Abstract. Air quality management in urban areas requires reliable and accurate computational methods for the forecasting of the concentration levels of pollutants. The Common Air Quality Index (CAQI) has been proposed by the European Environment Agency for assessing air quality in a harmonized way. We have evaluated the values of this index in Thessaloniki, Greece, in 2001-2003, using a wide range of Computational Intelligence (CI) models. We have applied Artificial Neural Networks (ANN) and Decision Trees (DT) for the forecasting of the CAQI, and we compared the results with those obtained via statistical regression models. An extensive number of computational experiments were performed, in which we evaluated the influences of (i) different model architectures, (ii) various input datasets, and (iii) the training methods of these models. Model sensitivity was analyzed, in terms of various modelling options. In total, the performance of 1118 different modelling options was investigated. The best of the evaluated models performed well in forecasting both the hourly and daily values of the CAQI. The use of model
ensembles (based on the same algorithms and structures), obtained by a k-fold cross-validation, resulted in more accurate forecasts than using individual models. However, the performance of various ANN and DT models was found to be dependent both on their internal structure and on the methods used for their training. The presented results are expected to be useful in developing and implementing operational air quality management and forecasting systems.

**Keywords:** air pollution, air quality index, artificial neural networks, decision trees, cross-validation, sensitivity analysis

1. **Introduction**

High levels of air pollutant concentrations are associated with a number of health problems and affect quality of life [1]. The legal and regulatory framework in Europe and elsewhere calls for the continuous monitoring of the concentrations of air pollutants, while it also dictates that authorities should be able to inform citizens about air quality for the next day. Two categories of forecasting models have been developed: (i) Deterministic numerical atmospheric dispersion models that are capable of simulating the physical and chemical processes affecting air pollution. They have the advantage of fully covering a geographical area of interest, but they require detailed input in terms of emission data, meteorology and physicochemical parameters, while they are computationally demanding. (ii) The second forecasting approach makes use of data-driven modelling, employing either statistical [2] or computational intelligence (CI) methods [3], [4]. However, very few studies have quantitatively inter-compared the performance of ANN models and deterministic modelling systems [5].

Due to the complicated nature of air pollution problems and the differences posed by local conditions (meteorology, emissions, urban morphology etc), the provision of generally understandable information on air quality is not straightforward. In order to inform the public as well as policy makers on air quality levels and their potential health effects, various air quality indices have been proposed. We make use of the Common Air Quality Index (CAQI) suggested by the European Environment Agency (EEA), as the parameter of interest that needs to be forecasted in order to provide with health related warnings and information to the general public. We have constructed
various operational forecasting models, with the aid of Computational Intelligence (CI) methods that include Artificial Neural Networks (ANNs) and Decision Trees (DT). In addition, we have employed cross-validation and sensitivity analysis in order to investigate the models’ performance, to provide with reliable and accurate results and finally to increase the forecasting performance in comparison to our previous study [6].

The aims of this study are (i) to evaluate the performance of various CI and statistical regression models in forecasting the CAQI, and (ii) to analyze the sensitivity of such models in terms of the model input data, the methods of training of the CI models, and the architecture of such models. The study material has been collected from the city of Thessaloniki in Greece, from 2001 to 2003.

2. Data and Methodology

2.1 Area of interest and datasets used

Thessaloniki is the second largest city of Greece, and the capital of the administrative Region of Central Macedonia. The greater urban area has a population of over 1,000,000 people with more than 400,000 vehicles being used daily. The city is characterized by high air pollution levels, especially in terms of Particulate Matter of mean aerodynamic diameter of 10 micrometers or less (respirable particles -PM$_{10}$, [7]). In the current study we use air quality concentrations as well as meteorological data resulting from the operation of relevant monitoring stations in various locations of the Greater Thessaloniki Area-GTA (www.airthess.gr). These data come from four monitoring stations (Figure 1) which are situated in the following areas-locations:

a) Agia Sofia (city centre area, characterized by traffic and mixed household – commercial activities).

b) Panorama (a suburb in the hilly area of the city, mostly used as a residential area).

c) Sindos (industrial-suburban area in the west of the city influenced by traffic).

d) Kordelio (a densely populated area close to the industrial area of the city, influenced by traffic).
Figure 1. The air quality monitoring stations in Thessaloniki that were used in this study. A = Agia Sofia, B = Panorama, C = Sindos, D = Kordelio.

These data include hourly concentration measurements of the following air pollutants: carbon monoxide (CO), nitrogen dioxide (NO₂), nitrogen monoxide (NO), ozone (O₃), respirable particles (PM₁₀), and sulphur dioxide (SO₂) and they are complemented by meteorological observations. Data correspond to the years 2001-2003. This time period was selected on the basis of its low percentage of problematic or missing data and for comparability with our previous study [6].

2.2 The Air Quality Index (CAQI)

The use of environmental indices comes as a result of scientific initiatives that combine the impacts of various pollutants in order to come up with an aggregated measure, that may be used for the description of the environmental (here atmospheric) quality [8]. Various air quality indices have been suggested by institutes and authorities in various countries [9], [10]. The EEA has proposed the Common Air Quality Index (CAQI, www.eyeonearth.org) to identify the atmospheric quality in respect of public safety and health.

The CAQI is defined in a way that allows for the calculation and the comparison of air quality levels on an hourly or daily basis. The CAQI has five categorical levels, corresponding to a range of values starting from 0 (very low) to >100 (very high), as presented in Figure 2. Two types of a CAQI are specified, an urban background index and a traffic (or roadside) related index, in order to better represent areas not directly
affected by traffic (the former) as well as areas mostly affected by traffic (the latter).

The urban background index takes into account three so-called main pollutants (NO₂, PM₁₀ and O₃) and two auxiliary pollutants (SO₂ and CO) while the traffic index takes into account two main pollutants (NO₂ and PM₁₀) and one auxiliary (CO). The use of the auxiliary pollutants in the calculation of the CAQI is decided on the basis of the availability of data. The scientific background of the CAQI as well as the way to calculate it is described in detail in [11].

The calculation method is defined in Table 1, where all concentrations are in μg/m³, except of CO where the 8-hour moving average value is used.

