AN EVALUATION OF THE UoWM MM5-SMOKE-CMAQ MODELING SYSTEM FOR WESTERN MACEDONIA

A. Syrakos[1], A. Sfetsos[1,2], M. Politis[1], G. Efthimiou[1], J.G. Bartzis[1]

[1]Environmental Technology Laboratory, Department of Engineering and Management of Energy Resources, University of Western Macedonia, Greece
[2] Environmental Research Laboratory, INT-RP, NCSR Demokritos, Greece

Abstract: The University of Western Macedonia (UoWM) has recently set up an operational system for predicting air quality in the Florina–Ptolemais–Kozani basin in Western Macedonia, Greece. This is a mountainous basin where around 65% of the total energy production of Greece takes place in lignite power plants. The concentrations of PM_{10} is the main air quality concern in the region, and the high particulate matter levels observed are attributed to the activities related to power generation. The system set up by the UoWM makes use of software such as MM5 for weather prediction, SMOKE for emissions processing, and CMAQ for air quality prediction. This paper describes a preliminary effort to evaluate the system by comparing its predictions against ground station observations for a selection of days of the 2007 summer period.

Key words: *MM5, CMAQ, PM_{10}, power plants, model evaluation, weather category.*

1. INTRODUCTION

The greater lignite basin of Bitola–Florina–Ptolemais–Kozani is where most of the electric power of Greece and the FYROM is produced. This is due to the abundance of lignite reserves in the region. The Greek Public Power Corporation (GPPC) operates 5 lignite power plants with a total power exceeding 4 GW, while a power plant near Bitola is the major lignite centre for the FYROM (675 MW), which covers 70% of the country's energy needs.

The basin is located in the northwest of Greece, with a mean elevation of 650 m above mean sea level (MSL), surrounded by mountains whose peaks exceed 2000 m MSL – see Figure 1. The main axis of the basin is oriented from northwest to southeast. More than 150,000 people live within the region of the basin, with the main towns being Kozani and Ptolemaida.

PM_{10} concentrations are the main concern in the region. The main PM sources are considered to be the power plant stacks, and the open-pit lignite mines (mining, transportation of soil and coal, movement of vehicles on unpaved roads). Although PM concentrations exhibit a falling trend in recent years due to actions taken by the PPC, such as the use of new electrostatic filters for the stacks and techniques for the control of fugitive dust emissions from the mines, their levels are still high and under certain meteorological conditions may exceed local and international standards (Triantafyllou, 2003).

The Laboratory of Environmental Technology of the UoWM, based in Kozani, is developing an operational system to monitor and predict air quality in the region. The system is connected to a network of monitoring stations which are operated by the PPC, the locations of some of which are shown in Figure 1. In addition the UoWM has its own PM monitoring station at the north edge of Kozani. For the prediction part of the system, the PSU/NCAR model MM5 (Grell *et al*, 1994) is set up to provide a meteorological prediction to the UNC/USEPA model CMAQ (Byun and Ching, 1999) which makes a 72-hour air quality prediction. The pollutant sources are provided to CMAQ by the UNC SMOKE model which processes a detailed emissions inventory for the region, prepared by the Environmental Technology Laboratory of the UoWM and NCSR Demokritos, Athens (Vlachogiannis *et al*, 2007). This emissions inventory is under continual improvement and apart from the power generation related sources it contains sources such as the traffic network, central heating, industry, and biogenic sources. The emissions inventory, and the CMAQ prediction, involves many pollutants but the present paper deals only with PM_{10}, since it is of most concern. In an attempt to evaluate the system, a number of days of the April–September 2007 period were selected, each considered representative of a specific weather category, and the predictions produced by the system were compared against measurements of the monitoring stations network. A similar study will have to be performed for the winter period.

2. SYSTEM SET-UP

MM5 (version 3.7.4) is set up to use 4 nested domains. Their horizontal dimensions are respectively, from coarser to finer, 39×39, 36×36, 54×54 and 72x72 cells, and the respective spatial resolutions are 54 km, 18 km, 6 km and 2 km. The finest 2×2 km domain covers the area shown in Figure 1. All domains consist of 30 vertical layers.

