Journal of Environmental Sciences Vol. 7, No. 1, pp. 66－73, 1995 ISSN 1001－0742

Regional air quality prediction models using the knowledge - based system approach

Du Min, Bai Naibin, Yao Guian

Research Center for Eco—Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China

Yuan Jiwen, Shen Baoling

Tianjin Environmental Monitoring Center, Tianjin 300191, China

Abstract—In order to predict of regional air quality, a knowledge - based system approach is presented. It illustrated by Gaussian puff model, the Gaussian plume model, the box model, the K - model and other special models. Using an approach which requires rules to choose the proper model. Each model in those cases where it is applicable were utilized. This system has been implemented in an expert system for regional air quality assessment to perform automated model selection and analysis. The results of a preliminary test using the air chemistry data from Tianjin Environmental Monitoring Station are also presented in this paper. Comparison with observed data was found to be favorable.

Keywords: expert system; air quality assessment; knowledge - based systems.

1 Introduction

While doing the qualitative prediction of the regional air quality, a model on air quality prediction must be used. However, only one model cannot be used to simulate the vary different conditions, so we have to utilize each model in those cases where it is most applicable. Therefore, it is important to select a suitable model from the existed model according to different conditions. At present, the model choice may only depend on the experience of the domain expert. Lam has used a knowledge - based system approach to select the watershed acidification models successfully (Lam, 1989). This paper presents an attempt to choose the most applicable model in many models of air quality prediction in regions by using the expert system approach(Hushon, 1987).

An implementation on a prototype expert system(EAS) has been developed, using the expert system shell "EXSYS", C and Fortran language. This system has the capabilities of calculating the numeric values of air quality index quantitatively, selecting the approapriate models based on decision rules and selecting parameters used in the selected models.

In this paper, we will discuss the construction of this system, model selection as well as model parameter selection knowledge base and demonstrate applications of the system to air quality assessment as a model selection tool.

2 Construction of coupled system

Regional air quality assessment generally include two different parts: numeral computation and logical reasoning. The purpose of numerical computation is providing a qualitative evaluation on air quality, such as calculating the assessment index of air quality using air pollution models. Logical reasoning include the selection of model and process algorithm. Based on the calculated results from fundamental data, conclusion can be deduced from facts step by step. Using expert system, logical reasoning can be implemented in computer. It is proved that both numerical calculation and logical reasoning play the important part in environmental quality assessment and it will be imperfect to use only one of them without another. If only using the expert system to process the symbolic reasoning without numerical computation, the obtained conclusion will be baseless and unacceptable, otherwise, only using the numerical calculation without expert system, the reasoning of problem - solving can not be made in computer. So, it is necessary to build up a coupled system to supplement each other which combine logical reasoning capability of expert system with the accuracy of qualitative numerical computation. Fig. 1 shows the procedure of air quality prediction.