![Figure 2. The CAQI five categorical levels.](image)

**Table 1.** The two types of CAQI (a background index and a traffic index) proposed by the EEA [11].

<table>
<thead>
<tr>
<th>CAQI Type</th>
<th>Index Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>SO₂ Index</td>
<td></td>
</tr>
</tbody>
</table>
\[
SO₂ \text{ Index} (\{SO₂\}) = \begin{cases} 
\frac{[SO₂]}{2} & 0 < [SO₂] < 100 \\
\frac{[SO₂]}{8} + 37.5 & 100 < [SO₂] < 500 \\
[SO₂] & [SO₂] > 500 
\end{cases}
\] |
| NO₂ Index | 
\[
NO₂ \text{ Index} (\{NO₂\}) = \begin{cases} 
\frac{[NO₂]}{2} & 0 < [NO₂] < 100 \\
\frac{[NO₂]}{4} + 25 & 100 < [NO₂] < 200 \\
\frac{[NO₂]}{8} + 50 & 200 < [NO₂] < 400 \\
\frac{[NO₂]}{4} & [NO₂] > 400 
\end{cases}
\] |
| O₃ Index | \(O₃ \text{ Index}(\{O₃\}) = ([O₃]\)^*5)/12\) |
| PM₁₀ Index | \(\text{PM}_{10} \text{ Index}(\{PM_{10}\}) = PM_{10}\) |
| CO Index | 
\[
CO \text{ Index} (\{CO\}) = \begin{cases} 
\frac{[CO]}{200} & 0 < [CO] < 5000 \\
\frac{[CO]}{100} - 25 & 5000 < [CO] < 10000 \\
\frac{[CO]}{400} + 50 & 10000 < [CO] < 20000 \\
\frac{[CO]}{200} & [CO] > 20000 
\end{cases}
\] |
Main Background Index = Max(NO\textsubscript{2}_Index, PM\textsubscript{10}_Index, O\textsubscript{3}_Index) \hspace{1cm} (6)

Auxiliary Background Index = Max(NO\textsubscript{2}_Index, PM\textsubscript{10}_Index, O\textsubscript{3}_Index, SO\textsubscript{2}_Index, CO\_Index) \hspace{1cm} (7)

Main Traffic Index = Max(NO\textsubscript{2}_Index, PM\textsubscript{10}_Index) \hspace{1cm} (8)

Auxiliary Traffic Index = Max(NO\textsubscript{2}_Index, PM\textsubscript{10}_Index, CO\_Index) \hspace{1cm} (9)

This CAQI has been computed for each of the four locations of interest, both using hourly and daily concentration values. The traffic CAQI was applied for Agia Sofia, Kordelio and Sindos, while the urban background CAQI for the Panorama station. A summary of the computed values is been presented in the results and discussions section (Table 2). The CAQI categorical levels have been defined in Figure 2.

**Table 2.** The number (and the corresponding percentage) of hourly and daily CAQI values, for each CAQI level and station.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hourly</td>
<td>Daily</td>
<td>Hourly</td>
<td>Daily</td>
</tr>
<tr>
<td>Very</td>
<td>3510</td>
<td>127</td>
<td>591</td>
<td>0</td>
</tr>
<tr>
<td>high</td>
<td>(14%)</td>
<td>(12%)</td>
<td>(7%)</td>
<td>(0%)</td>
</tr>
<tr>
<td>High</td>
<td>3706</td>
<td>168</td>
<td>618</td>
<td>0</td>
</tr>
<tr>
<td>(15%)</td>
<td>(16%)</td>
<td>(8%)</td>
<td>(0%)</td>
<td>(13%)</td>
</tr>
<tr>
<td>Medium</td>
<td>7414</td>
<td>438</td>
<td>1917</td>
<td>3</td>
</tr>
<tr>
<td>(30%)</td>
<td>(42%)</td>
<td>(24%)</td>
<td>(1%)</td>
<td>(26%)</td>
</tr>
<tr>
<td>Low</td>
<td>8264</td>
<td>304</td>
<td>3558</td>
<td>361</td>
</tr>
<tr>
<td>(34%)</td>
<td>(29%)</td>
<td>(45%)</td>
<td>(99%)</td>
<td>(34%)</td>
</tr>
<tr>
<td>Very</td>
<td>1666</td>
<td>18</td>
<td>1309</td>
<td>1</td>
</tr>
<tr>
<td>low</td>
<td>(7%)</td>
<td>(2%)</td>
<td>(16%)</td>
<td>(0%)</td>
</tr>
</tbody>
</table>

2.3 Methods employed in data analysis and for the forecasting of the CAQI

On the basis of our previous work on the same dataset [6], we make use of CI methods for the analysis of the data and for the forecasting of the CAQI. The methods employed are Artificial Neural Networks and Decision Trees.
ANNs are a family of algorithms that were advanced in the late ’80s, popularizing techniques like Multi-layer Perceptrons (MLP) [12] and Self-Organizing Maps (SOM) [13], which can be trained to successfully approximate virtually any continuous function [14]. Their advantages also include greater fault tolerance, robustness, and adaptability, especially compared to expert systems, due to the large number of interconnected processing elements that can be trained to learn from new patterns [15]. These features provide ANNs with the potential to model complex non-linear phenomena, like air pollution [16].

A Decision Tree is a hierarchical data structure implementing the divide-and-conquer strategy [17]. It is an efficient nonparametric method, which can be used for both classification and regression. DTs are essentially a map of the reasoning process in which a tree-like graph is constructed to explore options and investigate the possible outcomes of choosing the options. The reasoning process starts from a root node, transverses along the branches tagged with decision nodes, and terminates in a leaf node on the basis of criteria like Information Gain [18]. In this paper we used an ensemble of DTs, by employing the Bootstrap aggregation (bagging) meta-algorithm, in order to improve the performance of the models [19]. An ensemble aims at leveraging the performance of a set of models to achieve better prediction accuracy than that of the individual models. For this purpose, every model (tree) in the ensemble is developed by using an independently drawn bootstrap replica of the input data. This means that from the initial training set (consisting of $n$ observations), a number of $m$ new training sets (replicas) is created by randomly sampling the initial training set with replacement (this is called "bagging"). As a result, some observations are sampled (and thus included in the replicas) more than once, and others may not be selected at all. On average, about 63% of the rows of the initial data set will be included in the replica. The remaining 37% of the rows are called the “out of the bag” data (not included in this replica). In our paper, a DT model is developed on the basis of each one of the replicas, and the overall prediction performance is estimated on the basis of an ensemble of DTs.

