H14-123

EVALUATION OF WRF MODEL PERFORMANCES IN DIFFERENT EUROPEAN REGIONS WITH THE DELTA-FAIRMODE EVALUATION TOOL

M. Marcello Miglietta[1,2], P. Thunis[3], A. Pederzoli[3], E. Georgieva[3], B. Bessagnet[4], E. Terrenoire[4], and A.Colette[4]

[1]Institute of Atmospheric Sciences and Climate - National Research Council (ISAC-CNR), Lecce, Italy
[2]Institute of Ecosystem Study - National Research Council (ISE-CNR), Verbania Pallanza, Italy
[3]Institute for Environment and Sustainability - Joint Research Center (IES-JRC), Ispra, Italy
[4]National Institute for Industrial Environment and Risks (INERIS), Paris, France

Abstract: One-year (2006) WRF model simulations performed at a European scale with 30 and 10 km horizontal resolutions are evaluated against ECMWF-IFS forecasts (both at 0.2° and 0.5° horizontal resolution). The comparison with the wind speed observations from around 1200 surface stations is performed using the DELTA software, developed in the framework of FAIRMODE. Generally, ECMWF forecasts are pretty good in terms of different statistical indices over most of the domain, while WRF model simulations are less skillful, producing a larger bias and RMSE. Then, the statistical indices are evaluated in some specific areas, that is near Berlin, which is one of the urban areas where the model performance is better, and in the Alpine area, where the model skill is very low. Some considerations about the interpretation of such results are finally provided.

***Key words:** Numerical Weather Prediction, Model Validation, Benchmarking.*

INTRODUCTION

In order to bring together air quality modellers and users and to promote and support the harmonised use of modelling practices for the assessment of air quality by EU member countries, the “Forum for AIR quality MODElling”(FAIRMODE, http://fairmode.ew.eea.europa.eu/)” focuses on scientific research that will establish improved and validated modelling tools on which decision making will be based. The forum is a joint response action of the European Environment Agency (EEA) and the European Commission Joint Research Centre (JRC) towards the new Air Quality Directive (AQD, 2008 http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:152:0001:0044:EN:PDF).

In the framework of FAIRMODE, a quality assessment also focuses on meteorological models. Meteorology is indeed an important factor contributing to air quality, as it encompasses many atmospheric processes that control or strongly influence the evolution of emissions, chemical species, aerosols and particulate matter. Meteorological fields supplied to air-quality models are generally provided by Numerical Weather Prediction (NWP) models. NWP models solve numerically the physical equations describing atmospheric motions and processes and are usually divided into two major categories: Global Circulation Models (GCM) and Limited Area Models (LAM). Due to their coarse resolution, corresponding in the best cases to some tens of km, GCM cannot simulate accurately the weather in regional basins. The task of simulating the meteorological fields with better local accuracy is usually pursued by LAM, which integrate the model equations only on limited geographical domains, but with finer horizontal resolution.

Unfortunately, these modelling systems contain significant uncertainties which adversely affect air quality model simulations (e.g., Sistla et al., 1996). For this reason, it is extremely important to perform a preliminary evaluation of the meteorological input as a necessary step toward a better understanding of the air quality model performances. Criteria to identify conditions for which a meteorological forecast can be considered acceptable to be used as input into an air quality model have been proposed in the past for different parameters and different metrics (Emery et al., 2001).

In this work, one-year (2006) simulations at European scale, performed with different modelling tools, are compared with each other and with surface observations. The skill of the models is analysed using the evaluation tools included in the DELTA software, recently developed at JRC. Hereafter, after a preliminary description of DELTA, of the modelling tools and of the dataset of surface observations used for sake of comparison, an overview of some preliminary results is provided.

