SDS-WAS: Ensemble Prediction of Airborne Dust

G. García-Castrillo1, A. Callado1, E. Terradellas1 and P. Escribà1

1 State Meteorological Agency of Spain (AEMET), Barcelona, Spain

Abstract

The WMO SDS-WAS Regional Center for Northern Africa, Middle East and Europe has established a protocol to routinely exchange products from dust forecast models as the basis for model inter-comparison and forecast evaluation. Currently, 12 modeling groups provide daily forecasts of dust surface concentration (DSC) and dust optical depth (DOD) at 550 nm for a reference area intended to cover the main dust source areas in the region. The action involves forecasts up to 72 h with a 3-hour frequency.

Multi-model products are daily generated after bi-linearly interpolating all forecasts to a common grid mesh of 0.5 x 0.5 degrees. Centrality products (median and mean) are aimed at improving the forecasting skill of the single-model approach and spread products (standard deviation and range of variation) indicate whether forecast fields are consistent within the models, in which case there is greater confidence in the forecast.

Evaluation scores are routinely computed using aerosol optical depth retrievals provided by the AERONET network for 45 dust-prone stations. In a pilot study, forecasts of DSC have been compared with PM10 measurements performed by the Air Quality Control and Monitoring Network of the Canary Islands (Spain).

In this study, a one-month period has been selected to perform a deeper verification of the ensemble prediction system in order to evaluate its consistency and reliability. First, the ordinary deterministic verification of the different 12 models or members, as well as their median, has been carried out. Then, verification has been undertaken from a probabilistic point of view. This is a first step for the correct calibration of the system and the implementation of probabilistic forecast products as DSC and DOD EPSgrams. The study has been performed using the HARMONIE monitor deterministic and the HARP (Hirlam Aladin R-based package) probabilistic verification packages.

Keywords: *Ensemble prediction system, PM10, Dust optical depth, HIRLAM, HARP.*

1. Introduction

Forecasting severe weather events is a key objective for National Meteorological Services around the world. Due to the large amount of processes involved in those events and their non-linearity, a probabilistic approach is required. Ensemble prediction systems are a feasible framework and the most useful tool to improve such forecasts.

Ensemble prediction aims to describe the future state of the atmosphere from a probabilistic point of view. Multiple simulations are run to account for the uncertainty of the initial state and/or for the inaccuracy of the model and the mathematical methods used to solve its equations (Palmer et al., 2005). In particular, multi-model ensembles also represent a paradigm shift in which offering the best product to the users as a collective scientific community becomes more important than competing for achieving the best forecast as individual centres (Benedetti et al., 2014).

The World Meteorological Organization's Sand and Dust Storm – Warning Advisory and Assessment System (SDS-WAS, Terradellas et al., 2015) has the mission to improve the capacity of countries to produce and distribute to end-users accurate forecasts of the mineral dust content in the atmosphere. The SDS-WAS Regional Center for Northern Africa, Middle East and Europe (NA-ME-E) daily produces a poor-man ensemble (Atger, 1999) computed from the output of different models. Centrality products (median and mean) aim at improving the forecasting skill of the single-model approach. Spread products (standard deviation and range of variation) indicate whether the forecast fields are consistent within the contributing models, in which case there is greater confidence in the forecast (Terradellas et al., 2016).

The most relevant variables provided by dust prediction models are dust load, or alternatively dust optical depth (DOD), as a measure of the total dust contents in an atmospheric column, and dust surface concentration (DSC), as a measure of the dust contents near the ground. Other variables that are relevant for specific applications are dry and wet deposition or surface extinction.

The first problem of the forecast evaluation is the scarcity of suitable in-situ measurements, especially close of the main dust sources. The first option is the use of satellite products. However, satellite measurements are integrated over the atmospheric column and also over the different aerosol species. Another option is the use of ground-based photometric retrievals, but they present a similar problem. Initiatives to establish routine evaluation of dust predictions have been mainly focused on total-column DOD. The SDS-WAS Regional Center for NA-ME-E has set up and maintains a joint visualization and forecast evaluation (Terradellas et al., 2016), which currently involves 12 modeling systems and is based on AERONET (Holben et al., 1998; Dubovik and King, 2000) and MODIS retrievals (Kaufmann et al., 1997; Hsu et al., 2004). Other initiatives have been conducted in the framework of AeroCom (Huneeus et al., 2011), the Copernicus Atmosphere Monitoring Service (CAMS) (Eskes et al., 2015; Cuevas et al., 2015) and the International Cooperative for Aerosol Prediction (ICAP) (Sessions et al., 2015).