As for the physical parameterization, the following schemes are used:

- MRF PBL scheme.
- Grell cumulus parameterization on all domains except the finest, where no cumulus parameterization is used.
- "Simple ice" for the explicit moisture scheme.
- RRTM radiation scheme.

The boundary conditions are obtained from the output of the GFS model stored in the daily global repository of the National Center for Environmental Prediction (NCEP), USA.

For the purpose of this study, the boundary conditions and the initial conditions for the selected past dates were obtained from the NCEP "FNL Global Tropospheric Analyses" product available from http://dss.ucar.edu/datasets/ds083.2/. This product is in the form of GRIB files containing data covering the entire globe at 1×1 degree resolution (approx. 100 km), at 6-hour intervals, obtained by a combination of GFS model predictions and observations. However, in the MM5 runs data assimilation was *not* used to nudge the simulations towards observational data. This is expected to cause a slight degradation of the quality of the meteorological predictions.

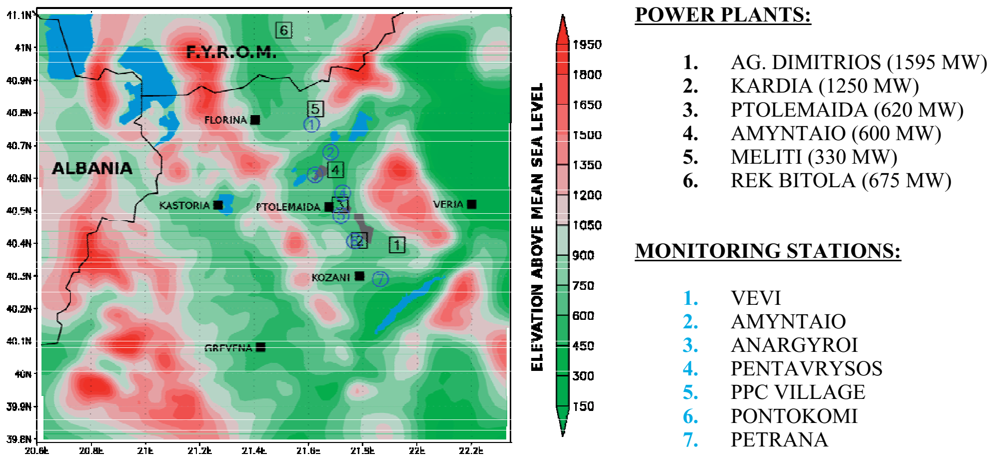

Figure 1. The working domain, and main towns. Numbered squares indicate the power plants, and numbered circles the PPC monitoring stations. The approximate area covered by the lignite mines is shown in grey.

On the other hand, CMAQ is set up to use only one domain, the finest one of the MM5 set-up, which covers the region shown in Figure 1. The grid spacing is the same as for MM5 (2×2 km) and the same vertical structure of 30 layers is retained. The power plant stacks emissions are spread among several layers, according to the plume rise prediction performed by SMOKE which uses the meteorological prediction provided by MM5. All other emissions are assumed to occur within the grid cells which are adjacent to the ground (i.e. within layer 1), which is about 35 m tall. Preliminary experiments showed no marked improvement of air quality predictions with increased resolution in the vertical direction near the ground.

The CMAQ setup includes using CB-IV as the chemical mechanism, and the "aero3" aerosol module.

Since boundary conditions are not available, concentrations at the lower levels of the domain boundaries were approximated using the measurements at monitoring stations near these boundaries, during days when the wind direction is from the boundary towards the interior of the domain.

For each of the selected dates, the simulation was performed starting one day in advance.

3. DAYS SELECTED

Each day selected corresponds to a different weather category. In previous studies concerning the region, such as Triantafyllou (2001) or Triantafyllou *et al* (2002), a number of weather types where identified when pollution episodes may occur. However, this approach is subjective and depends on the judgement of the meteorologist who must study the forecast and make a judgement as to which weather category it belongs to. For an operational system though it is more convenient that the weather types and the classification of forecasts into one of these types are done automatically using an algorithm.