DELTA

DELTA has been designed in the framework of the FAIRMODE activity to support modelling groups in their applications related to the implementation of the AQD. It proposes rapid diagnostics for meteorological and air quality model performances, focusing on the pollutants mentioned in the AQD and addressing all relevant spatial scales (from local to regional). It is an IDL-based off-line evaluation software; it works with modelled-observed data pairs at surface level, i.e. temporal series of modelled and monitored data at selected ground level locations. Model performances are assessed with respect to “criteria” representing the level of accuracy considered to be acceptable for regulatory applications. Both meteorological (scalars) and air quality data can be handled. Different statistical indicators and diagrams can be produced by the tool: they have been selected based on literature review, favouring the usage of composite diagrams. Stations can be grouped into clusters, which are based on the geographic characteristics (region, topography, distance from seashore, proximity to a urban area, …). A temporal analysis can be also performed to identify model’s strengths and weaknesses in terms of seasons and daytime.

DATA

Model results are compared with surface observations extracted from a set of more than 2000 meteorological (WMO) stations all over Europe. Observations have been provided by INERIS (French National Institute for Industrial Environment and Risks) and consist of three-hourly (SYNOP) or hourly data (METAR). After a preliminary screening of the stations, based on data availability and the need to cover the European domain as uniformly as possible (including some stations in northern Africa too), 1198 stations have been selected for the statistical analysis.

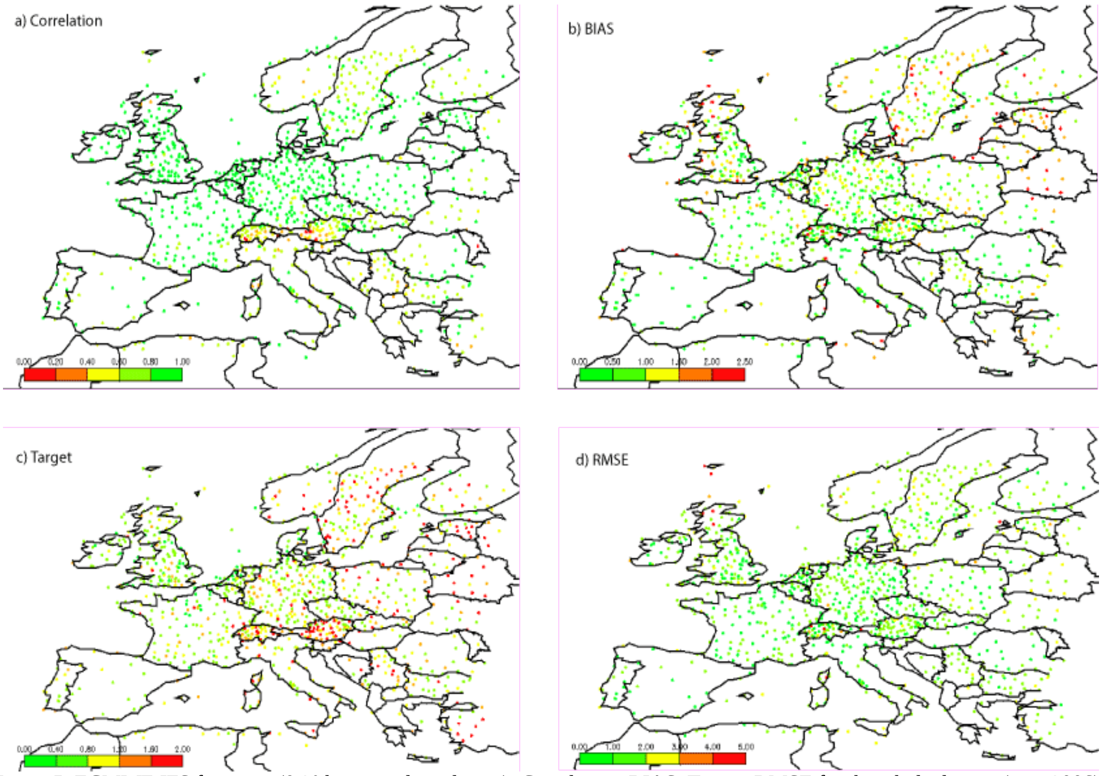

Figure 5. ECMWF-IFS forecasts (0.2° horizontal resolution): Correlation, BIAS, Target, RMSE for the whole dataset (year 2006).

