



HAL
open science

Contribution of statistical site learning to improve optical turbulence forecasting

C Giordano, A Rafalimanana, A Ziad, E Aristidi, Julien Chabé, Y
Fanteï-Caujole, C Renaud

► **To cite this version:**

C Giordano, A Rafalimanana, A Ziad, E Aristidi, Julien Chabé, et al.. Contribution of statistical site learning to improve optical turbulence forecasting. *Monthly Notices of the Royal Astronomical Society*, 2021, 504 (2), pp.1927-1938. 10.1093/mnras/staa3709 . hal-03266390

HAL Id: hal-03266390

<https://hal.science/hal-03266390v1>

Submitted on 9 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Contribution of statistical site learning to improve optical turbulence forecasting

C. Giordano^{1*}, A. Rafalimanana¹, A. Ziad¹, E. Aristidi¹, J. Chabé², Y. Fanteï-Caujole¹, C. Renaud¹

¹Université Côte d’Azur, Observatoire de la Côte d’Azur, CNRS, Laboratoire Lagrange, France

²Université Côte d’Azur, OCA, CNRS, IRD, Géoazur, 2130 route de l’Observatoire, 06460 Caussols, France

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

The forecast of the atmospheric and turbulence conditions above astronomical observatories is of interest for the community because it allows to plan observations with maximum efficiency, this is called the flexible scheduling. It can also be used to simulate a long term site testing to give local information useful for the conception of focal and post-focal instrumentation. We already have presented our forecasting tool in previous publications but in this paper, we will focus on the importance of using local measurements to improve the predictive turbulence model and to better consider local specificities of a given site, what we call site learning. For this study, we used local database provided by the Calern Atmospheric Turbulence Station, installed since 2015 at the Calern observatory. In addition, a set of several months of predictions to feed the turbulence model by taking into account daytime and nighttime conditions. This upgrade improves the quality of our forecasting by reducing the absolute bias between measurements and predictions from 25 to 50% for each layer of the C_n^2 , by 25% for the seeing and by 70% for the isoplanatic angle.

Key words: Turbulence – Atmospheric effects – Software: simulations – Methods: numerical, statistical, data analysis

1 INTRODUCTION

Since decades, the effects of atmospheric turbulence on degradation of astronomical observations are well known and characterized. Indeed, independently of the instrument size, the optical turbulence limits the maximal resolution reachable during the observations by deforming the incoming wavefront. This limited resolution is called the astronomical seeing.

From the beginning of the 90’s, the adaptive optics (AO) showed up to compensate these effects on the wavefront deformations by using deformable mirror to flatten it. We could think that the optical turbulence problem was solved with these systems, but in practice the complexity of both AO systems and turbulence phenomena are such that using AO only cannot compensate all the turbulence effects under all possible conditions. Moreover, AO is limited by its sky coverage, which is reduced by the anisoplanatism effects. This latter is characterized by the isoplanatic angle.

In practice, AO systems are dimensioned and built considering median optical conditions of a given site measured by local atmospheric instrumentations. This AO procedure leads to limitations when real time turbulence conditions are worse than median ones. Indeed, in presence of strong turbulent layers, or strong wind speed,

the AO correction will be degraded because of the impossibility of closing the AO loop or because the delay error due to a low coherence time becomes too large. Currently, these limits are not taken into account in observations plans, and they result in increasing the amount of out-of-constraint observations due to unforeseen changes in meteorological and optical turbulence conditions.

Coming from a pioneering work of Coulman et al. (1986), and followed by Bougeault et al. (1995), the turbulence forecasting and the flexible scheduling could bring a solution to this planning problem thanks to the use of meteorological and optical turbulence predictions to adapt the observation program to the real-time conditions and then to increase the observation efficiency.

Beyond the utility in observatories, the optical forecasting is of interest in all domains related to the optical propagation such as the optical communication, the laser telemetry, or other research fields. For example, for satellite to ground optical communications, having a prediction about the optical conditions could help to choose a ground station inside a network to optimize the quality and the bit error rate of optical communication signals.

Several models or tools exist to predict the meteorological conditions. Global circulation models (GCM) cover the entire Earth with a low horizontal resolution ($\Delta x > 10km$) and a low temporal resolution (3-6 hours) leading to obvious limits for flexible scheduling because precise optical forecasting needs higher fre-

* E-mail: cgiordano@oca.eu

quency (several minutes) and higher spatial resolution. To overcome these limits, the mesoscale models have been developed to reach better spatial resolution ($\Delta x \geq 0.1 \text{ km}$) and a personalized temporal resolution ($\Delta t \geq 1 \text{ minute}$). The two widely used mesoscale models are the Meso-Nh (Mesoscale Non Hydrostatic) model and the Weather Research and Forecasting (WRF) model. All these systems forecast or simulate meteorological conditions, but they don't give access to optical turbulence values such as the C_n^2 profiles and other related parameters. Therefore, a model, or an algorithm taking as inputs the meteorological values (pressure, temperature, wind speed, ...) to compute the optical turbulence conditions is requested.

Hach et al. (2012) have used the reanalysis data set provided by the NCEP/NCAR (National Centers for Environmental Prediction / National Center for Atmospheric Research) which are based on both GCM and available measurements. To deduce the C_n^2 they used equations developed by Tatarskii (1961). They found good agreements for meteorological parameters and encouraging results about the C_n^2 assessment within the free atmosphere.

Masciadri et al. (1999, 2017) (and references therein) used the Meso-NH (Mesoscale Non Hydrostatic) model to predict astroclimatic parameters from turbulent kinetic energy (TKE) assessment. This TKE method has also been used by Cherubini et al. (2008) with the MM5 model (Fifth-Generation Penn State/NCAR Mesoscale Model) and by Cherubini et al. (2011) with the WRF model. Mahalov & Moustou (2010) have also used the WRF system and a turbulence model similar to the one used by Hach et al. (2012). Later, in Mahalov et al. (2011) authors used also the WRF model to run a numerical study about the ability of WRF to well retrieve the mountain waves effects. They also showed the importance of thinner domain resolutions to have a precise assessment of meteorological effects.

We have chosen to use the WRF model which has been developed in the National Center for Atmospheric Research (NCAR) (Skamarock et al. 2019) coupled with an empirical turbulence model (Trinquet & Vernin 2007) developed from a statistical analysis of a large number of balloon radiosounding, this method will be called BDTM (Balloon Derived Turbulence Model). We have already used this BDTM method above La Palma island (Canaries Islands) (Giordano et al. 2013, 2014; Giordano 2014), hosting the Osservatorio del Roque de Los Muchachos.

In this paper, we present an extension of the work shown in Giordano et al. (2019) using the WRF model above the Calern observatory but constraining the BDTM method with measurements from the Calern Atmospheric Turbulence Station (CATS) (Ziad et al. 2019a). **A comparison between the different existing models will be done in the future. In this paper we only focus on the empirical one because it really depends on the local measurements.** Firstly, we detail our forecasting tool from the WRF model to the turbulence model and we present the configuration used in our case. Then we show how we use the local measurements from CATS station to feed our optical turbulence model in order to improve its efficiency and its accuracy, this is what we call the site learning (SL). After that, we analyze a large set of forecasts done between July 2019 and January 2020 to demonstrate the efficiency of the SL phase. Finally, we discuss our results and conclude.

2 FORECASTING AND TURBULENCE MODEL

2.1 WRF coupled to empirical turbulence model

As already presented in previous articles (Giordano et al. 2013, 2014; Giordano 2014) the WRF model is a mesoscale non-hydrostatic numerical model developed in the NCAR center (USA) allowing to forecast and/or simulate temporal evolution of meteorological conditions within a tridimensional domain. The main advantages of this model are its availability, its adaptability and its configurability. These points are really important because the divergence of specificities from one site to another makes impossible to have a universal configuration. The process of a WRF simulation follows three steps:

- (i) Definition of one or several tridimensional domains inside which simulations will be done.
- (ii) Initialization of simulations thanks to terrestrial data (orography, albedo, soil type, ...) and meteorological data coming from large organizations/laboratories/centers (GFS, ECMWF, ...).
- (iii) Running simulations for a given period to obtain predictions.

The meteorological parameters forecasted useful for an astronomical context are the absolute temperature, the wind velocity modulus and direction, the relative humidity, the cloud cover, and the precipitable water vapor. For the BDTM model, we also need the prediction of the potential temperature and the atmospheric pressure.