Many user communities are interested in the DSC (dust concentration in the air we breathe) rather than in the total column content. Therefore, evaluation of the predicted DSC is also necessary. Air quality monitoring networks are the main data providers for this purpose. They are common and with high spatial density in Europe, but very sparse and discontinuous close of the main source regions. The lack of observational data is particularly acute near the Sahara, the major dust source on Earth (Middleton and Goudie, 2001). In addition, the evaluation of dust forecasts using PM10 data has some drawbacks. On the one hand, the values of PM10 do not only reflect the mineral dust content in the atmosphere, but integrate the contribution of all airborne particles with aerodynamic diameter less than 10 µm, which may be of diverse origins (mineral dust, marine aerosol, anthropic pollution, etc.). On the other hand, dust prediction models provide the total content of mineral dust and, at least some of them, consider particles larger than 10 µm.

2. Observations and Forecasts

In the present study, DOD forecasts provided by different models and multi-model products are verificated for April 2016 using AERONET data. Verification is performed in spring, which is known to be the dustiest season in most parts of the geographical domain. Also, DSC forecasts are verificated using air-quality measurements in the Canary Islands. In this case, the verification is performed for December 2014 since Saharan dust outbreaks there occur near the ground normally in winter, but not in spring or summer. In both cases, verification is performed using the highest daily value both for observation and for prediction. This section describes the models and the observations involved in the study.

2.1. Observations

2.1.1. Aerosol Optical Depth

Direct-sun photometric measurements are a powerful tool that provides retrieval of column-integrated aerosol properties. In particular, AERONET is a comprehensive set of continental and coastal sites complemented with several sparsely distributed oceanic stations that provides large and refined data sets in near real-time (Holben et al., 1998; Dubovik and King, 2000). Retrievals from around 45 stations in Europe, Middle East and Northern Africa have been used here in the forecast verification (figure 1). A similar number of stations has been selected in each sub-region in order to prevent a part of the territory having more weight in the verification. In particular, level 1.5 of version 3 inversion products have been used. Level 1.5 data are cloud-screened, but the calibration correction has still not been applied.

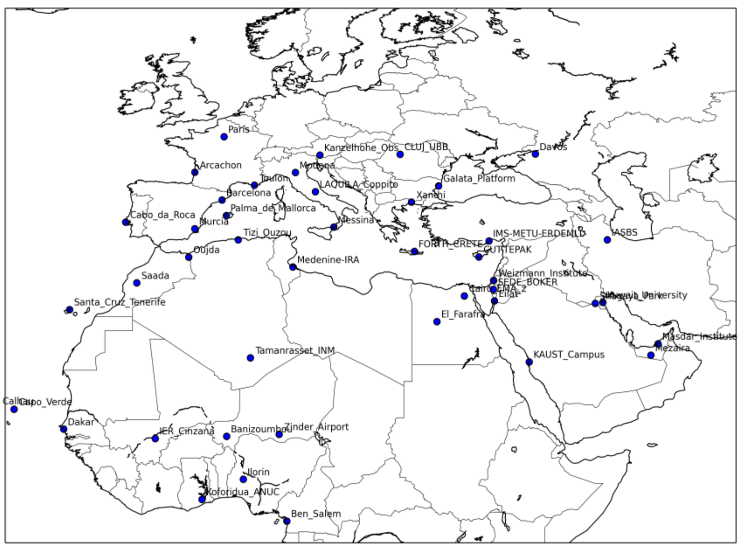

Figure 1. *AERONET Stations used in this work.*

To estimate the contribution of mineral dust to the total AOD, we have considered the coarse AOD yield by the spectral de-convolution algorithm described in O'Neill et al. (2003) that is part of the AERONET routine calculations. This algorithm yields fine (sub-micron) and coarse (super-micron) AODs at a standard wavelength of 500 nm.