For this study a number of different weather types have been identified using a methodology based on the subtractive clustering algorithm (Chiu, 1994) coupled with the "compactness and separation criterion" (Kim *et al*, 2001) for identifying the optimal number of clusters (Sfetsos *et al*, 2005). The algorithm uses GFS model forecasts and classifies a weather forecast for a particular day using the following variables: u and v wind components at 10 m AGL, u and v wind components at 500 hPa, temperature at 2 m AGL, relative humidity at 2 m AGL, and the mixing layer height (MLH). These variables are considered at three periods of the day: 00:00, 12:00 and 24:00, and they are considered only at the two vertical columns of GFS cells which cover the Ptolemaida–Kozani basin. To derive the weather categories, the "FNL Global Tropospheric Analyses" files from NCEP for the entire 2006 and 2007 were processed. Separate categories were derived for the summer (April–September) and winter (October–March) periods. In this paper we will only deal with the summer period.

Using this procedure, 11 weather categories were obtained for the summer period. Figure 2 shows the percentage of the number of days of each category for the summer periods of 2006 and 2007. Note that although category 8 is the largest category, it only appeared in 2006 and not in 2007 for which we have PM measurements data. Therefore this category is not treated in the following. The days shown in Table 1 each belong to a separate weather category. The mean wind speed and wind direction is shown for each day in the same table.

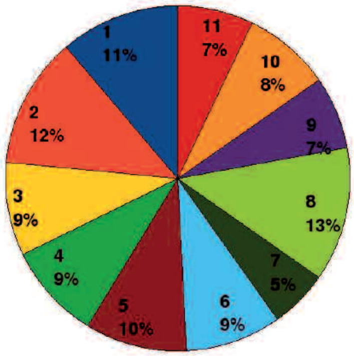

Figure 2. The percentage of the number of days of each summer weather category for 2006 and 2007.

4. METEOROLOGICAL PREDICTIONS

Table 1 shows a comparison between observation and model predictions using the mean bias (= $m_o - m_p$, see table caption) and root mean square error (RMSE) as statistical measures. The meteorological observations used in Table 1 come from the stations shown in Figure 1, which measure meteorological quantities on an hourly basis in addition to PM concentrations. Also the data from the UoWM meteorological station in Kozani were used, as well as data from two other stations (not shown in Figure 1), one in Florina at the northwest of the basin, and one at Kato Komi at the southeast.

For wind direction, the calculation of the mean bias and RMSE are not straightforward due to the problems associated with overlap from 0 and 360^o. The mean wind direction of the observations m_o is calculated by finding the means of the u and v wind components and m_o is the direction from which the resulting mean wind velocity vector blows. The mean wind direction of the predictions m_p is calculated likewise, and the mean bias is simply m_o–m_p. For RMSE, the difference between the hourly values of the observed and predicted wind directions is calculated so that it always lies between -180 and 180 degrees, and RMSE is the root mean square of these differences.

The values of the statistical measures shown in Table 1 are not much different, although slightly worse, than those summarized by Hanna & Yang (2001) for a number of studies. An improvement of these results can be done with the use of data assimilation, as was done in the test cases presented by Hanna and Yang (2001). It can be seen in Table 1 that in some cases the mean bias of the wind direction is relatively large, and this could cause the locations of the actual and predicted peak PM concentrations to differ significantly.

Table 1. Statistical measures for the meteorological predictions of each day. m_o = mean of observations, m_p = mean of predictions.

Category	**Date**	**Wind Speed (ms^{-1})**				**Wind Direction (deg)**			
		m_o	m_p	**mean bias**	**RMSE**	m_o	m_p	**mean bias**	**RMSE**
1	**28-07-2007**	*2.06*	*1.85*	*0.21*	*1.35*	*85.4*	*91.8*	*-6.4*	*64.5*
2	**04-05-2007**	*1.88*	*2.39*	*-0.52*	*1.68*	*123.7*	*103.3*	*20.4*	*76.4*
3	**29-05-2007**	*3.29*	*4.79*	*-1.51*	*3.04*	*233.1*	*253.9*	*-20.8*	*71.9*
4	**15-09-2007**	*3.13*	*3.24*	*-0.11*	*2.15*	*335.2*	*333.9*	*1.3*	*70.8*
5	**22-05-2007**	*2.71*	*3.53*	*-0.82*	*2.44*	*334.2*	*345.5*	*-11.3*	*80.6*
6	**21-09-2007**	*3.83*	*2.66*	*1.18*	*2.49*	*337.7*	*337.5*	*0.2*	*58.4*
7	**17-05-2007**	*2.67*	*2.78*	*-0.11*	*1.73*	*240.8*	*272.6*	*-31.8*	*72.4*
9	**13-05-2007**	*2.17*	*2.20*	*-0.03*	*1.75*	*73.1*	*32.7*	*40.4*	*84.9*
10	**26-06-2007**	*2.98*	*3.30*	*-0.32*	*1.95*	*261.7*	*266.6*	*-4.9*	*61.5*
11	**12-04-2007**	*2.07*	*3.12*	*-1.05*	*1.71*	*321.3*	*342.2*	*-20.9*	*89.23*