Once these parameters predicted, we use an empirical turbulence model deduced from a statistical analysis of a large number of radio-sounding balloons launched around the world and measuring both meteorological (pressure, temperature, wind speed and direction, relative humidity) and optical turbulence profiles (C_n^2) (Trinquet & Vernin 2007). The optical turbulence comes out from the presence of potential temperature gradient which can be shaken in presence of wind shear. Therefore, it makes sense to use these two parameters in the model to assess the optical turbulence. As presented in our previous papers (Giordano et al. 2013, 2014; Giordano 2014), the BDTM follows equation 1.

$$C_T^2(h) = \phi(h)\chi(h)s(h)^{1/2} \quad (1)$$

$$\chi(h) = \frac{\delta\theta}{\delta h} \quad (2)$$

$$s(h) = \left(\frac{dU^2}{dh} + \frac{dV^2}{dh} \right)^{1/2} \quad (3)$$

where the C_T^2 is the structure constant of temperature fluctuations, χ is the vertical gradient of the potential temperature θ , h is the altitude, U and V are the wind speed respectively in the direction West-East and South-North, and s is the horizontal wind shear. $\phi(h)$ is the vertical profile deduced from the statistical analysis of the balloons radio-sounding,

$$\phi(h) = \frac{\langle C_T^2(h) \rangle_m}{\langle \chi(h) \rangle_m \langle s(h) \rangle_m^{1/2}} \quad (4)$$

where $\langle \rangle_m$ is the median value.

Finally, from C_T^2 we can deduce the structure constant of the re-

fractive index fluctuations C_n^2 following the well-known Gladstone's formula,

$$C_n^2(h) = \left(\frac{80 \cdot 10^{-6} P(h)}{T(h)^2} \right)^2 C_T^2(h) \quad (5)$$

where $P(h)$ and $T(h)$ are respectively the vertical profiles of the atmospheric pressure and absolute temperature.

From the C_n^2 profile, we can deduce all optical parameters relevant for flexible scheduling (seeing, coherence time, isoplanatic angle, outer scale), as given by [Giordano et al. \(2012, 2013\)](#).

It is important to notice that this model has been deduced from non-convective case, i.e. $\frac{\delta\theta}{\delta h} \geq 0$. When there is a convective case, we use the theoretical formulation of the C_n^2 as given by [Coulman et al. \(1988\)](#) and used by [Hach et al. \(2012\)](#),

$$C_n^2(h) = 2.8 M^2 L_0^{4/3} \quad (6)$$

where M is the vertical gradient of the refractive index, given by,

$$M = -80 \cdot 10^{-6} \frac{P}{T\theta} \frac{\delta\theta}{\delta h} \quad (7)$$

and L_0 is the outer scale of the turbulence. These cases appear essentially in the first layer during daytime when the Sun warms the soil.

[Ziad \(2016\)](#) has point out the fact that the size of the outer scale is controversial (from 2m to more than 2km) and in the literature there is ambiguities between the local outer scale L_0 and the wavefront coherence outer scale \mathcal{L}_0 . Moreover, in our study equation 6 is mainly used during the daytime, where the turbulence is dominated by the surface layer conditions. As indicated by [Ziad \(2016\)](#), some authors indicated that L_0 is of the order of the height of this surface layer layer. Therefore, in our study we have found out that a value of $L_0 \approx 15m$ in equation 6 seems suitable to compute C_n^2 prediction more precisely in convective cases ($\frac{\delta\theta}{\delta h} < 0$).

This aforementioned empirical turbulence model, being deduced from in-situ measurements coming from different observatories or locations, doesn't take into account local specificities such as the roughness of the terrain, the local climatic conditions or other particularities. We propose in section 3 to use in-situ measurements from the CATS station on the Calern observatory to learn to the turbulence model the specificities of this site and then to reach better precision in forecasting.

2.2 Forecasted domains presentation

In this section, we briefly present the configuration used for our forecasting, and the input to inject in it. Table 1 summarizes this configuration. We chose to use 4 domains centered on the Calern observatory with an horizontal resolution going from 27km up to 1km, as shown in figure 1. Figure 2 shows the disposition of the vertical layers from ground up to the top of the domains ($\sim 20km$). Above 20km, the optical turbulence is negligible, so we haven't explored conditions above this altitude. We can see a highest resolution near the ground ($\Delta h = 0.015km$) and a lowest on the top ($\Delta h = 1.246km$). This is set to better consider ground layer turbulence which represents the main contribution to the total turbulent energy. This figure 2 shows also the PML's (Profiler of Moon Limb, see section 3) vertical resolution in function of the height.

To start a simulation or a prediction we need meteorological initial conditions that will be interpolated to our domain's grid

Table 1. Domains configuration used for the WRF forecasting.

Domains	D01	D02	D03	D04
Central latitude [deg]	43.7546 North			
Central longitude [deg]	6.9206 East			
$\Delta x \times \Delta y$ [km]	27 × 27	9 × 9	3 × 3	1 × 1
Δh [km]	0.015 ≤ Δh ≤ 1.246			
Δt [minute]	10			
Topographic resolution	0.927	0.927	0.927	0.09
$\Delta x = \Delta y$ [km]				
Meteorological input data	Global forecast system (GFS)			
	$\Delta x = \Delta y = 0.25deg$			

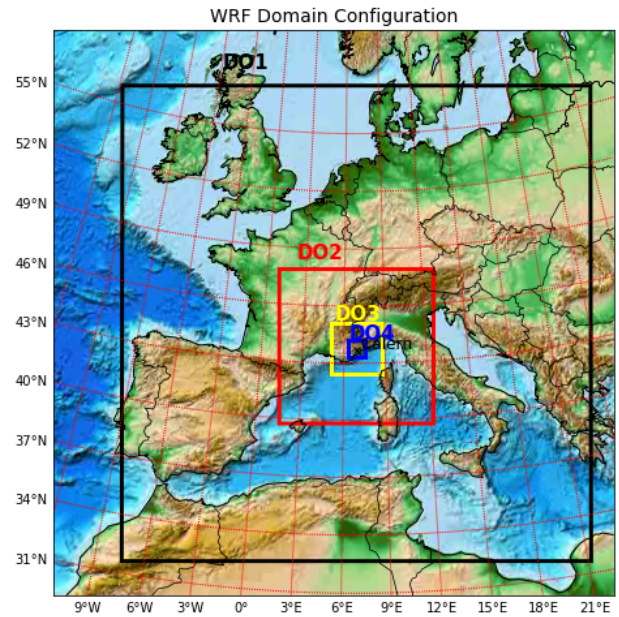


Figure 1. Spatial domains used for forecasting from $\Delta x = \Delta y = 27km$ (domain D01, in black) to $\Delta x = \Delta y = 1km$ (domain D04, in blue).

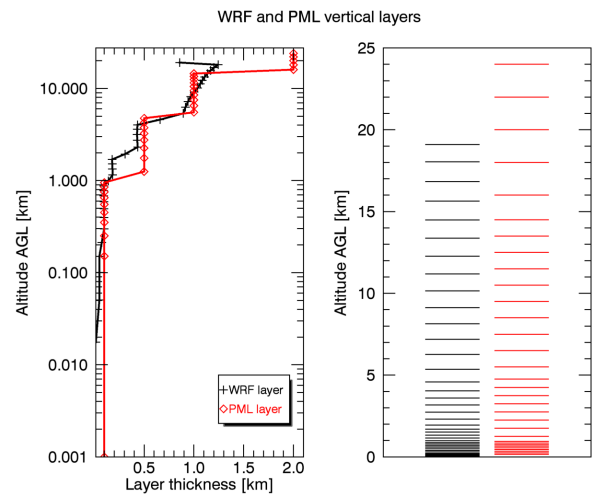


Figure 2. Vertical levels used in the WRF model (black) and in the PML instrument (red).

points. These data come from global forecasted or reanalysis system or center such as the ECMWF (European Centre for Medium-Range Weather Forecasts), the NCEP GFS (Global Forecast System) or others. In this study we used the GFS data, easily available and downloadable¹. These data have a horizontal resolution of $0.25 \times 0.25^\circ$ and a temporal resolution from 3 to 6 hours.

The orography has also a large importance for the meteorological forecasting because it has an influence on the circulation of the air mass over the soil, on the ground heating and on related effects which impact conditions within the first layers. To reach the best possible forecasts, we are convinced that using a well resolved topographic data is mandatory. Minimum default values used by WRF have a horizontal resolution of $30 \times 30 \text{ arcsec}$ corresponding to around $927 \times 927 \text{ m}$. Considering the complexity of the terrain above which observatories are installed (peak, plateau, mountains), we think this resolution is not fine enough and we decided to go further. Therefore, we found and implemented in WRF the topographic data coming from the Shuttle Radar Topography Mission (SRTM)². This latter has a horizontal resolution of $90 \times 90 \text{ m}$, which is far better than the default data.