2.1.2. Surface Concentration

The verification of DSC has been conducted for the Canary Islands. The archipelago suffers frequent intrusions of dust from the Sahara (i. e. Middleton and Goudie, 2001), with significant negative impacts, especially on air quality and health (Viana et al., 2002). Therefore, there is great interest in learning how the dust prediction models behave in the region. However, the complex orography of the islands, imperfectly represented in the models, especially in those with lower resolution, prevents a good simulation of the local variations of dust concentration and makes difficult a correct evaluation of the forecasts.

To quantify the contribution of mineral dust to PM10, the most reliable method is based on the chemical analysis of filters from gravimetric samplers (Rodrıguez et al., 2012). However, this is a very expensive and laborious technique, so it is difficult to apply routinely. As an alternative, the present work uses the coarse fraction of PM, defined as the difference PM10-PM2.5, as a proxy of the dust concentration.

We have selected five stations from the Canarian Air Quality Monitoring Network, operated by the regional government (table 1). As far as possible, the selection includes stations located away from urban centers, industrial parks and roads so that the contribution of anthropogenic particles in their records be small. Also, it has been intended that the location of the selected stations be representative of the different geographical areas of the archipelago (figure 2).

Table 1. *Air quality monitoring stations used in the study*

Site	Island	Measurement method
Costa Teguise	Lanzarote	TEOM
Polideportivo Afonso - Arucas	Gran Canaria (N)	Beta attenuation
Camping Temisas - Sta Lucía de Tirajana	Gran Canaria (S)	TEOM
Granadilla	Tenerife (S)	Scattering
Vuelta Los Pájaros - Santa Cruz de Tenerife	Tenerife (N)	Beta attenuation

Different continuous particle samplers are used in the network. They measure inertial mass (Tapered Element Oscillating Microbalance, TEOM), electron attenuation (Beta attenuation) or light scattering (scattering) of fine particles at a sampling rate of 1 hour. The reference (gravimetric) method to measure PM10 and PM2.5 consists of acquiring deposits over 24-hour periods on teflon membrane filters from air drawn at a controlled flow rate through the corresponding inlet. Then, a correction factor obtained through sampling campaigns has to be introduced to adjust the results to the reference method. The data used in the present study December 2014 had already been corrected by the network managers.

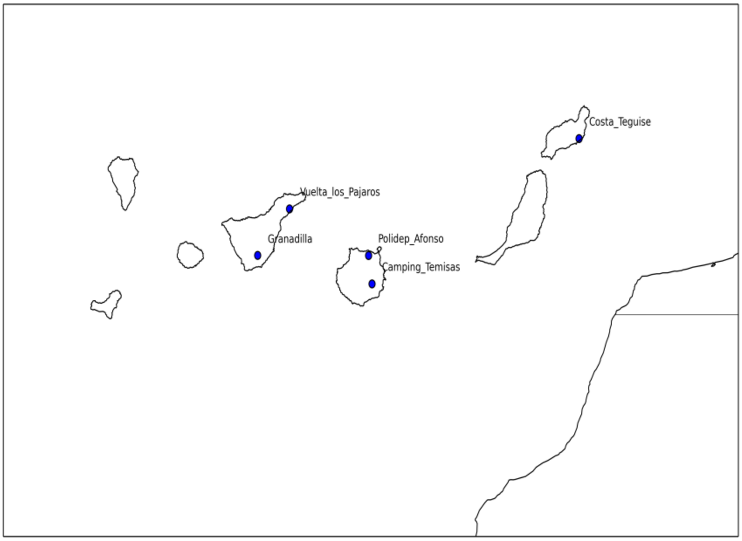

Figure 2. *Location of the air-quality monitoring stations used in the present study*

2.2. Forecasts

Daily predictions of DOD and DSC released by twelve dust prediction models have been considered in this work. The models have very different characteristics: there are global and limited-area models, some of them incorporate schemes of data assimilation, others do not. Their horizontal and vertical resolutions are diverse, as well as their meteorological drivers, parameterisation of the different steps of the dust cycle and physiographical databases of land use, soil texture, etc. The list of the models and their main characteristics are summarized in table 2.