5. PM_{10} CONCENTRATIONS

In this section some results will be shown concerning the prediction of PM_{10}. Statistical metrics will not be shown because the system is not yet capable of accurately predicting the concentrations, but further tuning is required. Some results will be shown concerning a subset of the selected dates. Unfortunately, for the hottest days of 26/06 and 28/07 the predictions greatly underestimate the measured PM concentrations. It turned out that the large PM concentrations were due to forest fires, so the choice of these dates was unfortunate.

Figure 3 shows mean PM_{10} concentration predictions for four of the selected dates: 12/04 when weak north winds are recorded in the basin, 04/05 when weak SW winds are recorded, and 15/09 and 21/09 when strong NE winds are observed in the basin in both cases. In all cases the wind direction seems to follow the main axis of the basin (NE–SW). Although the meteorology shown in Table 1 seems very similar for both days of September, the plots in Figure

3 are very different because the data of Table 1 apply only within the basin, where the flow direction follows the topography. One immediately notices from the Figures that the greatest PM10 concentrations appear in the vicinity of the lignite mines. In fact, preliminary numerical experiments where the stack emissions were varied showed that this had little effect on the PM_{10} concentrations. Therefore CMAQ predicts that the emissions from the lignite mines are more important than those of the power plants, despite the fact that the latter are estimated to be larger in magnitude.

