Expert Systems with Applications 36 (2009) 6389–6402

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Contour diagram fuzzy model for maximum surface ozone prediction

Zekai Şen [a], Abdüsselam Altunkaynak [a,*], Kadir Alp [b]

[a] İstanbul Technical University, Civil Engineering Faculty, Hydraulics Division, Maslak 34469, İstanbul, Turkey
[b] İstanbul Technical University, Civil Engineering Faculty, Environmental Engineering Division, Maslak 34469, İstanbul, Turkey

ARTICLE INFO

Keywords:
Contour diagrams
Fuzzy rules
Fuzzy sets
Meteorology
Non-parametric approach
Ozone concentrations
Vagueness

ABSTRACT

A contour diagram approach is presented for the identification of surface ozone concentration feature based on a set of rules by considering the meteorological variables such as the solar radiation, wind speed, temperature, humidity and rainfall. A fuzzy rule system approach is used because of the imprecise, insufficient, ambiguous and uncertain data available. The contour diagrams help to identify qualitative ozone concentration variability rules which are more general than conventional statistical or time series analysis. In the methodology, ozone concentration contours are based on a fixed variable as ozone precursor, namely, NO_x and as the third variable one of the meteorological factors. Such contour diagrams for ozone concentration variation are prepared for six months. It is possible to identify the maximum ozone concentration episodes from these diagrams and then to set up the valid rules in the form of **IF-THEN** logical statements. These rules are obtained from available daily ozone, NO_x and meteorological data as a first approximate reasoning step. In this manner, without mathematical formulations, expert maximum ozone concentration systems are identified. The application of the contour diagram approach is performed for daily ozone concentration measurements on European side of Istanbul city. It is concluded that through approximate reasoning with fuzzy rules, the maximum ozone concentration episodes can be identified and predicted without any mathematical expression.

1. Introduction

In general, volatile organic compounds and nitrogen oxides reductions are the preliminary conditions for the surface ozone concentration improvements. Meteorological variability adds significant factors for surface ozone concentration reductions control. Hence, the nitrogen oxide and the meteorological factors levels should be identified for effective and sustainable ozone concentration reduction strategies. Any national or local authority needs such changes for decisive conclusions and policies for sustainable control. Ozone concentration measurements with time at any station are embedded with various scales of fluctuations, trends and even shifts in addition to the random changes. In the literature, most often ozone data are processed through dynamic, stochastic, probabilistic or statistical models. However, any of these models require a set of restrictive assumptions, and their success depends on these assumptions. For instance, in any statistical assessment of ozone concentration, the set of assumptions includes the independence of the residuals, homoscadascity, i.e., constancy of the variance, linear dependence, normal (Gaussian) error distribution (Benjamin & Cornell, 1970). Unfortunately, these assumptions are overlooked in many practical applications and solely the curve fittings are adopted as basic models.

Garden and Dorling (2000a) have recently studied the maximum surface ozone concentrations in United Kingdom for identification of meteorologically adjusted trends by using artificial neural network modelling. This modelling technique takes into consideration the interrelationship of all the meteorological variables with the ozone concentration amounts without any explicit mathematical expression (Şen, Altunkaynak, & Özger, 2005). Besides, for the reliable application of the artificial neural network model, enough data set is necessary for many years. On the other hand, inherent temporal variations in ozone concentrations make the basic modelling equations as approximations whose values are conditioned on appropriate calibrations through numerous tuning parameters. Sometimes, models include many parameters, and hence they are not parsimonious which is a required property from the practical point of view (Box & Jenkins, 1973). Additional influence of meteorological variables provides large uncertainties embedded within the surface ozone data and these make it further difficult to obtain a good agreement between the model predictions and observed data (Chang & Suzio, 1995; Chen, Islam, & Biswas, 1998; Rao & Zurbenko, 1994; Seinfeld, 1988). Furthermore, such modelling approaches make the inherent natural variability of the ozone data to be filtered or smoothened out due to many assumptions. In order to avoid restrictive assumptions, in this paper as a first logical

* Corresponding author.
E-mail addresses: zsen@itu.edu.tr (Z. Şen), altunkay@itu.edu.tr (A. Altunkaynak).

0957-4174/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2008.07.050

step, the natural variabilities in the data are assessed prior to any modelling with qualitative simple diagrams and interpretations. Subsequently, on these qualitative knowledge an objective model is constructed through fuzzy logic and system for prediction possibilities.

Simpson, Olendzynski, Semb, Storen, and Unger (1977) have concluded, "It follows that long-term monitoring networks are essential if trends due to emission changes are to be detected. Models and statistical analysis will also be required to disentangle the various factors contributing to measured trends". There have been many statistical (Garden & Dorling, 2000b), time series and spectral analysis (Seabald, Treffeisen, Reimer, & Hies, 2000), stochastic approach (Simpson & Layton, 1983) and dynamic modelling (Chen et al., 1998; Şen, Koçak, & Tatlı, 2000) and the use of artificial neural networks for the surface ozone concentration prediction. On the other hand, the limitations of regression and neural network models are presented by Soja and Soja (1999). It is stated that the neural network model did not always perform better than the regression models.