Table 2. *Dust models involved in the study*

Model	Institution	Run time	Domain	Data assimilation
BSC-DREAM8b-v2	Barcelona Supercomputing Center	12 UTC	Regional	No
CAMS	ECMWF	00 UTC	Global	MODIS AOD
DREAM8-NMME-MACC	SEEVCCC	00 UTC	Regional	CAMS analysis
MetUM	Met Office	00 UTC	Global	MODIS AOD
NMMB/BSC-Dust	Barcelona Supercomputing Center	12 UTC	Regional	No
GEOS-5	NASA	00 UTC	Global	MODIS
NGAC	NCEP	00 UTC	Global	No
EMA REG CM4	Egyptian Meteorological Authority	00 UTC	Regional	No
DREAMABOL	ISAC	00 UTC	Regional	No
NOA WRF-CHEM	National Observatory Athens	12 UTC	Regional	No
FMI-SILAM	FMI	00 UTC	Global	No
LOTOS-EURO	TNO	00 UTC	Regional	MODIS AOD

The collected models have different run times (00 or 12 UTC, as shown in Table 2). In order to set the start time of the multi-model products at 00 UTC, we consider the previous-day runs of the models starting at 12 UTC.

A problem we have to deal with is the eventual lack of availability of models since most of them are not run in an operational mode. Daily availability of models is shown in Figure 3. We have decided not to name the models in the Figures in this study.

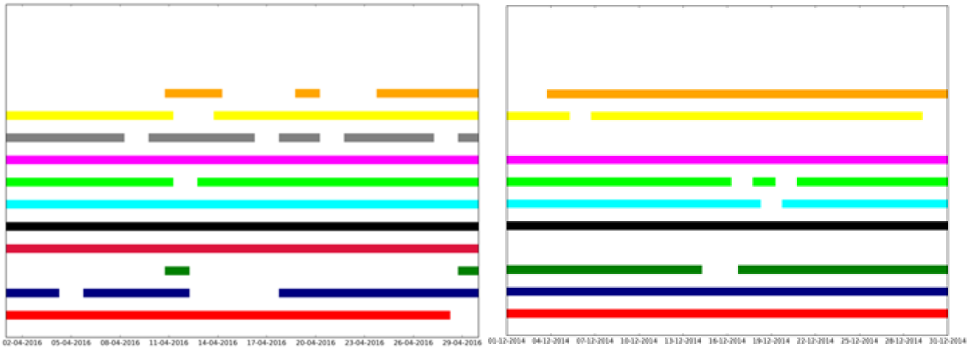

Figure 3: *Model availability for April 2016 (left) and December 2014 (right).*

3. Experimental Method

In this Section we describe the method followed to verify the dust forecasts. First, we proceed with a deterministic verification, which give us information about the quality of the different models. One of the objectives of this step is to determine which models will be part of the ensemble. The next step is the probabilistic verification, which gives us information about the quality and the consistency of the ensemble. Consistency means the degree to which the forecast corresponds to the forecaster's best judgement about the situation, based upon his/her knowledge base (Murphy, 1993). The consistency is related to the need for calibration and this is one key point of this work. If the ensemble has bad consistency, we will introduce a calibration to adjust the ensemble to the observation. All this process has the purpose of supplying value products for the end-users.

We have used two packages to verify DOD and DSC ensembles. As Spanish Meteorological Agency (AEMET) belongs to HARMONIE consortium for developing mesoscale Numerical Weather Prediction (NWP) modeling, we have used HARMONIE packages for deterministic and probabilistic verification. These packages are monitor and HARP (Hirlam Aladin R-based package) (HARMONIE wiki page).

4. Results

4.1 Dust optical depth

Figure 4 shows the mean bias and standard deviation of daily maximum DOD for lead times of 12 and 36 hours in April 2016. The plot only contains the eight models that will be part of our ensemble. We have removed three models for different reasons: two of them because of their sparse availability and the other because the scores were very far from those of the rest of models. The plot has been built after comparison of DOD forecasts with AERONET retrievals from 35 stations.