3 SITE LEARNING FOR TURBULENCE MODEL UPGRADE

To optimize the forecasting above a given site, the turbulence model has to reflect local specificities of this location. Therefore, we need to improve the initial BDM model by injecting in-situ measurements of C_n^2 , χ and s in equation 4 and 5. This phase is called "site learning" (SL) because the upgraded model learns peculiarities of the given site. **This site learning is particularly well adapted to empirical models which depend only on experimental measurements. The models coming directly from theory such as the one from Tatarskii (equation 6) cannot be updated by statistics on a given site, and the SL method cannot be applied to them. Therefore, in this paper we do not focus on other models but only on the way to improve the one presented in section 2.1 by taking into account local specificities.**

Since 2015, the Calern Atmospheric Turbulence Station (CATS) (Chabé et al. 2016; Ziad et al. 2018, 2019a) is fully operational (see figure 3), and measures continuously atmospheric conditions thanks to a complementary set of instruments:

- A weather station measuring ground meteorological parameters: wind speed modulus and direction, pressure, temperature, relative humidity.
- An all-sky camera measuring during the night the cloud cover.
- A Generalized Differential Image Motion Monitor (GDIMM) (Aristidi et al. 2014, 2018, 2019b) measuring all integrated optical parameters: seeing, isoplanatic angle, coherence time, outer scale.
- A Profiler of Moon Limb (PML) (Ziad et al. 2013; Blary et al. 2014; Aristidi et al. 2019a; Chabé et al. 2020) measuring vertical profiles of the C_n^2 during daytime on the Sun limb and during nighttime on the Moon limb.

The CATS database is the reason why we have chosen to test the SL method above the Calern observatory.

The main limit is that we have only access to C_n^2 profiles and



Figure 3. Calern Atmospheric Turbulence Station (CATS). On the left the GDIMM installed on a 4m tower, on the right the PML instrument observing the Sun limb.

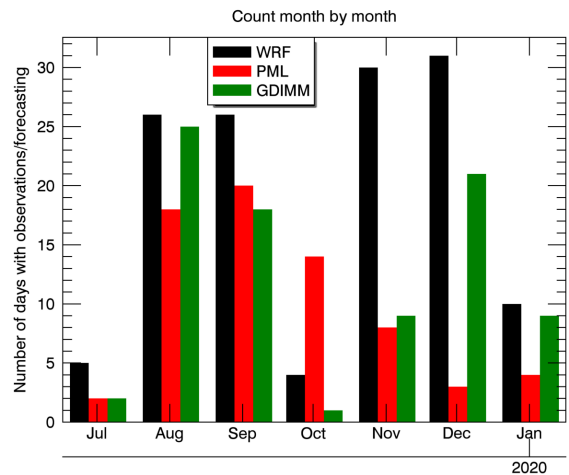


Figure 4. Distribution of data from each source in number of days per month, starting from July 2019. The black, red and green colors represent respectively WRF, PML and GDIMM. We only count the number of days where there was at least one measurement or forecast.

not to C_T^2 , χ and s profiles, therefore it is impossible to compute ϕ with measurements only. To overcome this lack of data, we decided to use WRF outputs to compute χ and s where measurements are not available. This method assumes that WRF is precise enough in the forecasting of the meteorological profiles.

We have launched, in July 2019, an automatic forecast running each day and covering the next 48h to collect a large WRF database. These predictions are also uploaded to the CATS website³ to support observers in Calern observatory (see Appendix A). On the period from July 2019 to January 2020, we have covered more than 130 days of forecasting allowing comparisons with 69 days where PML was running and 116 days where GDIMM was operational. Figure 4 shows the distribution of forecasted and measurements along the covered period, from July 2019 to January 2020. In this figure, we count only the number of days where there were observations or forecasting and not the number of measurement points.

Using these CATS statistics and this WRF dataset, we can recompute the ϕ parameter by replacing, in equation 4, the values of the $\langle C_T^2 \rangle_m$, $\langle \chi(h) \rangle_m$, and $\langle s(h) \rangle_m$ by those coming from in-situ measurements and forecasting. The C_T^2 profile is computed by inverting equation 5 and using C_n^2 profile from PML's outputs. P

¹ <https://rda.ucar.edu/>

² <http://srtm.csi.cgiar.org/>

³ <https://www.oca.eu/fr/cats-predictions>

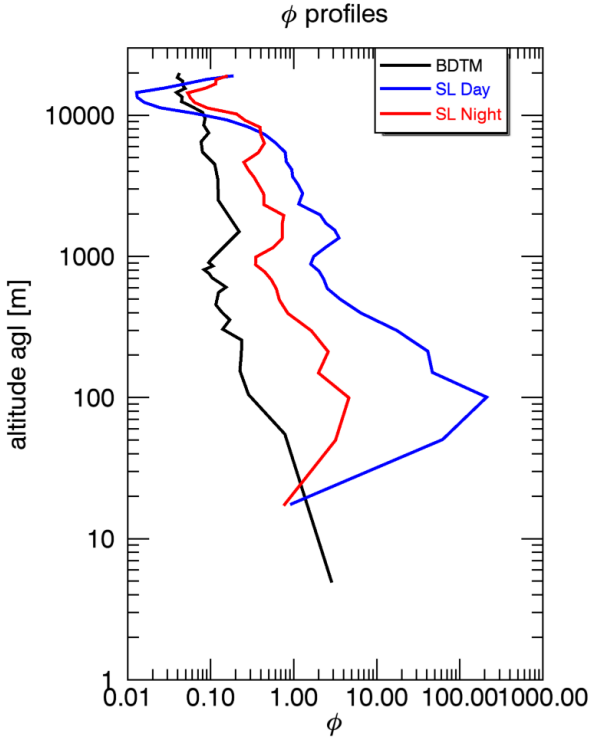


Figure 5. Vertical profiles of the ϕ parameters (see equation 4) for the BDTM model (black) and for the SL one considering daytime (blue) and nighttime (red) conditions apart.

and T profiles come from both the CATS ground weather station and the WRF output.

Another aspect that we took into account is that daytime turbulent conditions are more intense than nighttime ones. The BDTM model has been computed only from nighttime radiosoundings. Indeed, the radiosoundings balloons don't work during the daytime due to probes saturation. It is then obvious that BDTM model cannot reflect the daytime turbulence because of the large difference between the two regimes. Therefore, for the SL method, we have decided to compute two $\phi(h)$ profiles separating daytime and nighttime conditions. Figure 5 shows the BDTM $\phi(h)$ profile and the SL ϕ profile for both daytime and nighttime conditions. The difference between them are significant, especially during daytime where ϕ is larger than during the night.

In section 4.2, we present the effect of the upgrade of the turbulence forecasting computed using the WRF configuration detailed in section 2.2.

4 RESULTS

4.1 WRF meteorological forecasting accuracy

As explained in section 2.1 and 3, the turbulence forecasting is based on the use of meteorological prediction from the WRF model. It is then worthwhile to evaluate the WRF accuracy for meteorological forecasting. On Calern observatory, there are two different weather stations (WS), one integrated in CATS and another one which be-

longs to the observatory (OWS). They measure the wind speed modulus and direction, the temperature, the atmospheric pressure and the relative humidity at ground level, with a temporal resolution of about one minute. All measurements (meteorological and turbulence) have been averaged over a 10 minutes window to have the same resolution as WRF. We have used them to assess the accuracy of the WRF forecasting at ground level. Table 2 shows the Pearson correlation coefficient (PCC), the bias and the absolute bias for the pressure, the temperature, the wind speed modulus and the relative humidity. Figure 6 shows the corresponding 2D histograms. We have also added in this table 2 the comparison between the two weather stations.

The pressure is the best predicted parameter because it depends essentially on the altitude, which is invariable. The values computed between the two WSs and those computed between WRF & WS are comparable. The correlation coefficient is above 99% and the absolute bias is around 1.5hPa.

The temperature is also well predicted with a correlation coefficient above 95% (99% for WS comparisons). The absolute bias and its dispersion are around 3 times higher than for WSs comparison. This effect is visible on figure 6 where the cloud of points is thicker.

The wind speed is an important parameter because it mixes and shakes atmospheric layers and it generates dynamical turbulence and optical turbulence (if the temperature is heterogeneous). Its correlation coefficient reaches 75% (92% for WSs comparison) and its bias is of 1.7m/s which is a good result considering the resolution of our domain (1km). Indeed, the wind speed is really sensitive to the orography and to the ground effects and therefore the accuracy of its forecasting at ground level greatly depends on the spatial resolution of the domain and of the terrestrial data used as input.