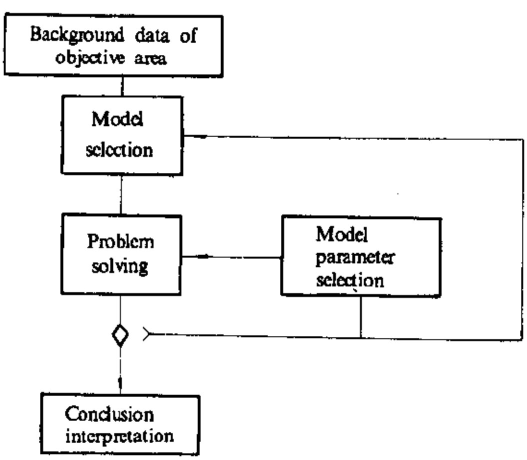

Fig. 1 The procedure of air quality prediction

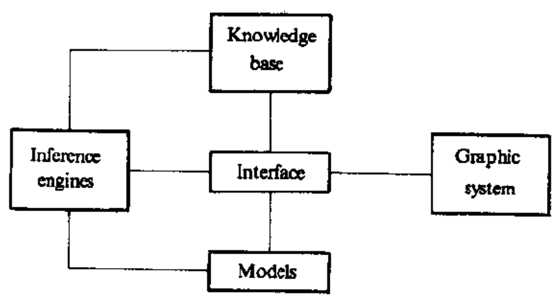

Fig. 2 Diagram of coupled system

First, user select one model supplied by the system according to model of domain experts to predict pollutant concentrations of each point in objective area based on measured data of pollutants. Next, they can compare the predicting result with the actual measured data. If deviations are within the error range which experts are allowed, the system will take this model to be a applicable one, else it has to try another model and repeat the procedure above. Actually, this procedure has two kinds of components. One is numerical calculation which can be demonstrated by computing program of model and another is heuristic experience knowledge which can be shown by production rules. The latter can be used in selecting model, adjusting parameters and analyzing computational results. On the basis of function, several relative independence sub - knowledge bases have been built up, such as model selection knowledge base, parameter selection knowledge base and conclusion interpretation knowledge base. A total of twenty - seven production rules are included in them. The knowledge bases are managed using the expert system shell "EXSYS". This system also include multi - variable statistic program package and a GEO - EAS configuration system. The statistic program package is compiled with Fortran -77 language. The computed values by the statistic programs are delivered to the graphic system, the Kriging display and contour plot of pollutant concentrations are output in LQ -1600 or LQ -2500 printer. Fig. 2 shows the diagram of coupled system.

3 Model selection knowledge of air quality prediction

It is well know that several models are often to make air quality prediction, for instance, Gaussian Puff model, Gaussian Plume model, Box model and K model. There are also other models such as Smoke model and complex terrain which are applicable in special conditions. Each model is suitable to specific conditions(Wispelaere, 1981, 1983; Weber, 1982).

Domain experts usually take into account of three factors in selecting air quality predicting models. The three factors are as follows: (1) Conditions in objective area. It means considering the natural conditions of objective area, such as size, terrain and meteorology and whether or not the selected models is applicable to the conditions; (2) acquisition of monitoring data. It means considering source of original data and parameter required by model. For example, if the mean values of pollutant concentrations of point source and area source are available easily; (3) environment of system. It means considering the hardware environment such as consumption of CPU time, storage capacity and weather or not the selected model can be run correctly in such condition.

Practically, for these problems of what scale should be used to describe the three factors above and which threshold parameter is needed to distinguish different cases for one factor, experts' opinion may be different according to their different experinces. Referring to the experience with the air quality assessment of Tianjin and the other successful experience of air quality assessment in some northern cities of China such as Beijing, Taiyuan and Shenyang and so on, we propose seven factors and twenty parameters which should be considered in model selection. Table 1 shows the selected parameters of air quality prediction model.

Table 1 actually include the thumb of rule used in model selection. For example, rules used to select Gaussian and Box model can be expressed as follows:

If: meteorological condition is general, and wind speed is stable, and the area size less than 20 km^2, and prevailing direction of wind is stable, and underlying surface is flat, and not only the mean value of concentration is needed, and omit the demands of computer. Then Box model can be used.

If: Meteorological condition is general, and wind speed is great than 1m/s, and the area size is great than 20 km^2, and prevailing direction of wind is stable, and underlying surface is flat, and the mean value of concentrations for point and area source is needed, and omit the demands of computer. Then Gaussian plume model can be used.

Table 1 Environmental air quality prediction model selection

Models	Conditions of objective area						
	Air stable condition	Mean value of size, wind speed, m/s	km^2	Prevailing direction of wind	Underlying surface condition	Requirement for monitoring data	Requirement for computer
Gaussian plume mode	Normal	>1.0	<20	Stable	Flat	Point source or area source	—
Gaussian	Normal	<1.0	—	—	—	—	—
puff model	Normal	>1.0	>20	—	—	—	General
Box model	Normal	—	<20	Stable	Flat	Area source	—
K model	Normal	>0.1	<20	Unstable	—	—	—
	Normal	>0.1	>20	—	—	—	Special
Mountain model	Normal	>0.1	<20	Unstable	Complex	—	—
Smoke	Worst	—	—	—	—	—	—

After the model is chosen, the basic equation of the model should be expressed in their specific forms such as point source, line source or area source equation. Perhaps, some difficult problems can not be solved while using the selected model in practice so that a new suitable model will be chosen once again. Therefore, model selection is an iterating and optimizing procedure. In addition, a large objective area can be divided into several sub-objective areas according to their different conditions. Different model can by applied in each sub-objective area. In this way, the comprehensive prediction of air quality in the large area will be more accurate.