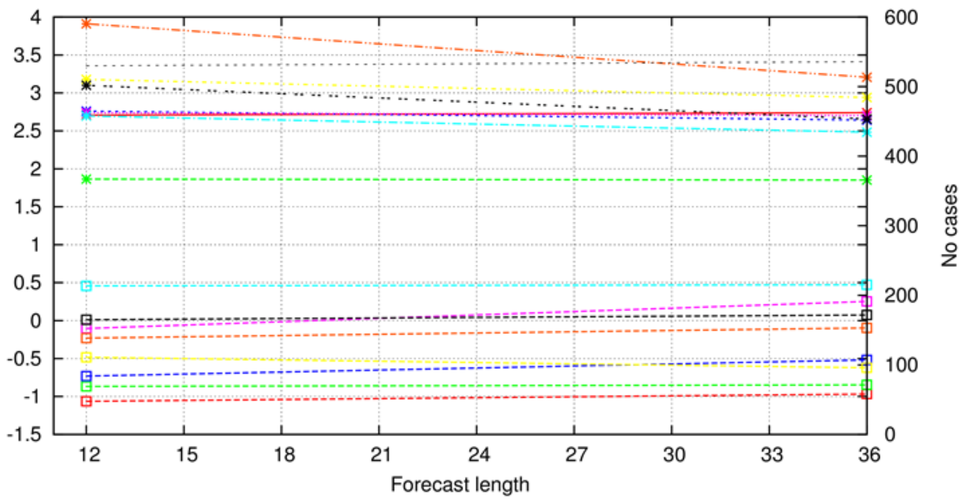

Figure 4. *Deterministic verification of daily maximum DOD for lead times of 12 and 36 hours in April 2016 (* represents standard deviation and □ bias Each color corresponds to a different model.)*

.

We present two of the most common methods to determine the consistency of an ensemble. The first one is based on the Rank Histogram or Talagrand diagram (Talagrand et al, 1997; Hamill, 2001). If the ensemble forecast is consistent, the Rank Histogram will be flat. Deviations from a uniform distribution denote lack of consistency. Figure 5 shows the Rank Histogram of our ensemble for forecast times of 12 and 36 hours. Both plots show a slightly descending tendency, indicating that there is a small positive bias, which means over-prediction of the ensemble.

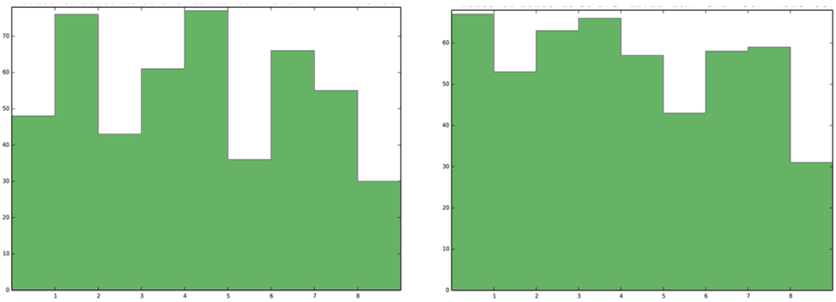

Figure 5: *Rank Histogram of daily maximum DOD in April 2016 for 35 stations. The left plot corresponds to a lead time of 12h and the right one to 36h*

A second method to assess the quality of an ensemble is based on the reliability and sharpness diagrams. The reliability diagram groups the forecasts into bins according to the issued probability of exceeding a specific threshold (horizontal axis). The frequency with which the event was observed to occur for this sub-group of forecasts is then plotted against the vertical axis. For perfect reliability the forecast probability and the frequency of occurrence should be equal, and the plotted points should lie on the diagonal

The sharpness diagrams show the frequency with which the event has been predicted with different levels of probability. Forecast systems that are capable of predicting events with probabilities different from the observed event frequency are said to have 'sharpness'. Diagrams for forecast systems with little sharpness would exhibit a frequency peak near the climatological frequency. So the ideal ensembles would present a U-shape, in which case the ensemble perfectly predicts an event or discards it (Hamill, 1997).

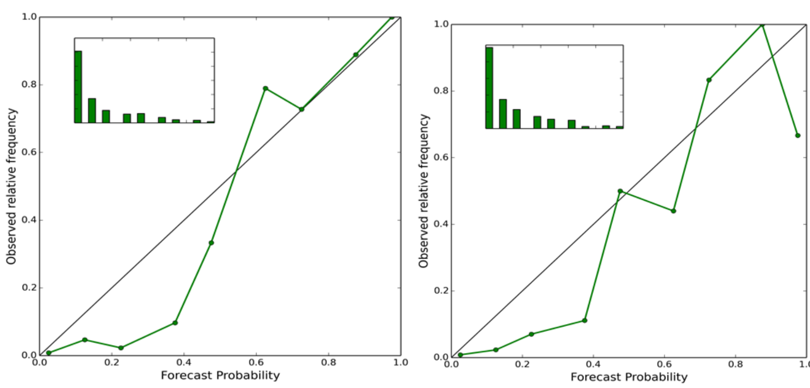

Figure 6*. Reliability (large plot) and sharpness (small plot) diagrams of daily maximum DOD in April 2016 and a threshold of 0,5. The left plot corresponds with a leadtime of 12h and the right one of 36h*