The above two CI methods are used in this paper and their comparison and analysis also included Linear Regression (LR) models that were developed for the same dataset in our previous paper [6]. LR models may take into account the persistence of
the air quality system, and have been proven successful especially if their input data includes lagged values, i.e., values determined before the time in which the forecast is made (this time is denoted by $T$ in the following paragraph) [7].

In order to forecast hourly and daily CAQIs, three input datasets were used for the models, as follows:

- **Dataset 1** (lagged index values) includes only hourly or daily lagged CAQI values determined during the previous day (day $T-1$, $T-2$, … , $T-10$), or the previous hours (hours $T-1$, $T-2$, … , $T-10$). The CAQI values were thus calculated with the aid of one up to 10 lagged CAQI values, leading to 10 different model results for hourly and daily values. Dataset 1 consists of 1042 daily records and 23671 hourly records.

- **Dataset 2** (concentration and meteorological values at time $T$) includes concentration values of CO, SO$_2$, and O$_3$ and six meteorological values, as follows: (1) temperature, (2) dew point, (3) humidity, (4) sea level pressure, (5) wind direction and (6) wind speed. Dataset 2 consists of 1043 daily records and 23672 hourly records (plus one record in comparison to Dataset 1, as the latter includes only lagged values).

- **Dataset 3** (lagged index values and dataset 2) includes hourly or daily lagged (previous) CAQI values determined during the previous day (day $T-1$, $T-2$, … , $T-5$), or the previous hours (hours $T-1$, $T-2$, … , $T-5$), separately for the two types of index, and the above mentioned air quality and meteorological values. The CAQI values were thus calculated with the aid of one up to 5 lagged CAQI values, leading to 5 different model results for hourly and daily values. Dataset 3 consists of 1042 daily records and 23671 hourly records.

Input data records for each hour or day were pre-processed for the use in the forecasting algorithms (ANNs and DTs). The first step of the pre-processing procedure was to select all records that include all the required input values.

The next step was to use an interpolation method to calculate all missing values [20]. In addition, the wind direction values were encoded by using the formula $WD=1+sin(\theta+\pi/4)$, in order to replace the cyclic nature of this variable with a linear one. Moreover, as ANNs require numerical input, nominal values were converted to
numerical values for the CAQI, as described in Table 3. The specific numerical values were selected, because of the hyperbolic tangent sigmoid transfer function \((htstf)\), which was used to normalize the input data in the case of ANNs and DTs. This function receives values between -1 and 1, and because we have to cover the full range of these values, our mapping of the 5 nominal values is done by assigning to the lower and higher nominal values the relevant lower and higher values of the \(htstf\), and by assigning equidistant numerical values (i.e. -0.5, 0, and 0.5) to the three other nominal values of the CAQI. Table 4 presents the way that the continuous numerical values produced as model outputs (forecasts) are mapped back to nominal CAQI values. For this mapping the numerical values were divided in five equal in range groups (Figure 3), again because we have to be able to properly map back any numerical value to the nominal CAQI value.

**Table 3.** Conversion table from CAQI levels (nominal values) to CAQI numerical values.

<table>
<thead>
<tr>
<th>CAQI Level Nominal Value</th>
<th>Numerical Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Low</td>
<td>-1</td>
</tr>
<tr>
<td>Low</td>
<td>-0.5</td>
</tr>
<tr>
<td>Medium</td>
<td>0</td>
</tr>
<tr>
<td>High</td>
<td>0.5</td>
</tr>
<tr>
<td>Very High</td>
<td>1</td>
</tr>
</tbody>
</table>

**Table 4.** Conversion table from CAQI numerical values to CAQI levels (nominal values).

<table>
<thead>
<tr>
<th>Numerical Value ((n))</th>
<th>CAQI Level Nominal Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n\leq-0.6)</td>
<td>Very Low</td>
</tr>
<tr>
<td>(n&gt;-0.6 \text{ AND } n\leq-0.2)</td>
<td>Low</td>
</tr>
<tr>
<td>(n&gt;-0.2 \text{ AND } n&lt;0.2)</td>
<td>Medium</td>
</tr>
<tr>
<td>(n\geq0.2 \text{ AND } n&lt;0.6)</td>
<td>High</td>
</tr>
<tr>
<td>(n\geq0.6)</td>
<td>Very High</td>
</tr>
</tbody>
</table>
**Figure 3.** The mapping of the continuous CAQI numerical values to nominal CAQI values.

The performance of the models is estimated with the aid of statistical indicators like the Index of Agreement (IA or d) [21], the Cohen’s Kappa and the Percentage of Agreement (PA), which are defined as follows:

\[
IA = 1 - \frac{\sum_{i=1}^{n} |P_i - O_i|^2}{\sum_{i=1}^{n} (P_i - \bar{O})^2 + (O_i - \bar{O})^2}
\]  

(1)

where \(P_i\) and \(O_i\) are the predicted and observed concentration values respectively. The index of agreement is used for the estimation of the accuracy of models for the forecasting of CAQI values, and ranges between 0 and 1, where the value of 1.0 corresponds to a perfect agreement. Actually, an IA equal to 0.4 represents a random correlation between two sets of values having the same overall range [22].

The Cohen’s Kappa or Critical Success Index (CSI) verifies how well the events of interest (CAQI levels) are predicted and it is unaffected by the number of correct forecasts. Its calculation is based on the Eq. 2, where \(a, b, c\) and \(d\) are calculated as shown in the Table 5. Positives or negatives refer to the occurrence or not of the event of interest (i.e correct classification of the CAQI), respectively. The CSI applied in this study does not use the "d" events (true negatives). As a consequence, the CSI is tending to give poorer scores for rare events, but it takes into account both false negative events and false positive events. Therefore, the CSI can be characterized as a more balanced score [23].
The PA is used in order to calculate the percentage of the time that the forecasted CAQI levels were identical with the observed CAQI levels and it is expressed as a numerical value ranging from 0% up to 100%.

### Table 5. The meaning of parameters $a$, $b$, $c$ and $d$ of the CSI formula.

<table>
<thead>
<tr>
<th>Event forecasted</th>
<th>Event Observed</th>
<th>a: true positives</th>
<th>b: false positives</th>
<th>c: false negatives</th>
<th>d: true negatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### 2.4 Cross-Validation

Cross Validation (CV) is a popular method applied in order to measure the predictive performance of a statistical model [24]. In CV, the available dataset is divided into two segments, one is used to teach or train a model and the other is used to validate the performance of the model. There are several types of cross-validation methods like the Holdout method, k-fold cross-validation, and leave-one-out cross-validation [25].