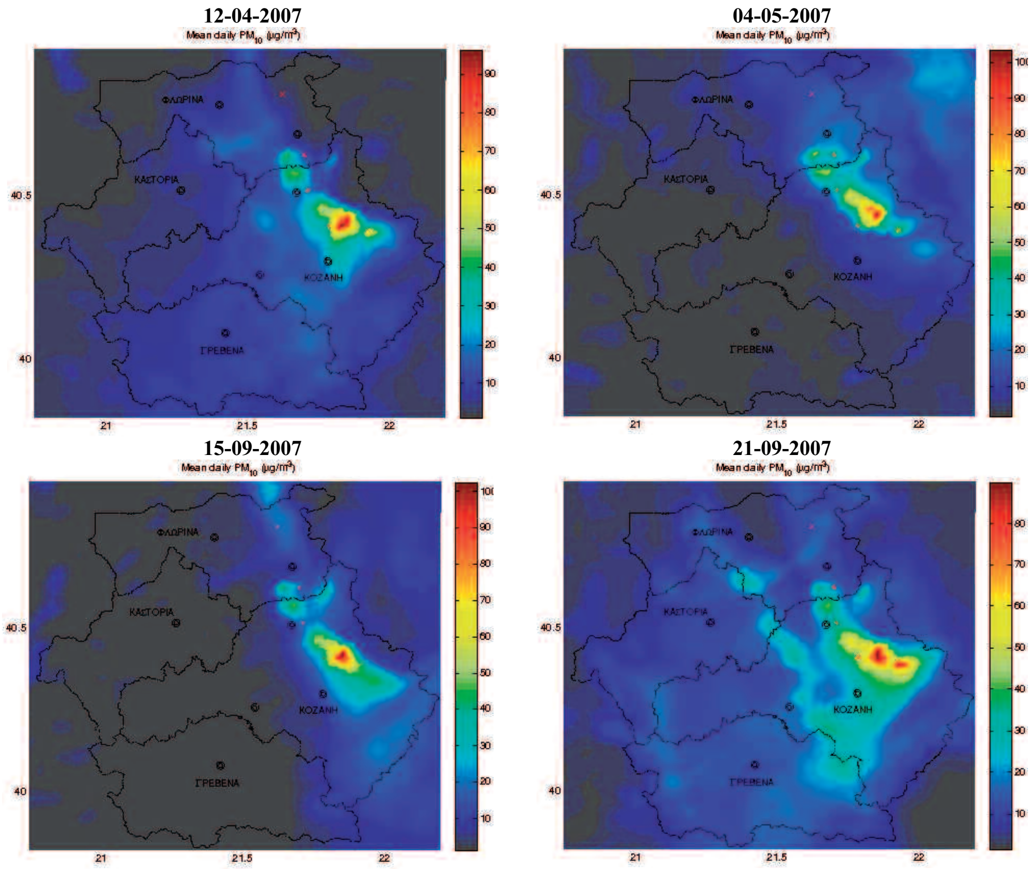

Figure 3. Predicted mean daily PM_{10} concentrations at ground level. The locations of the power plants are marked with a red x. Also shown are the main towns and the prefecture boundaries for West Macedonia.

Figure 4 shows the daily average concentrations at each station (missing data means that the station was not functioning that day). Although the concentrations predicted by CMAQ are in most cases not very far from the observed concentrations, it seems that it fails to accurately follow the concentration distribution along the basin (the stations are placed approximately from north to south along the x-axis of Figure 4). However one must keep in mind that some of the stations are very close to each other, a few grid cells apart, yet the observed concentrations vary significantly. This may suggest that there are components of the concentrations which vary at a spatial scale too small for CMAQ to capture, especially near the mines. CMAQ seems to underestimate the concentrations in general, so the emissions inventory may need some modification. In particular, for both the power plant stacks and the lignite mines a static inventory based on annual emissions was used, whereas in reality the emissions vary according to the power demands which are heavier during the summer. Also, the only fugitive dust source that is included in the inventory is the lignite mines. Although this is the most important source, there are other sources the effect of which cannot be approximated by simply modifying the boundary and initial conditions. This can be seen in Figure 5 which shows cases where CMAQ successfully predicts the temporal variation of the PM_{10} concentration at the UoWM station in the north of the town of Kozani but the minima of the predicted distributions are very low. One can also see this clearly in Figure 3 which shows that the model predicts large areas with near-zero concentrations. Figure 5 and others for other days (not shown) are encouraging in the sense that CMAQ compares relatively well to the measurements of the UoWM and Petrana stations, which are the ones closest to Kozani, the largest town in the region (about 60,000 inhabitants) and the location of most interest. One may also notice in Figure 4 that the station at Vevi in the north records high concentrations in general, something which is verified also during the days not shown in Figure 4. However, CMAQ seems to fail to predict this behaviour. It is quite likely that the emissions inventory needs refinement in that area.

6. CONCLUSIONS

This paper provided an overview of a preliminary evaluation of an air quality prediction system set up at the UoWM to monitor the air quality at the heavily industrialised area of northern Macedonia in Greece. The results are encouraging although further work is needed to refine the emissions inventory, and possibly improve the quality of meteorological predictions. The evaluation was based on categorising the weather into different types using an algorithm. Further work is also necessary to establish any connections between this weather categorization and pollution patterns.