It is the main purpose of this paper to expose first of all qualitative statements based on contour diagram patterns. Subsequently, a fuzzy approximate reasoning approach is proposed for the prediction of surface ozone. The identification of linguistic rules is needed prior to any modelling study for the reliability of the model developed. This is a sort of non-parametric model assessment of the ozone concentration variability with NO_x emissions and meteorological variables. Herein, NO_x measurements are invariably considered in the ozone contour diagrams because it is the main ozone-producing agent. The basic monthly fuzzy rules are derived for the maximum ozone concentration prediction. The methodology is applied for ozone concentration measurements in İstanbul together with the ozone precursor NO_x and meteorological variables.

2. Fuzzy sets and rules

Zadeh (1965) has proposed the use of imprecise information through fuzzy sets for quantitative evaluation of the available data. Instead of two-valued, i.e. bivalent logic of scientific approaches and solutions, he suggested the use of infinite valued logical statements for ambiguous, vague, imprecise and uncertain information and knowledge assessments. His approach does not include any scientific assumption or mathematical formulation based on conservation principles and body laws as in analytical modelling but rather simple logical relationships that can be obtained between different input and output variables. These are rather in the form of **IF-THEN** logical statements. For instance, a very simple fuzzy rule may be stated as.

IF the solar radiation is high **THEN** the ozone concentration is big (1)

This statement as a rule relates the input solar radiation variable to output ozone concentration variable linguistically in a vague manner. In any problem the number of such rules may be numerous depending on the nature of the problem concerned. In the above statement the part between the **IF** and **THEN** words is referred to as the premise or condition section which should include all the independent variables of the problem. However, the part after **THEN** is called consequent of the fuzzy rule and it includes the prediction variable, which is the ozone concentration in this paper. In the fuzzy rule the words like "high" and "big" are the atomic and uncertain i.e. fuzzy words that should be quantified through membership functions as stated by Zadeh (1965). Although there are different ways of deducing the fuzzy set membership functions based on intuition, expert view, regression, artificial neural networks, genetic algorithms and other objective methods as stated by Ross (1995), these are time consuming and the first two are not data dependent (Altunkaynak & Şen, 2007; Şen & Altunkaynak, 2004). Consequently their uses require expertise and detailed methodological techniques. It is therefore, preferred in this paper to develop a straight forward, quantitative and objective approximate reasoning procedure through contour diagrams by considering ozone concentration and NO_x variables as basis in addition to a third variable which is taken as one of the meteorological variables.

3. Data

Daily ozone concentrations are measured automatically at Saraçhane station on the European side of İstanbul city (Fig. 1). The greater İstanbul municipality, Department of Environmental Protection Authority measures this station. Especially, during summer, ozone concentrations increase within the city mostly due to the traffic volume and high temperature. Ozone data are measured using O_3 41 M sensor produced by the Environmental Inc. Commercial. Ozone concentration measurements are taken during heavy traffic density especially, in the morning and evening hours. The concurrent daily meteorological data are obtained from the State Meteorological Works Department of Turkey. The distance between the ozone concentration measurement site and the nearest meteorology station at Sarıyer is not more than 15 km (see Fig. 1). This station is chosen such that all the required meteorological variables are measured daily including maximum and minimum temperatures, total daily sunshine duration, total solar irradiation, mean and maximum daily wind speeds, relative humidity and rainfall amounts. In general, wind direction persistence is from the north and the frontal rainfall patterns that are effective in İstanbul city originate from the Balkan Peninsula in south-eastern Europe. For the application of the methodology developed in this paper, daily ozone concentration measurements are taken at Saraçhane station for 1999. The relatively low ozone values indicate a direct influence of traffic emission.

4. Ozone concentration contour diagrams

In general, contour diagram (CD) can be regarded as the full range mapping of ozone concentration variations with two effecting variables. This method can also be referred to as the graphical three-dimensional regression approach, which has not been adopted in this paper. A representative CD pattern is shown in Fig. 2. This simple procedure can be explained as follows:

(i) Consider daily ozone concentrations as the independent variable that will be interpreted on the basis of two dependent variables.
(ii) Since NO_x is one of the main precursors for ozone generation, it will be always considered as one of the dependent variables on the vertical axis with the second dependent variable from the set of meteorological variables on the horizontal axis. Hence, ozone concentration is considered as dependent and NO_x and one of the meteorology factors as independent variables.
(iii) In this manner, equal ozone lines are drawn and the resulting diagram provides the change of ozone concentrations with full ranges of NO_x and meteorological variable concerned.

On the basis of the CD, it is possible to interpret the whole variation ranges of ozone concentrations based on two of the dependent variables. Since there are no parameters involved in the procedure, it can be referred to as a non-parametric approach. Rather than numerical crisp parameter values as in any classical