4 The knowledge base of environmental air quality prediction model selection

4.1 Context tree

Fig. 3 shows the context tree of environmental air quality predicting model selection.

4.2 Rule set

Based on the context tree, the production rules are compiled and also shown in Fig. 3. The functions of each rule in Fig. 3 are as follows:

Rule 1 : choose the stability of air; Rule 2: choose the condition of wind speed; Rule 3: choose the evaluating size of objective area; Rule 4: choose the condition of prevailing wind direction; Rule 5: choose the condition of underlying surface; Rule 6: determine pollutant concentration requirements; Rule 7: determine the demand of computer.

4.3 Probability of conclusion

The probability of conclusion are obtained from the scores given by experts.

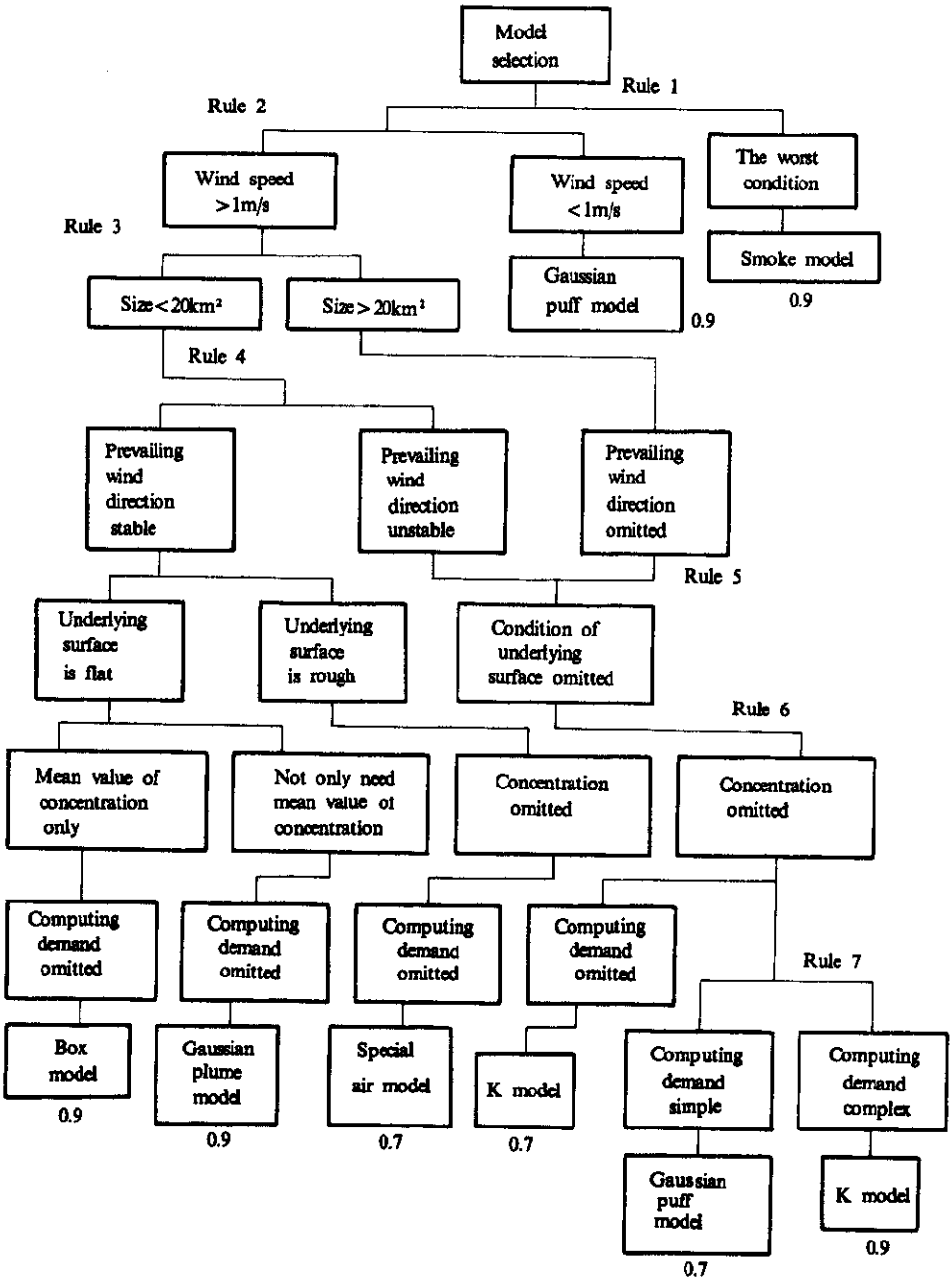

Fig. 3 The context tree of air quality model selection

5 The knowledge base of parameter selection for environmental air quality predicting model

After the model has been chosen and specific equations are formed according to the condition of objective area, the next work is to select parameters which are used by the model. The accuracy of model prediction highly depends on the reasonable selection of parameters. The selection of parameters concern with experience. Generally, first some parameters are selected and comported, then they are adjusted by comparing the computed and observed values. This procedure will be repeated until the result is satisfied. The procedure can be cut down greatly by expert, so that the efficiency can be raised. Based on the experience of experts, four sub - knowledge bases of model parameter selection have been built up which include Gaussian puff model, Gaussian plume model, Box model and K model. For short, only a context tree of parameter selection for Gaussian plume model is shown in Fig. 4.

6 Applications