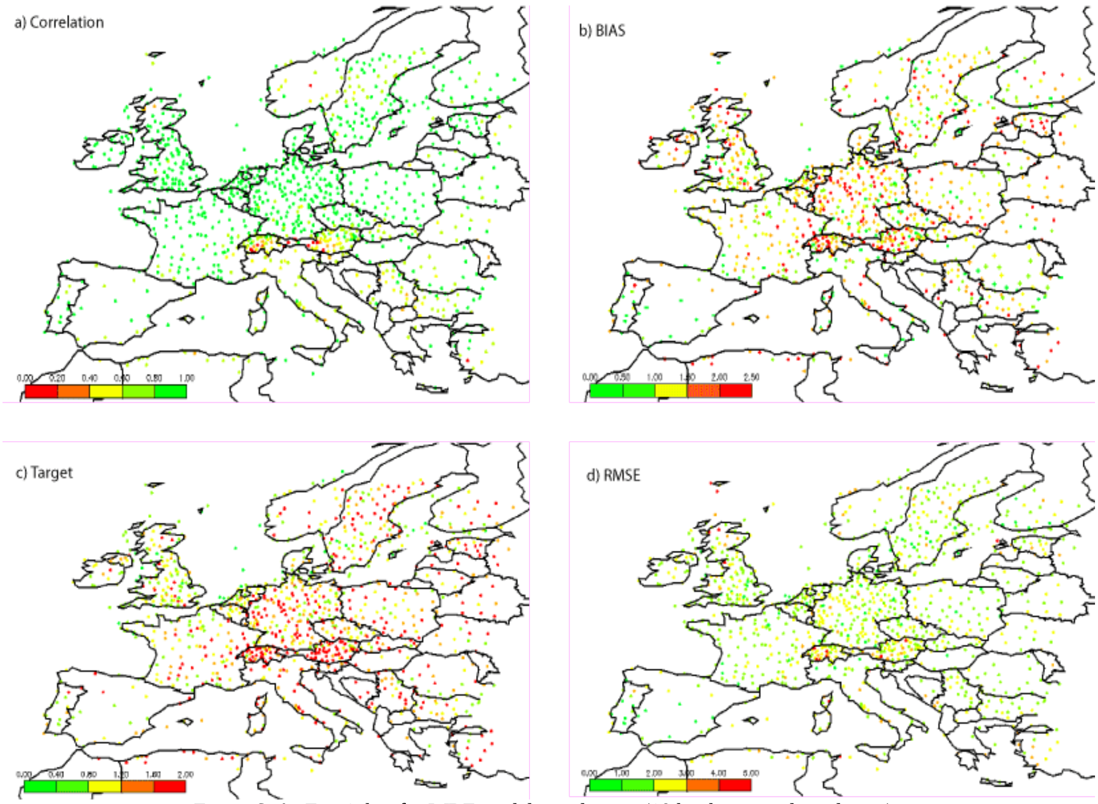

Figure 6. As Fig. 1, but for WRF model simulations (10 km horizontal resolution).

A GCM and a LAM are considered for sake of comparison. Both the modelling system outputs are used as input for air quality models. The GCM is the Integrated Forecast System (*IFS*) of the European Centre for Medium-Range Weather

Forecasts (*ECMWF*). For 2006, the model was implemented with a spectral resolution TL512 (about 0.5°) before 1st February and TL799, corresponding approximately to 0.25°, afterward. The model fields have been provided by the Norwegian Meteorological Institute, that uses the 3-hourly analysis as initial and boundary conditions for the Unified EMEP (*European Monitoring and Evaluation Programme)* model (http://www.emep.int/). The meteorological fields are ingested into the EMEP model with horizontal resolutions of 0.2° and 0.5° after a bilinear interpolation.