In this study the k-fold CV is used to measure the predictive performance of the models being developed. K-fold CV divides the dataset into k (equal) subsets. Each time, one of the k subsets is used as the test set (i.e. the set against which the model performance is estimated) and the other k-1 subsets are used as a training set. The k results from the folds (k times) are averaged for producing a single estimation. In this way an ensemble of models is actually produced, as the same ANN is employed in every training phase, but its weights are calculated separately each time. The advantage of this method is that the results are not influenced by the way the data are divided. In this study we use the 10-fold cross-validation for the development of ANN models, as it is the one most commonly applied [26], and proved to provide with better results than other similar techniques like bootstrapping [26, 27].
For the application of the ANN backpropagation algorithm, we made use of the following procedure: For each one of the input datasets, the 10-fold CV method was employed. Every dataset was thus divided into 10 subsets, each one of the 10-1 subsets was used for training, and the last subset was used for validation and testing (actually half was used for the validation and half for the final testing and evaluation). The error on the validation subset was monitored during the training process: when the network begins to overfit the data, the error on the validation subset typically begins to rise [28]. On this basis, when the validation error increases, training can be stopped, and the weights and biases at the minimum of the validation error are retained. The error of the test subset was not used during the training, but was used for comparing the performance of different models. Thus, if the error of the test subset reached a minimum at a significantly different iteration number than the validation set error, this was considered as an indication of poor division of the data set and the procedure was repeated.

In the case of the DT models, we used the 10-fold cross validation also, where the 10-1 folders were involved in the training phase and the remaining folder was used for testing. We used the bootstrap aggregation meta-algorithm for each input subset of data resulting by the 10-fold cross validation, in order to grow different number of trees.

2.5 Sensitivity Analysis

The model predictions can be highly dependent upon the structure of the model, as well as uncertainties in the input data. Such model behaviour can be investigated through sensitivity analysis, which seeks to determine the variation in model output as a function of variations in input variables and parameters (i.e., forward sensitivity analysis). Sensitivity analysis could provide information on the input factors that are mainly responsible for the output uncertainties and thus for the model performance [29].

The simplest form of (forward) sensitivity analysis is to vary one input or parameter value in the model by a given amount, and examine the impact on the predictions. This is known as one-way sensitivity analysis (or one-at-a-time), since only one
A comprehensive analysis of the neural network system should include a sensitivity analysis to determine whether and under what circumstances the combination of a specific ANN and of the actual training data may result in an ill-conditioned system that will require further refinement [30]. In the current paper, we used the one-way sensitivity analysis in order to examine (a) different ANN architectures and (b) different numbers of decision trees grown independently. In our previous study [6], we did not examine alternative architectures, but we used only one type of ANN architecture that included one hidden layer, where the number of neurons was the same as the number of the input parameters, so we now wanted to further investigate the topology of the model and improve its performance.

In the first case (a), we examined two important topology parameters: the increase of the hidden layers and the increase of the neurons per layer. We used two different empirical methods in order to generate the different architectures (neurons per layer), having as a “starting point” the number of the input parameters (N-Input). We used the number of the hidden layers as one of the parameters of the ANN architecture, and the number of neurons per hidden layer as an additional parameter. Thus, in Method 1, the number of neurons was the same in all hidden layers. In Method 2, the number of neurons was different for each hidden layer and more specifically it increased as the number of hidden layers also increased. Both methods were designed in order to generate multiple architectures, where they will be different in one parameter at a time. The way that the number of neurons was decided per method, is explained with the aid of Eqs. 3 and 4.

\[
\text{Method 1: } \begin{align*}
\text{Number of neurons} & = N-\text{Input} + ((N-\text{Input}) \div 2) \times \text{Neuron Multiplier} \\
(\text{same for all hidden layers}) & \\
\text{Method 2: } & = N-\text{Input} + ((N-\text{Input}) \div 2) \times \text{Hidden Layer Number} \times \text{Neuron Multiplier}
\end{align*}
\]
The “Neuron Multiplier” was an additional parameter that was used in order to increase the number of neurons and to thus generate different ANN architectures. The “Hidden Layer Number” was used in order to increase the number of neurons per hidden layer. For example, the number of neurons of the third hidden layer when Method 2 is used with five (5) input parameters and with Neuron Multiplier equalled two (2), should be: \(5 + \left(\frac{5}{2} \times 3 \times 2\right) = 20\) neurons. It should be noted that the calculated number of neurons (by using Method 1 and 2) was rounded up.

By combining the results of the two methods, with maximum three hidden layers and with a Neuron Multiplier ranging from 2 to 4 (increased by one), we came up with 15 different ANN architectures per input parameter set (N-input) that were generated in order to perform the sensitivity analysis. Table 6 summarizes the construction parameters used for the 15 different architectures, while Table 7 presents an example of the different architectures generated on the basis of 5 input parameters, with maximum three hidden layers and with a neuron multiplier ranging from 2 to 4 (increased by one). The hidden layer columns in Table 7 report the number of neurons per layer.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>From</th>
<th>To</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Hidden Layers</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Neuron Multiplier</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Methods</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 6. Construction of the 15 different architectures.

<table>
<thead>
<tr>
<th>Architecture ID</th>
<th>N-Input(s)</th>
<th>Neuron Multiplier</th>
<th>Hidden Layer 1</th>
<th>Hidden Layer 2</th>
<th>Hidden Layer 3</th>
<th>Generated by Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>-</td>
<td>10</td>
<td>-</td>
<td>-</td>
<td>1 and 2</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>15</td>
<td>15</td>
<td>20</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>13</td>
<td>13</td>
<td>-</td>
<td>-</td>
<td>1 and 2</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>13</td>
<td>13</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 7. An example of the 15 different architectures with 5-input (parameters).
In the second category of computational algorithms (b), four different numbers of decision trees were tested (50, 100, 150 and 200). The initial number of trees (50) was chosen on the basis of previously published studies [31]. For both model types, we did not investigate any additional number of model topologies (ANN architectures and number of DTs), due to the requirements in computational time and physical memory of the machine (RAM). As an indication of the latter, it should be noted that the computational (CPU) time required for this study was approx. 64 days while the computational experiments required at least 4GB of RAM in order to be completed (for a machine with Intel(R) i3 Core (2.1 GHz) processor and 4GB of RAM).

