Simulation of virtual PM10 monitoring network for urban area (Case study: Berlin)

Hamid Taheri Shahraiyni[1,2], Sahar Sodoudi[1], Ulrich Cubasch[1], Andreas Kerschbaumer[3]

1-Institut für Meteorologie, Freie Universität Berlin, Carl-Heinrich-Becker-Weg 6-10, 12165 Berlin, Germany.
2- Faculty of Civil Engineering, Shahrood University, Shahrood, Iran.
3- Senate Department for Urban Development, Berlin, Germany

Abstract

PM10 is the particulate matters with less than 10 µm in diameter. Sometimes the PM10 concentration in outdoor air in Berlin exceeds from EU-Limit value and this pollutant offend the individual health. The major effects of PM10 on human health are: problem in respiratory systems, heart decease, cancer and premature death.

Modeling of spatial distribution of hourly PM10 concentration is the main step toward the PM10 control and management. It can specify the regions with high PM10 concentration in the city. Therefore it is possible to inform people about the air pollution condition in different parts of city in hourly time scale by internet or on mobile phone. The accuracy of spatial distribution modeling depends mainly on the number of air pollution monitoring stations but many cities (e.g. Berlin) have no dense monitoring network with enough number of stations for high accuracy modeling, because installation, operation and maintenance of a dense monitoring network is very expensive.

In this study, a new method is proposed for the simulation of virtual PM10 monitoring network in Berlin to overcome this problem.

Now, there are 13 monitoring stations in Berlin that measure hourly PM10 concentration and hereinafter these stations are named 'real stations'. In some years ago, hourly PM10 has been measured in some other points of Berlin concurrent with real stations. These points are considered as virtual station positions and hourly PM10 is simulated for these

virtual stations. The concurrent hourly PM10 data for each virtual station and the real stations are gathered. These PM10 data are time dependent and for the generation of time independent variables, the ratios of different pairs of real stations (e.g. real station1/real station2, real station2/real station3,...) were calculated and these ratios were utilized as input variables to a knowlegde-based modeling system. In addition, the ratio of PM10 concentration in the virtual station and one of the real stations (virtual_ratio) was calculated and utilized as the output variable of modeling system. The generated input-output database was divided randomly to two sub-databases for train and test of modeling system and it was trained and tested using the train and test sub-databases, respectively.

This knowlegde-based modeling system is a fuzzy modeling system and it is composed of three different parts. The first part of modeling system is a new feature selection method, that we have developed. This feature selection not only can determine the important variables for simulation but also it can determine the optimum number of fuzzy membership functions for the important variables for the fuzzy simulation. Therefore, when the ratios of hourly PM10 concentrations are introduced to this feature selection method, it determines the important ratios for the simulation of virtual station and then determines the optimum number of fuzzy membership functions for the important ratios. The output of this feature selection method is imported to the second part of modeling system, which is a single-output Sugeno-type fuzzy inference system. This fuzzy system similar to other fuzzy systems generates a rule-based for modeling and in the Sugeno-type fuzzy inference system the output of each rule is a linear combination of input variables and the final output is the weighted average of the all of the rules. The first part of modeling system only determines the number of membership functions for each important ratio, the second part adjust the shape of membership functions and provide the other initial values for the third part of modeling system and the outputs of second part are utilized as inputs to the third part of modeling system. The third part is a ANFIS (Adaptive Network-based Fuzzy

Inference System) . ANFIS is a well known modeling system that is a fuzzy inference system, implemented in the framework of an adaptive networks and it utilizes a hybrid learning procedure. The output of the ANFIS is the virtual_ratio.

The virtual_ratio is simulated by training and testing of this knowledge-based modeling system. Then, PM10 concentration in the virtual station can be estimated using the developed modeling system. Consider that the real stations in Berlin measure hourly PM10 data today. The measured hourly PM10 data are converted to the ratios of PM10 in the real stations and introduced to the developed modeling system, then the ratio of hourly PM10 in virtual station and one of the real station is estimated automatically by developed modeling system. The estimated virtual_ratio is multiplied to the PM10 concentration in the real station and consequently the PM10 concentration in the virtual station is calculated.

This simulation procedure was performed for each virtual station individually. Finally, many virtual stations were simulated and a dense hourly PM10 monitoring stations was created for Berlin by the combination of real and virtually simulated stations. The results of simulation of virtual stations using this new modeling system shows very high accuracy for the estimation of hourly PM10 concentration in the virtual stations. In addition, knowledge-based modeling methods (e.g. ANFIS, Neural networks) often need to appropriate initial input variables parameters and values of parameters for the appropriate simulations, hence simulation procedure is iterated by different initial values and different combination of input variables to achieve satisfactory results. Therefore, the simulation of many virtual stations with knowledge-based modeling techniques such as ANFIS is very time consuming. But the developed modeling system in this study is able to automatically perform the feature selection and initial value determination. Consequently simulation of many virtual stations using this modeling system is very fast, easy and straightforward. Each virtual station is equal to a free of charge real station, hence using the idea of this study, it is possible to develop high density air pollution monitoring network for the urban areas with minimum cost. In addition, it is possible to calculate the hourly spatial variations of pollutants with high accuracy using the developed dense monitoring network in the urban area.