The LAM considered here is the Weather and Research Forecasting model (WRF), version ARW-3.2 (Skamarock et al., 2008). The model output for 2006 has been provided by INERIS, where the inner domain fields are used as input for CHIMERE chemistry-transport model (http://www.lmd.polytechnique.fr/chimere/). Twenty-eight vertical levels are used; the model is implemented in a two-way nesting technique, with two domains covering Europe with a horizontal resolution of 30 km (150 x 140 grid points) and 10 km (361 x 328 grid points), respectively. 365 daily simulations are performed in restart mode, using the 6-hourly ECMWF ERA-INTERIM reanalysis as large scale fields for analysis nudging (four-dimensional data assimilation). Nudging is active both on the external and the internal domain, for temperature, humidity and wind fields, in the PBL too. The parametrization schemes selected for the simulations are shown in Tab. 1.

Table 4. WRF model parametrization schemes on the two grids.

PHYSICS	**Parametrization schemes**
Microphysics	WSM 5-class
Longwave radiation	RRTMG
Shortwave radiation	RRTMG
Land-Surface	NOAH
PBL	YSU
Cumulus	KAIN-FRITSCH

RESULTS

In the present study, only surface wind speed is taken into consideration. As a first step, the geographic distribution of the value of four different statistical indices is discussed. Figure 1 shows the results for ECMWF analysis (0.2° horizontal resolution). Correlation coefficient is pretty good over all the domain, being above 0.8 in most of the stations. Few areas can be identified where the correlation is lower, mainly concentrated in the Alpine region, where the value of the coefficient is locally below 0.2. Bias is generally below 1 m s^{-1}, however some stations mainly in Finland, Eastern Europe, Scotland and near the Alps show wind speed overestimation larger than 1.5 or even 2 m s^{-1}. The target diagram represents the normalized (by the standard deviation of the observation) RMSE. A value of the target indicator smaller than unity can be considered as reasonable (Jolliff et al., 2009). Several stations do not verify such criteria, especially in the eastern side of the domain, Scandinavia and in the Alpine area. About RMSE, where a value lower than 2 m s^{-1} can be considered acceptable (Emery et al., 2001), a large majority of the stations satisfy this condition.

The results for WRF model (10 km horizontal resolution) are shown in Fig. 2. While correlation is almost identical to the large scale fields, a larger bias can be observed in several stations distributed over all the domain, with several stations reporting overestimations larger than 2 m s^{-1}. Similarly, the WRF model forecasts have less skill in terms of target, with just a few stations showing the target indicator smaller than unity; also, several stations do not satisfy any longer the criterion RMSE < 2 m s^{-1}.

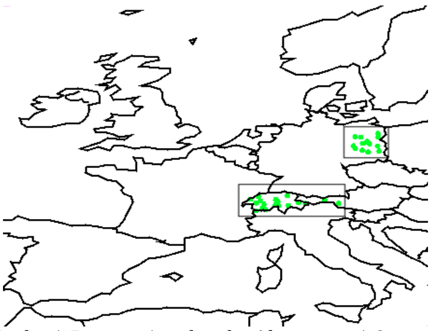

Figure 7. Stations near Berlin (15 stations) and in the Alpine area (19 stations) considered in Figures 4 and 5.

In order to discuss these results further, we analyse the statistical indices in some specific areas. First, we consider 15 stations near Berlin (see Fig. 3), which is one of the urban areas where the model performance is better. The results for the four different model implementations are shown. The Taylor diagram (Fig. 4a) shows that while the correlation coefficient is high and very similar for all the simulations and the stations in the area, the ECMWF standard deviation is much closer to the observations than WRF, which produces larger values compared to observations ($\sigma_M/\sigma_O > 1$). In the scatterplot (Fig. 4b), the ECMWF data are closer to the diagonal of the plot, while the WRF model data are shifted above. This indicates that the mean wind speed of the ECMWF data is closer to the observations, as the WRF model adds about 1 m s^{-1} to the ECMWF average wind speed in most of the stations, thus producing a larger bias. However, the WRF model data (especially those at 10 km) show a larger inter-variability among the stations. The two implementations of ECMWF data are almost identical; similarly, the WRF model runs do not differ significantly among each other, apart from three stations where the average wind speed for the WRF model at 10 km (but not for the implementation at 30 km) is slightly smaller than the ECMWF data. The soccer diagram is a synthetic way to analyse if both the criteria on bias (< 0.5 m s^{-1}) and RMSE (< 2 m s^{-1}) are satisfied simultaneously (the box delimited with dashed lines). It is apparent (Fig. 4c) that the ECMWF results fall mostly in such a box and are more evenly distributed around the no-bias vertical line, while the WRF model data show an apparent overestimation of the wind speed, as almost all of the points fall on the right side of the box. The data are mainly distributed along a straight line, suggesting that the RMSE is significantly larger in WRF than in ECMWF data mainly as a consequence of the larger bias. Finally, the target diagram (Fig. 4d) shows that for the ECMWF data nearly 100% of the stations fall inside the circle of radius = 1, while for the WRF model it occurs only for about 60% of the stations. Also, there is no significant improvement moving from the coarser to the finer model run, apart from a few stations.

