# AN INVESTIGATION OF PROBLEM OF ARTIFICIAL NEURON NETWORKS APPLICABILITY FOR SOLUTION OF PROBLEM OF FORECASTING OF GAS DISPERSION IN ATMOSPHERE

Igor Shostak, Valentina Davidenko

*Abstract. The problems of artificial neuron networks (ANN) applicability for forecasting of hazardous pollutants (HP) discharge into the atmosphere are considered in this article. It is spoken in detail about existing approaches' and methods' advantages and disadvantages. The fundamental scheme of ANN usage for modelling is offered. The scope of its applicability is described.*

*Keywords: emergency, artificial neuron networks, air pollution, gas dispersion.*

*ACM Classification Keywords: H.3.4 Systems and Software*

*Conference: The paper is selected from XVth International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

## Introduction

The problem of gas dispersion modelling is of interest not only in science, but also in practical life, especially if this problem has deal with emergency discharge of HP into the atmosphere due to emergency situation at transportation time or in permanent placements (stores, chemical plants and so on). These situations in turn increase probability of HP affection of civilian population or possible damage of special resources.

The problem of provision of emergency services with precise information about forecasting of HP emission in the atmosphere for efficient salvage operations.

## Problem Statement

In terms of mathematical description the problem of gas dispersion modelling comes to evaluation of function (1) with known initial parameters.

$$F(t_n, \{Met_n ... Met_0\}, \{x_n, y_n, z_n\}, \{t_0, Met, S, \{x_0, y_0, z_0\}\}) = Con_n, \quad (1)$$

where $t_n$ – interest time at which we want to know gas concentration $Con_n$;

$\{x_n, y_n, z_n\}$ – location coordinates of interest place in which we want to know gas concentration $Con_n$;

$\{Met_n ... Met_0\}$ – description of meteorological parameters in a period *[0..n]* time units;

$$\{Met_i = \{Met_{ij}\}, \quad (2)$$

where $Met_{ij}$ – meteorological parameters which are get at time moment *i* from meteorological station *j*;

$$Met_{ij} = \{Temp_{ij}, Humid_{ij}, WS, WD\}, \ j=1,k, \quad (3)$$

where $Temp_{ij}$ – set of temperatures measured from different heights;

$Humid_{ij}$ – relative air humidity;

*WS* – set of wind speeds measured on different heights;

*WD* – set of wind directions measured on different heights;

k – amount of meteorological stations;

*S* – emission characteristic parameter:

$$S = \{Type, Vol\_All, Press_0, Temp_0, Vol\_Ex\}, \quad (4)$$

where *Type* – gas's type;

*Vol_All* – total gas volume in the source;

*Vol_Ex* – released gas volume ( $Vol\_Ex \leq Vol\_All$ );

$Press_0$ – gas pressure inside the source at the release moment;

$Temp_0$ – gas temperature inside the source at the release moment.

However the problem of evaluation of gas concentration in some fixed location with certain radius is greatly more actual in the real practice.

The problem of modelling for short-term periods is considered. So having these conditions we can account meteorological parameters to be constant for further calculations.

## Analysis of Existing Approaches

Existing models and approaches which describe gas dispersion in the atmosphere don't have a satisfactory accuracy when its deal with open space. And they are almost useless for forecasting and further building of the most precise picture of HP dispersion into the urban regions.

In some cases when gas dispersion process is described by physical laws expressed through mathematical formulas it becomes impossible to take into account all physical and statistical laws. And it leads to simplify to that view which can be programmed (e.g. Navier-Stokes equations) but accuracy is extremely decreased. In other cases Gaussian plume model and its modifications are used but it give to us only a very approximate picture what is going on, because shape of potential gas dispersion region is determined as ideal regular curve.

The Lagrange and Eurlian model-based algorithms give more precise results. The main idea of these methods consists in division of gas emission into many elementary boxes with individual characteristics. But these models don't sufficiently take into account complexity of that processes which occur at dispersion time, especially in cases concerning with complex obstacles (e.g. 3D location between arbitrarily situated buildings).

## Task Decision

So assigned task has following characteristics:

- nonlinearity;
- accounting of huge count of factors (meteorological conditions, geometrical parameters of landscape and different objects);
- dynamics of process and parameter variability.