The relative humidity is the most difficult parameter to measure precisely, and therefore, the 63% of correlation reached with our predictions is a good result even considering the dispersion shown in figure 6 which is 3 times higher than for the WSs comparison. However, the relative humidity has almost no impact in the visible and near-infrared observations. It is important because of the condensation on the mirrors of the telescopes.

The meteorological parameters being rather well predicted our assumption detailed in section 3 to use both prediction and measurements in the site learning method is correct, and we can now study the effect of this SL method.

4.2 Site learning impact analysis

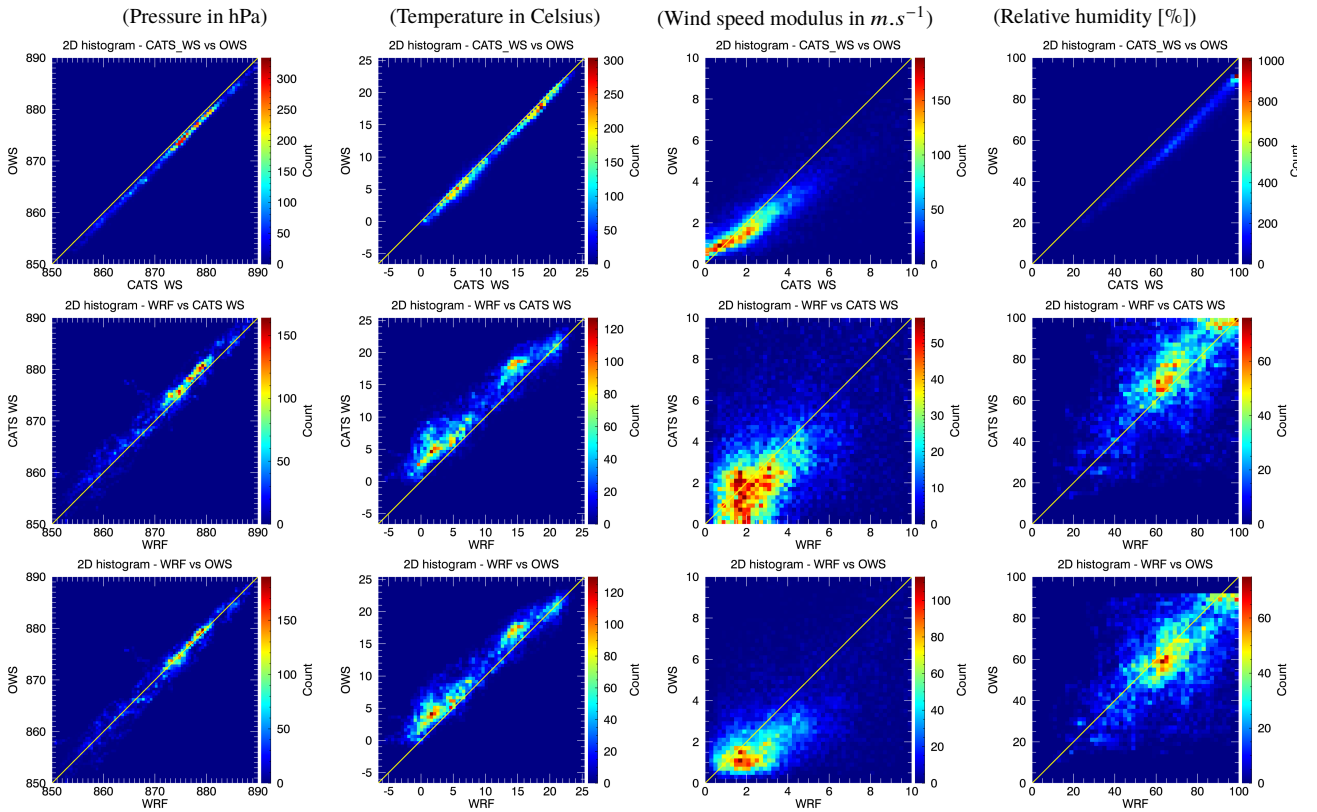
In this section we present the effects of the site learning phase used to optimize the turbulence model (see section 3).

Firstly, figures 7 and 8 show an example of the seeing and isoplanatic angle evolution measured by the CATS station instruments and predicted with WRF using the both BDTM and SL methods. We can see on these few samples the improvements brought by the SL phase. To go further, a statistical analysis over a larger sample is presented below.

The first parameter for which we have statistically analyzed the effect of the site learning is the vertical profile of the C_n^2 . Indeed, it is the main parameter characterizing the optical turbulence and used for AO system dimensioning and site testing. The in-situ data set comes from the PML instrument which is the only one able to measure the C_n^2 profiles. Figure 9 shows the C_n^2 vertical profiles measured by the PML and forecasted by the WRF model coupled with both BDTM and SL methods. The associated colored surfaces represent, for each layer, the interval between the first and third quar-

Table 2. Statistical comparisons between meteorological measurements from OCA and CATS weather stations (WS), and forecasting from the WRF model.

Statistical number	Comparison sources	Pressure	Temperature	Wind speed modulus	Relative humidity
Correlation PCC	CATS WS vs OCA WS	0.998	0.994	0.916	0.963
	WRF vs OCA WS	0.976	0.963	0.757	0.625
	WRF vs CATS WS	0.974	0.954	0.755	0.629
Bias	CATS WS vs OCA WS	-1.28 ± 0.54 hPa	-0.74 ± 0.73 °	-0.74 ± 1.53 m/s	-9.25 ± 5.34 %
	WRF vs OCA WS	-0.15 ± 1.82 hPa	1.98 ± 1.95 °	-0.98 ± 2.04 m/s	-7.12 ± 17.08 %
	WRF vs CATS WS	1.15 ± 1.92 hPa	2.73 ± 2.14 °	-0.22 ± 2.38 m/s	2.13 ± 17.36 %
Relative Bias [%]	CATS WS vs OCA WS	-0.15 ± 0.06	-0.26 ± 0.26	-20.18 ± 41.79	-12.72 ± 7.34
	WRF vs OCA WS	-0.02 ± 0.21	0.70 ± 0.69	-24.81 ± 51.85	-10.10 ± 24.24
	WRF vs CATS WS	0.13 ± 0.22	0.97 ± 0.76	-5.67 ± 60.53	3.03 ± 24.63
Absolute Bias	CATS WS vs OCA WS	1.29 ± 0.53 hPa	0.84 ± 0.61 °	1.10 ± 1.29 m/s	9.61 ± 4.64 %
	WRF vs OCA WS	1.26 ± 1.32 hPa	2.25 ± 1.63 °	1.66 ± 1.54 m/s	13.94 ± 12.17 %
	WRF vs CATS WS	1.70 ± 1.45 hPa	2.90 ± 1.90 °	1.73 ± 1.66 m/s	12.73 ± 11.99 %
Relative abs. Bias [%]	CATS WS vs OCA WS	0.15 ± 0.06	0.29 ± 0.21	30.15 ± 35.28	13.22 ± 6.39
	WRF vs OCA WS	0.14 ± 0.15	0.80 ± 0.58	42.10 ± 39.14	19.78 ± 17.26
	WRF vs CATS WS	0.19 ± 0.17	1.03 ± 0.68	43.89 ± 42.07	18.06 ± 17.02

**Figure 6.** Two-dimensional histograms of meteorological values measured by two different weather stations and predicted by WRF. From top to bottom we show histograms for respectively comparisons between the two weather stations, the OCA station (OWS) and WRF and the CATS station (CATS WS) and WRF. From left to right are shown respectively ground atmospheric pressure, ground temperature, ground wind speed modulus and ground relative humidity.

tile. The daytime and nighttime conditions have been separated to evaluate separately the two regimes. The PML profiles have been interpolated to the altitude of the WRF outputs. This interpolation has been done conserving the integrated energy of the turbulence over the thickness of each layer to keep the same seeing values. Regarding these curves, the SL method brings a significant improvement to the forecasts. Indeed, the BDTM model (green) underestimate the C_n^2 more than the SL one, and even considering first-third quartile interval, there is no overlapping. Considering the SL curve (blue),

the agreement is very good during nighttime, but there remains a small bias for daytime curves.

