17th European Symposium on Computer Aided Process Engineering – ESCAPE17
V. Plesu and P.S. Agachi (Editors)

Risk of Gaseous Release Assessment Based on Artificial Intelligence Methods

Calin Ioan Anghel[a], Alexandru Ozunu[b]

[a]*Department of Chemical Engineering, Faculty of Chemistry and Chemical Engineering, University Babes-Bolyai, Cluj-Napoca, canghel@chem.ubbcluj.ro*
[b]*Department of the Environmental Physics and Chemistry, Faculty of the Environmental Science, University Babes-Bolyai, Cluj-Napoca, aozunu@chem.ubbcluj.ro,*

Abstract

Based only on current pollutant measured concentrations and atmospheric parameters the paper presents a novel procedure able to predict pollutant emission concentrations and to estimate the risk of pollution. Instead of deterministic or probabilistic methods, cumbersome regression analysis or physical models, a *minimax decision procedure based on support vector machine in a minimax* approach implemented in *MATLAB* object oriented language, was utilised. This procedure can perform highly complex mappings on nonlinearly related data, inferring subtle relationships between inputs and outputs. Numerical experiments were reported to gaseous emissions of pollutant sulphur dioxide from a thermo power station smokestack.

Keywords: pollutant emissions, artificial intelligence, minimax decision procedure, predicted concentration, risk assessment.

1. Introduction

Monitoring and forecasting of air quality parameters are important topics of atmospheric safety assessment and environmental quality. The risk assessment can be achieved based on various deterministic or probabilistic methods and models. Inferring subtle non-linear relationships between inputs and outputs artificial intelligence techniques have the ability to investigate new phenomenon cases where the information cannot be easily accessed theoretically or explicitly relational physics. Today these techniques became popular in environmental engineering domains. The paper focuses only on predictions of pollutant concentrations and risk of pollution. The procedure casts both regression (numerical values as outputs) and classification problems (class labels as outputs) into a unified technique [2]. According most usually accepted definitions, the risk will be defined as the probability that at a given location, a hazard index exceeds a critical limit value. The gaseous centreline pollutant concentration of sulphur dioxide, $C(x_i)$, at different locations along the downwind direction was chosen as hazard index. Because only the likelihood of a particular state may be of interest, the "risk" of pollution from a pollutant source will be evaluated based on the principle of *limit state function* [1] according to a critical concentration, C_R, representing the value of 'dangerous concentration'. This procedure is opportune al least in the following situations: (1) to verify the data reported by the industry and the compliance with stated regulations, (2) to monitor the emissions and to estimate the trend of pollution and (3) as a method for preliminary analyse regarding the pollution risk of pollutant emissions. The procedure works without needing theoretically dispersion models and dependencies or directly explicitly relational informations.

2. Problem Statement, background

In supports vectors machine *in a minimax* approach [3] into a binary classification problem of $\mathbf{z}$ random vectors, with $\mathbf{z_1}$ and $\mathbf{z_2}$ denoting random vectors from each of two classes as $\mathbf{z}_1 \in \text{Class 1}$ and $\mathbf{z}_2 \in \text{Class 2}$, a hyperplane that separates the two separable classes of points,

$$H(\mathbf{w}, b) = \left\{\mathbf{z} \middle| \mathbf{w}^T \cdot \mathbf{z} = b\right\}, \text{ where } \mathbf{w} \in \mathbf{R}^{\mathbf{d}} \setminus \{0\} \text{ and } b \in \mathbf{R} \qquad (1)$$

with maximal probability in respect to all distributions having mentioned means $\bar{\mathbf{z}}_1, \bar{\mathbf{z}}_2$ and covariance matrices $\sum\mathbf{z}_1, \sum\mathbf{z}_2$, will be determined by:

$$\max_{\alpha, \mathbf{w} \neq 0, b} \alpha \quad s.t. \begin{cases} \inf\limits_{\mathbf{z}_1\left(\bar{\mathbf{z}}_1, \Sigma\mathbf{z}_1\right)} Probability\left\{\mathbf{w}^T \cdot \mathbf{z}_1 \geq b\right\} \geq \alpha \\ \inf\limits_{\mathbf{z}_2\left(\bar{\mathbf{z}}_2, \Sigma\mathbf{z}_2\right)} Probability\left\{\mathbf{w}^T \cdot \mathbf{z}_2 \leq b\right\} \geq \alpha \end{cases}. \qquad (2)$$

The term α represents lower bound of the correct classification of future data, $w\in R^d \setminus \{0\}$ is the outward normal vector of the hyperplane, b offset of the hyperplane from origin. In high-dimensional problems, the kernel trick is used. Mapping original input data points into a high dimensional features space through a non-linear mapping function 'Φ' which is usually unknown, a linear classifier-surface between the two classes corresponds to a non-linear decision into original input space. Recently [4-5] working with "kernel" trick *minimax probability machine regression* implemented regression into feature space as:

$$\hat{\mathbf{y}} = \hat{f}(\mathbf{z}) = \sum_{i=1}^{N} \beta_i K(\mathbf{z}_i, \mathbf{z}) + b_k \tag{3}$$

Here $K(\mathbf{z}_i, \mathbf{z}) = \Phi(\mathbf{z}_i) \cdot \Phi(\mathbf{z})$ is so-called kernel function, N represents the number of learning examples (points), β_i are weighting coefficients and 'b_k' offset of the minimax regression model, obtained from the learning data.

3. Paper approach

3.1. Methodology

The starting point for application is to establish the *limit state function*:

$$LSF = C_R - C(x_i) \text{ in real space, or } LSF = C_R - C(u_i) \text{ in reduced space} \tag{4}$$

The first order approximation for the risk of pollution is the probability that the current values of limit state function will exceed critic zero value given by the following expression:

$$P_f = Prob\left\{\mathbf{u} \in U^d \,\middle|\, LSF(u) \le 0\right\} = \Phi(-\beta) \tag{5}$$

where $\Phi(...)$ is the standard normal cumulative density function, and β the perpendicular distance from the optimal separating hyperplane to the origin of reduced space [1]. To ensure robustness the experimental database is selected and divided in a random manner into different training and test sets. The experiments are performed cyclic "k" times inside multiple "n" random trials. The performance of predictions was evaluated based on the following:

- empirical correlation factor R_C and relative error RE of predictions as:

$$\mathrm{RE} = \left|Y_{predicted} - Y_{test}\right| \Big/ \left|Y_{predicted}\right| . \tag{6}$$

- a simple equivalent linear dependency between predicted and test values:

$$Y_{predicted} = a \cdot Y_{test} + b . \tag{7}$$

Table 1. Values of input parameters

Inputs	$(SO_2)_O$ [ppm]	$(W_O)_O$ [m/s]	PA [hPa]	TA [^OC]	UA [%]	W [m/s]	SO_2 [ppm]
Range	22÷180	0.119÷ 3.426	9.518÷ 9.827	-6.5÷8.2	51÷98	0.5÷4	0.048÷ 2.072
Mean value	63.627	1.621	9.717	-0.064	74.324	1.725	0.538
Standard deviation	36.248	0.692	0.070	3.874	13.626	1.120	0.441

Notes ***(1)*** Values $(\ldots)_O$ were measured at the pollutant source (smokestack)
(2) The others values were measured at 10 km location along the downwind direction

Table 2. The main results reported to *model data*.

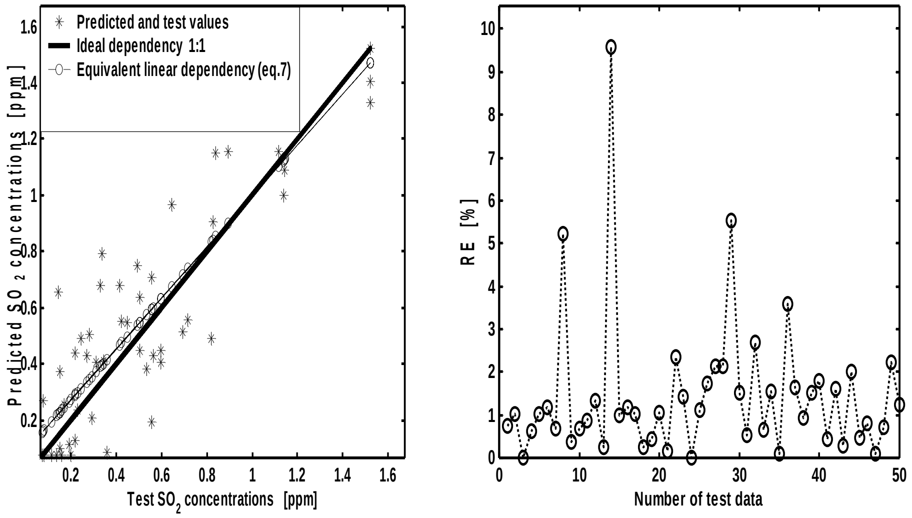
Parameters	***Minimax decision results***
Test pollutant concentration of sulphur dioxide [ppm]	0.073÷1.525
Predicted pollutant concentration of sulphur dioxide [ppm]	0.072÷1.522
Empirical correlation factor (R_C)	0.898
Relative error (RE) range [%]	0.000÷9.580
Coefficients of equivalent linear dependency Eq. (7)	a = 0.906 b = 0.091
Probabilistic risk of pollution for pollutant sulphur dioxide emissions	0.829
Test set accuracy [%]	94.736
Lower bound on correct classification of future data [%]	93.229

Note-Critical concentrations of pollutant emissions as C_R=0.133 ppm for sulphur dioxide are reported to human health limit according to Romanian O.M.A.P.M – 462/2002, 541/2003.

3.2. *Case study*

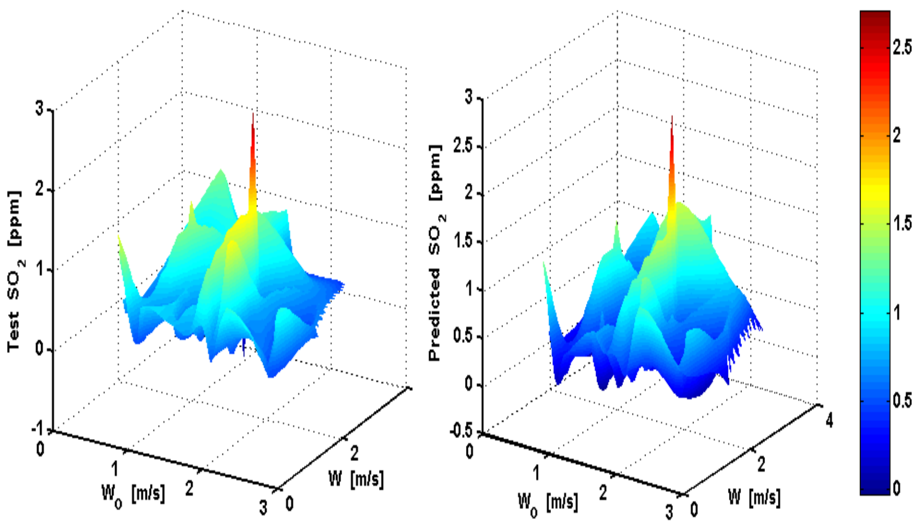
The air quality database includes pollutant concentration of sulphur dioxide $(SO_2)_O$, SO_2 and meteorological parameters as, wind speed (W_O,W), atmosphere pressure (PA), outdoor temperature (TA), relative air humidity (UA). The pollutant database (Table 1) was collected in two distinct areas: $(SO_2)_O$, W_O at the top of a smokestack in a thermo power station and the others at 10 km location along the downwind direction. To limit the computing time a recent statement [6] was extended to limit the cyclic random trials *k,n* = 50…100.

3.3. Results & discussions



a. **b.**

Fig. 1 The performance of procedure for sulphur dioxide predictions
a. Dependency between predicted / test values; b. Variation of relative errors;



a. **b.**

Fig. 2 Dependencies between sulphur dioxide concentrations and wind speeds.
a. Dependencies for test values; b. Dependencies for predicted values;

The correspondences between, predicted and test SO_2 pollutant concentrations, are in a reasonable agreement (Fig. 1a). The value of empirical correlation coefficients $R_C = 0.898$ and the relative percent errors reported to *model data* (Fig. 1b) reveal a good performance of the procedure and the capacity to predict data with accuracy. The procedure allows to assess the effect of each concentration input over other examined individually and to gain dependencies which may be very difficult or impossible to do in practice. The dependencies were done for both test and predicted data within a confidence interval of 95%. Based on predictions as illustrated in Figure 2, qualitatively and quantitatively interactions can be emphasized. Probabilistic risk of pollution of pollutant sulphur dioxide release over 0.829 value and test/predicted values of pollutant sulphur dioxide emissions, reveal that basically the thermo power station works in a wrong way and pollutant emissions represents a real environmental risk at the considered location of measure.

4. Conclusions

The paper presents a novel procedure able to predict the variations of pollutant sulphur dioxide concentration along the downwind direction and to estimate the risk of pollution. Therefore *minimax decision procedure* provides a promising alternative for pollutants concentrations forecast. Used with an appropriate policy related to environmental protection it could lead to a drop from industrial pollutant emissions and hence to a decrease in the atmospheric levels of primary pollutants. The application reveals the opportunity of such approaches based on artificial intelligence methods to verify the data reported by the industry, to monitor the pollutant emissions, or environmental risk assessment. Numerical experiments reveal the accuracy and predictive power of this procedure and the opportunity of such assessments for engineers concerning with pollution and risk management.

References

1. C.I. Anghel and Al. Ozunu, Chem.Pap. 59 (6b) (2005) 469.
2. C.I. Anghel and Al. Ozunu, Chem.Pap. 60(6) (2006) 410.
3. G.R.G. Lanckriet, E.L. Ghaoui, C. Bhattacharyya, and M.I. Jordan J. Machine Learning Res. 3 (2002) 555.
4. T.R. Strohmann, G.Z. Grudic, Advances in Neural Information Processing Systems, 15 (2003) 869.
5. G.Z. Grudic, Probabilistic regression using basis function models, Report CU-CS-975-04,Colorado, (2004). http://www.cs.colorado.edu/»grudic/software.
6. A. Bordes, S. Ertekin, J. Weston, and L. Bottou, J.Machine Learning Res. 6 (2005) 1579.