. 5]. Table D11 is the contingency table between MESO-NH and DIMM for the 42 nights (over the total of 89 nights of the validation sample) related to summer. Table D12 is the contingency table between MESO-NH and DIMM for the 47 nights (over the total of 89 nights of the validation sample) related to winter. Table D13 is the contingency table between MESO-NH and MASS for the 21 nights (over the total of 48 nights of the validation sample) related to summer. Table D14 is the contingency table between MESO-NH and MASS for the 27 nights (over the total of 48 nights of the validation sample) related to winter. Tables D13 and D14 have been calculated using obviously the data corrected as in Figs 5(f) and (g).

As we did for the seeing, we focused our attention on the most crucial $POD_i$ for the $\tau_0$. Differently from the seeing, the most critical and interesting POD for $\tau_0$ is $POD_3$ that is the probability to forecast the $\tau_0$ that is larger than the second tertile of the cumulative distribution. When the $\tau_0$ is large, an AO system can be run at low temporal frequency. This not only allows for better AO performances but it also facilitates the achievement of scientific programmes that could not be completed otherwise. Looking at contingency tables concerning the model (Tables D11–D14), we observe that we obtain, in all cases, very good values of $POD_3$. $POD_3 = 70.7$ per cent for MESO-NH-DIMM (summer time, Table D11) and 65.6 per cent for MESO-NH-DIMM (winter time, Table D12). $POD_3 = 71.9$ per cent for MESO-NH-MASS (summer time, Table D13) and 78.2 per cent for MESO-NH-MASS (winter time, Table D14). Also the values of $POD_1$ that indicates the probability to detect a $\tau_0$ weaker than the first tertile, is very satisfactory and included in the range [48–69] per cent depending of the instrument we take as a reference.

For what concerns the discrepancy/bias between MASS and DIMM at present, we can say that Masciadri et al. (2014) proved that

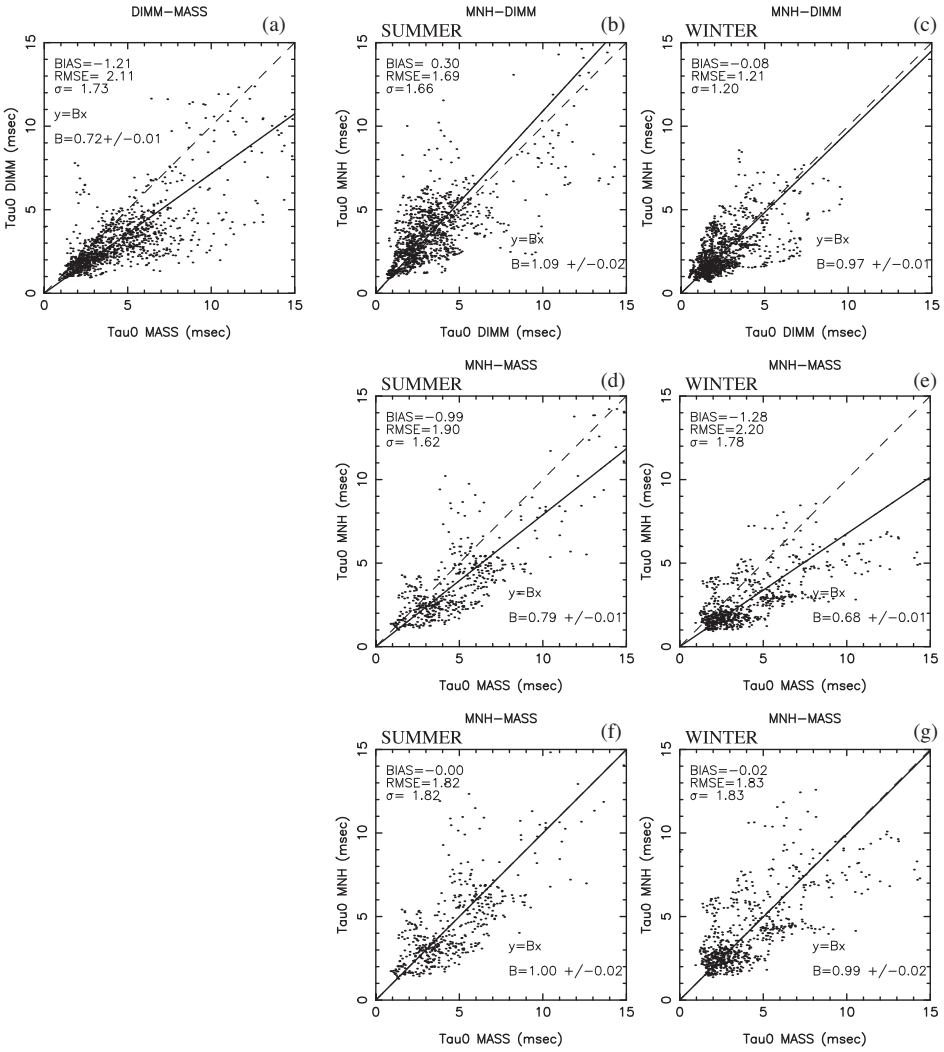


**Figure 5.** Scattered plots (validation sample) of the wavefront coherence time $\tau_0$ between (a) MASS and DIMM – 48 nights in 2010/1011 (there are no MASS measurements in 2007); (b) MESO-NH and DIMM in summer – 42 nights in 2010/2011; (c) MESO-NH and DIMM in winter – 47 nights in 2010/2011; (d) MESO-NH and MASS in summer – 21 nights in 2010/2011; (e) MESO-NH and MASS in winter – 27 nights in 2010/2011; (f) the same as (d) but with the MASS corrected using the coefficient of the regression line from (d); and (g) the same as (e) but with the MASS corrected using the coefficient of the regression line from (e). For the model performances, the key panels to take into account are (b), (c), (f) and (g). See extended discussion in the text.

$\tau_0$ from MASS appears reliable. Only when values are large (larger than 5 ms), the instrument shows a tendency to slightly overestimate $\tau_0$. On the other side, the DIMM is not a vertical profiler and for this reason is not an instrument particularly suitable to quantify $\tau_0$, a parameter strongly dependent on the vertical stratification of $C_N^2$ and wind speed. The method used to retrieve $\tau_0$ from the DIMM has been proposed by Sarazin & Tokovinin (2002) but the same authors affirm that the solution is just based on approximations and has to be considered with the suitable precautions.

### 5.3 Isoplanatic angle – $\theta_0$

In the case of $\theta_0$, we could compare the ASTRO-MESO-NH estimates with measurements coming from the DIMM and the MASS. However, as we have done for the seeing, we present only comparison model/measurements obtained in summer time because they are the only useful results. This is because $\theta_0$ require a separate calibration in summer and winter time as well as the seeing $\varepsilon$.

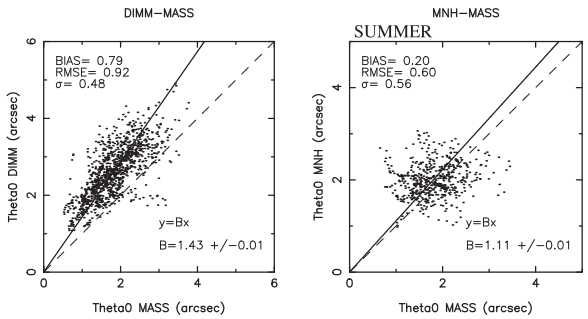


**Figure 6.** Scattered plots of the isoplanatic angle $\theta_0$ between DIMM and MASS (left) and between MASS and MESO-NH (right).

#### 5.3.1 *Validation sample*

Fig. 6 shows the scattered plots of $\theta_0$ between DIMM and MASS on the subsample of 44 nights over the 53 nights in 2010/2011 in which there were simultaneous measurements (left) and between

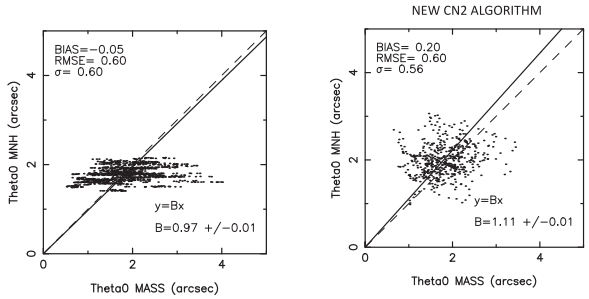


**Figure 7.** Scattered plots of the isoplanatic angle $\theta_0$ between MASS and MESO-NH before (left) and after (right) the implementation of the new $C_N^2$ algorithm.

MASS and the model for the summer sample corresponding to 21 nights over the total sample of 48 nights in 2010/2011 (right).

