Short - Term PM10 Concentration Forecast Modelling in the MARQUIS-Service

Achim Lohmeyer¹, Ingo Duering², Tilo Hoffmann², Daniel Nicklaß³, Christine Soergel¹

Abstract

The information system MARQUIS will provide to the public short-term air quality forecasts for selected areas in Europe. Basis for the forecasts are the current data gained by the air quality monitoring stations in the region and forecasts of the meteorological data. For the service concerning PM10, there is a need to do the forecast in the morning, forecasting the daily mean of the PM10 concentration for the current and the following days.

The paper presents the examined methods for this short term PM10 forecast modelling at two open country monitoring stations in Germany on the basis of the results of EURAD (classical emission and dispersion modelling based on the area of Europe), Machine Learning, Multiple Linear Regression and Neural Networking. Advantages and disadvantages of the methods are discussed and some of the validations and of the future plans are presented.

Additionally, the paper presents the status of the ProFet-System based operational modelling for three vehicle traffic dominated monitoring stations, using a Multiple Linear Regression model. The results of ProFet are foreseen for the real time information of the public and the triggering of measures for ad hoc reduction of vehicle traffic induced PM10 concentrations. Some of the validations and of the future plans are presented for ProFet as well.

1. Introduction

The public interest to gain environmental information is growing constantly and the possibilities to access and to provide such information is increasing. One of the possibilities for a service in this field is covered by the project MARQUIS (Multimodal Air Quality Information System for General Public), supported in the framework of the e-Content Program of the EC. Among the services offered by MARQUIS is the provision of forecasts of particle (PM10) concentrations for the current and the following days. To realize the forecasts, adequate short-term forecast models had to be selected and provided. The paper describes this procedure and its success for two open country locations.

Another application for this kind of forecasting models resulted in the development of the model ProFet, aimed to operationally forecast the PM10 concentrations at monitoring stations situated adjacent to three roads with heavy traffic. The forecasts are the basis for the information of the public and serve for triggering of actions to influence vehicle traffic with the aim to avoid exceedance of the daily mean limit value for PM10 concentrations.

¹ Ingenieurbuero Lohmeyer GmbH & Co. KG, An der Rossweid 3, D-76229 Karlsruhe
² Ingenieurbuero Lohmeyer GmbH & Co. KG, Mohrenstraße 14, D-01445 Radebeul, email: achim.Lohmeyer@lohmeyer.de, tel. nr.: 0721-625100, internet: http://www.lohmeyer.de
³ Institut für Rationelle Energiewirtschaft und Rationelle Energieanwendung, Universität Stuttgart
2. Forecast models for MARQUIS

2.1 Objective

The relevant air quality limit value for PM10 concentration is the daily mean. In the framework of MARQUIS, therefore, in the morning of a given day, a forecast of the daily mean of the PM10 concentration was needed for this and the following days. This forecast had to be executable on the basis of operationally available input data and preferably on the basis of existing models.

For the user of the MARQUIS-service, the following information was provided:

a) the latest monitoring data and
b) the forecast, however, not the absolute concentrations but the trend, i.e. whether the concentration is expected rather to increase, to stay as it is or to decrease. The provision of a trend instead of an absolute value seemed to be easier because of the expected uncertainties of the forecasts.

2.2 Available input data

As results of already existing forecasts and as input data for forecasting models, the following data were available for the application within the German State of Baden-Württemberg:

- The forecast of EURAD (see Section 2.3.1) for 12 h, 36 h und 60 h in advance within a 25 km x 25 km grid
- The results of the 36-station based concentration and meteorology monitoring network of the State of Baden-Württemberg
- The forecasts of the meteorological conditions in Baden-Württemberg, provided on a 22 km x 22 km grid by the model HIRLAM (Niska et al. 2005) run by the Finnish Meteorological Institute, a partner within the MARQUIS consortium.

2.3 Results

To provide the forecasts, four different procedures were tested. These tests are exemplarily described in the following sections for the monitoring station Freudenstadt and (partially) Erpfingen. These stations are located at a sufficient distance from agglomerations and therefore expected to be representative for larger areas. The results for stations within cities close the major vehicle roads are presented in Section 3.

2.3.1 EURAD

EURAD executes a classical Europe-covering emission and dispersion modelling and provides the results on maps in the internet (www.eurad.uni-koeln.de/). See also (Jakobs et al. 2005). Additionally, EURAD kindly provides since April 2005 the numerical data of Baden-Württemberg for MARQUIS.