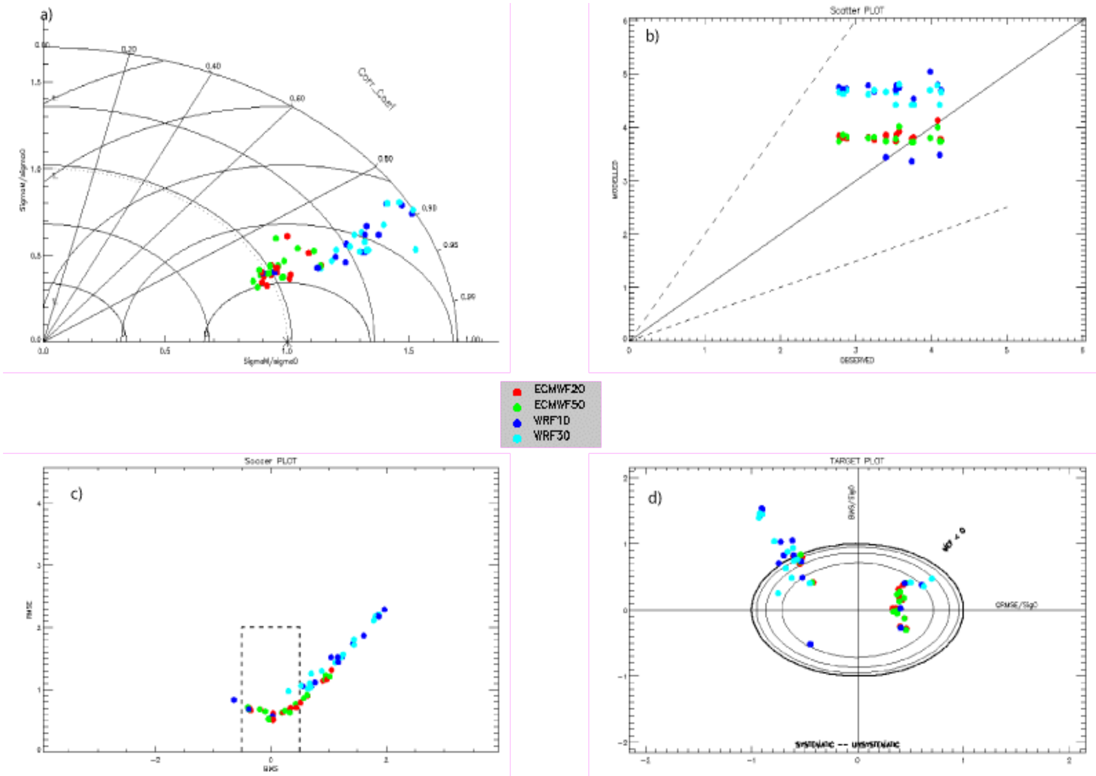

Figure 8. Statistical plots for WRF model simulations and ECMWF analysis in Berlin area (15 stations): a) Taylor diagram; b) Scatterplot; c) Soccer plot; d) Target Plot.

