International Journal of Information Systems for Crisis Response and Management
Volume 9 • Issue 3 • July-September 2017

# Hybrid Unsupervised Modeling of Air Pollution Impact to Cardiovascular and Respiratory Diseases

Lazaros Iliadis, Civil Engineering Department, School of Engineering, Democritus University of Thrace, Xanthi, Greece

Vardis-Dimitris Anezakis, Department of Forestry, School of Agriculture and Forest Sciences, Democritus University of Thrace, Orestiada, Greece

Konstantinos Demertzis, Civil Engineering Department, School of Engineering, Democritus University of Thrace, Xanthi, Greece

Georgios Mallinis, Department of Forestry, School of Agriculture and Forest Sciences, Democritus University of Thrace, Orestiada, Greece

## ABSTRACT

During the last few decades, climate change has increased air pollutant concentrations with a direct and serious effect on population health in urban areas. This research introduces a hybrid computational intelligence approach, employing unsupervised machine learning (UML), in an effort to model the impact of extreme air pollutants on cardiovascular and respiratory diseases of citizens. The system is entitled Air Pollution Climate Change Cardiovascular and Respiratory (APCCCR) and it combines the fuzzy chi square test (FUCS) with the UML self organizing maps algorithm. A major innovation of the system is the determination of the direct impact of air pollution (or of the indirect impact of climate change) to the health of the people, in a comprehensive manner with the use of fuzzy linguistics. The system has been applied and tested thoroughly with spatiotemporal data for the Thessaloniki urban area for the period 2004-2013.

## KEYWORDS

Fuzzy Chi-Square Test, Fuzzy Logic, Meteorological Factors, Morbidity, Mortality, Photochemical Cloud, Self-Organizing Maps, Smog Cloud, Thessaloniki, Unsupervised Machine Learning

## INTRODUCTION

The increase of primary air pollutants (CO, NO, $NO_2$, $SO_2$) or secondary ones ($O_3$), has caused serious degradation in the quality of life of urban areas' residents. Moreover, changes in the heating methods of Greek houses due to the financial crisis, has influenced the concentration of Particulate Matter (PM) in the cities. Extended exposition of the urban population to high concentrations of pollution, increase the percentages of morbidity and mortality due to Cardiovascular (CARD) and Respiratory (RES) problems. Especially, people who live in areas with high levels of air pollution are phasing not only risks of cardiological and respiratory problems, but they are also risking narrowing the arteries and specifically the carotid one. This fact increases the possibility of stroke due to low levels of brain oxygen. Patients with severe disease history and young children or elder population (sensitive groups)

DOI: 10.4018/IJISCRAM.2017070102

are more vulnerable to atmospheric pollution (APO) and they should avoid transportation in the city during the days with high concentration of pollutants.

The combination of meteorological conditions and APO levels plays a major role for the determination of Morbidity and Mortality Risk index (MOMORI). The fluctuation of the MOMORI in an urban center, is mainly influenced by the meteorological conditions that favor the development of smog (CO, $SO_2$, $PM_x$) during the winter and the development of Photochemical Cloud (PHOC) in the summer ($NO_x$, $O_3$). The topography of an area, the hours of the population's employment in external activities and the percent of the people who live near industrial zones are crucial for the level of risk estimation.

The analysis and continuous monitoring of the APO levels as well as timely forecasting of the conditions that can cause high concentrations, result to the impose of preemptive actions and thus to an effective management of the problem and to the estimation of the impact on related diseases.

This research paper presents the *Air Pollution and Climate Change Cardiovascular and Respiratory Modeling* (APCCCR) hybrid intelligent system. APCCCR considers the effect of atmospheric pollution parameters to Cardiovascular-Respiratory hospitalization incidents and it determines the interdependencies between them in the wider Thessaloniki urban area.

The system is developed in two discrete phases. The first one uses the UML *Self Organizing Maps* algorithm (SOM) to cluster the values of the involved features and to determine the meteorological values that directly affect the pollutants' concentrations which have a serious effect on the considered diseases. The second phase uses the Fuzzy Chi-Square Test (FUCS) to determine the interdependency between the parameters in a rational and comprehensive mode by using proper Linguistics. This is achieved by fuzzifying the P-Value of the Chi-Square test. This process produces Linguistics that express Low, Medium or High dependency by employing fuzzy Membership functions (FMF).

The FUCS application is performed for each cluster in order to determine which atmospheric parameters determine the level of the hospital treatment incidents in the prefecture of Thessaloniki. The testing of the APCCCR was based on a comparative performance analysis between four UML algorithms namely: Self-Organizing Maps, Expectation Maximization, Sequential Information Bottleneck, and Simple K-Means.

Wide use of this approach can enforce the mechanisms of civil protection authority by acting as a means of warning the public hospitals regarding the days of bad meteorological conditions that favor high pollutants' concentrations.

## Literature Review-Related Work

To the best of our knowledge, the Sequential Information Bottleneck algorithm has not been used in modeling and assessment of environmental risks. The Expectation Maximization algorithm, Self-Organizing Maps and Simple k-Means have been used for the classification of meteorological and air quality data. The following lines present some cases of classification met in the literature.

(Hernawati, Insani, Bambang, Nur Hadi, & Sahid, 2017) used an unsupervised SOM approach. They considered data related directly or indirectly to pollution (e.g. demographic and social data, air pollution water and soil pollution levels) as well as the geographical situation of each province.

(Štrbová, Štrba, Raclavská, & Bilek, 2018) used SOM to find association between PM concentrations, elevation, selected meteorological variables, and GPS location coordinates.

(Cortina-Januchs, Quintanilla-Dominguez, Andina, & Vega-Corona, 2012) used a Multilayer Perceptron Neural Network (MPNN) to make the prediction of pollutant concentrations for the next hour. A database used to train the ANN based on historical time series of meteorological variables and air pollutant concentrations of $SO_2$. Before the prediction, Fuzzy C-Means (FCM) and k-Means Clustering (k-MC) algorithms were employed in order to find relationship among pollutant and meteorological variables.

The EM, SOM and SKM Algorithms have been used in the literature to cluster and correlate the cardiovascular and respiratory (CARE) health problems with air pollution.

# The Past as the Future of Emergency Preparedness and Management

Murray Turoff, Starr Roxanne Hiltz, Connie White, Linda Plotnick, Art Hendelaand Xiang Yao (2011). *Crisis Response and Management and Emerging Information Systems: Critical Applications (pp. 12-28).*
www.irma-international.org/chapter/past-future-emergency-preparedness-management/53984