Tables D15 and D16 are the contingency tables for $\theta_0$ between MASS and DIMM for the 44 nights in 2010/2011. Table D17 is the contingency table between MESO-NH and MASS, during summer (21 nights over 48 nights) for $\theta_0$. We observe that, in a similar way as for $\tau_0$, the MASS and DIMM distribution of points in Fig. 6 presents a not negligible bias. However, we know that a study from Masciadri et al. (2014) proved that the $\theta_0$ from MASS is well correlated to that of GS (at least on a sample of 14 nights) and $\theta_0$ estimates from MASS can be considered reliable. No biases have been observed between MASS and GS. We think therefore that, for the same arguments/reasons presented in the case of $\tau_0$, we can reasonably think that the problems are for $\theta_0$ in DIMM measurements (and not in MASS). The method proposed by Sarazin & Tokovinin (2002) to estimate $\theta_0$ and $\tau_0$ with a DIMM is based on an approximation and cannot be considered very accurate (as said by the same authors). We concentrated here therefore only on a comparison model/MASS measurements. Observing the scattering plots of Fig. 6, we note that the two figures have a comparable value for $\sigma$ (0.48 arcsec versus 0.56 arcsec) even if measurements are better correlated among them than ASTRO-MESO-NH model versus MASS.[11] The cloud of points in the panel on the left is indeed better elongated along the regression line than panel on the right.

On the other side, it is possible to observe that the model has visibly improved its performances for the reconstruction of $\theta_0$ using the new algorithm of the $C_N^2$ (equation 7). Fig. 7 shows indeed the distribution before and after the modification of the algorithm. It is evident that before the modification of the algorithm the model had serious problems in well reconstructing the spatio-temporal variability of the turbulence in the high part of the atmosphere during the night. $\theta_0$ is indeed mainly driven by the turbulence in the free atmosphere. $\theta_0$ covers now almost on the same interval of values observed with the MASS (typically [1,3] arcsec).[12] Even if there is still space for improvements in the algorithm, we have now comparable dispersion of model values with respect to measurements.

If we analyse the contingency tables and the values of $\mathrm{POD}_i$ (Table D17), we observe that they are somehow weaker than those of the seeing $\varepsilon$ and $\tau_0$. However, the most interesting $\mathrm{POD}_i$ from an astronomical point of view for $\theta_0$ is $\mathrm{POD}_3$, that is the probability that the model detects a $\theta_0$ larger than the second tertile. $\mathrm{POD}_3$ permits us to identify temporal windows with large $\theta_0$ and those

[11] We recall that $\sigma$ is not affected by the bias.

[12] We recall that measurements and model outputs are both treated with a moving average of 1 h. Raw variability is larger in both measurements and simulations.

**Table 2.** Typical contingency table using the climatological median value (as N is undefined, we display the per cent only).

| Climatology whatever N. of nights | | OBS par. < X1 | OBS X1 < par. < X2 | OBS par. > X2 |
|---|---|---|---|---|
| Median value | par. < X1 | 0 | 0 | 0 |
| | X1 < par. <X2 | 33 | 33 | 33 |
| | par. > X2 | 0 | 0 | 0 |

PC = 33 per cent; EBD = 0 per cent.
$\mathrm{POD}_1 = 0$ per cent; $\mathrm{POD}_2 = 100$ per cent; $\mathrm{POD}_3 = 0$ per cent.

intervals are favourable to ground-based observations supported by GLAO, MCAO and WFAO systems. Our system proves therefore to be useful in particular for the Adaptive Optics Facility (AOF – Kuntschner et al. 2012; Arsenault et al. 2014; Madec et al. 2016), the ESO AO system for the unit UT4 of the VLT that will use an adaptive secondary and it is composed of the two modules GALACSI and GRAAL. Our estimate of $\mathrm{POD}_3$ is of the order of 58.8 per cent, therefore quite satisfactory and almost the double of the 33 per cent of the random case. $\mathrm{POD}_1$ is still weak and this tells us that there is still space for improvements (even if $\mathrm{POD}_1$ is definitely less critical from the point of view of the service mode).

### 5.4 Discussion

We report here a few considerations to complete the analyses presented.

(i) The use of $\mathrm{POD}_i$ (much more than PC that is associated with model performances as a whole) can support us to quantify a concrete score of success calculated at posteriori on a rich statistical sample of estimates. We observed that if we take into account all the $\mathrm{POD}_i$ ($\mathrm{POD}_1$, $\mathrm{POD}_2$ and $\mathrm{POD}_3$) we find sometimes very satisfactory results, other times results are more modest. If we limit us to the important $\mathrm{POD}_i$ for each astroclimatic parameter, i.e. $\mathrm{POD}_1$ for the seeing $\varepsilon$ and $\mathrm{POD}_3$ for $\theta_0$ and $\tau_0$, results are always very satisfactory and, in all cases, the relative PODs are much better than 33 per cent. In many cases, they are two times better and in some cases even more than this. Our results indicate therefore that the impact on the service mode is concrete. In the previous session, we quantified that.

What about the impact of our forecast method with respect to other strategies? We cite here just a couple of alternative approaches. The first method is to have no system at all for the forecast that corresponds to the random case. In that case, all the $\mathrm{POD}_i$ are equal to 33 per cent. Our system is visibly better than that. To appreciate the difference we would obtain using as a forecast method the climatological median (of whatever astroclimatic parameter), the reader can have a look at the contingency table shown in Table 2. The method of the climatological median provides a probability to detect the parameter we intend to study of 100 per cent for values between the first and second tertiles but a probability of 0 per cent for values smaller or larger than the first and second tertiles. These latter intervals are, however, the most interesting one from an astronomical point of view. The real critical temporal windows in which we are interested are those characterized by a very weak seeing during which to select most challenging programmes, for example for the search and characterization of exo-planets or, in alternative, those temporal windows with a very large $\tau_0$ that can permit us to use the AO system in their best configuration with the best performances permitting the best sky coverage or, in alternative, the temporal windows with the largest $\theta_0$ that guarantee us very

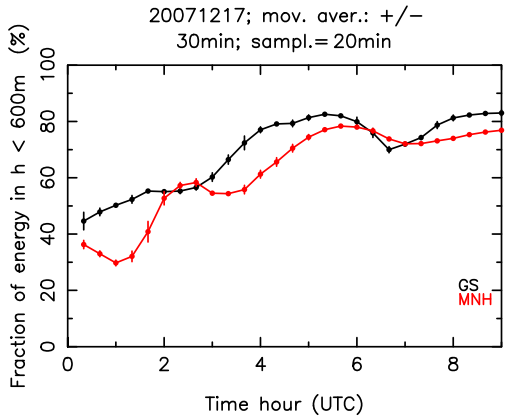


**Figure 8.** Temporal evolution of the fraction of turbulent energy in the first 600 m with respect to the total turbulence energy on the whole 20 km as measured by the GS and as estimated by the ASTRO-MESO-NH model during one night of the PAR2007 site testing campaign.

weak turbulence conditions in the free atmosphere. Those are the best conditions to use WFAO systems on very large field of view for science cases such as, for example, resolved stellar population studies and photometry in globular clusters (Fiorentino et al. 2016). If we trust in the climatology median, we will be more efficient in detecting the right time in which the parameter we are studying is close to the median climatology but we will never identify those precious temporal windows close to the extremes values. The ASTRO-MESO-NH model is visibly much more effective in this respect. Other methods might be envisaged such as the prediction for persistence that assumes that the value of a parameter remains equal to itself. In reality, one might imagine many different versions of such a method: (a) the value at the beginning of the night remains the same for the whole night, (b) the seeing remains the same on intervals of time inside the duration of the night, (c) others. We planned to study these cases in the near future when we will have a larger statistics and when a more clear strategy of the Service Mode will be defined.[13] It is worth saying that, in general, an atmospheric model can be more effective when we want to put in evidence some changes with respect to the previous conditions. Under conditions of perfect stability, an atmospheric model is less useful.

(ii) An interesting parameter that can be estimated by the ASTRO-MESO-NH model is the fraction of the turbulent energetic budget contained in the first 600 m with respect to the turbulent energetic budget included in the whole 20 km ($J_{BL}/J_{TOT}$). Fig. 8 shows the temporal evolution of $J_{BL}/J_{TOT}$ during the whole night of PAR2007 campaign as measured by the GS and as reconstructed by the model. In the figure, it is visible that the model well reconstructs the growing of $J_{BL}/J_{TOT}$ during the night. This figure of merit is extremely useful for the observations to be done with whatever WFAO system (MCAO, GLAO, MOAO, LTAO). In the context of the MOSE project, we performed this analysis in perspective to an application to AOF at the VLT (Kuntschner et al. 2012; Madec et al. 2016).

(iii) In general, model performances tend to decrease if it is initialized and forced with 'forecasts' instead of 'analyses' coming from the GCM. The level of the deterioration depends in general on the forecast delay. However, we could prove (Masciadri, Lascaux & Fini 2015) that, for all the atmospheric parameters and for the forecast delay we selected for the operational configuration, the PODs obtained with the ECMWF forecasts are weakly worse than those obtained with the ECMWF analyses, and sometime they present the same value. The impact seems therefore weak and therefore comfortable. Due to the fact that the $C_N^2$ depends on the atmospheric parameters, we deduce that the same conclusion is valid for the $C_N^2$.

(iv) In the analysis we presented, it appears evident that the necessity of rich statistical sample of measurements taken with different instruments is fundamental to progress in this research. Also we observed that the uncertainty in quantifying the astroclimatic parameters with measurements is not negligible in some cases. We observed for example that the $\sigma$ between GS and DIMM (PAR2007 campaign) was of 0.30 arcsec. It has therefore no sense to pretend a better result than this for a model.

(v) Results obtained in this study indicate that it is very important to be able to run routinely and simultaneously above all observatories as many as possible different instruments based on independent principles.

## 6 HIGH VERTICAL RESOLUTION $C_N^2$ PROFILES

All results and analyses presented so far implied the use of a model configuration having 62 vertical levels covering the whole atmosphere (roughly 20 km above the ground). The vertical resolution of the ASTRO-MESO-NH model is of the order of 600 m above 3.5 km above the ground (i.e. in the central and high part of the atmosphere). It is evident that the higher the model vertical level number, the higher is the computing effort. A $C_N^2$ with a resolution of ∼600 m in the central part of the atmosphere can be considered a good baseline because it is enough for most of applications of operations in service mode and in AO context. Also we note that the model configuration we selected is suitable to perform nightly automatic operational forecasts using a relative cheap hardware, in other words to implement an operational forecast system that does not require super clusters. However, for most sophisticated AO systems, typically the WFAO, i.e. GLAO, MCAO, LTAO, MOAO it is extremely important to know $C_N^2$ profiles with vertical resolutions as high as 150 m (Costille & Fusco 2012). The tomographic error is, indeed, strongly dependent on the $C_N^2$ profiles and a lack of knowledge of the vertical stratification with a sufficiently high vertical resolution can induce large tomographic errors. These AO systems are currently tested and implemented on the present-generation facilities, and they are planned to be implemented on the ELTs in the next decades. For example at the E-ELT the near-infrared spectrograph HARMONI (Thatte et al. 2016) will work with a LTAO system; MAORY (Diolaiti et al. 2016) is a MCAO system that will feed the imager MICADO (Davies et al. 2016) and the multi-object spectrograph MOSAIC (Hammer et al. 2016) will be feed by a MOAO system. The wider is the field of view, the higher is the resolution with which one has to know the vertical stratification of the turbulence in advance. The most demanding AO system is the MOAO that requires the highest resolution (see a summary in Masciadri et al. 2013c).

For this reason, we therefore put a great effort in trying to improve the vertical resolution of the $C_N^2$ in perspective of an implementation in an operational forecast system. In Fig. 9, we present the fundamental results we achieved. On the top we have the temporal evolution of the $C_N^2$ along a night obtained by the ASTRO-MESO-NH model having 62 vertical levels (left) and 173 levels (right) both with the new $C_N^2$ algorithm. On the bottom is reported the average $C_N^2$

[13] This will be discussed with the Science Operation Division of ESO and the ESO MOSE Board.

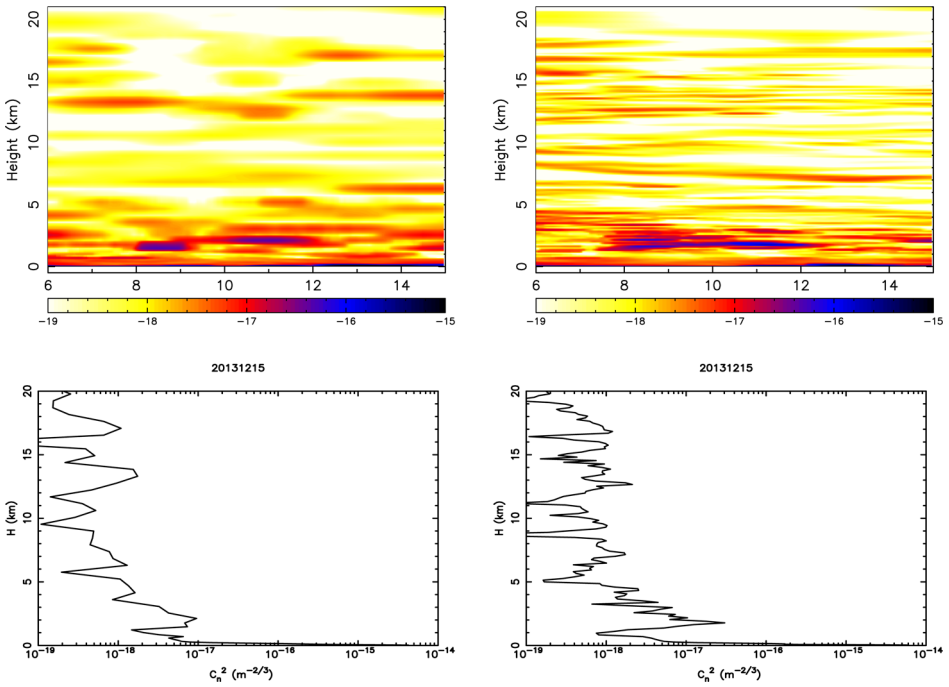


**Figure 9.** Top: temporal evolution of the vertical $C_N^2$ profiles simulated by the ASTRO-MESO-NH model with different model vertical resolutions [62 levels (left), 173 levels (right)] related to the night 15/12/2013. Bottom: corresponding average $C_N^2$ profile related to the $C_N^2$ temporal evolution on the top.

of the night related to the respective simulations. It is well visible that the simulations done with 173 vertical levels provide a $C_N^2$ with a much larger number of turbulent layers. Also the turbulent layers are much thinner than those resolved by the 62 level configurations. This result was possible thanks to the new $C_N^2$ algorithm presented in this paper (see Section 7) that permits us to better represent the spatio-temporal variability of the vertical stratification of the OT in the 20 km above the ground. The $C_N^2$ shown in Fig. 9 proves the ability of our technique in reconstructing a vertical stratification of the OT with a resolution that opens crucial and exciting new perspectives in the field of the WFAO. We calculated, indeed, that we should be able to use this model configuration also for an operational model configuration. Since the model calibration depends on its configuration, we planned to calibrate the model with this configuration in the near future. In the meanwhile, this $C_N^2$ profile has been used for a preliminary study aiming to test the performances of HARMONI (Neichel et al. 2016) and for a more extended study aiming to address the ultimate limitations of the WFAO technique.[14]

## 7 CONCLUSIONS

In this paper, we presented the main results we obtained on the model performances in forecasting the OT above Cerro Paranal in the context of the MOSE project. We analysed the three main integrated astroclimatic parameters as originally planned: seeing ($\varepsilon$), isoplanatic angle ($\theta_0$) and wavefront coherence time ($\tau_0$) that are fundamental in the AO optimization and planning of ground-based observations. The ASTRO-MESO-NH model has been first calibrated using measurements taken with a GS. The model performances have been quantified using measurements coming from samples of nights completely independent from the calibration one. The results we obtained on the validation samples are summarized in a nutshell.

(i) The wavefront coherence time ($\tau_0$) is the parameter that, at present time, is predicted with the best score of success. All the $\mathrm{POD}_i$ ($i = 1$, 2 and 3) are satisfactory and comparable to those obtained among different instruments, in some cases even better. From an astronomical point of view, the most interesting $\mathrm{POD}_i$ for $\tau_0$ is $\mathrm{POD}_3$ that is the probability to detect a $\tau_0$ larger than the second tertile of the cumulative distribution of measurements. We are therefore interested in the temporal windows in which the $\tau_0$ is large. $\mathrm{POD}_3$ on the validation sample is included in the range [72 per cent–78 per cent] if we considered the most reliable instrument as a reference, i.e. the MASS. The range includes the analysis done in different seasons: summer and winter. The samples we are dealing about include 42 nights (summer) and 47 nights (winter). $\mathrm{POD}_1$ is included in the range [48 per cent–69 per cent] (slightly smaller than $\mathrm{POD}_3$ but still very good) using as a reference the same instrument. Another not negligible positive element for $\tau_0$ is that this score of success has been obtained with just one model calibration valid for the whole year (i.e. not a calibration for summer and one for winter) and we analysed data in summer as well as in winter. It is highly probable that this score of success would further improve with a seasonal model calibration. The ability in detecting a large $\tau_0$ in advance is extremely useful for the following reasons. A large $\tau_0$ permits an AO system to run at low temporal frequency and to use faint stars as guide stars. A low temporal frequency permits indeed to better sample light coming from stars that, in the presence of a smaller $\tau_0$, cannot be used as guide stars. A large $\tau_0$ represents an advantage for the WFAO because the number of available guide stars, in general, increases (i.e. the sky coverage increases) and because a higher Strehl ratio (SR) is reachable and more ambitious scientific projects can be carried out. The advantages are also for the classical SCAO configuration (not only for WFAO) because it permits to observe fainter stars reaching a higher SR.