3. Results and discussion

We have made use of the data and methods described in chapter 2, and we have firstly identified the CAQI value ranges per station (Table 2). It is therefore evident that Agia Sofia is the station in the city centre that demonstrates the highest percentage of values in the CAQI classes “High” and “Very High” [6]. For this reason, this station location was selected as the one for the development of CAQI forecasting models. Thus, air quality and meteorological data used for ANNs and DT models described hereafter, originate from this location and are compared to the results produced by our previous study [6].

3.1 CAQI Forecasting with the aid of Regression Models

In our previous work we used Linear Regression Models (LRM) [6], in order to forecast the CAQI values and the CAQI levels, and we briefly report here on the findings, for reasons of comparability. The current study focused on the sensitivity
analysis in the forecasting procedure, by using ANNs and DTs, in order to improve their performance.

3.1.1 Forecasting of the numerical hourly and daily CAQI values

We used the LRM as the reference method for the development of the CAQI forecasting models. For this reason lagged values of the CAQI (a total of up to 10 values, Dataset 1) were tested. Results indicate that the hourly CAQI numerical values for Agia Sofia can be forecasted by a LRM when employing only one lagged hourly value (IA = 0.92). The daily CAQI numerical values were not as successfully forecasted (IA = 0.76). The Cohen’s Kappa Index is not applied in this case, as such an Index makes sense only in categorical forecasts (nominal values). It should also be noted that the numerical values forecasted here ranged between 0 and up to over 100, according to Fig. 2.

3.1.2 Forecasting of the nominal hourly and daily CAQI levels

The calculations were repeated with the same method for the CAQI levels (nominal values). Overall results are presented in Tables 8 and 9. Both hourly and daily CAQI levels are not forecasted in a satisfactory manner with the aid of LRM, one possible reason being the mapping of the CAQI between arithmetic and nominal values. This can be seen via their lower IA (reaching 0.5 for daily values) as well as via their low Cohen’s Kappa Index (Table 10), calculated on the basis of the numerical values that the CAQI receives according to Table 3.

Table 8. Observed vs. predicted hourly CAQI levels for Agia Sofia, using linear regression models. Each row sums 100%.

<table>
<thead>
<tr>
<th>Observed hourly CAQI levels</th>
<th>Predicted CAQI levels</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Very low</td>
</tr>
<tr>
<td>Very low</td>
<td>65.31</td>
</tr>
<tr>
<td>Low</td>
<td>7.14</td>
</tr>
<tr>
<td>Medium</td>
<td>0.23</td>
</tr>
<tr>
<td>High</td>
<td>0.17</td>
</tr>
<tr>
<td>Very High</td>
<td>0.03</td>
</tr>
</tbody>
</table>
Table 9. Observed vs predicted daily CAQI levels for Agia Sofia, using linear regression models. Each row sums 100%.

<table>
<thead>
<tr>
<th>Observed daily CAQI levels</th>
<th>Predicted CAQI levels</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Very low</td>
<td>Low</td>
</tr>
<tr>
<td>Very low</td>
<td>23.53</td>
<td>35.29</td>
</tr>
<tr>
<td>Low</td>
<td>5.75</td>
<td>62.94</td>
</tr>
<tr>
<td>Medium</td>
<td>0.64</td>
<td>33.26</td>
</tr>
<tr>
<td>High</td>
<td>0.00</td>
<td>12.87</td>
</tr>
<tr>
<td>Very High</td>
<td>0.00</td>
<td>3.94</td>
</tr>
</tbody>
</table>

Table 10. Results of the nominal hourly and daily CAQI level predictions via linear regression [6].

<table>
<thead>
<tr>
<th></th>
<th>Index of agreement (IA)</th>
<th>Cohen’s Kappa Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hourly CAQI levels</td>
<td>0.65</td>
<td>0.53</td>
</tr>
<tr>
<td>Daily CAQI levels</td>
<td>0.50</td>
<td>0.28</td>
</tr>
</tbody>
</table>

3.1.3 Forecasting of the numerical 24-hourly values to calculate the daily CAQI levels

The fact that the forecast of the numerical hourly CAQI values reached an IA equal to 0.92 led us to investigate the forecasting of the daily CAQI levels on the basis of forecasted 24 hourly values of the next day. For this purpose, the hourly CAQI values of the previous day were used in the following manner: for the forecasting of the first hourly CAQI value for the next day, the 24 hourly values of CAQI of the previous day (observed) were used. For the forecasting of the second hourly CAQI value of the next day, the value of the previous hour (forecasted), as well as the 23 hourly values of the previous day (observed) were used, and so forth. Via this procedure, the hourly CAQI values were forecasted and then we calculated the daily average CAQI value, based on these values.

Table 11 presents the comparison of daily CAQI forecasting performance via LRM at the station of Agia Sofia, by using lagged observed numerical values and by (equally) using 24 hourly forecasted values. The hourly CAQI numerical values for the Agia
Sofia station achieved an IA = 0.70, on the basis of LR. The performance was decreased in comparison to the performance of the model that uses ten observed lagged values, which reached an IA = 0.92. On the other hand, the performance concerning the forecasting of daily CAQI levels was slightly increased when using 24 forecasted lagged values with an IA=0.51 (in comparison to 0.50). This slight difference in the performance may be attributed to the use of a higher percentage of observations for the forecasting of the hourly CAQI values for the first hours of the next day, thus allowing information of the previous day to “penetrate” the target day of the forecast.

### Table 11. The IA for the forecasted numerical daily CAQI via linear regression [6].

<table>
<thead>
<tr>
<th>Forecast by using:</th>
<th>Hourly CAQI Values</th>
<th>Daily CAQI levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 observed lagged values</td>
<td>0.92</td>
<td>0.50</td>
</tr>
<tr>
<td>24 predicted lagged values</td>
<td>0.70</td>
<td>0.51</td>
</tr>
</tbody>
</table>

#### 3.2 CAQI Forecasting with the aid of ANNs and DTs

In order to develop ANN and DT models, data were firstly normalized by applying the hyperbolic tangent sigmoid transfer function, which was also applied on the hidden layers. On the output layers the linear transfer function was used. This is a common structure for function approximation-numerical value prediction (or regression) problems. For the training phase the Levenberg-Marquardt Backpropagation algorithm [32] was implemented. We also applied Principal Component Analysis (PCA) [7] to the data in order to change their dimensionality, and thus help the ANN to find patterns into the data. It should also be noted that none of the features (input parameters) were excluded from the training phase.