Figure 6 shows the reliability (large plot) and sharpness (small plot) diagrams for a threshold value of 0,5 and forecast times of 12 and 36 hours. The sharpness diagrams present a normal shape, as most of the ensembles. Regarding the reliability diagrams, the plots are not far from the diagonal. However, it can be mentioned that when the ensemble predicts a low probability of ocurrence, the observed frequency is smaller than expected. This means that our ensemble is over-predictive for some models (left region of the diagram).

4.2 Dust Surface Concentration

Figure 7 shows the mean bias and standard deviation of daily maximum DSC for lead times of 12 and 36 hours in December 2014. The plot contains the nine available models in this period; in this case any of the models has been removed. The plot has been built after comparison of DSC forecasts with PM data from 5 stations.

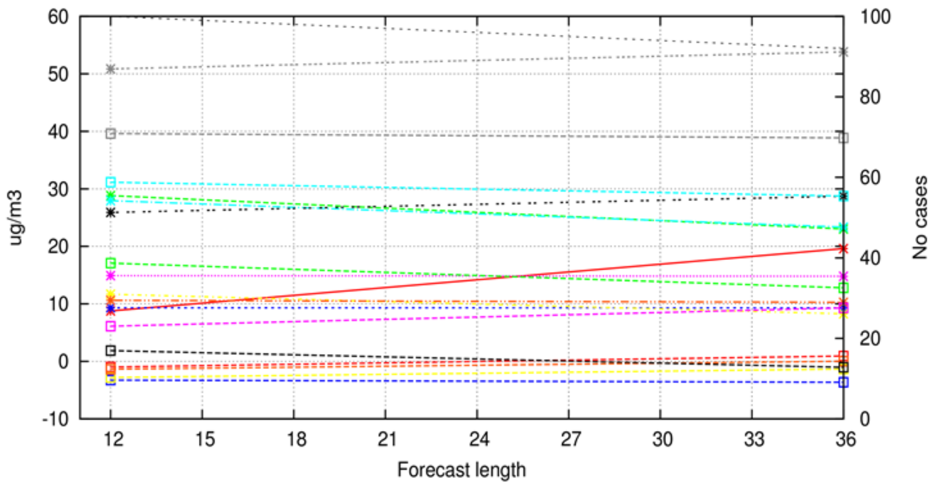

Figure 7. *Deterministic verification of daily maximum DSC for lead times of 12 and 36 hours in December 2014 (* represents standard deviation and □ bias. Each color corresponds to a different model.)*

The Rank Histogram for DSC (see Figure 8) has a different shape in comparison with that for DOD. In this case, most observations fall in the central bins of the ensemble and the plot shows a dome shape. It means that the ensemble spread too large and the reason is that the models yield too scattered results.

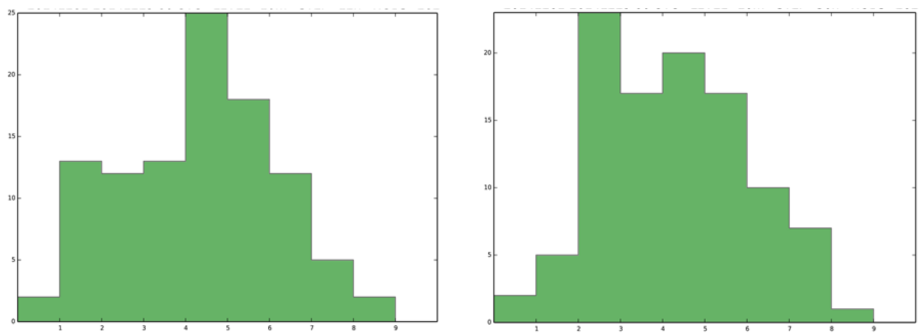

Figure 8. *Rank Histogram of daily maximum DSC in December 2014 for 5 stations. The left plot corresponds to a lead time of 12h and the right one to 36h*

Figure 9 shows the reliability and the sharpness diagrams for a threshold of 50 µg/m^3. In this case, results are not satisfactory. On the one hand, sharpness is scarce. On the other hand, most points in the reliability diagram lie far from the diagonal, with important over-forecast in the left half of the plot.