[14] The study has been presented by Thierry Fusco et al. at SPIE Conference, Adaptive Optics Systems V, 2016, 9909-273.

(ii) The total seeing ($\varepsilon$) revealed to be sensitive to the seasonal model calibration. This means that the model needs a twofold calibration: one for summer and one for winter. We provided therefore in this study only the score of success for the summer time because the calibration sample belongs to this season. However, the goal of this study was to prove the feasibility of the OT forecast; therefore, results we obtained fit perfectly with objectives. We will be able to use the measurements of the site testing campaign planned for Paranal in the near future with a Stereo-SCIDAR (Osborn et al. 2016) to calibrate the model for the winter time. In the case of the seeing, the most interesting POD from a scientific point of view is $POD_1$. The probability to detect the $\varepsilon < 1$ arcsec is in the range [62 per cent–72 per cent]. The probability to detect the $\varepsilon <$ of the first tertile is in the range [48 per cent–53 per cent]. The range of values in this case have the following meaning: the most conservative values (62 per cent and 48 per cent) have been obtained with the sample of 42 nights (summer) in the 2010/2011 years. The most optimistic values have been obtained with the subsample of the 18 nights (summer) of 2007 (the same year in which the calibration sample belongs). This indicates that the identification of a sample of measurements used for the calibration is an important issue in the process. We envisage an as extended as possible sample of nights for that in the future. The instrument used as a reference is in this case the DIMM. The advantage in detecting in advance temporal windows with weak $\varepsilon$ is obvious. The best seeing conditions guarantee the highest SR of an AO system and the most challenging scientific programmes such as the search and the characterization of extra-solar planets would definitely gain from this condition. Besides we note that the selection of the temporal windows with excellent seeing will be even more critical for the AO in the visible (Close 2016) that is much more dependent on the conditions of the turbulence than the AO in the near-infrared. This new typology of AO represents the new frontier of ground-based observations.

(iii) The isoplanatic angle ($\theta_0$) revealed to be sensitive to the seasonal model calibration as is the case for the seeing too. We provide therefore in this study only the score of success of the model related to the summer time for the same reason described for the seeing. From a scientific point of view, for $\theta_0$ the most interesting $POD_i$ is $POD_3$ that is the probability to detect a $\theta_0$ larger than the second tertile. We obtained a very promising 58 per cent for $POD_3$ on the validation sample using, as a reference, the MASS. The sample we are dealing about includes 21 nights (summer) of (2010/2011). No measurements with MASS are available for 2007. For completeness, we also note that, at present, the probability to detect an isoplanatic angle smaller than the first tertile ($POD_1$) is not sufficiently good and we plan to undertake further investigations to improve this score of success. At the same time, we highlight the fact that, from a scientific point of view, in the context of the service mode $POD_3$ is definitely more interesting than $POD_1$. We think therefore that, at present, our forecasting system is definitely able to provide useful information for the service mode. A large value of the $\theta_0$ is particularly useful for whatever kind of WFAO (GLAO, LTAO, MCAO, MOAO). If we are able to detect temporal windows with not so much turbulence in the free atmosphere (large $\theta_0$), we should be able to give a crucial input for the service mode. Observations supported by AO systems of extended and crowded fields such as clusters would definitely benefit from this condition if it would be known in advance.

(iv) The general commentary is that, as expected, model performances in reconstructing the astroclimatic parameters are not as good as in reconstructing the atmospheric parameters [treated in Masciadri et al. (2013a) and Lascaux et al. (2013, 2015)]. However, if we look at the model performances in reconstructing values of $\varepsilon$, $\tau_0$ and $\theta_0$ in the most critical and interesting ranges of values for each of them, model performances are very good. Quantitative results indicate that the ASTRO-MESO-NH model performances in forecasting the OT have definitely a fundamental impact on the service mode of top-class telescopes.

(v) We proved to be able to forecast $C_N^2$ profiles with a vertical resolution as high as 150 m. As discussed in the paper, such an achievement definitely opens interesting new perspectives for the development of the most sophisticated AO systems planned for the next-generation ground-based telescopes.

In terms of perspectives, we intend to continue our studies/work on the $C_N^2$ algorithm in order to improve the model performances. There is still space to improve the mixing length definition in stable regime, and this is a research line that we will follow with attention. The access to new measurements of the stratification of the turbulence ($C_N^2$ ) that are planned to be done above Paranal will be important for a calibration of the model on the whole year. A richer sample of measurements will be also useful to enlarge the typology of analysis. Our kind of research will definitely take advantage from the fact to have as many as possible different reliable instruments running simultaneously above a site because this permits us to better constrain the model but also to better quantify the uncertainty with which we can estimate the turbulence. In perspective to an operational implementation, we planned, in agreement with the ESO Science Operational division and the ESO MOSE Board, some more dedicated studies on the impact of the forecast on the service mode.

## ACKNOWLEDGEMENTS

This study is co-funded by the ESO contract: E-SOW-ESO-245-0933 (MOSE Project). We are very grateful to the ESO Board of MOSE (Marc Sarazin, Pierre-Yves Madec, Florian Madec and Harald Kuntschner) for their constant support to this study. Part of measurements used in this study were part of the PAR2007 site testing campaign. A great part of simulations are run on the HPCF cluster of the European Centre for Medium Weather Forecasts (ECMWF) – Project SPITFOT.

## REFERENCES

Abahamid A., Jabiri A., Vernin J., Benkhaldoun Z., Azouit M., Agabi A., 2004, A&A, 416, 1193
Arakawa A., Messinger F., 1976, GARP Tech. Rep., 17, WMO/ICSU, Geneva, Switzerland
Arsenault R. et al., 2014, Proc. SPIE, 9148, 914802
Asselin R., 1972, Mon. Weather Rev., 100, 487
Assemat F. et al., 2004, Proc. SPIE, 5237, 211
Avila R., Cuevas S., 2009, Opt. Express, 17, 10926
Beckers J. M., 1988, ESO Conference and Workshop Proceedings, Vol. 30, Very Large Telescopes and their Instrumentation. European Southern Observatory, Garching, p. 693
Bougeault P., Lacarrère P., 1989, Mon. Weather Rev., 117, 1872
Cherubini T., Businger S., 2013, J. Appl. Meteorol. Climatol., 52, 498
Cherubini T., Businger S., Lyman R., 2011, in Masciadri E., Sarazin M., eds, Optical Turbulence – Astronomy Meets Meteorology. Imperial College Press, London, p. 196
Close L., 2016, Proc. SPIE, 9909, 99091E
Costille A., Fusco T., 2012, Proc. SPIE, 8447, 844757
Coulman C. E., Vernin J., Coqueugniot Y., Caccia J. L., 1988, Appl. Opt., 27, 155