This system has been applied to Tianjin and the air quality prediction has been made in the regional area. According to the circumstance of Tianjin, Gaussian plume model was chosen by the system automatically. Several fundamental data including concentrations of SO_2, meteorologic data and pollutant source data are inputted in the system which come from the yearbooks of EPA of Tainjin form 1980 to 1990. Then calculation is made using selected model. Histogram , Kriging display and contour plot for concentrations of SO_2 are generated by the GEO - EAS system. The model selection results is shown in Table 2.

Table 2 The model selection condition and result from 1980 to 1990 in winter

Condition	Conclusion	Probability
Stability of air: general	Gaussian	9/10
Mean value of wind: 2. 7m/s	plume model	
Size of area: <20km^2		
Prevailing wind direction: Northern, stable		
Underlying surface: flat		
Required data of concentrations: point and area source		
Computational process demand: general		

Tianjin First Power Plant is the biggest power station in Tianjin. It is the main emission source of air pollutants in downtown area of Tainjin. In order to reduce the concentrations of air pollutants in downtown area, the plant built up one taller stack instead of seven lower stacks in the ends of 1980s. We use Gaussian plume model to calculate the concentrations of SO_2 of 1981 and that of 1990 respectively. The predicted results are presented as follows: (1) In the winter of 1990, the mean value of concentrations for SO_2 from the 99% of downtown area is up to standard of China EPA. This value is better than that of 1981. (2) Points with higher concentrations of SO_2 have moved from downtown area to the suburbs of Tianjin during the decade. (3) The highest concentration of SO_2 caused by coal combustion of the power plant is reduced from 0. 321 ppm in 1981 to 0. 184 ppm in 1990, the latter approaches the standard of China EPA.