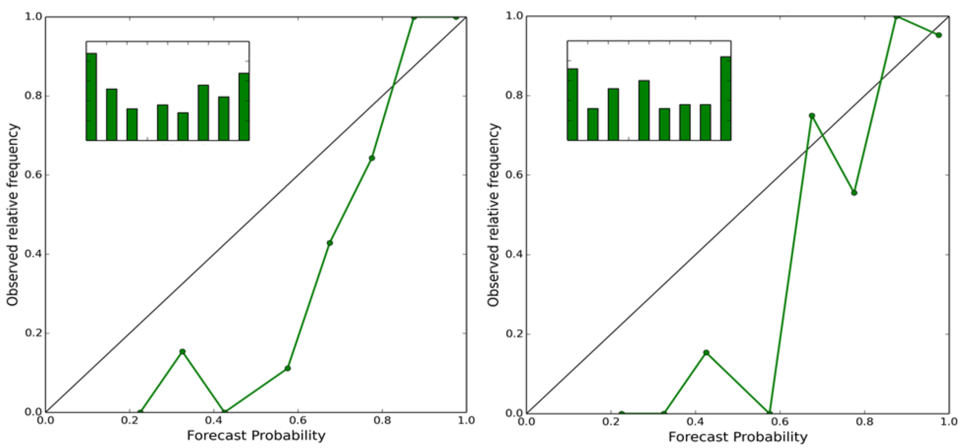

Figure 9. *Reliability diagram of daily maximum DSC in December 2017 and a threshold of 50 µg/m^3. The left plot corresponds with a leadtime of 12h and the right one of 36h .*

5. Conclusions and open issues

In the present work, we assess the skill of an ensemble prediction system to forecast DOD and DSC using probabilistic and deterministic verification.

For DOD, our ensemble is built from seven models. The Rank Histogram presents a slightly descending tendency, denoting a small over-prediction. The sharpness diagram presents a typical shape for this parameter. Finally, the reliability diagram denotes over-forecasting in the region of low probabilities. So, we have a relatively good ensemble in terms of consistency, sharpness and reliability.

For DSC, it is important to bear in mind that conclusions have limited significance, since evaluation has been performed with only five stations from the Canary Islands. In this case our ensemble has been built from nine dust models. In the probabilistic verification we find a too large spread, over-forecasting in low-medium probabilities and poor sharpness. In short, we have a worse ensemble than for DOD.

From these results we can conclude that DOD is much more predictable from a probabilistic point of view than DSC with our system. Finally, bias correction and other post-processing tools like statistical calibration could potentially increase the quality of these ensembles.

References

Atger, F. (1999). The skill of ensemble prediction systems. Monthly Weather Review, 127(9), 1941-1953.

Benedetti, A., Baldasano, J. M., Basart, S., Benincasa, F., Boucher, O., Brooks, M. E., ... & Jones, L. (2014). Operational dust prediction. In Mineral Dust (pp. 223-265). Springer Netherlands.

Callado, A., Escribà, P., García-Moya, J. A., Montero, J., Santos, C., Santos-Muñoz, D., & Simarro, J. (2013). Ensemble forecasting. In Climate change and regional/local responses. InTech.

Cuevas, E., Camino, C., Benedetti, A., Basart, S., Terradellas, E., Baldasano, J. M., ... & Berjón, A. (2015). The MACC-II 2007–2008 reanalysis: atmospheric dust evaluation and characterization over northern Africa and the Middle East. Atmos. Chem. Phys, 15(8), 3991-4024.

Dubovik, O., & King, M. D. (2000). A flexible inversion algorithm for retrieval of aerosol optical properties from Sun and sky radiance measurements. Journal of

Geophysical Research, 105(D16), 20673-20696.