Cuxart J., Bougeault P., Redelsperger J.-L., 2000, Q. J. R. Meteorol. Soc., 126, 1
Dali Ali W. et al., 2010, A&A, 524, A73
Davies R. et al., 2016, Proc. SPIE, 9908, 99081Z
Diolaiti E. et al., 2016, Proc. SPIE, 9909, 99092D
Fiorentino G. et al., 2016, Proc. SPIE, 9909, 990906
Foy R., Labyerie A., 1985, A&A, 152, L29
Gal-Chen T., Somerville C. J., 1975, J. Comput. Phys., 17, 209
Giordano C., Vernin J., Vázquez-Ramió H., Muñoz-Tuñón C., Varela A. M., Trinquet H., 2013, MNRAS, 430, 3102
Hagelin S., Masciadri E., Lascaux F., 2011, MNRAS, 412, 2695
Hammer F. et al., 2016, Proc. SPIE, 9908, 990824
Johnston D. C., Welsh B. M., 1994, J. Opt. Soc. Am. A, 11, 394
Johnston R., Dainty C., Wooder N., Lane R. G., 2002, Appl. Opt., 41, 6768
Kornilov V., Tokovinin A., Vozyacova O., Zaitsev A., Shatsky N., Potanin S. F., Sarazin M. S., 2003, Proc. SPIE, 4839, 837
Kuntschner H., Amico P., Kolb J., Madec P. Y., Arsenault R., Sarazin M., Summers D., 2012, Proc. SPIE, 8448, 844808
Lafore J.-P. et al., 1998, Ann. Geophys., 16, 90
Lascaux F., Masciadri E., Stoesz J., Hagelin S., 2009, MNRAS, 398, 1093
Lascaux F., Masciadri E., Hagelin S., 2010, MNRAS, 403, 1714
Lascaux F., Masciadri E., Hagelin S., 2011, MNRAS, 411, 693
Lascaux F., Masciadri E., Fini L., 2013, MNRAS, 436, 3147
Lascaux F., Masciadri E., Fini L., 2015, MNRAS, 449, 1664
Liu L. Y., Giordano C., Yao Y.-Q., Vernin J., Chadid M., Wang H.-S., Yin J., Wang Y.-P., 2015, MNRAS, 451, 3299
Madec P. Y. et al., 2016, Proc. SPIE, 9909, 99090Z
Masciadri E., Egner S., 2006, PASP, 118, 1604
Masciadri E., Jabouille P., 2001, A&A, 376, 727
Masciadri E., Vernin J., Bougeault P., 1999a, A&AS, 137, 185
Masciadri E., Vernin J., Bougeault P., 1999b, A&AS, 137, 203
Masciadri E., Vernin J., Bougeault P., 2001, A&A, 365, 699
Masciadri E., Avila R., Sanchez L. J., 2002, A&A, 382, 378
Masciadri E., Avila R., Sanchez L. J., 2004, Rev. Mex. Astron. Astrofis., 40, 3
Masciadri E., Lascaux F., Lombardi G., Vázquez-Ramió H., 2012, MNRAS, 420, 2399
Masciadri E., Lascaux F., Fini L., 2013a, MNRAS, 436, 1968
Masciadri E., Lascaux F., Fini L., 2013b, MOSE: MOdeling Sites ESO – Report Phase A, E-TRE-INA-245-0001
Masciadri E. et al., 2013c, in Esposito S., Fini L., eds, Proceedings of the Third AO4ELT Conference, Adaptive Optics for Extremely Large Telescope. Firenze, Italy, p. 107, Available at: http://ao4elt3.sciencesconf.org/, id.107
Masciadri E., Lombardi G., Lascaux F., 2014, MNRAS, 438, 983
Masciadri E., Lascaux F., Fini L., 2015, MOSE: MOdeling Sites ESO – Report Phase B, E-TRE-INA-245-0009.1
Neichel B. et al., 2016, Proc. SPIE, 9909, 990909
Noilhan J., Planton S., 1989, Mon. Weather Rev., 117, 536
Osborn J., Wilson R., Butterley T., Morris T., Dubbeldam M., Dérie F., Sarazin M., 2016, Proc. SPIE, 9909, 99091N
Redelsperger J. L., Sommeria G., 1981, Bound.-Layer Meteorol., 21, 509
Rigaut F., 2002, in Vernet E., Ragazzoni R., Esposito S., Hubin N., eds, Proc. ESO Vol. 58, Beyond Conventional Adaptive Optics. European Southern Observatory, Garching, p. 11
Roddier F., 1981, Progress in Optics, Vol. 19. North-Holland Publishing Co., Amsterdam, p. 281
Sarazin M., Roddier F., 1990, A&A, 227, 294
Sarazin M., Tokovinin A., 2002, in Vernet E., Ragazzoni R., Esposito S., Hubin N., eds, Proc. ESO Vol. 58, Beyond Conventional Adaptive Optics. European Southern Observatory, Garching, p. 321
Stein J., Richard E., Lafore J.-P., Pinty J. P., Asencio N., Cosma S., 2000, Meteorol. Atmos. Phys., 72, 203
Tatarskii V. I., 1971, The Effects of the Turbulent Atmosphere on Wave Propagation. Israel Program for Scientific Translations, Jerusalem
Thatte N. A. et al., 2016, Proc. SPIE, 9908, 99081X
Tokovinin A., Vernin J., Ziad A., Chun M., 2005, PASP, 317, 395
Vázquez Ramió H. et al., 2008, The Messenger, 132, 29
Vernin J., 2002, in Vernin J., Benkhaldoun Z., Muñoz-Tuñón C., eds, ASP Conf. Ser. Vol. 266, Astronomical Site Evaluation in the Visible and Radio Range. Astron. Soc. Pac., San Francisco, p. 2
Ye Q. Z., 2011, PASP, 123, 113

## APPENDIX A: $C_N^2$ ALGORITHM

A few differences have to be signalled with respect to the $C_N^2$ algorithm of Masciadri et al. (1999a). The main goal of this paper is to show the model performances and we expressly by pass extended analyses on the testing of different algorithms that lead us in using this algorithm for this analysis. For clarity, we report here some information that in the context of this paper plays a secondary role but that just clarifies differences in the $C_N^2$ algorithm with respect to Masciadri et al. (1999a). The mixing length in this paper (equation 4) is different from that of Bougeault & Lacarrère (1989) reported in Masciadri et al. (1999a). This is due to the fact that, more recently, the MESO-NH code assumed this new formulation[15] for mainly two reasons (Masson, private communication): (1) the new expression better fits with the diagnostic of the surface layer in convective regime in LES configuration, (2) the new formulation permits us to have a linear dependence on the limit of the surface layer where $z_{\rm down} \ll z_{\rm up}$ and the mixing length of the turbulence is proportional to the distance from the ground. Point (1) is mainly irrelevant for our application.

We point out also a difference in equation (1) with respect to that of a precedent used (Masciadri et al. 1999a). In this paper, we used, indeed, the factor

$$\left(\frac{80\times10^{-6}\times P(z)}{T(z)\times\theta(z)}\right)^2 \tag{A1}$$

instead of

$$\left(\frac{80\times10^{-6}\times P(z)}{T(z)^2}\right)^2. \tag{A2}$$

Equation (A2) (also called Gladstone's relation) is found in the literature (Roddier 1981; Vernin 2002) and assumes that the atmosphere is in hydrostatic equilibrium and the gradient of temperature follows the adiabatic approximation. However, it has been proven by Tatarskii (1971) that equation (A1) is more general and formally correct and can be obtained replacing $T$ with the potential temperature $\theta$ that, differently from the temperature, is a conservative quantity. In some parts of the atmosphere (typically in the high part of the atmosphere), the difference cannot be negligible. In reality, the difference between equations (A1) and (A2) has negligible impact in our context because the model has to be calibrated and the calibration makes the impact on the model outputs irrelevant (Masciadri et al. 2015). We prefer, however, to use equation (A1) because it is more general and formally correct. This expression can be found, correctly used, in other studies (for example Coulman et al. 1988; Abahamid et al. 2004; Cherubini & Businger 2013). Therefore, it is not to be considered as a novelty but just a warning to avoid in the future a not suitable use of Gladstone's relation.

## APPENDIX B: CORRELATION COEFFICIENT

The statistical estimators used for this analysis are bias, RMSE and regression line passing by the origin. We decided not to include the correlation coefficient (cc) because, after a detailed analysis, we

[15] See Scientific Documentation MESO-NH at http://mesonh.aero.obs-mip.fr/mesonh52

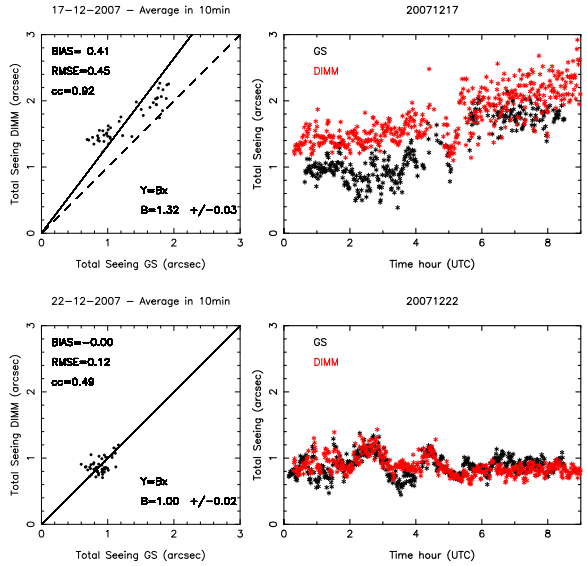


**Figure B1.** Example of the temporal evolutions of the total seeing (right) for two different nights between GS and DIMM and the corresponding values of bias, RMSE and cc (left).

realized that the cc is not an estimator as reliable as others and, in our context, it can lead us to misleading conclusions with respect to our scientific goal. Just as an example Fig. B1 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, looking at Fig. B1 (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). What is important for us is how much close or far to observations are the values calculated by the model during the night. 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 misleading conclusions.