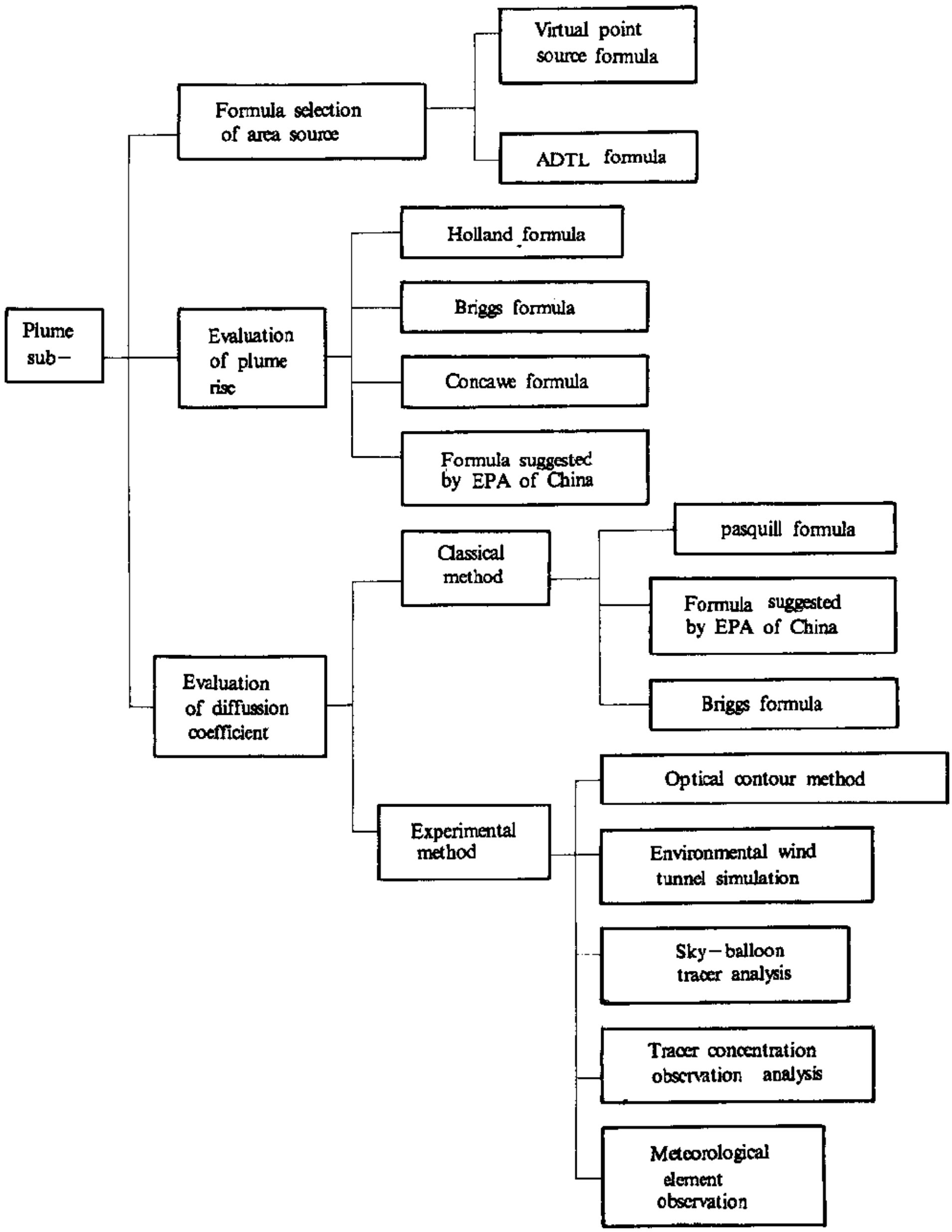

Fig. 4 The context tree of parameter selection for Gaussian plume model

The predictions above conform to the actual monitoring results. This shows that using taller stack instead of lower stacks is an effective way to reduce the concentrations of SO_2 of big point source in downtown (Hao, 1989). If a box model is selected, result could not be obtained because the stack altitude is not included in the box calculation. Otherwise, when the K Model is selected, although a result could be obtained, the cost of calculation is too expensive.

References

Hushon JM. Environmental Science and Technology, 1987; 21(9): 838

Hao Jiming. Air pollution control engineering. Beijing: High - Education Press. 1989

Lam DL. Ecological Modelling. 1989; 47(2): 131

Wispelaere CD. Air pollution modelling and its application I. New York and London: Plenum Press. 1981

Weber E. Air pollution assessment methodology and modelling. New York and London: Plenum Press. 1982

Wisplaere CD. Air pollution modeling and its application II. New York and London: Plenum Press. 1983

(Received December 4, 1992)