Geophysical Research Abstracts, Vol. 8, 02642, 2006
SRef-ID: 1607-7962/gra/EGU06-A-02642

Advanced regression method for prediction of NOx pollutant concentrations in urban air

K. Li (1), A. Ogle (2), J. Peng (1), B. Pizzileo (1)

1) School of Electronic, Electrical Engineering and Computer Science

Queen's University Belfast, Ashby Building, Stranmillis Road, Belfast BT9 5AH, UK

(k.li@qub.ac.uk, j.peng@qub.ac.uk, bpizzileo01@qub.ac.uk)

(2) ATU, QUESTOR centre, David Keir Building, Queen's University Belfast

Stranmillis Road, Belfast, BT9 5AG, UK

(a.ogle@qub.ac.uk)

As a major pollutant in the air that affects human health, the nitrogen oxide (NO_x) emission problem has received a lot of public attentions and academic researches in the past decade. Vehicle exhaust and other combustion emissions (including household heating systems, industrial combustors of different scales, etc) are the main sources of NO_x in urban air. Most NO_x is emitted in the form of nitric oxides (NO), but most of it is ultimately converted to NO_2 by reaction with ozone (O_3) in the atmosphere. Both for the air quality forecasting and for the development of control strategy and policy, it is important to identify the factors that control NO_x concentrations and to develop a function (model) to predict the NO_x concentration.

There exist two different general approaches in developing the prediction model. The first approach is to develop atmospheric diffusion models, and the second is the regression models. The first approach requires detailed NO_x emission data distributed over the studied area which is usually very difficult to obtain, and the modelling process is computationally quite demanding. The second approach is to develop a regression model, which is perhaps more computationally efficient, however heavily depends on the modelling method and the historic data quality and richness.

Among various regression models, linear-in-the-parameters structure is perhaps one

of the most popular forms that are used in time-series prediction, nonlinear system modelling and identification, signal processing and pattern recognition. Examples of such model structure include the linear or nonlinear AutoRegressive model with eXogenous inputs (ARX model/NARX), B-spline neurofuzzy network, Volterra neural networks and radial basis function (RBF) networks, etc. One problem with the identification of linear-in-the-parameters models is that a very large pool of model terms has to be considered initially, from which a useful model is then generated based on the parsimonious principle, of selecting the smallest possible model, in terms of size, which explains the data. In the linear regression field, this problem is referred to as the subset selection. Given a model selection criterion, exhaustive search of all possible models is too expensive to implement. For example, exhaustive search of the optimal model with 20 candidate terms involves 2.43×10^{18} search paths. Part of the problem is referred to as the curse of dimensionality in the literature.

Among various subset selection approaches, the forward stepwise is perhaps the only possible method in the literature when a very large term pool and a large database are presented. Despite the great efficiency of forward stepwise methods, the Achilles heel is that the final model is not optimal. In other word, a good model can be easily missed using such approaches. To overcome this problem, a two-stage algorithm is proposed in this paper. The main objective is to improve the compactness of the model which is obtained by the forward model selection methods, while retaining their computational efficiency. The proposed algorithm first generates an initial model using a forward stepwise algorithm. The model is then reviewed at the second stage, aiming to improve its compactness. In detail, the significance of each selected term is reviewed at the second stage and all insignificant ones are replaced, resulting in an optimised compact model with significantly improved performance.

This paper uses the proposed method to model and predict NO_x emissions in the urban air in Belfast of Northern Ireland. The historic emission data, traffic and weather information are used to predict NO_x pollutant concentrations. Various model types, including the ARX model, Nonlinear ARX model, and RBF neural networks are developed and the performances of these models are then compared in terms of their long-term prediction capabilities.