ADAPTIVE MODEL FOR SIMULATION OF ATMOSPHERIC POLLUTION

Jana Pazúriková
Master Degree Programme (2.), FIT BUT
E-mail: xpazur00@stud.fit.vutbr.cz

Supervised by: Radim Dvořák
E-mail: idvorak@fit.vutbr.cz

Abstract: Air pollution harms the environment and human welfare. Computer models and their simulation are useful tools for deeper understanding of processes behind as they quite accurately represent the dispersion and transformation of pollutants with advection diffusion equation or by other concepts. However, current models give valid results only to constrained cases of initial conditions. The general model combining the several specific models which is able to change according to input parameters and improve with training is proposed. The adaptiveness of the system is provided by decision tree as data structure with information for selection and combination process and genetic algorithm as optimization method for adjusting the tree to fit training data. The implementation and testing are ongoing, preliminary results given.

Keywords: air pollution, modelling and simulation, ADE, genetic algorithm, artificial intelligence

1 INTRODUCTION

Air pollution harms the environment and human welfare. In order to solve existing and prevent future problems, the deep understanding of air pollution's processes is necessary. Computer models use various mathematical and statistical concepts to find the relation between input parameters and measured concentrations of the pollutant after some time in given spatial scale of interest [Bui01]. Input parameters include source properties, such as its location, shape, emission rate, exit velocity and temperature of pollutant, weather conditions, mainly wind velocity, air turbulence and ambient temperature, and pollutant characteristics, its physical phase and chemical reactivity [Val08].

Current models usually use advection-dispersion equation (ADE) for calculating the expected mean and standard deviation of pollutant's concentrations [SJ05]. Results are calculated by numerical discrete approximation of ADE with methods as finite volume approximation or method of lines. Another approach is to use one of known analytical solutions, but they are valid only under specific initial conditions. Naturally, these models are most accurate for situations they were designed for.

However, the model which is able to adapt to input parameters, select appropriate specific models, combine them into one system and calculate the results could offer more general perspective. The choosing and combining these specific systems are two non-trivial tasks and due to desired adaptive behaviour of the system artificial intelligence methods might succeed. Proposed solution uses two of them - a genetic algorithm and a decision tree. The decision tree contains information about which models to select according to input characteristics and how to combine them into one general model which calculates concentrations. Genetic algorithm adjusts this decision tree by changing the information included and adding new one so that it fits best to training data.

The article follows with the description of design and gives preliminary results of testing and validation.

2 SOLUTION DESIGN

Proposed adaptive model of atmospheric pollution is to our knowledge the first attempt to select and combine specific models according to input parameters, however, few systems of blending two models together were made by [HPM08, ch.14]. The solution applies decision tree and genetic algorithm to the tasks of combining models and adjusting the system.

The model works from user's point of view like every other air pollution model. User inserts input variables, system calculates the results in temporal and spatial area of interest. However, the calculation starts with building the combined model with information from the decision tree. The concentrations are then computed with this combined model. The main feature of the proposed model, its adaptiveness, requires a bit different behaviour. In training mode, after the input of training data which consists of input variables and measured concentrations, the system's results are obtained as in user mode. The genetic algorithm compares the calculated and measured results and change decision tree accordingly.

The decision tree contains all the information about the process of specific models's selection and combination. The leaves contain implemented specific models, nodes (except leaves) represent conditions such as *point source* or *constant wind speed* required for model validity or recommended application to certain input properties, e.g. Lagrangian model is *suitable for long temporal scales* [Bui01, p.8], and threshold values. The leaf links to the model which should be used if the conditions in the nodes of the path from root are satisfied. Moreover, all conditions of the decision tree and the extent to which they need to be satisfied (value *1.0* - must be/validity requires, *0.7* - should be/it is recommended, *0.5* - might be/does not matter, *0* - can not be/would cause invalidity) are stored in data structure for each implemented model. The simple example of decision tree is in Figure 1. The nodes contain conditions about temporal and spatial scale and choose one model according to values of those variables given by the user.