ANN can theoretically cope with this task (as follows from problem definition). And it is one of all ANN's purposes [1]. It is also declared in [2] about possibility of application of artificial intelligence techniques, especially about ANN for decision of similar tasks. It is necessary to formulate the task in terms of ANN and properly prepare data for possibility of ANN application.

Figure 1 shows scheme of gas dispersion in 2D flat ground which includes following peculiarities:

- the area is divided into conventional square regions - the cells;

- coordinate system is usually relative with center in the locate of gas emission source with information S (4);
- meteorological stations with information Met (2).

It is obvious that all meteorological parameters are not known exactly at any point of space under consideration. In order to determine approximate parameter value an interpolation of existing data is used.

- gas concentration level is determined in all corners of space cells, i.e. any cell is characterized by concentration data C:

$$C_{ij} = \{M_{ij}, Con_{ij}\},\ i=1,n,\ j=1,m, \qquad (5)$$

where $M_{ij}$ – interpolated environment parameters at time *i* for cell *j*;

$Con_{ij}$ – concentration level at time *i* in cell *j*;

*m* – current amount of cells;

*n* – amount of time units.

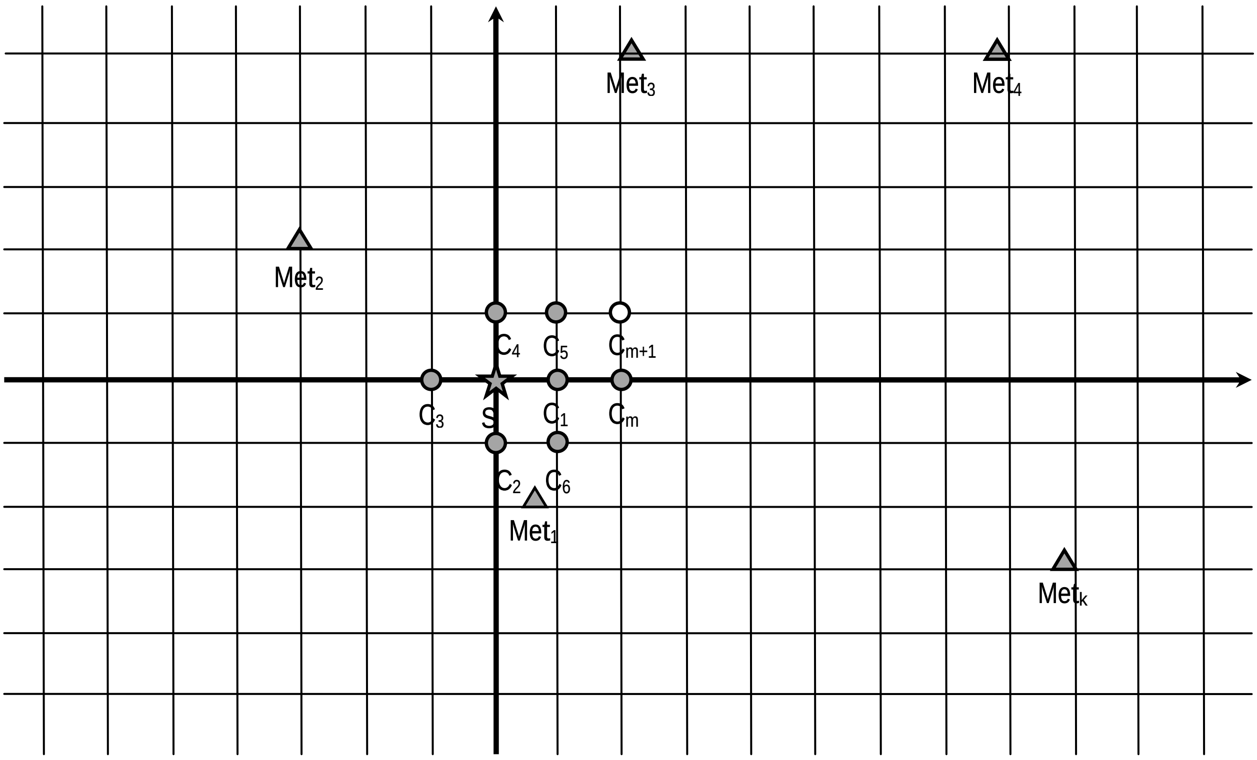


Fig. 1. Gas dispersion general scheme

– gas emission source;

– meteorological station;

– locations with determined concentration;

– location with undetermined concentration.

