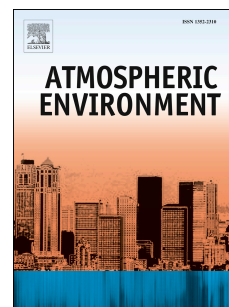


# Accepted Manuscript

A novel regression imputation framework for Tehran air pollution monitoring network using outputs from WRF and CAMx models

Hossein Shahbazi, Sajjad Karimi, Vahid Hosseini, Daniel Yazgi, Sara Torbatian



PII: S1352-2310(18)30365-0

DOI: [10.1016/j.atmosenv.2018.05.055](https://doi.org/10.1016/j.atmosenv.2018.05.055)

Reference: AEA 16046

To appear in: *Atmospheric Environment*

Received Date: 27 September 2017

Revised Date: 22 May 2018

Accepted Date: 26 May 2018

Please cite this article as: Shahbazi, H., Karimi, S., Hosseini, V., Yazgi, D., Torbatian, S., A novel regression imputation framework for Tehran air pollution monitoring network using outputs from WRF and CAMx models, *Atmospheric Environment* (2018), doi: 10.1016/j.atmosenv.2018.05.055.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## A novel regression imputation framework for Tehran air pollution monitoring network using outputs from WRF and CAMx models

Hossein Shahbazi<sup>1</sup>, Sajjad Karimi<sup>2</sup>, Vahid Hosseini<sup>1</sup>, Daniel Yazgi<sup>3</sup> and Sara Torbatian<sup>4</sup>

<sup>1</sup> Department of Mechanical Engineering, Sharif University of Technology

<sup>2</sup> Department of Electrical Engineering, Sharif University of Technology

<sup>3</sup> Institute of Geophysics, University of Tehran

<sup>4</sup> Tehran Air Quality Control Company

### Abstract

Missing or incomplete data in short or long intervals is a common problem in measuring air pollution. Severe issues may arise when dealing with missing data for time-series prediction schemes or mean analysis. This study aimed to develop a new regression imputation framework to impute missing values in the hourly air quality data set of Tehran and enhance the applicability of Tehran Air Pollution Forecasting System (TAPFS). The proposed framework was designed based on three types of features including measurements of other stations, WRF and CAMx physical models. In this framework, elastic net and neuro-fuzzy networks were efficiently combined in a two-layer structure. The framework was applied on Tehran's air pollution monitoring network. The hourly imputing results of the suggested method were seen to be superior to existing methods according to statistical criteria such as *RMSE*, *MAE* and *R*-values. Average *R*-values of 0.88, 0.73, 0.76 and 0.81 were obtained for *O<sub>3</sub>*, *NO*, *PM<sub>2.5</sub>* and *PM<sub>10</sub>*, respectively. The measurements of other stations had the main predictive power with a modest increase for the two physical models. The benefit of the models was somewhat higher for stations on boundaries of monitoring network. In addition, the central stations had better performance than the boundary stations and an approximately 0.05 increase was obtained in average *R*-value.

**Key words**— urban areas, regression imputation, elastic networks, ANFIS, air pollutants, concentration retrieval, Tehran Air Pollution Forecasting System.

### 1 Introduction

Air pollution is a major environmental and human health concern worldwide, in particular in cities where pollutant sources are concentrated [1]. Apart from its major environmental risks to human health, it has adverse impacts on the ecosystem. The use of fossil fuels in most industries, transport and energy production can be considered as the major source of atmospheric pollution [2]. One of the primary tools to assess air-pollution patterns is through continuous monitoring of ambient pollution levels. To accomplish this, numerous physiochemical methods have been developed and air quality monitoring (AQM) station networks have been established worldwide [3]. Missing or incomplete data is a common problem in AQM pollutant measurements [1]. Data can be missing in long chunks due to a critical failure or in short intervals due to, for example, calibration, maintenance or temporary power outages. Major causes for missing air pollutant data include monitor fluctuation and errors, power outage, computer system crashes and filter changes [2].

Aside from the reasons associated with missing data, discontinuities pose a significant obstacle for time-series prediction schemes, which generally require continuous data as a condition for their use. Incomplete data may lead to three major issues. The first is data loss, which consequently leads to loss of efficiency. Second, there are several complications related to data handling, computation and analysis, due to the irregularities in data structure and impossibility of using standard softwares. Third, and most importantly, the results may be biased due to systematic differences between

Download English Version:

<https://daneshyari.com/en/article/8863594>

Download Persian Version:

<https://daneshyari.com/article/8863594>

[Daneshyari.com](https://daneshyari.com)