Figure 2.3.1.1 shows the comparison of the forecast for a given day with the monitored values for the monitoring stations Freudenstadt (FDS) and Erpfingen (ERP). As mentioned above, the stations are located far from polluters and the concentrations at the two stations do not differ significantly. Figure 2.3.1.1 contains the mean value at the two stations to nevertheless damp local peculiarities, not contained in the EURAD model and to also the problem of comparing mean values calculated for a 25 km x 25 km grid to

---

4 As far as Germany was concerned, MARQUIS used the data from the monitoring network of Baden-Württemberg
monitored data at a fixed position. Figure 2.3.1.1 shows partly good agreement (April to mid June), partly underestimations (end of June, July and beginning of September).

Figure 2.3.1.1: Comparison of daily means of EURAD-PM10-concentration-forecast for a given day with the arithmetic mean of the concentrations monitored at the two open country monitoring stations Erpfingen and Freudenstadt for April to October 2005

2.3.2 K-Nearest-Neighbour procedure

M. Giereth (Lohmeyer et al., 2007) also examined a classification scheme on the basis of the k-Nearest-Neighbour (kNN)-Algorithm (Mitchell, 1997; Russel et al., 2003). To do so, a procedure, originally developed for ozone forecasts, was adapted for PM10 forecasting and evaluated. The kNN-procedure is based on the comparison of patterns for monitored time series for the parameter to be forecasted and the time series of the relevant input parameters. The series are inspected for constellations that are similar to the given constellation of all parameters and the forecast of the input parameters. Basis were the monitored data for the past 4 years for PM10 and the meteorological data wind speed and wind direction, temperature, global radiation and precipitation. Given the day \( t \), a forecast was provided for the following day \( t+1 \). Figure 2.3.2.1 (top) displays the forecasted PM10 daily mean in comparison to the monitored data at station Erpfingen.

The comparison in Figure 2.3.2.1 (top) looks good at the first glance. But evaluating the correctness of the trend to be forecasted – as foreseen for use within MARQUIS –, the obvious time lag of the forecasted time series leads to some incorrect predictions. An overview of the distribution of correct or incorrect predictions of the trends is displayed in Figure 2.3.2.1 (bottom). The values +1 (-1) indicate a correct (incorrect) prediction of the trend. In case the PM10 concentration stays within an interval \( \Delta \) (here 5.0 \( \mu g/m^3 \)), the result is “concentration stays as it is”. In case the increase (decrease) is larger than \( \Delta \), the result is “concentration will increase (decrease)”.

In order to improve the forecast, a further optimization of the comparison of the patterns is needed. Possible approaches are the use of additional input parameters as well as the evaluation of other machine learning procedures such as Naive Bayes or Support Vector Machines.
2.3.3 Multiple linear regression

For MARQUIS furthermore the application of a forecasting model based on multiple linear regression was evaluated. It was helpful to be able to do this on the basis of the experience gained by the FCT of the University of Lisbon, a partner within the MARQUIS consortium (Neto et al., 2005).

The development of the model was based on the monitoring data for the years 2001-2004; the data for 2005 were used for validation purposes. As input parameters were available: wind speed and direction, precipitation, temperature (max, min, mean), global radiation, atmospheric pressure, and the PM10 con-
centrations (daily means) as well as the PM10 concentration (hourly mean) for the given day at 6 o’clock a.m.. At 7 a.m., a forecast was provided for the daily mean of the given day and the following day.

A linear approach to the modelling function was used. For the determination of the constants, the statistics package SYSTAT version 11 of the SYSTAT Software Inc. (www.Systat.com) was used. Figure 2.3.3.1 (top) provides the comparison between the forecasted and monitored concentrations; the bottom part displays the performance concerning the forecast of the trend. At the given day, the trend for the next day could correctly be forecasted for 71 % of the cases for this period of time.

Figure 2.3.3.1: Comparison of forecasted (multiple linear regression) and monitored concentrations at monitoring station Freudenstadt (top) and performance of the trend forecast for January to December 2005. \( \Delta \) is here 0.5 \( \mu g/m^3 \).

2.3.4 Neural Network

In the framework of MARQUIS, also first trials with neural networking have been carried out by Nicklaß (Lothmeyer et al. 2007). The available input parameters were the same as for the multiple linear regression modelling, cf. Section 2.3.3. The training and the validation data set was also the same. The type of neural
network used in the experiment was the multilayer-perceptron network, which is one of the most popular types of neural networks.