Eskes, H., Huijnen, V., Arola, A., Benedictow, A., Blechschmidt, A. M., Botek, E., ... & Engelen, R. (2015). Validation of reactive gases and aerosols in the MACC global analysis and forecast system. Geoscientific model development, 8(11), 3523-3543.

Hamill, T. M. (2001). Interpretation of rank histograms for verifying ensemble forecasts. Monthly Weather Review, 129(3), 550-560.

Hamill, T. M. (1997). Reliability diagrams for multicategory probabilistic forecasts. Weather and forecasting, 12(4), 736-741.

HARMONIE wiki page. https://hirlam.org/trac/wiki/HarmonieSystemDocumentation

Holben, B. N., Eck, T. F., Slutsker, I., Tanre, D., Buis, J. P., Setzer, A., ... & Lavenu, F. (1998). AERONET—A federated instrument network and data archive for aerosol characterization. Remote sensing of environment, 66(1), 1-16.

Hsu, N. C., Tsay, S. C., King, M. D., & Herman, J. R. (2004). Aerosol properties over bright-reflecting source regions. IEEE Transactions on Geoscience and Remote Sensing, 42(3), 557-569.

Huneeus, N., Schulz, M., Balkanski, Y., Griesfeller, J., Prospero, J., Kinne, S., ... & Diehl, T. (2011). Global dust model intercomparison in AeroCom phase I. Atmospheric Chemistry and Physics, 11(15).

Murphy, A.H., 1993: What is a good forecast? An essay on the nature of goodness in weather forecasting. Wea. Forecasting, 8, 281-293.

Kaufman, Y. J., Tanré, D., Remer, L. A., Vermote, E. F., Chu, A., & Holben, B. N. (1997). Operational remote sensing of tropospheric aerosol over land from EOS moderate resolution imaging spectroradiometer. Journal of Geophysical Research: Atmospheres, 102(D14), 17051-17067.

Middleton, N. J., & Goudie, A. S. (2001). Saharan dust: sources and trajectories. Transactions of the Institute of British Geographers, 26(2), 165-181.

Palmer, T. N., Shutts, G. J., Hagedorn, R., Doblas-Reyes, F. J., Jung, T., & Leutbecher, M. (2005). Representing model uncertainty in weather and climate prediction. Annu. Rev. Earth Planet. Sci., 33, 163-193.

Pérez, C., Nickovic, S., Baldasano, J. M., Sicard, M., Rocadenbosch, F., & Cachorro, V. E. (2006). A long Saharan dust event over the western Mediterranean: Lidar, Sun photometer observations, and regional dust modeling. Journal of Geophysical Research: Atmospheres, 111(D15).

O'Neill, N. T., Eck, T. F., Smirnov, A., Holben, B. N., & Thulasiraman, S. (2003). Spectral discrimination of coarse and fine mode optical depth. Journal of Geophysical Research: Atmospheres, 108(D17).

Rodríguez, S., Alastuey, A., & Querol, X. (2012). A review of methods for long term in situ characterization of aerosol dust. Aeolian Research, 6, 55-74.

Sessions, W. R., Reid, J. S., Benedetti, A., Colarco, P. R., da Silva, A., Lu, S., ... & Brooks, M. E. (2015). Development towards a global operational aerosol consensus: basic climatological characteristics of the International Cooperative for Aerosol Prediction Multi-Model Ensemble (ICAP-MME). Atmos. Chem. Phys, 15, 335-362.

Talagrand, O. (1999). Evaluation of probabilistic prediction systems. In Workshop proceedings" Workshop on predictability", 20-22 Oc. 1997, ECMWF, Reading, UK.

Terradellas, E., Nickovic, S., and Zhang, X. (2015): Airborne dust: a hazard to human health, environment and society, WMO Bulletin, 64, 42–46.

Terradellas, E., Basart, S., & Cuevas, E. (2016). Airborne dust: From R&D to Operational forecast. AEMET, Madrid, NIPO: 281-16-007-3; WMO, Geneva, WMO/GAW Report No. 230; WMO/WWRP No. 2016-2.

Viana, M., Querol, X., Alastuey, A., Cuevas, E., & Rodrıguez, S. (2002). Influence of African dust on the levels of atmospheric particulates in the Canary Islands air quality network. Atmospheric Environment, 36(38), 5861-5875.