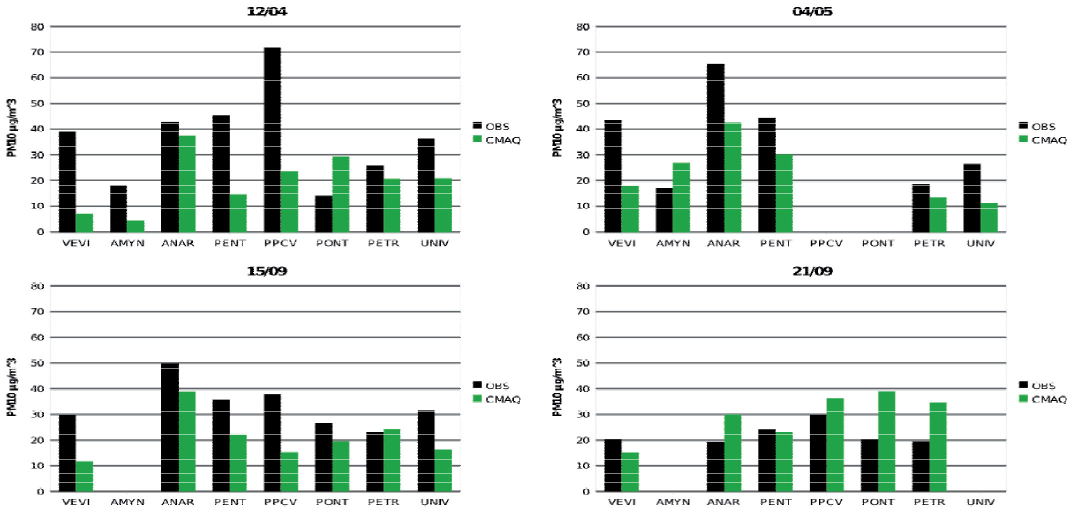

Figure 4. Observed and predicted daily average PM_{10} concentrations at each station.

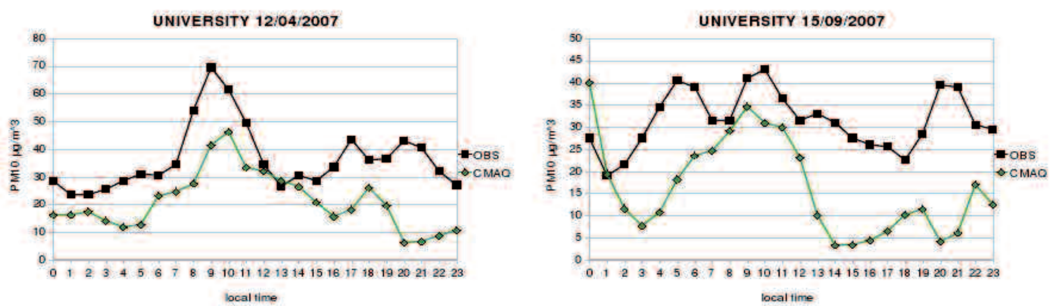

Figure 5. Observed and predicted PM10 concentrations at the UoWM station at the north edge of the town of Kozani.

REFERENCES

Byun, D.W. and J. K.S. Ching, 1999: Science algorithms of the EPA models3 community multiscale air quality (CMAQ) modeling system. *EPA/600/R-99/030.*

Chiu, S. L., 1994: Fuzzy model identification based on cluster estimation. *J. of Intelligent Fuzzy Systems*, **2**, 267-278.

Grell, G. A., J. Dudhia, and D. R. Stauffer, 1994: A Description of the Fifth-Generation Penn State/NCAR Mesoscale Model (MM5). *NCAR Technical Note, NCAR/TN-* ***398****+STR.*

Hanna, S. R. and R. Yang, 2001: Evaluations of mesoscale models simulations of near-surface winds, temperature gradients, and mixing depths. *J. Applied Meteorology*, **40**, 1095-1104.

Kim, D. J., Y. W. Park, D. J. Park, 2001: A novel validity index for determination of the optimal number of clusters. *IEICE Transactions on Information Systems*, E84-D, 281-285.

Sfetsos, A., D. Vlachogiannis, N. Gounaris, A. K. Stubos, 2005: On the identification of representative samples from large data sets, with application to synoptic climatology. *Theor. and Appl. Climatology*, **82**, 177-182.

Triantafyllou, A.G., 2001: PM_{10} pollution episodes as a function of synoptic climatology in a mountainous industrial area. *Env. Poll.*, **112**, 491-500.

Triantafyllou, A.G., E. S. Kiros, G. Evagelopoulos, 2002: Respirable particulate matter at an urban and nearby industrial location: Concentrations and variability and synoptic weather conditions during high pollution episodes. *J. Air & Waste Manage. Assoc.*, **52**, 174-185.

Triantafyllou, A.G, 2003: Levels and trend of suspended particles around large lignite power stations. *Env. Monit. and Assess.*, **89**, 15-34.

Vlachogiannis, M, A. Sfetsos, N. Gounaris, J.G Bartzis, 2007: Compilation of Air emission inventories for Western Macedonia, Greece. *Urban Air Quality 2007*, Cyprus.