## APPENDIX C: DEFINITIONS OF CONTINGENCY TABLES, PC, POD AND EBD

Table C1 is an example of a generic 3 × 3 contingency table where the observations and simulations are divided into three categories delimited by two thresholds. PC, $POD_i$ and EBD can be defined using *a*,*b*,*c*,*d*,*e*,*f*,*g*,*h*,*i* (number of times in which an observation and a simulation fall inside each category) and *N* (the total events). The percentage of correct detection PC is defined in equation (C1) where PC = 100 per cent is the best score; the probability to detect the value of a parameter inside a specific range of values ($POD_i$) is given by equations (C2)–(C4) where $POD_i$ = 100 per cent is the best score. The EBD probability is given by equation (C5) where EBD = 0 per cent is the best score. For a total random prediction and in the case of a 3 × 3 contingency table, we have $a = b = \cdots = i = N/9$ and PC = $POD_i$ = 33 per cent and EBD = 22.2 per cent,

**Table C1.** Generic 3 × 3 contingency table.

| Intervals | | OBSERVATIONS 1 | 2 | 3 | Total |
|---|---|---|---|---|---|
| | 1 | *a* (hit 1) | *b* | *c* | *a*+*b*+*c* 1 (Model) |
| MODEL | 2 | *d* | *e* (hit 2) | *f* | *d*+*e*+*f* 2 (Model) |
| | 3 | *g* | *h* | *i* (hit 3) | *g*+*h*+*i* 3 (Model) |
| | Total | *a*+*d*+*g* 1 (OBS) | *b*+*e*+*h* 2 (OBS) | *c*+*f*+*i* 3 (OBS) | $N = a+b+c+d+e+f+g+h+i$ Total of events |

$$\mathrm{PC} = \frac{a+e+i}{N} \times 100; 0 \text{ per cent} \leq \mathrm{PC} \leq 100 \text{ per cent} \quad \text{(C1)}$$

$$\mathrm{POD}(\text{event}_1) = \frac{a}{a+d+g} \times 100; \quad 0 \text{ per cent} \leq \mathrm{POD} \leq 100 \text{ per cent} \quad \text{(C2)}$$

$$\mathrm{POD}(\text{event}_2) = \frac{e}{b+e+h} \times 100; \quad 0 \text{ per cent} \leq \mathrm{POD} \leq 100 \text{ per cent} \quad \text{(C3)}$$

$$\mathrm{POD}(\text{event}_3) = \frac{i}{c+f+i} \times 100; \quad 0 \text{ per cent} \leq \mathrm{POD} \leq 100 \text{ per cent} \quad \text{(C4)}$$

$$\mathrm{EBD} = \frac{c+g}{N} \times 100; 0 \text{ per cent} \leq \mathrm{EBD} \leq 100 \text{ per cent.} \quad \text{(C5)}$$

## APPENDIX D: CONTINGENCY TABLES – RESULTS

To facilitate the reading of the paper, we gathered in this appendix all the contingency tables useful to summarize and discuss results presented in Section 5.

**Table D1.** Sample (A): contingency tables for the total seeing, in summer (42 nights in 2010 and 2011) – validation sample (case 1). Rounded values of the first and second tertiles of the climatology have been used as thresholds.

| ε – SUMMER 42 nights over 89 (arcsec) | | DIMM ε < 0.88 | 0.88 < ε < 1.17 | ε > 1.17 |
|---|---|---|---|---|
| MNH | ε < 0.88 | 181 | 175 | 125 |
| | 0.88 < ε < 1.17 | 129 | 106 | 83 |
| | ε > 1.17 | 66 | 95 | 168 |

Total points = 1128; PC = 40.34 per cent; EBD = 16.93 per cent.
$POD_1$ = 48.14 per cent; $POD_2$ = 28.19 per cent; $POD_3$ = 44.68 per cent.

**Table D2.** Sample (A): contingency tables for the total seeing, in summer (42 nights in 2010 and 2011) – validation sample (case 2).

| ε – SUMMER 42 nights over 89 (arcsec) | | DIMM ε < 1 | 1 < ε < 1.4 | ε > 1.4 |
|---|---|---|---|---|
| MNH | ε < 1 | 353 | 209 | 77 |
| | 1 < ε < 1.4 | 177 | 77 | 43 |
| | ε > 1.4 | 39 | 48 | 105 |

Total points = 1128; PC = 47.43 per cent; EBD = 10.28 per cent.
$POD_1$ = 62.04 per cent; $POD_2$ = 23.05 per cent; $POD_3$ = 46.67 per cent.

**Table D3.** Sample (B): contingency tables for the total seeing, in summer (18 nights in 2007) – validation sample (case 1). Rounded values of the first and second tertiles of the climatology have been used as thresholds.

| $\varepsilon$ – SUMMER 18 nights over 36 (arcsec) | | DIMM $\varepsilon < 0.86$ | DIMM $0.86 < \varepsilon < 1.20$ | DIMM $\varepsilon > 1.20$ |
|---|---|---|---|---|
| MNH | $\varepsilon < 0.86$ | 86 | 87 | 52 |
| | $0.86 < \varepsilon < 1.20$ | 64 | 61 | 34 |
| | $\varepsilon > 1.20$ | 12 | 14 | 76 |

Total points = 486; PC = 45.88 per cent; EBD = 13.17 per cent.
$POD_1$ = 53.09 per cent; $POD_2$ = 37.65 per cent; $POD_3$ = 46.91 per cent.

**Table D4.** Sample (B): contingency tables for the total seeing, in summer (18 nights in 2007) – validation sample (case 2).

| $\varepsilon$ – SUMMER 18 nights over 36 (arcsec) | | DIMM $\varepsilon < 1$ | DIMM $1 < \varepsilon < 1.4$ | DIMM $\varepsilon > 1.4$ |
|---|---|---|---|---|
| MNH | $\varepsilon < 1$ | 177 | 87 | 38 |
| | $1 < \varepsilon < 1.4$ | 58 | 34 | 14 |
| | $\varepsilon > 1.4$ | 9 | 18 | 51 |

Total points = 486; PC = 53.91 per cent; EBD = 9.67 per cent.
$POD_1$ = 72.54 per cent; $POD_2$ = 24.46 per cent; $POD_3$ = 49.51 per cent.

**Table D5.** Contingency tables for the total seeing, between DIMM and GS, using GS as a reference – calibration sample (20 nights).

| $\varepsilon$ 20 nights (arcsec) | | GS $\varepsilon < 0.97$ | GS $0.97 < \varepsilon < 1.24$ | GS $\varepsilon > 1.24$ |
|---|---|---|---|---|
| DIMM | $\varepsilon < 0.97$ | 108 | 45 | 5 |
| | $0.97 < \varepsilon < 1.24$ | 39 | 60 | 14 |
| | $\varepsilon > 1.24$ | 21 | 63 | 149 |

Total points = 504; PC = 62.90 per cent; EBD = 5.16 per cent.
$POD_1$ = 64.29 per cent; $POD_2$ = 35.71 per cent; $POD_3$ = 88.69 per cent.

**Table D6.** Contingency tables for the total seeing, between DIMM and GS, using DIMM as a reference – calibration sample (20 nights).

| $\varepsilon$ 20 nights (arcsec) | | DIMM $\varepsilon < 1$ | DIMM $1 < \varepsilon < 1.42$ | DIMM $\varepsilon > 1.42$ |
|---|---|---|---|---|
| GS | $\varepsilon < 1$ | 120 | 52 | 13 |
| | $1 < \varepsilon < 1.42$ | 48 | 103 | 77 |
| | $\varepsilon > 1.42$ | 0 | 13 | 78 |

Total points = 504; PC = 59.72 per cent; EBD = 2.58 per cent.
$POD_1$ = 71.43 per cent; $POD_2$ = 61.31 per cent; $POD_3$ = 46.43 per cent.

**Table D7.** Contingency tables for the total seeing, between MESO-NH and GS – calibration sample (20 nights).

| $\varepsilon$ 20 nights (arcsec) | | GS $\varepsilon < 0.97$ | GS $0.97 < \varepsilon < 1.24$ | GS $\varepsilon > 1.24$ |
|---|---|---|---|---|
| MODEL | $\varepsilon < 0.97$ | 138 | 115 | 15 |
| | $0.97 < \varepsilon < 1.24$ | 18 | 35 | 48 |
| | $\varepsilon > 1.24$ | 12 | 18 | 105 |

Total points = 504; PC = 55.16 per cent; EBD = 5.36 per cent.
$POD_1$ = 82.14 per cent; $POD_2$ = 20.84 per cent; $POD_3$ = 62.50 per cent.

**Table D8.** Contingency tables for the total seeing, between MESO-NH and DIMM – calibration sample (20 nights).

| $\varepsilon$ 20 nights (arcsec) | | DIMM $\varepsilon < 1$ | DIMM $1 < \varepsilon < 1.42$ | DIMM $\varepsilon > 1.42$ |
|---|---|---|---|---|
| MODEL | $\varepsilon < 1$ | 156 | 102 | 27 |
| | $1 < \varepsilon < 1.42$ | 17 | 60 | 70 |
| | $\varepsilon > 1.42$ | 4 | 14 | 79 |

Total points = 529; PC = 55.76 per cent; EBD = 5.86 per cent.
$POD_1$ = 88.13 per cent; $POD_2$ = 34.10 per cent; $POD_3$ = 44.89 per cent.

**Table D9.** Contingency table of $\tau_0$ between MASS and DIMM, using the DIMM as a reference (48 nights in 2010/2011) – validation sample.

| $\tau_0$ 48 nights (ms) | | DIMM $\tau_0 < 2.07$ | DIMM $2.07 < \tau_0 < 3.26$ | DIMM $\tau_0 > 3.26$ |
|---|---|---|---|---|
| MASS | $\tau_0 < 2.07$ | 182 | 15 | 8 |
| | $2.07 < \tau_0 < 3.26$ | 167 | 139 | 16 |
| | $\tau_0 > 3.26$ | 46 | 240 | 370 |

Total points = 1183; PC = 58.41 per cent; EBD = 4.56 per cent.
$POD_1$ = 46.08 per cent; $POD_2$ = 35.28 per cent; $POD_3$ = 93.91 per cent.

**Table D10.** Contingency table of $\tau_0$ between MASS and DIMM, using the MASS as a reference (48 nights in 2010/2011) – validation sample.

| $\tau_0$ 48 nights (ms) | | MASS $\tau_0 < 2.68$ | MASS $2.68 < \tau_0 < 4.87$ | MASS $\tau_0 > 4.87$ |
|---|---|---|---|---|
| DIMM | $\tau_0 < 2.68$ | 364 | 215 | 65 |
| | $2.68 < \tau_0 < 4.87$ | 20 | 170 | 226 |
| | $\tau_0 > 4.87$ | 11 | 9 | 103 |

Total points = 1183; PC = 53.85 per cent; EBD = 6.42 per cent.
$POD_1$ = 92.15 per cent; $POD_2$ = 43.15 per cent; $POD_3$ = 26.14 per cent.

**Table D11.** Contingency table of $\tau_0$ between MESO-NH and DIMM, in summer – validation sample.

| $\tau_0$ – SUMMER 42 nights over 89 (ms) | | DIMM $\tau_0 < 2.20$ | DIMM $2.20 < \tau_0 < 3.58$ | DIMM $\tau_0 > 3.58$ |
|---|---|---|---|---|
| MNH | $\tau_0 < 2.20$ | 178 | 61 | 13 |
| | $2.20 < \tau_0 < 3.58$ | 138 | 139 | 95 |
| | $\tau_0 > 3.58$ | 54 | 169 | 261 |

Total points = 1108; PC = 57.17 per cent; EBD = 6.05 per cent.
$POD_1$ = 48.11 per cent; $POD_2$ = 37.67 per cent; $POD_3$ = 70.73 per cent.

**Table D12.** Contingency table of $\tau_0$ between MESO-NH and DIMM, in winter – validation sample.

| $\tau_0$ – WINTER 47 nights over 89 (ms) | | DIMM $\tau_0 < 1.67$ | DIMM $1.67 < \tau_0 < 2.48$ | DIMM $\tau_0 > 2.48$ |
|---|---|---|---|---|
| MNH | $\tau_0 < 1.67$ | 271 | 187 | 37 |
| | $1.67 < \tau_0 < 2.48$ | 118 | 116 | 107 |
| | $\tau_0 > 2.48$ | 31 | 116 | 275 |

Total points = 1258; PC = 52.62 per cent; EBD = 5.40 per cent.
$POD_1$ = 64.52 per cent; $POD_2$ = 27.68 per cent; $POD_3$ = 65.63 per cent.

**Table D13.** Contingency table of $\tau_0$ between MESO-NH and MASS, in summer – validation sample.

| $\tau_0$ – SUMMER 21 nights over 48 (ms) | | MASS $\tau_0 < 3.27$ | MASS $3.27 < \tau_0 < 5.36$ | MASS $\tau_0 > 5.36$ |
|---|---|---|---|---|
| | $\tau_0 < 3.27$ | 111 | 70 | 7 |
| MNH | $3.27 < \tau_0 < 5.36$ | 49 | 52 | 38 |
| | $\tau_0 > 5.36$ | 1 | 38 | 115 |

Total points = 481; PC = 57.80 per cent; EBD = 1.66 per cent.
$POD_1$ = 68.94 per cent; $POD_2$ = 32.50 per cent; $POD_3$ = 71.87 per cent.

**Table D14.** Contingency table of $\tau_0$ between MESO-NH and MASS, in winter – validation sample.

| $\tau_0$ – WINTER 27 nights over 48 (ms) | | MASS $\tau_0 < 2.45$ | MASS $2.45 < \tau_0 < 4.19$ | MASS $\tau_0 > 4.19$ |
|---|---|---|---|---|
| | $\tau_0 < 2.45$ | 112 | 86 | 6 |
| MNH | $2.45 < \tau_0 < 4.19$ | 94 | 108 | 45 |
| | $\tau_0 > 4.19$ | 28 | 40 | 183 |

Total points = 702; PC = 57.41 per cent; EBD = 4.84 per cent.
$POD_1$ = 47.86 per cent; $POD_2$ = 46.15 per cent; $POD_3$ = 78.20 per cent.

**Table D15.** Contingency table of $\theta_0$ between MASS and DIMM, using the MASS as a reference (44 nights in 2010/2011).

| $\theta_0$ 44 nights (arcsec) | | MASS $\theta_0 < 1.53$ | MASS $1.53 < \theta_0 < 2.04$ | MASS $\theta_0 > 2.04$ |
|---|---|---|---|---|
| | $\theta_0 < 1.53$ | 47 | 1 | 0 |
| DIMM | $1.53 < \theta_0 < 2.04$ | 157 | 45 | 17 |
| | $\theta_0 > 2.04$ | 168 | 325 | 354 |

Total points = 1114; PC = 40.04 per cent; EBD = 15.08 per cent.
$POD_1$ = 12.63 per cent; $POD_2$ = 12.13 per cent; $POD_3$ = 95.42 per cent.

**Table D16.** Contingency table of $\theta_0$ between MASS and DIMM, using the DIMM as a reference (44 nights in 2010/2011).

| $\theta_0$ 44 nights (arcsec) | | DIMM $\theta_0 < 2.26$ | DIMM $2.26 < \theta_0 < 2.94$ | DIMM $\theta_0 > 2.94$ |
|---|---|---|---|---|
| | $\theta_0 < 2.26$ | 354 | 329 | 177 |
| MASS | $2.26 < \theta_0 < 2.94$ | 13 | 38 | 156 |
| | $\theta_0 > 2.94$ | 5 | 4 | 38 |

Total points = 1114; PC = 38.60 per cent; EBD = 16.34 per cent.
$POD_1$ = 95.16 per cent; $POD_2$ = 10.24 per cent; $POD_3$ = 10.24 per cent.

**Table D17.** Contingency table of $\theta_0$ between MASS and MESO-NH (21 summer nights in 2010/2011).

| $\theta_0$ – SUMMER 21 nights over 48 (arcsec) | | MASS $\theta_0 < 1.55$ | MASS $1.55 < \theta_0 < 2.01$ | MASS $\theta_0 > 2.01$ |
|---|---|---|---|---|
| | $\theta_0 < 1.55$ | 16 | 18 | 12 |
| MNH | $1.55 < \theta_0 < 2.05$ | 88 | 80 | 61 |
| | $\theta_0 > 2.05$ | 70 | 75 | 100 |

Total points = 520; PC = 37.69 per cent; EBD = 15.77 per cent.
$POD_1$ = 9.20 per cent; $POD_2$ = 46.24 per cent; $POD_3$ = 57.80 per cent.

## APPENDIX E: OT TEMPORAL EVOLUTION

Figs E1–E3 show the temporal evolution of the OT developed on the whole atmosphere (∼20 km) during all 20 nights of the PAR2007 site testing campaign. ASTRO-MESO-NH model estimates are compared to observations done with a GS and a DIMM.