We have computed for each layer the Pearson's correlation coefficient, the relative mean bias and the relative mean absolute bias between predictions and measurements for daytime and nighttime conditions (Fig. 10). The C_n^2 values being spread over decades ($10^{-18} \lesssim C_n^2 \lesssim 10^{-13} m^{-2/3}$), we compute our statistical analysis on the decimal logarithm of the C_n^2 . From this figure 10 we deduced several interesting points:

- The correlation coefficient is improved by the SL method but it

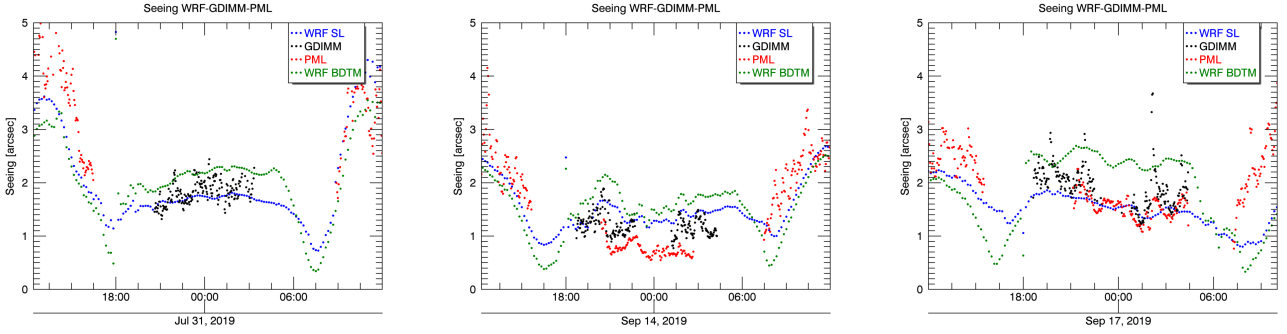


Figure 7. Examples of seeing evolutions measured with GDIMM (black) and PML (red) and forecasted with WRF coupled to BDTM model (green) and coupled to the SL model (blue).

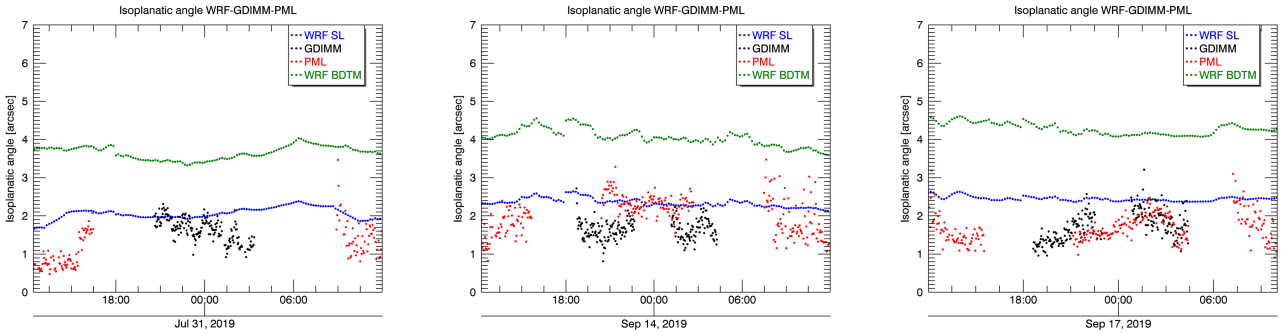


Figure 8. Example of isoplanatic angle evolutions measured with GDIMM (black) and PML (red) and forecasted with WRF coupled to BDTM model (green) and coupled to the SL model (blue).

remains low especially around 100m. This parameter can be difficult to interpret because it is largely impacted by divergent agreement between the two data sets, even if they are rare.

- SL method is more accurate than the BDTM one for the C_n^2 forecasting for both daytime and nighttime. In particular, the relative absolute bias has been improved by 25 to 50%.

- The predictions are more precise during the night. Indeed, the nighttime bias is 2 to 3 times lower than the daytime one.

- WRF with both turbulence models has more difficulties to well predict the turbulence within the first 500m of the atmosphere. The bias is maximum at an altitude of 100m. The difficulties to forecast ground effects (orography, soil type, ocean, ...) could be the origin of these discrepancies. Therefore, it is important to optimize forecasting within the first hundreds of meters by searching the best possible physical and micro-physical configuration of WRF.

- There is almost no difference between SL and BDTM methods in the first layer during the day. This is because during the day, in this layer, there is sometimes convection. In such case, the model used is the one from Tatarskii's formula (see 2.1) which is independent of the site learning phase. We are currently working on a way to inject site learning either in Tatarskii method or in the BDTM model even with convective gradient of the potential temperature.

From the C_n^2 profile, we deduced all turbulence parameters, and we compared them with measurements obtained with both GDIMM and PML. Then, we have selected only common sequences between CATS measurements and WRF forecasting. From these points, we have computed the same statistical parameters as in table 2 for the WRF & PML, WRF & GDIMM and PML & GDIMM comparisons.

Table 3 shows these statistics for the seeing forecasting and measurements separating daytime and nighttime conditions. It appears that the SL method improves the precision of the seeing forecasting. The correlation coefficient between the WRF-SL method and GDIMM measurements has been improved by a factor 300% (almost no improvement of the PCC considering PML data), but as aforementioned, the Pearson correlation coefficient is really sensitive to divergent values, even if they are rare. The absolute bias between the forecasting and both GDIMM and PML has been improved by about 25%. The bias becomes then closer to the one computed between GDIMM & PML seeing measurements. This last comparison is important because we can see that the error brought by the forecasting method gets closer to the error visible in instrumental cross-comparison.

Figures 11 to 13 show the 2D histograms of the scattered plot for the seeing considering both BDTM and SL methods and for instrumental cross comparison. The improvements shown in table 3 are well visible on these figures, especially regarding the highest density of point closer to the slope line $y = x$ with the SL method. They also show a larger dispersion in the scatter plot when forecasting is used with respect to instrumental cross-comparison (figure 13).

To have another comparison, we show the change brought by the SL method on the isoplanatic angle forecasting (table 4). Figure 14 shows the related 2D histograms between WRF forecasted and PML measurements and between GDIMM and PML. In Ziad et al. (2019b) the authors have shown that measuring the isoplanatic angle with PML and GDIMM gives more or less different

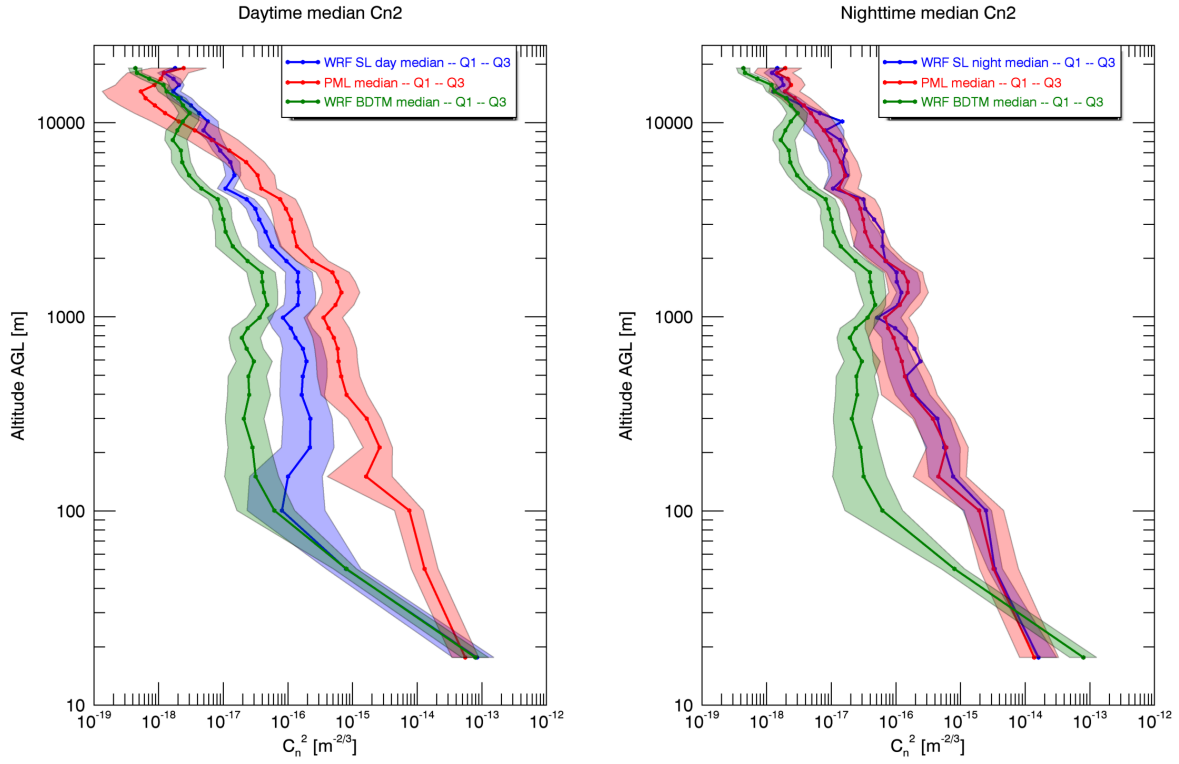


Figure 9. Median vertical profiles of the C_n^2 measured by PML (red) and predicted by WRF using either the BTDM (green) and SL (blue) methods. The site learning method considers daytime and nighttime conditions separately. The colored surfaces show the interval between the first and the third quartiles at each altitude. From left to right are plotted respectively the median for both daytime and nighttime data.