The 10-fold cross-validation was used in ANN and DT models in order to measure the predictive performance of the models, as described in the previous chapter. In our previous work [6] all ANN models had one hidden layer and the number of neurons was the same as the number of the input parameters. In order to perform sensitivity analysis in the ANN models, we now used 15 different architectures, depending on the number of the input parameters of each model (Table 7). In the DT models, four...
different numbers of trees were used (50, 100, 150 and 200) in order to perform sensitivity analysis.

The computational experiments of [6] were repeated in order to perform sensitivity analysis and develop the forecasting models for the CAQI levels by using ANNs and DTs. The basic experiments (indicated as models) and the different datasets that were used for each model, are the ones presented in points 1 to 4 listed below:

1. Hourly and daily CAQI numerical values (from 0 up to over 100) were forecasted via ANNs, in order to calculate the hourly and daily CAQI levels (according to fig. 2). Tables 12 and 13 present the forecasting results for the different datasets for daily and hourly CAQI levels respectively, indicated as Model 3 (when Dataset 1 was used), Model 4 (when Dataset 2 was used) and Model 5 (when Dataset 3 was used).

2. Hourly and daily NO$_2$ and PM$_{10}$ numerical values were forecasted via ANNs, in order to calculate the hourly and daily CAQI levels by using the formulas of Table 1. As input to the ANN, Dataset 2 was used. Tables 12 and 13 present the forecasting results of different datasets for daily and hourly CAQI levels (indicated as Model 6). The values at columns H.L.x in Tables 12 and 13 indicate the number of neurons for the hidden layer x.

3. Hourly and daily CAQI levels were forecasted via ANNs. In order to convert these numerical values (here the output of the ANNs was between -1 and 1) to CAQI levels (nominal values), the logical expressions of Table 4 were used. Table 12 and 13 present the forecasting results of different datasets for daily and hourly CAQI levels, indicated as Model 7 (when Dataset 1 was used), Model 8 (when Dataset 2 was used) and Model 9 (when Dataset 3 was used).

4. The hourly and daily CAQI levels were forecasted with the aid of DTs. Table 14 presents the forecasting results of different datasets for daily and hourly CAQI levels, indicated as Model 10 (when Dataset 1 was used), Model 11 when Dataset 2 was used) and Model 12 (when Dataset 3 was used).

As a result of the aforementioned work, a total 1118 different models were developed and trained for daily and hourly data (559 for each one respectively). Table 15 presents in detail the 559 models that were tested for daily and hourly data (thus leading to a total of 1118 models).
Table 12. Comparison of the best forecasting performance of all ANN Forecasting Models for daily CAQI levels.

<table>
<thead>
<tr>
<th>Basic Model</th>
<th>Number of Input parameters</th>
<th>Architecture ID</th>
<th>H.L. 1</th>
<th>H.L. 2</th>
<th>H.L. 3</th>
<th>Architecture Generated by Method</th>
<th>Percentage Agreement</th>
<th>Cohen's Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>1 and 2</td>
<td>0.50</td>
<td>0.23</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>6</td>
<td>23</td>
<td>-</td>
<td>-</td>
<td>1 and 2</td>
<td>0.56</td>
<td>0.35</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>1</td>
<td>20</td>
<td>-</td>
<td>-</td>
<td>1 and 2</td>
<td>0.48</td>
<td>0.22</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>5</td>
<td>18</td>
<td>27</td>
<td>36</td>
<td>2</td>
<td>0.54</td>
<td>0.33</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>8</td>
<td>8</td>
<td>12</td>
<td>-</td>
<td>2</td>
<td>0.50</td>
<td>0.26</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>1</td>
<td>18</td>
<td>-</td>
<td>-</td>
<td>1 and 2</td>
<td>0.55</td>
<td>0.36</td>
</tr>
<tr>
<td>9</td>
<td>11</td>
<td>6</td>
<td>28</td>
<td>-</td>
<td>-</td>
<td>1 and 2</td>
<td>0.48</td>
<td>0.23</td>
</tr>
</tbody>
</table>

Table 13. Comparison of the best forecasting performance of all ANN Forecasting Models for hourly CAQI levels.

<table>
<thead>
<tr>
<th>Basic Model</th>
<th>Number of Input parameters</th>
<th>Architecture ID</th>
<th>H.L.1</th>
<th>H.L.2</th>
<th>H.L.3</th>
<th>Architecture Generated by Method</th>
<th>Percentage Agreement</th>
<th>Cohen's Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2</td>
<td>13</td>
<td>6</td>
<td>10</td>
<td>-</td>
<td>2</td>
<td>0.63</td>
<td>0.50</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>7</td>
<td>23</td>
<td>23</td>
<td>-</td>
<td>1</td>
<td>0.50</td>
<td>0.31</td>
</tr>
<tr>
<td>5</td>
<td>12</td>
<td>1</td>
<td>24</td>
<td>-</td>
<td>-</td>
<td>1 and 2</td>
<td>0.62</td>
<td>0.49</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>4</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>1</td>
<td>0.50</td>
<td>0.31</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>10</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>2</td>
<td>0.62</td>
<td>0.50</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>2</td>
<td>18</td>
<td>18</td>
<td>-</td>
<td>1</td>
<td>0.50</td>
<td>0.32</td>
</tr>
<tr>
<td>9</td>
<td>12</td>
<td>1</td>
<td>24</td>
<td>-</td>
<td>-</td>
<td>1 and 2</td>
<td>0.62</td>
<td>0.49</td>
</tr>
</tbody>
</table>

Table 14. Comparison of the best forecasting performance of all DT forecasting models for daily and hourly CAQI levels.

<table>
<thead>
<tr>
<th>Basic Model</th>
<th>Number of Trees</th>
<th>Number of Input parameters</th>
<th>Percentage Agreement</th>
<th>Cohen's Kappa</th>
<th>Number of Trees</th>
<th>Number of Inputs</th>
<th>Percentage Agreement</th>
<th>Cohen's Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>150</td>
<td>9</td>
<td>0.56</td>
<td>0.32</td>
<td>150</td>
<td>9</td>
<td>0.64</td>
</tr>
<tr>
<td>---</td>
<td>----</td>
<td>-----</td>
<td>---</td>
<td>------</td>
<td>------</td>
<td>-----</td>
<td>---</td>
<td>------</td>
</tr>
<tr>
<td>11</td>
<td>200</td>
<td>9</td>
<td>0.69</td>
<td>0.51</td>
<td></td>
<td>100</td>
<td>9</td>
<td>0.60</td>
</tr>
<tr>
<td>12</td>
<td>50</td>
<td>14</td>
<td>0.59</td>
<td>0.37</td>
<td></td>
<td>100</td>
<td>12</td>
<td>0.65</td>
</tr>
</tbody>
</table>

Table 15. The 559 models that were tested for daily and hourly data.