The algorithm of determining of gas concentration in any point of space has similar nature of graph breadth-first search algorithm, i.e.:

1. At first concentrations in the cells which are the nearest to the release source are determined.

2. A concentration in the cell is determined by using concentration data from neighbor cells and meteorological conditions interpolated from parameters measured in the weather stations (Fig.2). This step is performed by means of ANN.

The ANN's structure is multi-layer feed-forward perceptron. The learning process is implemented on data sampling which is received from different experiments. The measurements of meteorological parameters and gas concentration level are made in different places of the space when experiments are carried out [3].

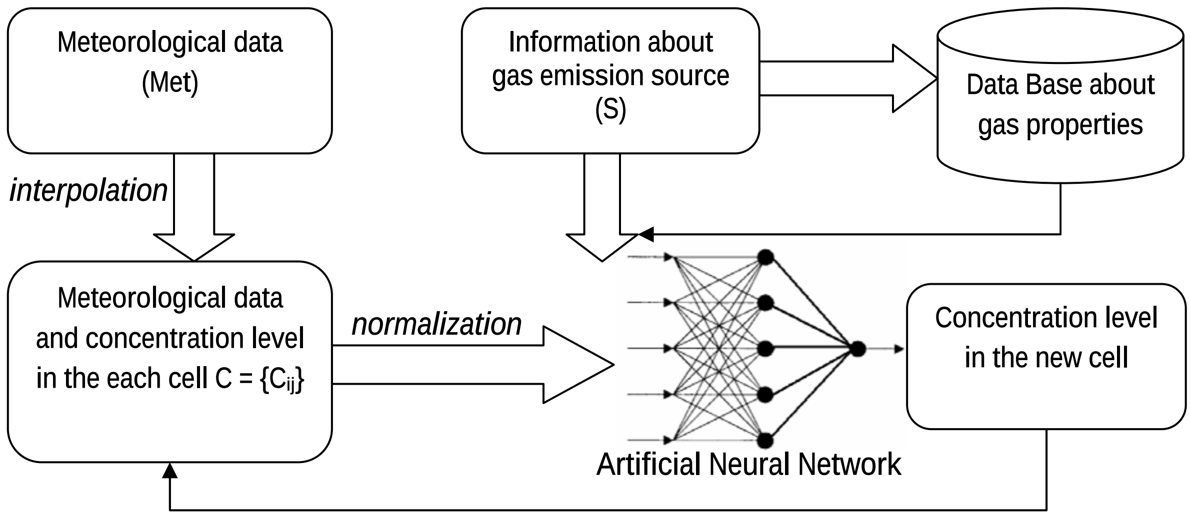


Fig. 2. General scheme of algorithm of computation of concentration in the cell

The advantages of proposed approach:

- accounting of many factors with different natures in one system;
- an ability to find decisions in nonlinear dependences.

The disadvantages of proposed approach:

- this approah needs a lot of time to analyse whole picture;
- there is an ANN's teaching problem: it is hard to receive precise information about different data.

## Proposed approach testing

Proposed approach was tested on base of data from open source files [5].The Kincaid field experiments were performed in 1980-1981. This investigation was a part of the EPRI Plume Model Validation and Development Project. A numerous reports described these experiments are comprehensive source of information about gas dispersion process. The Kincaid power plant is situated in Illinois, USA and is surrounded by flat farmland with some lakes. The power plant has a 187 m stack with a diameter of 9 m. During the experiment, $SF_6$-gas was released from the stack. The most meteorological measurements were taken from a "Central Site" located around 650 m east of the Kincaid plant. The radiosonde data supplied on diskette are routine data from the station Peoria, 120 km north of the source.

So as stated above it is necessary to represent research area in the grid format. This performance claims two stages:

1. To choose a step – grid cell size. In fact this is distance in kilometers.
2. To calculate average concentration in each cell.

A step parameter has been chosen according to the following rule: the amount of cells with nonzero concentration must be maximal and the amount of cells with zero concentration because this would make ANN teaching easier. As a result of the test carried out on base of data extracted on May 7, 1980 at 11 a.m. is Tab.1.

Tab. 1. The “useful” cells amount dependence on step parameter

| Step size, km | An amount of cells with average concentration $0<con_i<10$ | An amount of cells with average concentration $10<=con_i<100$ | An amount of cells with average concentration $100<=con_i<1000$ | Total amount of cells with nonzero average concentration |
|---|---|---|---|---|
| 1 | 7 | 9 | 9 | 25 |
| 0.5 | 6 | 13 | 15 | 34 |
| 0.1 | 7 | 19 | 21 | 47 |
| 0.05 | 7 | 19 | 21 | 47 |
| 0.01 | 7 | 19 | 21 | 47 |

So, optimal step size is 0.1 km.

After preliminary described above the set of data which specifies concentration distribution for each cell of investigated area at certain time is generated. The grid with 0.1 km step for situation on May 7, 1980 at 10 a.m. is depictured on Fig.3. This figure shows that cells with nonzero concentration is not situated consecutive but have disordered arrangement. Gas monitors were located rarely and this explains previous fact, but it doesn’t mean that unmeasured concentration is equal to zero. But lack of information about precise data doesn’t allow to use these cells for analysis. It is offered the following decision of this problem: not only neighbouring cells concentrations have to be included into the learning sample for ANN but also distances between cells and analysis cell (is bold encircled in Fig.3) have to be included too.

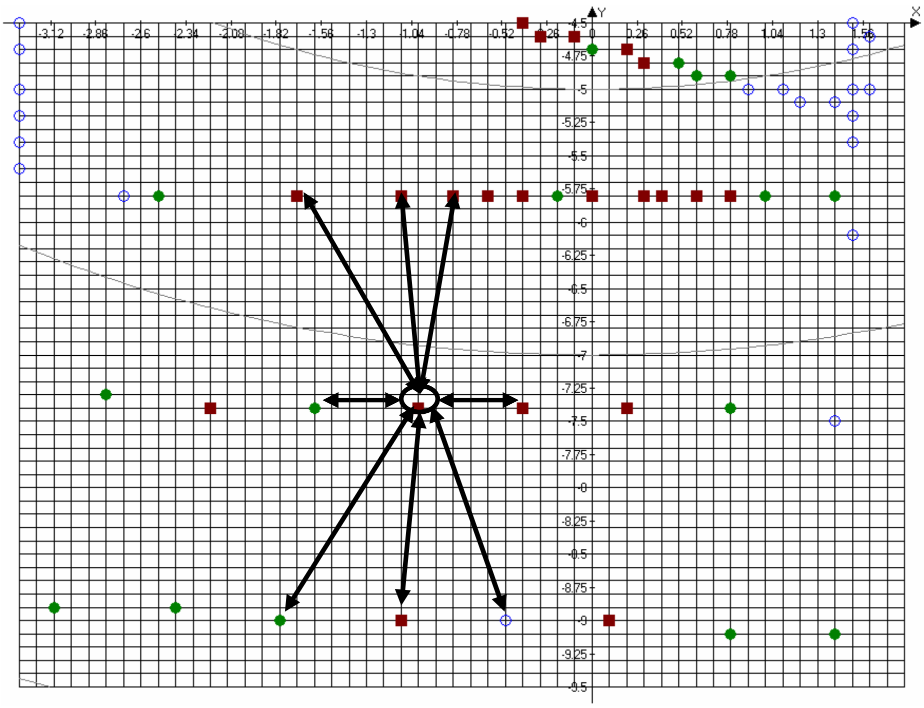


Fig.3. SF6 gas concentration distribution on May 7, 1980 at 10 a.m.

empty circles – cells with average concentration $0< con_j <10$;

filled circles – cells with average concentration $10<= con_j <100$;

filled squares – cells with average concentration $100<= con_j <1000$.

The analysis of eight nearest cells is performed for each cell taking into account distances. So, the learning sample LernSamp has following structure (6):

$$LearnSamp = \{learn_i\},\ i = 1..k, \quad (6)$$

where $learn_i$ – training pattern;

k – amount of training patterns.

$$learn_i = \{Con_i, Dist_i, WS_i, WD_i, Temp_i\}, \quad (7)$$

where $Con_i = \{con_j\}$ – concentration levels in the eight nearest cells, j=1..8;

$Dist_i = \{dist_j\}$ – distances between eight nearest cells and analysis cell, j=1..8; $WS_i, WD_i, Temp_i$ is described above (3).

As stated above ANN is multi-layer feed-forward perceptron. As a result of tests carried out network's structure was determined: first (input) layer consists of 20 neurons (19 of input data + 1 addition neuron); first hidden layer consists of 4 neurons, second hidden layer – 6 neurons. Output layer consists of single element which output result determines a grade of membership to classes with low (<10), middle (<100) or high (<1000) concentrations. Hyperbolic tangent function (8) is chosen to be activation function for ANN testing.

$$\varphi(v) = a \cdot th(b \cdot v), (a, b) > 0 \quad (8)$$

As a result of ANN's teaching the average learning error was minimized to 0.007 value (Fig.4).

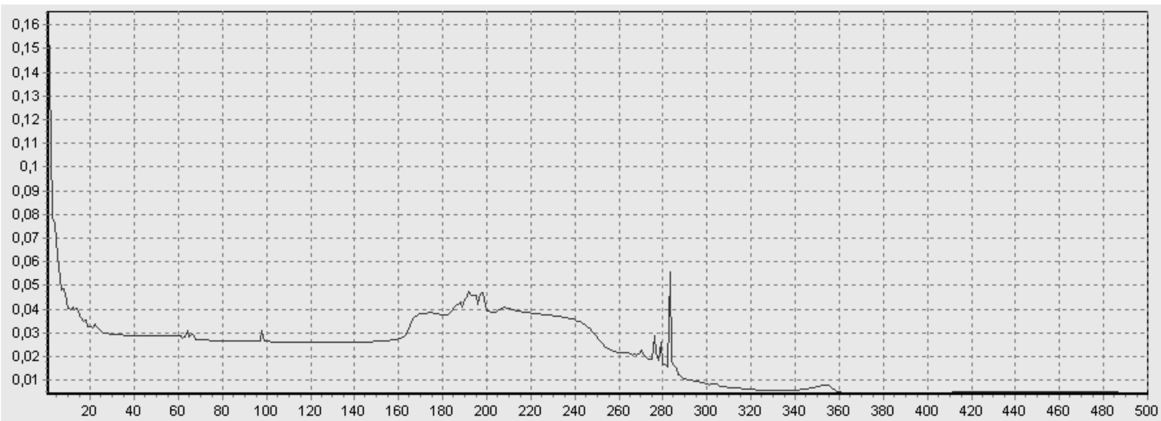


Fig. 4. The average learning error – learning epoch curve

## Conclusion

Proposed approach can be applied for gas dispersion modelling in case of accidental releases of hazardous pollutants. Unfortunately, there is no research works about modelling such situations in Ukraine. It leads to impossibility to estimate model efficiency in circumstances of certain climatic zone. Nevertheless estimations made on basis of data [3] confirm acceptability and applicability of chosen approach of concentration evaluation. For the further development of this approach it is necessary:

- to carry out additional research of dependence of effective ANN teaching from meteorological parameters;
- to give a qualitative and quantitative assessments of the algorithm by means of data received experimentally.

## Acknowledgements

The paper is published with financial support by the project ITHEA XXI of the Institute of Information Theories and Applications FOI ITHEA Bulgaria www.ithea.org and the Association of Developers and Users of Intelligent Systems ADUIS Ukraine www.aduis.com.ua.

## Bibliography

[1] Simon Haykin Neural Networks: A Comprehensive Foundation, 2nd edition – Upper Saddle River, NJ, USA: Prentice Hall PTR, 1998. – 842p.

[2] Sven-Erik Gryning, Ekaterina Batcharova Air pollution modeling and its application XIII – 2000. – 815p.

[3] Thompson, R. S. (1993): Building Amplification Factors for Sources Near Buildings – A Wind-Tunnel Study. Atmospheric Environment Part A – General Topics 27, 2313-2325.

[4] M.J. Borysiewicz, M.A. Borysiewicz Atmospheric dispersion modelling for emergency management CoE MANHAZ, Institute of Atomic Energy

[5] Olesen, H.R.2005: User's Guide to the Model Validation Kit. National Environmental Research Institute, Denmark. 72pp. – Research Notes from NERI no. 226.

## Authors' Information

*Valentina Davidenko –post-graduate student, National Aerospace University named after N.E. Zhukovsky "KhAI", 61070, Chkalova str. 17, Kharkov, Ukraine; e-mail: Valyuxa@ukr.net*

*Igor Shostak – professor in Software Engeneering depr., National Aerospace University named after N.E. Zhukovsky "KhAI", 61070, Chkalova str. 17, Kharkov, Ukraine; e-mail: iv_shostak@rambler.ru*