The comparison of the forecasted and monitored PM10 concentrations is displayed in Figure 2.3.4.1 (top). The correlation coefficient is 0.81, 76% of the trends (as determined for the multiple linear regression procedure) could be forecasted correctly (Figure 2.3.4.1, bottom). Thus the result of the trial is encouraging.

2.3.5 Comparison of the models presented and their performance

At the present stage of the development of the models, within MARQUIS, i.e. for the special needs of MARQUIS, the multiple linear regression and the neural network seemed to be the most appropriate methods. But it is to be noted that EURAD and the k-nearest-neighbour procedure contain certain possibilities for improvement (Jakobs, 2005).

Figure 2.3.4.1: Comparison of forecasted (multilayer-perceptron neural network) and monitored concentrations at monitoring station Freudenstadt (top) and performance of forecast of trend for January to December 2005. ∆ is here 0.5 µg/m³.
When comparing the multiple linear regression method with neural networks, additionally to the above mentioned facts, some further differences might also be taken into account:

a) Behaviour during high concentration episodes
   The correct performance of the models is especially important during periods with high concentrations. Figure 2.3.5.1 focuses on the period with the highest concentrations in Freudenstadt in the year 2005. No significant difference can be seen at the first glance for this period for days with daily mean concentrations above 25 µg/m³: measurement 8 days; multiple linear regression 7 days; MLP 9 days.

b) Ability to provide insight into reasons for trend
   The multiple linear regression model is able to provide a physical insight into the reasons for an increase or decrease of the forecasted concentrations and the influence of single parameters, whereas the neural network model is not able to provide this.

c) Performance
   Concerning the performance of the models presented, except EURAD, it needs to be noted, that the development of the models as well as their evaluation are currently based on monitored time series for the meteorological data and not on meteorological forecasts. Therefore, the performance of the models will be different (most probably worse) than displayed so far. This problem currently exists, because no meteorological forecast time series were available in the development phase. Now, they are available and the effect is under examination.

Figure 2.3.5.1: Comparison of forecasted (multiple linear regression model and multilayer-perceptron neural network model) and monitored concentrations during period of highest concentrations in the year 2005 for the open country monitoring station FDS

3. Forecast model ProFet for vehicle traffic dominated monitoring stations
   The experience gained within MARQUIS for the open country stations was used to set up a forecasting system for the vehicle traffic dominated monitoring station Halle, Magdeburg and Wittenberg of the Luf\-\überwachungssystem Sachsen-Anhalt (LUESA). The aim of the system is to provide PM10 forecasts for
public real time information and for triggering of measures to reduce vehicle traffic induced PM10 concentrations at times where limit value exceedance for the daily mean concentration might occur.

The multiple linear regression method was used. The model was subdivided into 4 cases: winter working day, summer working day, winter weekend and summer weekend. For the station Halle, a 4-year time series (2002 to 2005) was available for development. In order to include long range transport on the most reasonable physical basis, 2 versions were developed: one including EURAD forecasts for the regional background, another working without EURAD. Presently, comparisons are executed to decide whether there is an increase in performance including EURAD and whether this increase is worth the additional cost for the provision and use of EURAD (Hoffmann, 2007).

Figure 3.1 displays the forecasts and the measurements for the second half of the year 2006. Concentrating, for example, on the period between August and October 2006 it can be observed that 13 exceedance episodes of the limit value for the daily mean (50 µg/m³) were monitored. Including the EURAD forecast, ProFet forecasted correctly 9 episodes; 7 were forecasted but did not occur; 4 occurred but were not forecasted. Thus, ProFet forecasted 9+7 exceedance episodes. 9 of them occurred; this is a hit rate of 56 %.

The results discussed are based on a version of ProFet, in which the model development was done on a basis of monitored meteorological data, but where the validation was done using data forecasted by the German Weather Service (DWD). This is one of the shortcomings, which is presently eliminated in ProFet (Hoffmann, 2007).

Figure 3.1: Comparison of ProFet forecasts for the actual day, done at 8.30 o’clock a.m. to values at station Halle-traffic

4. Conclusions

The work done for MARQUIS and ProFet is not yet finished; not all easily realizable improvements are carried out. But it can be seen that with a reasonable effort encouraging results can be obtained in this field of short term air pollution concentration forecasting.
Bibliography