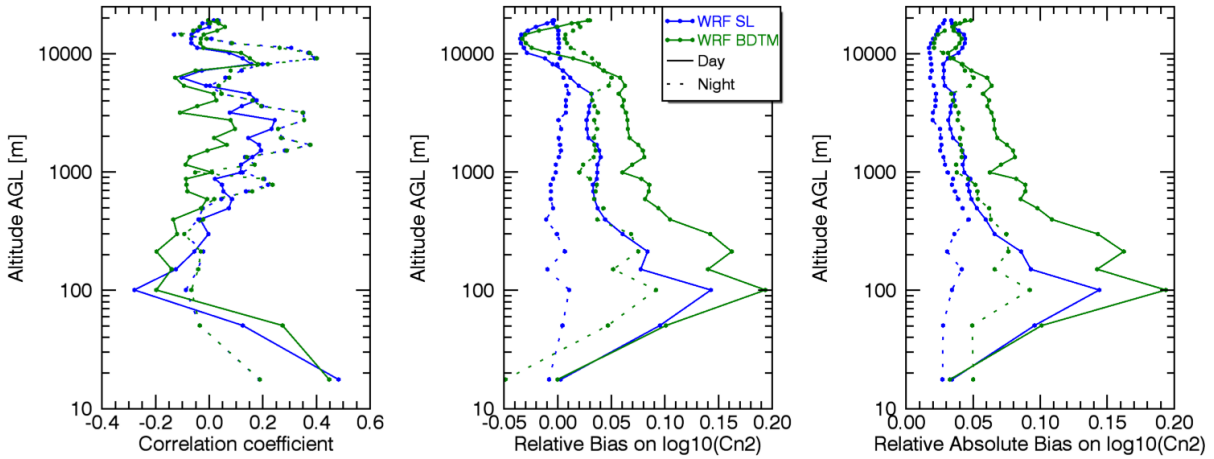


Figure 10. From left to right are shown the vertical profiles of the Pearson correlation coefficient, the relative bias and the relative absolute bias between WRF forecasting and PML measurements. Values are computed layer by layer on the $\log_{10}C_n^2$. The blue curves correspond to the SL method and the green ones correspond to the BDTM method. The solid lines and dashed lines show respectively daytime and nighttime analysis.

Table 3. Statistical comparisons between seeing measurements from GDIMM and PML, and forecasting from the WRF model coupled to the BDTM model and to the SL model which separating daytime and nighttime conditions.

Statistical number		PML vs GDIMM	WRF vs PML			WRF vs GDIMM		
			BDTM	SL model	Improv. [%]	BDTM	SL model	Improv. [%]
# common points		437	1424			2335		
Correlation PCC	day	-	0.67	0.65	-3	-	-	-
	night	0.82	0.20	0.21	5	0.06	0.24	300
Bias [arcsec]	day	-	0.73 ± 0.65	0.41 ± 0.68	44	-	-	-
	night	-0.20 ± 0.28	-0.29 ± 0.65	-0.14 ± 0.54	52	-0.28 ± 0.75	-0.10 ± 0.60	64
Relative Bias [%]	day	-	43.82 ± 38.63	20.48 ± 34.20	53	-	-	-
	night	-14.04 ± 19.46	-18.52 ± 41.26	-10.09 ± 38.17	46	-16.13 ± 43.86	-6.26 ± 39.08	61
Absolute Bias [arcsec]	day	-	0.80 ± 0.57	0.61 ± 0.51	24	-	-	-
	night	0.27 ± 0.21	0.57 ± 0.42	0.44 ± 0.34	23	0.64 ± 0.49	0.47 ± 0.38	27
Relative abs. Bias [%]	day	-	47.65 ± 33.78	30.46 ± 25.71	36	-	-	-
	night	18.78 ± 14.94	36.30 ± 26.95	31.11 ± 24.28	14	37.08 ± 28.43	30.67 ± 25.01	17

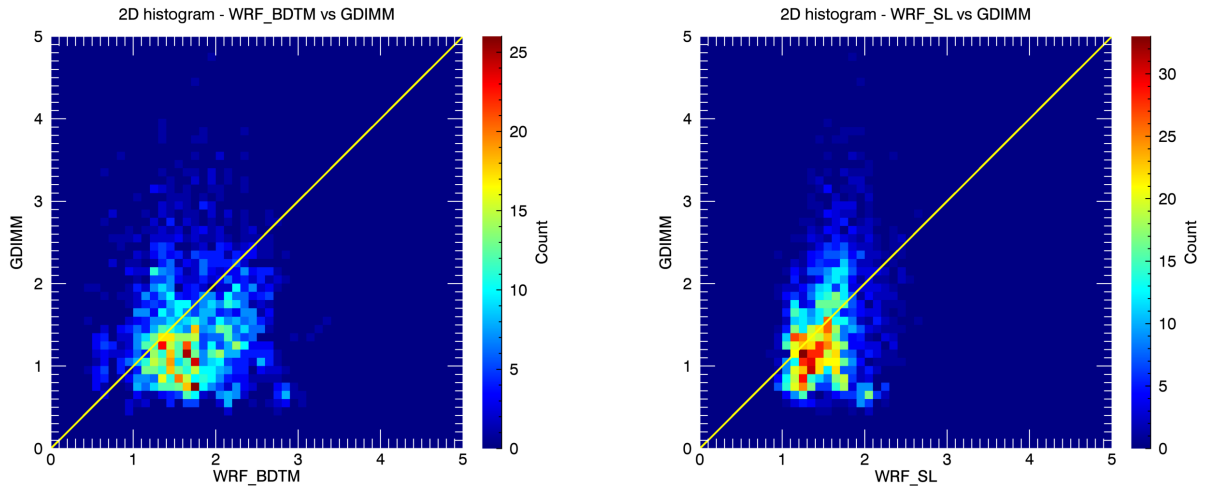


Figure 11. 2D histogram computed for common data between GDIMM seeing and respectively WRF seeing forecasted with BDTM model (left) and SL model (right). The yellow curve is the slope line $y = x$

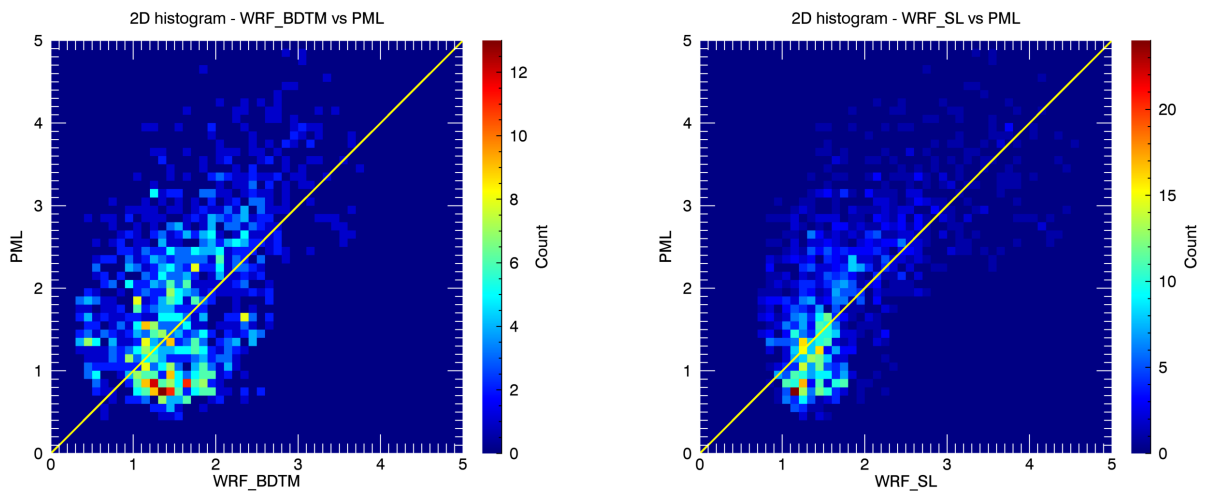


Figure 12. 2D histogram computed for common data between PML seeing and respectively WRF seeing forecasted with BDTM model (left) and SL model (right). The yellow curve is the slope line $y = x$

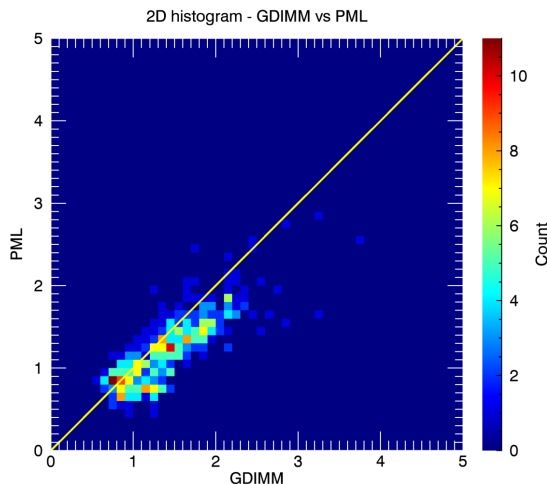


Figure 13. 2D histogram computed for common data between GDIMM and PML seeing. The yellow curve is the slope line $y = x$

results due to errors on the different instrumental techniques used; in figure 14 left we obtain the same results. We can also notice that the dispersion of the instrumental scatter plot is larger for the isoplanatic angle than for the seeing. Regarding table 4 and figure 14 the improvement brought by the SL method is more important for the isoplanatic angle than for the seeing for both daytime and nighttime conditions. Indeed, the absolute bias has been improved by 70-75% (25% for the seeing) to reach values comparable with instrumental bias only (PML vs GDIMM). Considering together figure 9 and table 4, we can deduce that this improvement is mainly due to the agreement between PML and forecasting curves within the highest layer. Indeed, the isoplanatic angle is greatly sensitive to the turbulent energy contained at high altitude (see formula in Giordano et al. (2012, 2013)).

5 DISCUSSION AND CONCLUSION

In this paper, we have presented an analysis of the forecasting tool we have used above the Calern Observatory from July 2019 to January 2020. We have used the WRF model coupled to an empirical turbulence model. Indeed, until now (Giordano et al. 2013, 2014; Giordano 2014) we used only the BDTM model computed from the radio-sounding analysis done in Trinquet & Vernin (2007). This model gives global tendencies, but it cannot retrieve fine effects due to local specificities of each site. We have then decided to upgrade the BDTM model by means of a site learning phase using a local database large enough to compute an optimized version of the turbulence model called SL method. The CATS station installed at the Calern Observatory since 2015 has a large database containing measurements of both C_n^2 vertical profiles and ground meteorological parameters. Therefore, we have decided to run our SL method on this site following the methodology described in section 3 and considering separately daytime and nighttime conditions because of the differences between the two regimes.

In section 4.1, we have compared WRF meteorological forecasting with the two weather stations installed on the Calern observatory, and we have found a very good agreement especially for the pressure and the temperature. The wind speed is also well predicted but with a higher dispersion certainly due to the sensitivity of this parameter to the ground effect and to the resolution of our domain.

In section 4.2, we have compared the performances of the SL method with the BDTM method and with the instrumental biases between GDIMM & PML. Our predictions cover more than 130 days with a temporal resolution of 10 minutes. We have shown the significant improvement brought by the SL method for the C_n^2 profiles, the seeing and the isoplanatic angle for both daytime and nighttime conditions. As an example, on the absolute bias, the forecasting accuracy has been improved by a factor varying from 25% to 80%.

The site learning phase is of a real interest because it will allow in the future to use forecasting in a real flexible scheduling tool minimizing errors and upgrading model day after day by learning the site specificities thanks to local measurements.

Currently the SL method considers daytime and nighttime conditions separately. We plan to refine the model to include other effects (season, Sun altitude...). For example, in figure 7 one can see that during the day-night and night-day transitions, the forecasting doesn't retrieve well the real turbulence conditions. Therefore, consider this period apart in SL method would improve the accuracy of our forecasting.

Nevertheless, we have to point out the fact that the surface layer ($h < 500m$) properties are poorly predicted as shown on figure 10. This could be due to ground effects which are particularly difficult to predict in circulation models because of the high resolution theoretically requested. Indeed, we have shown that even if the ground weather conditions are well predicted, it subsists some dispersion in particular on the wind speed prediction which is used in turbulence model. Moreover, currently, in the absence of meteorological profiles measurements, the SL method depends on the weather profiles coming from WRF to compute the $\phi(h)$ parameters. Therefore, the forecasting errors in these layer inject uncertainties within the SL model. To overpass these limits, we are currently studying different methods:

- The effects of the WRF configuration on meteorological results. A paper in preparation will present a comparison between in-situ meteorological profiling measurements and WRF forecasts using different configurations.
- We have to explore the possibility to refine the resolution of

Table 4. Statistical comparisons between isoplanatic angle measurements from GDIMM and PML, and forecasting from the WRF model coupled to the BDTM model and to the SL model separating daytime and nighttime conditions.

Statistical number		PML vs GDIMM	WRF vs PML			WRF vs GDIMM		
			BDTM	SL model	Improv. [%]	BDTM	SL model	Improv. [%]
# common points		437	1424			2335		
Correlation PCC	day	-	0.12	0.27	125	-	-	-
	night	0.50	0.55	0.49	-11	0.44	0.38	-14
Bias [arcsec]	day	-	-2.05 ± 0.66	-0.44 ± 0.58	79	-	-	-
	night	0.34 ± 0.51	-1.75 ± 0.49	-0.19 ± 0.46	89	-1.99 ± 0.50	-0.53 ± 0.45	73
Relative Bias [%]	day	-	-55.43 ± 7.85	-21.23 ± 27.89	62	-	-	-
	night	20.63 ± 31.04	-48.08 ± 13.50	-9.11 ± 21.96	81	-56.49 ± 14.35	-25.87 ± 21.78	54
Absolute Bias [arcsec]	day	-	2.05 ± 0.65	0.63 ± 0.38	69	-	-	-
	night	0.51 ± 0.35	1.75 ± 0.49	0.40 ± 0.29	77	1.99 ± 0.50	0.60 ± 0.35	70
Relative abs. Bias [%]	day	-	55.48 ± 17.68	30.00 ± 18.11	46	-	-	-
	night	30.56 ± 21.31	48.12 ± 13.37	19.42 ± 13.71	60	56.49 ± 14.35	29.21 ± 17.04	48

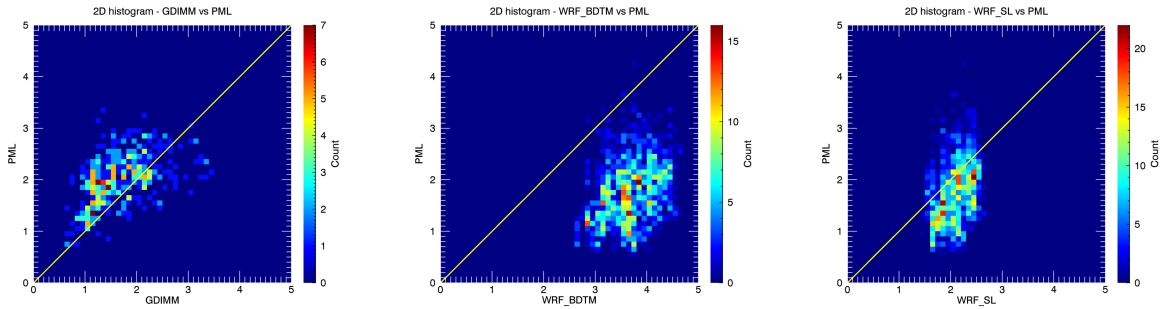


Figure 14. 2D histogram of the isoplanatic angle computed for common data between GDIMM & PML (left), PML & WRF coupled with BDTM model (middle) and SL model (right). The yellow curve is the slope line $y = x$.

our model (vertically and horizontally) to go closer to the scale of the ground phenomena and then to better model them.