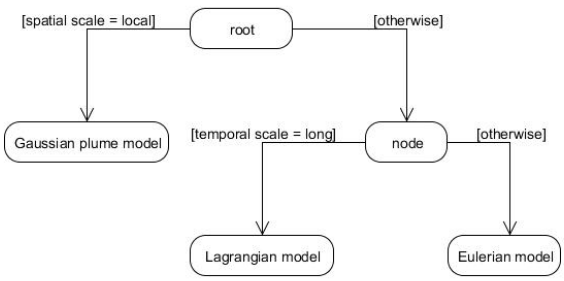

Figure 1: Example of the decision tree.

The combining of models into one general system uses the decision tree with information necessary for decisions about selection. The building itself is explained by following example. The new industrial stack starts to release continuously non-reactant gas with a few particles of heavy particulate matter. On the right side of the factory, the terrain starts to rise and form a mountain, on the left side the city lies in plane. The user wants to know how the pollution in the area of 15 km will look in two days if it does not rain. At least two models can be used, first simple tilted plume model for heavy particles, second Lagrangian or Gaussian model for gas. If the system has also model for mountainous terrain it can be included and calculate the results for right side. The transient area's results can be calculated as weighted mean of results from two areas. In the case of two or more models suitable for the same area, linear combination of specific models's concentrations is given as the result.

Genetic algorithm makes the system adaptive by optimizing the decision tree structure. Generally, the algorithm iteratively evaluates the population of individuals and builds new one with operations of mating and mutation. Chromozome representation of a tree consists of nodes's list in preorder. For every node, it includes its id number, split threshold and binary flag set if the threshold can be changed. Population is small, less than 10 individuals and only the best one is used for mutation. Basic mutation operations are altering the threshold and inserting new node into the tree. The need to keep validity conditions in the path from root to the specific model, mating and mutation operations over trees as in genetic programming can not be applied. Cost function is root mean squared error between measured concentrations and results given by combined model built according to a tree from population. The best found tree is then saved and used afterwards in user mode.

3 IMPLEMENTATION AND EXPERIMENT'S RESULTS

So far, five specific models have been implemented. They were all tested on artificial input parameters. Gaussian model and analytical solution for continuously emitting point source were validated on Copenhagen experiment's data, see table 1, as it fits their validity conditions (point source, mostly constant wind speed, mostly constant wind direction, constant emission rate). Combined model uses $0.9 * c_{gaussian} + 0.1 * c_{pointSource}$ to calculate results, coefficients were determined by the genetic algorithm using all runs of Copenhagen experiment [GLRNLD98].

	Gaussian model	Cont. point source model	combined model
NMSE	0.141	0.328	0.146

Table 1: Average of normalized mean squared error of four runs of Copenhagen experiment

4 CONCLUSIONS

We presented the adaptive model for simulation of atmospheric pollution. It uses decision trees and genetic algorithms to select and combine various specific models of pollution into one. Such general model can modify its behaviour according to input parameters and learn with training. The completion of implementation and further testing are yet to be done.

REFERENCES

[Bui01] Peter Builtjes. Major twentieth century milestones in air pollution modelling and its application. In Sven-Erik Gryning and Francis Schiermeier, editors, *Air Pollution Modeling and Its Application XIV*, pages 3–16. Springer US, 2001.

[GLRNLD98] S.E. Gryning, E. Lyck, Roskilde. Wind Energy Risø National Lab., and AtmosphericPhysics Dept. *The Copenhagen tracer experiments: Reporting of measurements*. RISØ-R–1054. 1998.

[HPM08] Sue Ellen Haupt, Antonello Pasini, and Caren Marzban. *Artificial Intelligence Methods in the Environmental Sciences*. Springer, 2008.

[SJ05] Scott A. Socolofsky and H. Jirka, Gerhard. *CVEN 489-501: Special Topics in Mixing and Transport Processes in the Environment, lecture notes*. Texas A&M University, 5 edition, 2005. Available online <https://ceprofs.civil.tamu.edu/ssocolofsky/cven489/Book/Book.htm>.

[Val08] D. Vallero. *Fundamentals of Air Pollution, 4th edition*. Elsevier, 2008.