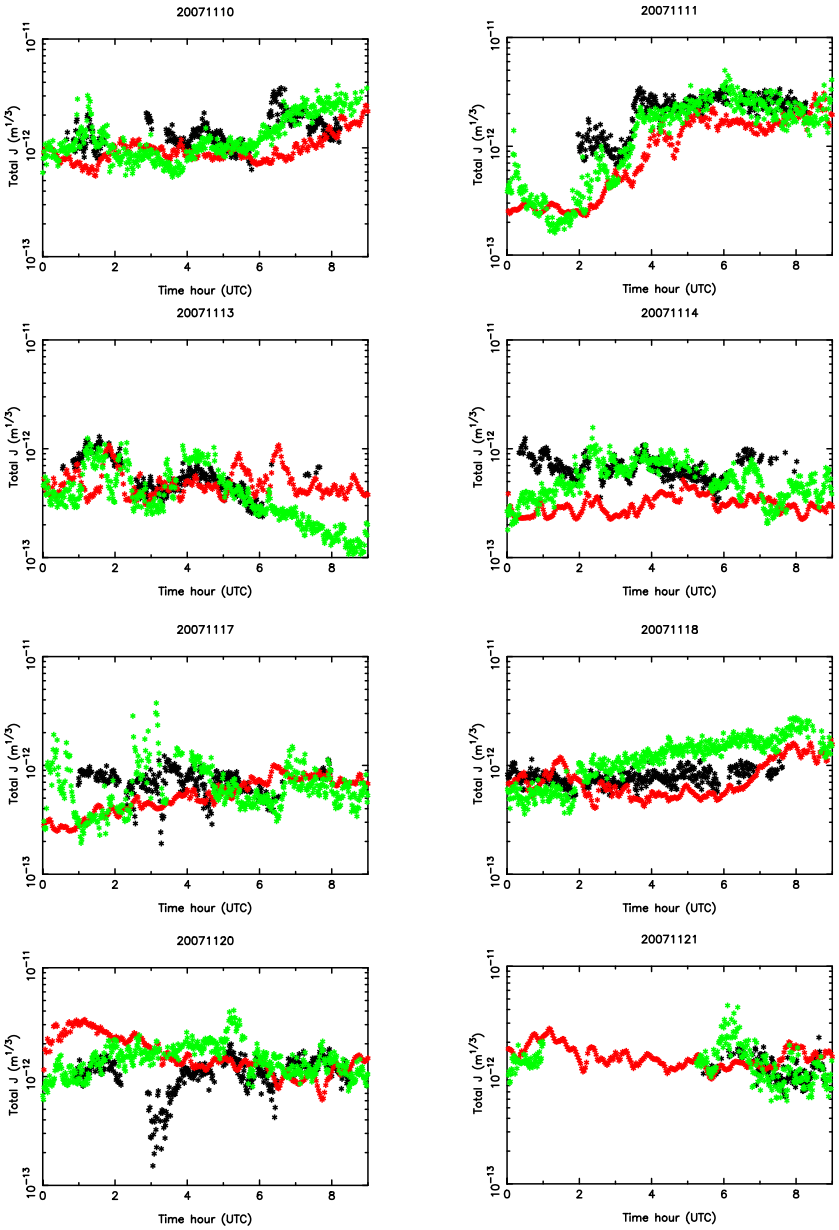


**Figure E1.** Temporal evolution of $J$ (equation 11) for all the 20 nights of the PAR2007 site testing campaign: GS (black), DIMM (green) and ASTRO-MESO-NH model (red).

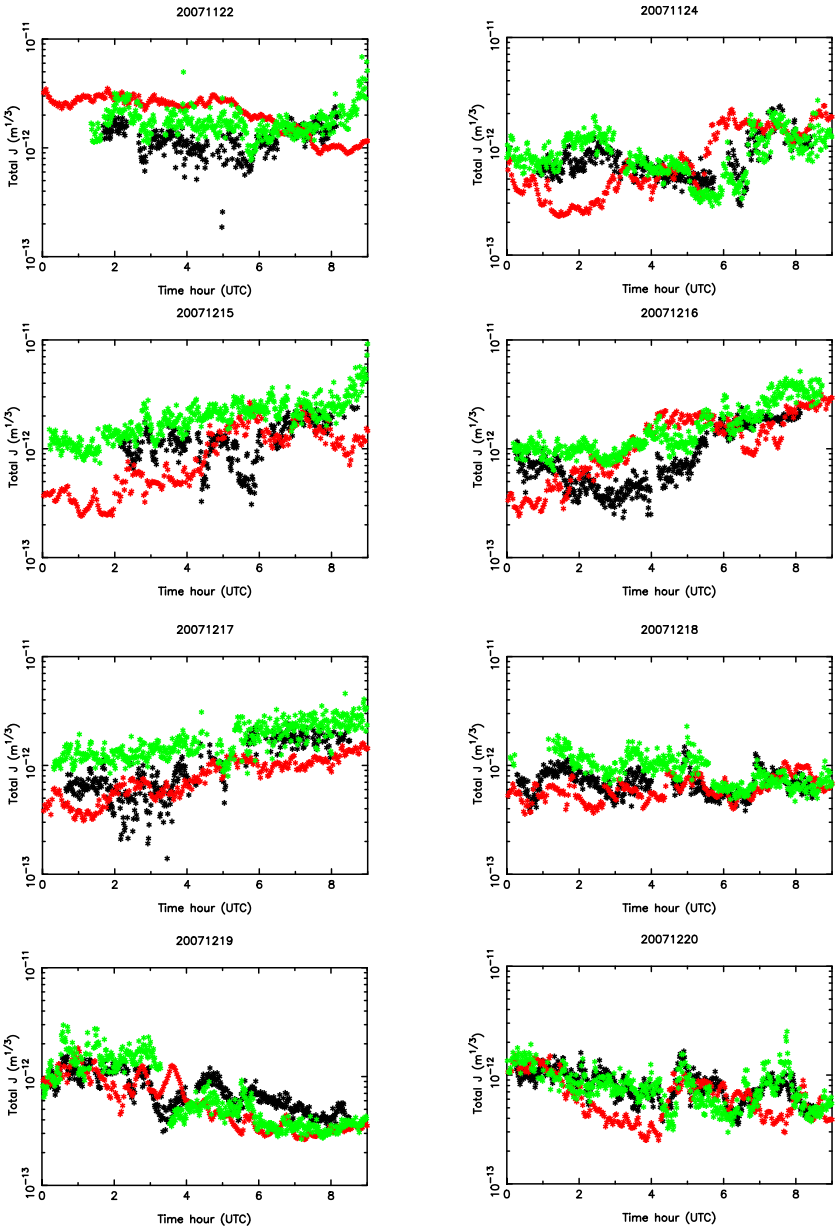


**Figure E2.** As Fig. E1 – *continued*.

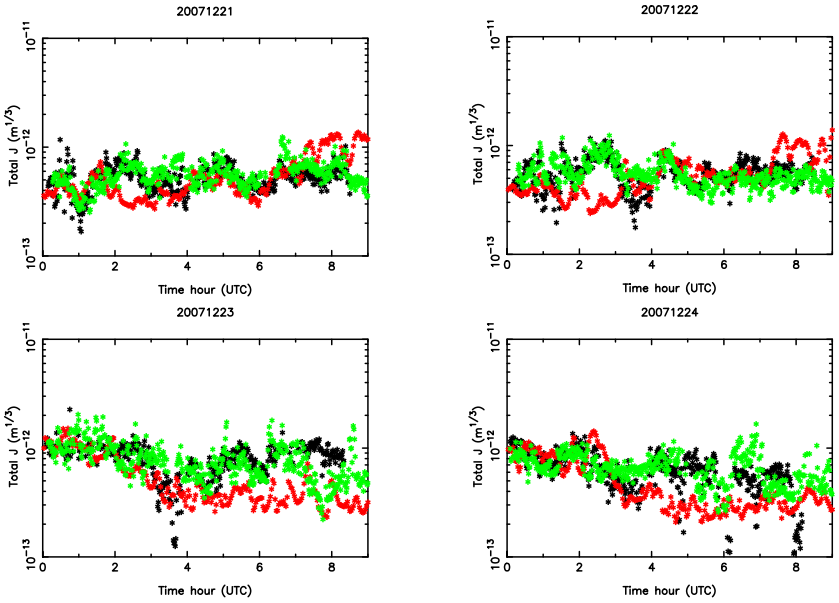


**Figure E3.** As Fig. E1 – *continued*.

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