- We have to search for finest input data (meteorological, topographic, ...) to initialize our simulations in the best way and then to have more precise results

- We are studying the possibility of acquiring instruments for in-situ measurements of the meteorological and optical conditions within the first 500m of the atmosphere. The instrument is currently in preparation and we hope to run first measurements during the Summer-Autumn 2020. Having vertical profiles in this portion of the atmosphere will allow us to improve again the turbulence model (see section 3) in this particular zone by using in-situ measurements instead of WRF forecasting (see 2.1).

- We will interest about the WRF data assimilation feature. This tool would allow us to inject local measurements directly within the WRF simulations to add a constraint and to improve weather prediction, and then to improve the SL method and the turbulence forecasting precision.

On the other hand, we are currently studying a new method based on the statistical or machine learning algorithms allowing us to use the large database acquired above Calern Observatory to run short term predictions without using circulation and/or turbulence model. This method could be complementary to the one presented in this paper. It could be used to correct the WRF forecasts by injecting real time measurements in the statistical learning tool to upgrade the predictions for the next hours.

Finally, it is important to notice that the studies presented here are directly dependent of the site measurements. Therefore, they give specific models usable at a given location. In practice this

learning phase must be done above each site interested by optical forecasting. This implies that each site/observatory interested by turbulence forecasting has to be equipped with specific turbulence monitors during enough time to optimize the forecasting. **In the future, we will study the effects of the different existing models allowing to deduce turbulence parameters from meteorological parameters. This work could point out the benefits and drawback of each model with respects to measurements but it needs an important study to master other meteorological and turbulence models.**

ACKNOWLEDGEMENTS

We would like to thank the Centre National d'Études Spatiales CNES for the financing of this activity. We gratefully thank the NCEP/NCAR for the availability of the WRF tool and for providing the GFS meteorological data. We thank the SRTM for giving access to the high resolution topographic data. We would also like to thank the Mesocentre SIGAMM of the Observatoire de la Côte d'Azur for hosting our model and for their help in all the installation process. **Finally we thank the reviewers for their help in improving this paper.**

DATA AVAILABILITY STATEMENTS

The data underlying this article cannot be shared publicly due to intellectual properties rights. Partial data could be shared on reasonable request to the corresponding author.

REFERENCES

- Aristidi E., Fantei-Caujolle Y., Ziad A., Dimur C., Chabé J., Roland B., 2014, A new generalized differential image motion monitor. p. 91453G, [doi:10.1117/12.2056201](https://doi.org/10.1117/12.2056201)
- Aristidi E., Fantei-Caujolle Y., Chabé J., Renaud C., Ziad A., Ben Rahhal M., 2018, in Proc. SPIE. p. 107036U ([arXiv:1811.07310](https://arxiv.org/abs/1811.07310)), [doi:10.1117/12.2313352](https://doi.org/10.1117/12.2313352)
- Aristidi E., Ziad A., Fantéi-Caujolle Y., Chabé J., Giordano C., Renaud C., Lantéri H., 2019a, in Proceedings of the sixth AO4ELT Conference.
- Aristidi E., Ziad A., Chabé J., Fantéi-Caujolle Y., Renaud C., Giordano C., 2019b, *MNRAS*, **486**, 915
- Blary F., Ziad A., Borgnino J., Fantéi-Caujolle Y., Aristidi E., Lantéri H., 2014, Monitoring atmospheric turbulence profiles with high vertical resolution using PML/PBL instrument. p. 91453H, [doi:10.1117/12.2056415](https://doi.org/10.1117/12.2056415)
- Bougeault P., de Hui C., Fleury B., Laurent J., 1995, *Appl. Opt.*, **34**, 3481
- Chabé J., Ziad A., Fantéi-Caujolle Y., Aristidi E., Renaud C., Blary F., Marjani M., 2016, The Calern atmospheric turbulence station. p. 99064Z, [doi:10.1117/12.2232383](https://doi.org/10.1117/12.2232383)
- Chabé J., Aristidi E., Ziad A., Lantéri H., Fantéi-Caujolle Y., Giordano C., Renaud C., 2020, *Applied Optics*, submitted
- Cherubini T., Businger S., Lyman R., Chun M., 2008, *Journal of Applied Meteorology and Climatology*, **47**, 1140
- Cherubini T., Businger S., Lyman R., 2011, An operational perspective for modeling optical turbulence. Seeing Clearly: The Impact of Atmospheric Turbulence on the Propagation of Extraterrestrial Radiation.. S. Businger and T. Cherubini, Eds., VBW Publishing, pp 165–182
- Coulman C. E., Andre J.-C., Lacarrere P., Gillingham P. R., 1986, *PASP*, **98**, 376
- Coulman C. E., Vernin J., Coqueugnot Y., Caccia J. L., 1988, *Appl. Opt.*, **27**, 155
- Giordano C., 2014, PhD thesis, Université de Nice Sophia Antipolis, Nice, France
- Giordano C., Vernin J., Chadid M., Aristidi E., Agabi A., Trinquet H., 2012, *PASP*, **124**, 494
- Giordano C., Vernin J., Vázquez Ramió H., Muñoz-Tuñón C., Varela A. M., Trinquet H., 2013, *MNRAS*, **430**, 3102
- Giordano C., Vernin J., Trinquet H., Muñoz-Tuñón C., 2014, *MNRAS*, **440**, 1964
- Giordano C., Rafalimanana A., Ziad A., Aristidi E., Chabé J., Fantei-Caujolle Y., Renaud C., 2019, in Proceedings of the sixth AO4ELT Conference.
- Hach Y., Jabiri A., Ziad A., Bounhir A., Sabil M., Abahamid A., Benkhal-doun Z., 2012, *MNRAS*, **420**, 637
- Mahalov A., Moustauoui M., 2010, *Laser & Photonics Rev.*, **4**, 144
- Mahalov A., Moustauoui M., Grubišić V., 2011, *Atmospheric Chemistry & Physics*, **11**, 5123
- Masciadri E., Vernin J., Bougeault P., 1999, *A&AS*, **137**, 185
- Masciadri E., Lascaux F., Turchi A., Fini L., 2017, *MNRAS*, **466**, 520
- Skamarock W. C., et al., 2019, Technical report, A Description of the Advanced Research WRF Version 4. NCAR
- Tatarskii V. I., 1961, *Wave Propagation in Turbulent Medium*. Dover Publication
- Trinquet H., Vernin J., 2007, *Environmental Fluid Mechanics*, **7**, 397
- Ziad A., 2016, in Proc. SPIE. p. 99091K, [doi:10.1117/12.2231375](https://doi.org/10.1117/12.2231375)
- Ziad A., et al., 2013, in Esposito S., Fini L., eds, Proceedings of the Third AO4ELT Conference. p. 126, [doi:10.12839/AO4ELT3.21355](https://doi.org/10.12839/AO4ELT3.21355)
- Ziad A., Chabé J., Fantei-Caujolle Y., Aristidi E., Renaud C., Ben Rahhal M., 2018, in Proc. SPIE. p. 107036L, [doi:10.1117/12.2313386](https://doi.org/10.1117/12.2313386)
- Ziad A., Chabé J., Fantei-Caujolle Y., Aristidi E., Renaud C., Giordano C., 2019a, in Proceedings of the sixth AO4ELT Conference.
- Ziad A., Aristidi E., Chabé J., Borgnino J., 2019b, *MNRAS*, **487**, 3664

APPENDIX A: CALERN FORECASTING WEBSITE

Since 2015, the CATS station ([Ziad et al. 2019a](https://arxiv.org/abs/2019a)) is running above the Calern observatory, and measures in real time all information about the local conditions of the observatory both for meteorological and optical turbulence parameters. All these measurements are uploaded in real time on a website ⁴ accessible for all users. It displays the following parameters:

- Meteorological: pressure, temperature, wind speed, relative humidity, cloud cover, precipitation
- Optical turbulence: C_n^2 , seeing, isoplanatic angle, scintillation

All these conditions are useful for observers of the observatory, especially for the metrology station MéO and the C2PU (Centre Pédagogique Planète et Univers) instruments which use in particular the seeing information.

In this framework we have decided to prepare the same tool but using forecasting data. Since July 2019, one forecasting is launched each day and covers 48 hours of predictions. The results are uploaded on the following website: <https://www.oca.eu/fr/cats-predictions>. This tool is used as a complement to the CATS station for observers because they are able to plan their own observations regarding 24h or 48h forecasting. The available parameters are:

- Meteorological: pressure, wind speed, temperature, relative humidity, precipitable water vapor, cloud cover.
- Optical turbulence: C_n^2 profile, seeing, isoplanatic angle.

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.

⁴ <https://cats.oca.eu>