<table>
<thead>
<tr>
<th>Data</th>
<th>Used by</th>
<th>Models Developed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset 1 (1 to 10 lagged values)</td>
<td>Model 3</td>
<td>150</td>
</tr>
<tr>
<td>Dataset 2</td>
<td>Model 4</td>
<td>15</td>
</tr>
<tr>
<td>Dataset 3 (1 to 5 lagged values)</td>
<td>Model 5</td>
<td>75</td>
</tr>
<tr>
<td>Dataset 2</td>
<td>Model 6</td>
<td>15</td>
</tr>
<tr>
<td>Dataset 1 (1 to 10 lagged values)</td>
<td>Model 7</td>
<td>150</td>
</tr>
<tr>
<td>Dataset 2</td>
<td>Model 8</td>
<td>15</td>
</tr>
<tr>
<td>Dataset 3 (1 to 5 lagged values)</td>
<td>Model 9</td>
<td>75</td>
</tr>
<tr>
<td>Dataset 1 (1 to 10 lagged values)</td>
<td>Model 10</td>
<td>40</td>
</tr>
<tr>
<td>Dataset 2</td>
<td>Model 11</td>
<td>20</td>
</tr>
<tr>
<td>Dataset 3 (1 to 5 lagged values)</td>
<td>Model 12</td>
<td>4</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>559</strong></td>
</tr>
</tbody>
</table>

3.3 Comparison of results

The results obtained are reported in Tables 12, 13 and 14 which present the best performance of all forecasting Models for daily and hourly CAQI levels of the current study and the best forecasting performance models of our previous study [6].

3.3.1 Models for daily CAQI levels

The best forecasting performance of daily CAQI levels by using ANNs (56% PA and Cohen’s Kappa=0.35) was achieved when Model 4 was used (calculate the CAQI levels by forecasting the CAQI values via ANNs, with Dataset 2). In most cases simple ANN architectures (with one hidden layer) achieved the best forecasting performances. The best forecasting performance for daily CAQI levels (PA=69% and Cohen’s Kappa=0.51), was achieved when DT Model 11 with 200 trees was used, Table 14 (forecast directly the CAQI levels by using the Dataset 2).
3.3.2 Models for hourly CAQI levels

The best forecasting performance of hourly CAQI levels by using ANNs (63% PA and Cohen’s Kappa=0.50) was achieved when Model 3 was used (calculate the CAQI levels by forecasting the CAQI values via ANNs, with Dataset 1). In contrast to the results achieved for the forecasting of daily CAQI levels, in the case of hourly levels complex architectures (with more than one hidden layer) were the ones that achieved the best forecasting performance (71% of the studied cases). The best forecasting performance of the hourly CAQI levels (PA=65% and Cohen’s Kappa=0.53) was achieved when DT Model 12 with 100 trees was used (Forecast directly the CAQI levels by via DTs, with Dataset 3. The same forecasting performance was achieved by the LRM s of our previous study [6].

3.3.3 Sensitivity analysis and comparison with our previous study

Table 16 presents the average performance of the ANN forecasting Models for daily and hourly CAQI levels for each different ANN architecture. Each ANN architecture has a unique ID and is characterized by a different parameter distinguishing it from other architectures. The results reported in Table 16 support our previous finding that simple ANN architectures (with one hidden layer) achieve better forecasting performance for the daily CAQI levels. Contrary to this finding, the results for the hourly CAQI levels suggest that more complex architectures lead to better forecasting performances.

<table>
<thead>
<tr>
<th>Architecture ID</th>
<th>Architecture Generated by Method:</th>
<th>Daily CAQI levels</th>
<th>Hourly CAQI levels</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Percentage Agreement Average</td>
<td>Cohen's Kappa Average</td>
</tr>
<tr>
<td>1</td>
<td>1 and 2</td>
<td>0.46</td>
<td>0.20</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.43</td>
<td>0.16</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0.43</td>
<td>0.16</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0.43</td>
<td>0.15</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>0.43</td>
<td>0.16</td>
</tr>
</tbody>
</table>
Method 1: Increase the number of Hidden Layers from the previous topology.

Method 2: Increase the Number of Neurons from the previous topology.

Method 1 and 2: Increase the Number of Neurons from the previous “1 and 2” Topology.

Table 17 presents the average performance of the DT Forecasting Models for daily and hourly CAQI levels for each different number of trees. Results indicate that an increment at the number of the trees will not affect in general the forecasting performance, but it may improve the accuracy and our confidence of the forecasting results. It worthies mentioning that the best forecasting performance was achieved with 200 trees (the maximum number of trees being investigated) in the case of the daily CAQI levels and with 100 trees in the case of the hourly CAQI levels.

Table 17. The average performance of the DT forecasting models for daily and hourly CAQI levels for each different number of trees.

<table>
<thead>
<tr>
<th>Number of Trees</th>
<th>Daily CAQI levels</th>
<th>Hourly CAQI levels</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Percentage Agreement</td>
<td>Cohen's Kappa Average</td>
</tr>
<tr>
<td>50</td>
<td>0.500</td>
<td>0.255</td>
</tr>
<tr>
<td>100</td>
<td>0.484</td>
<td>0.230</td>
</tr>
<tr>
<td>150</td>
<td>0.495</td>
<td>0.256</td>
</tr>
<tr>
<td>200</td>
<td>0.505</td>
<td>0.249</td>
</tr>
</tbody>
</table>
When comparing the results obtained in this study with the ones of our previous study [6] we should take into account that in the latter we used only one type of ANN architecture that included one hidden layer, where the number of neurons was the same as the number of the input parameters (depending on the Dataset that was used, i.e. 1-10 inputs for Dataset 1, 9 inputs for Dataset 2 and 10-14 inputs for Dataset 3)). In the current study sensitivity analysis was used to further investigate the topology of the model and improve its performance. Additionally, the 10-fold cross-validation was used in order to measure the predictive performance of the models being developed.