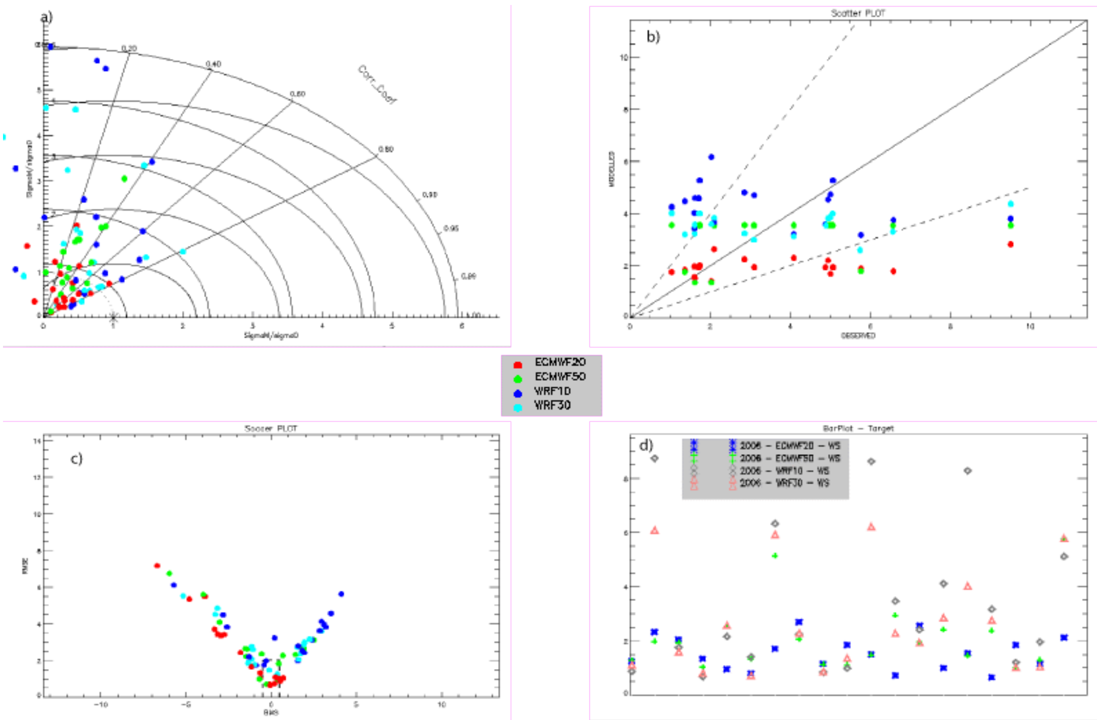

Figure 9. As Fig. 4, but for the 19 stations in the Alpine area whose altitude above 1000 m. In panel d), target=RMSE/σ_O is shown.

The skill of the models becomes worse when we consider other groups of stations, especially those near the orography. In these cases, the inappropriate representation of the mountains can negatively affect the results both for the large scale and the mesoscale models. Figure 5 shows the statistics of the 19 stations in the Alpine region whose height is above 1000 m. The Taylor diagram (Fig. 5a) shows that the correlation coefficient for this group of stations is much smaller than in Fig. 4a, in some stations the modelled wind speed is even anti-correlated with the observations. The WRF model overestimates significantly the standard deviation of the observations, and in some cases σ_M/σ_O is close to 6, the finer the resolution is the higher is the ratio. For the ECMWF data, σ_M is closer to σ_O although the points are more spread than in Fig. 4a. The scatterplot in Fig. 5b shows that the ECMWF analyses underestimate the wind speed when observed values are greater than 2 m s^{-1}, as a nearly constant value of 2 m s^{-1} is simulated in the stations of the area. The 0.5° horizontal resolution data are better than the finer resolution data: this is probably because the interpolation includes points farther from the stations where the modelled wind speeds are on average higher. The wind overestimation of the WRF model compared to ECMWF forecasts (that is larger for the 10 km resolution run) somehow corrects the negative bias of the ECMWF data for observed wind speed above 3 m s^{-1}. Figure 5c shows that only for a limited number of stations the models satisfy both conditions on bias and RMSE, and that the errors of both models are high for most of the stations. Differently from Fig. 4c, the WRF model predictions are more uniformly distributed around the no-bias vertical line, while ECMWF data show a clear underestimation. As a consequence, the value of the target=RMSE/σ_O is pretty high (Fig. 5d).

DISCUSSION

The purpose of the analysis performed in the present study is two-fold: on the one side, the DELTA-FAIRMODE tool is tested using comprehensive meteorological datasets including one-year complete simulations; on the other side, the outputs of a GCM (ECMWF-IFS) and a LAM (WRF model), used as input for air quality models, are analysed in order to provide a comparative statistical analysis. The main result of the present paper is that the skill of the GCM is generally better than that of the LAM in terms of different statistical indices. This conclusion suggests different considerations. First, a more comprehensive study should require the inclusion into DELTA of the ERA-INTERIM fields used to force WRF. In this way, we could determine how much of the worse statistical scores can be attributed to the large scale forcing. This important point is left for future work. Then, it is important to stress that these results refer to the specific implementation of the WRF model chosen here. This is just one among the hundreds of implementations that could be obtained by combining the different parametrisation schemes. A different implementation could significantly modify the value of some statistical indices. We leave for a future study the possibility to explore, at least partially, such sensitivity.

Also, statistical scores should be properly interpreted. Our results show the known tendency of WRF and other models to overestimate wind speed at the surface (Mass and Ovens, 2011). However, a worse skill does not mean that a LAM does not produce an additional value with respect to a GCM. Mass et al. (2002) indicate that, for precipitation, the traditional statistical indices should be taken in consideration with caution. For example, a slightly wrong timing and location of a heavy rain event, although predicted with the exact amount, is worse in statistical terms than the forecast of a GCM, which predicts a smaller rainfall in the correct, coarser grid box, but is definitely much more skillful for a forecaster. Similar considerations can be applied to the wind field: a LAM provides more details (the finer resolution the more detailed the flow patterns), which a GCM cannot produce, although the wrong timing and location can negatively affect the statistical indices. Finally, the present study is based on surface data and a more comprehensive analysis should include 3D fields too.

ACKNOWLEDGEMENTS

The authors thank EMEP for providing surface stations data.

REFERENCES

Emery, C., E. Tai and G. Yarwood, 2001: Enhanced Meteorological Modeling and Performance Evaluation for Two Texas Ozone Episodes. Report to the Texas Natural Resources Conservation Commission, prepared by ENVIRON, International Corp, Novato, CA.

Jolliff J. K., J. C. Kindle, I. Shulman, B. Penta, M. A. M. Friedrichs, R. Helber and R. A. Arnone, 2009: Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment. Journal of Marine Systems, **76**, 64-82.

Mass, C. F., D. Ovens, K. Westrick and B. A. Colle, 2002: Does Increasing Horizontal Resolution Produce More Skillful Forecasts?. *Bull. Amer. Meteor. Soc.*, **83**, 407–430.

Mass, C. F. and D. Ovens, 2011: Fixing WRF's high speed wind bias: a new subgrid scale drag parameterization and the role of detailed verification. *Preprints*, 91st *AMS Annual Meeting*, Seattle, WA, [available online at http://ams.confex.com/ams/91Annual/webprogram/Paper180011.html].

Seaman, N.L., *2000*: *Meteorological Modeling for Air-Quality Assessments*. *Atmospheric Environment*, **34**, 2231-2259.

Sistla, G., N. Zhou, W. Hao, J.-K. Ku, S.T. Rao, R. Bornstein, F. Freedman and P. Thunis, 1996: Effects of uncertainties in meteorological inputs on Urban Airshed Model predictions and ozone control strategies. *Atmospheric Environment*, **30**, 2011-2025.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. Duda, X.-Y. Huang, W. Wang and J. G. Powers, 2008: A Description of the Advanced Research WRF Version 3 NCAR Technical Note.