Table 18 presents detailed information for the best forecasting methods of the current as well as of the previous study. In basic model 9 (of the Table 18) the CAQI levels (which are encoded to values that range from -1 to 1) were forecasted directly. In basic models 1, 3 and 4 (of the same Table) the CAQI levels are calculated from the forecasted CAQI values (which ranges from 0 to higher than 100). It is therefore evident that when cross-validation was used at the model with the best forecasting performance for daily data resulting from our previous study [6], the performance decreased from 61% PA (Model 9°2) to 44% PA (Model 9°1). On the other hand, when cross-validation was used in the model with the best forecasting performance for hourly data (that resulted from our previous study [6]) the performance did not change. The fact that the performance was decreased for the forecasting of the daily CAQI and not for the hourly one, suggests that the results of our previous study for the daily data were not as accurate as we would expect. Thus, cross validation should be used in all times, in order to increase the credibility and validity of the forecasting models.

From Table 18, it is evident that the forecasting PA of the daily levels was increased by 25% (when cross-validation is used). In the case of hourly CAQI Levels the PA was equal in both studies (65%). From the same Table we can observe the difference in the performance of the models. For example, basic model 4 has an observed mean value of 65.32, because the CAQI levels were calculated from the forecasted CAQI values (which range from 0 to higher than 100 as we can see in Figure 2 and in Eq. 8). On the other hand, Model 9°1 has an Observed Mean value of 0.03, because (a) the CAQI levels (which are encoded to values that range from -1 to 1 as we can see in
Table 3 and Figure 3) were directly forecasted, and (b) the tangent sigmoid transfer function was used.

**Table 18.** Detailed information for the best model performance of the current and our previous study [6].

<table>
<thead>
<tr>
<th>CAQI levels</th>
<th>Daily</th>
<th>Hourly</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Computational Intelligence Method</strong></td>
<td>ANN</td>
<td>DT</td>
</tr>
<tr>
<td><strong>Basic Model</strong></td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>Observed mean</td>
<td>65,32</td>
<td>-</td>
</tr>
<tr>
<td>Predicted mean</td>
<td>65,61</td>
<td>-</td>
</tr>
<tr>
<td>Observed standard deviation</td>
<td>28,28</td>
<td>-</td>
</tr>
<tr>
<td>Predicted standard deviation</td>
<td>23,74</td>
<td>-</td>
</tr>
<tr>
<td>Normalized mean difference (NMD)</td>
<td>0,04</td>
<td>-</td>
</tr>
<tr>
<td>Root mean square error (RMSE)</td>
<td>22,22</td>
<td>-</td>
</tr>
<tr>
<td>RMSE systematic</td>
<td>3,08</td>
<td>-</td>
</tr>
<tr>
<td>RMSE unsystematic</td>
<td>5,24</td>
<td>-</td>
</tr>
<tr>
<td>Mean absolute error (MAE)</td>
<td>15,11</td>
<td>-</td>
</tr>
<tr>
<td>Mean absolute percentage error (MAPE)</td>
<td>0,25</td>
<td>-</td>
</tr>
<tr>
<td>Mean bias error (MBE)</td>
<td>-0,28</td>
<td>-</td>
</tr>
<tr>
<td>Correlation coefficient (R)</td>
<td>0,66</td>
<td>-</td>
</tr>
<tr>
<td>Index of agreement (IA)</td>
<td>0,79</td>
<td>-</td>
</tr>
<tr>
<td>Percentage Agreement</td>
<td>0,56</td>
<td>0,69</td>
</tr>
<tr>
<td>Cohen's Kappa</td>
<td>0,35</td>
<td>0,51</td>
</tr>
</tbody>
</table>

*1: The basic model 9 of our previous study [6] applied with cross-validation.

*2: The basic model 9 of our previous study [6].

*3: The basic model 1 of our previous study [6] applied with cross-validation.

*4: The basic model 1 of our previous study [6].
4. Conclusions and further research directions

There are scarce research results concerning the performance of air quality forecasting models in terms of combined environmental pressure indicators, such as the CAQI [33]. Moreover, few studies have addressed the influence of the characteristics, the architecture and the training process of ANN and DT models over their accuracy concerning air quality forecasting. Our study has therefore evaluated and inter-compared the performance of a wide variety of ANNs, DTs, and statistical regression models on the basis of the following scenarios: (a) the estimation of the CAQI levels by forecasting the CAQI values, (b) the estimation of the CAQI levels by forecasting individual pollutants and (c) the direct forecasting of the CAQI levels. In order to investigate these scenarios, three different datasets were employed for daily and hourly data. For the different datasets and the different scenarios, 15 ANN architectures and 4 different DT model ensembles were used, indicated as different models. Overall, a total of 1118 different models were developed and trained for daily and hourly data (559 for each one respectively).

Cross-validation and sensitivity analysis were used in order to perform a number of computational experiments to test the performance of different ANN model architectures, with the aim to evaluate their applicability and reliability concerning CAQI forecasting. The obtained results and the accompanied sensitivity analysis demonstrated that in order to predict daily CAQI levels, simple ANN architectures (with one hidden layer) should be used. On the other hand, in order to predict hourly CAQI levels, more complex architectures (with more than one hidden layer) should be employed. We estimate that the reasons for these differences lie in the aggregated nature of the daily CAQI levels, that smoothes-out any short term perturbations associated with the temporal profile of air pollutants, thus allowing for simpler ANN architectures to map the any “knowledge” interwoven in the dataset under investigation that is related to the daily CAQI. It should be mentioned that the number of inputs of the training data affect the size of hidden layers, and consequently the ANN architecture (number of hidden layers and number of neurons per layer). The ANN architecture has an impact on the learning time and the generalization capabilities of the network and thus its performance (the ability of the neural network to produce an output for an input data that was not part of the training set).
It was also found that the DT models outperform the ANN models concerning their forecasting ability. When hourly CAQI levels were forecasted the best performance was achieved by a DT model (PA=65%, equal to the performance of a LRM developed on the basis of our previous study [6]). In addition, it became evident that cross-validation should be used in order to increase the credibility and validity of the results.

The presented results suggest that the performance of various ANN and DT models depends both on their internal structure and on the methods used for their training. Using model ensembles obtained by a k-fold cross-validation resulted in more accurate forecasts than using individual models. These results are expected to be useful in developing and implementing operational air quality management and forecasting systems. From the algorithms point of view, it would be interesting to further investigate the characteristics and capabilities of the DT models, as well as the performance of other classifiers (e.g. Learning Vector Quantization (LVQ) networks, Vector Autoregressive (VAR) Models).

References:


