CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Institutional Repository of the Freie Universität Berlin

Atmospheric Pollution Research 6 (2015) 904-915

Atmospheric Pollution Research
www.atmospolres.com

The development of a dense urban air pollution monitoring network

Hamid Taheri Shahraiyni [1,2], Sahar Sodoudi [1], Andreas Kerschbaumer [3], Ulrich Cubasch [1]

[1] Institut für Meteorologie, Freie Universität Berlin, Carl–Heinrich–Becker–Weg 6–10, 12165 Berlin, Germany
[2] Faculty of Civil Eng., Shahrood University, Shahrood, Iran
[3] Senate Department for Urban Development and the Environment, Berlin, Germany

ABSTRACT

The importance of air pollution monitoring networks in urban areas is well known because of their miscellaneous applications. At the beginning of the 1990s, Berlin had more than 40 particulate matter monitoring stations, whereas, by 2013, there were only 12 stations. In this study, a new and free–of–charge methodology for the densifying of the PM_{10} monitoring network of Berlin is presented. It endeavors to find the non–linear relationship between the hourly PM_{10} concentration of the still–operating PM_{10} monitoring stations and the shut–down stations by using the Artificial Neural Network (ANN), and, consequently, the results of the shut–down stations were simulated and re–constructed. However, input–variables selection is a pre–requisite for any ANN simulation, and hence a new fuzzy–heuristic input selection has been developed and joined to the ANN for the simulation. The hourly PM_{10} concentrations of the 20 shut–down stations were simulated and re–constructed. The mean error, bias and absolute error of the simulations were 27.7%, –0.03 (μg/m^3), and 7.4 (μg/m^3), respectively. Then, the simulated hourly PM_{10} concentration data were converted to a daily scale and the performance of ANN models which were developed for the simulation of the daily PM_{10} data were evaluated (correlation coefficient >0.94). These appropriate results imply the ability of the developed input selection technique to make the appropriate selection of the input variables, and it can be introduced as a new input variable selection for the ANN. In addition, a dense PM_{10} monitoring network was developed by the combination of both the re–constructed (20 stations) and the current (12 stations) stations. This dense monitoring network was applied in order to determine a reliable mean annual PM_{10} concentration in the different areas in Berlin in 2012.

Keywords: Artificial neural networks, input variable selection, urban areas, Berlin, PM_{10} monitoring network

Corresponding Author:
Hamid Taheri Shahraiyni
☎ : +49-30-83854366
🖷 : +49-30-83871160
✉ : hamid.taheri@met.fu-berlin.de

Article History:
Received: 24 November 2014
Revised: 05 April 2015
Accepted: 06 April 2015

doi: 10.5094/APR.2015.100

1. Introduction

There are some major objectives for the development of air pollution monitoring networks in urban areas. The importance and application of air pollution monitoring networks have been well–known from the 1960s, and the main reported objectives and uses of air pollution monitoring in the literature are:

- Planning for the appropriate urbanization and land use development (WHO, 1977; Trujillo–Ventura and Ellis, 1991; Chen et al., 2006);
- Evaluation of the exposure of people to air pollution and consequently its effects on human health, and the protection of the public health (Darby et al., 1974; Hougland and Stephens, 1976; Ott, 1977; WHO, 1977; Modak and Lohani, 1985; Trujillo–Ventura and Ellis, 1991; Kanaroglou et al., 2005; Lozano et al., 2009; Ferradas et al., 2010; Zheng et al., 2011; Pope and Wu, 2014);
- Quantifying the effects of the emission sources (e.g., power plants) on air pollution (Leavitt et al., 1957; Seinfeld, 1972; Pope and Wu, 2014);
- The control and management of urban air pollution (Hougland and Stephens, 1976; WHO, 1977; Van Egmond and Onderdelinden, 1981);
- The evaluation of air pollution control programs and strategies (Seinfeld, 1972; WHO, 1977; Zheng et al., 2011);
- The initial assessment of air pollution condition, e.g., the determination of the mean concentrations of air pollutants in urban areas in different time scales (i.e., hourly, daily) (Goldstein and Landovitz, 1977; WHO, 1977; Shannon et al., 1978);
- Time series analysis for the determination of the trends of air pollutants (WHO, 1977; Trujillo–Ventura and Ellis, 1991; Pope and Wu, 2014);
- Investigation of the compliance of the concentrations of air pollutants with air quality standards (Hougland and Stephens, 1976; Ott, 1977; WHO, 1977; Van Egmond and Onderdelinden, 1981; Modak and Lohani, 1985; Chen et al., 2006; Ferradas et al., 2010; Pope and Wu, 2014)
- The evaluation and validation of the mechanistic models describing the spatio–temporal emission, transport and transformation of air pollutants (WHO, 1977; Van Egmond and Onderdelinden, 1981; Trujillo–Ventura and Ellis, 1991; Zheng et al., 2011);
- The spatial and knowledge–based modeling of air pollutants (Shannon et al., 1978; Modak and Lohani, 1985; Briggs et al., 1997; Briggs et al., 2000; Lozano et al., 2009; Taheri Shahraiyni et al., 2015);
- The determination of critical air pollution conditions and notification to the people affected (at risk) and to the relevant organizations (Seinfeld, 1972; WHO, 1977; Trujillo–Ventura and Ellis, 1991; Chen et al., 2006).

Many studies have been performed on the air pollution monitoring network design. Geostatistical techniques has been widely used for the calculation of the local spatial representativity of each monitoring station and the determination of the location of monitoring stations based upon the minimization of the estimation variance (Trujillo–Ventura and Ellis, 1991; Kanaroglou et

© Author(s) 2015. This work is distributed under the Creative Commons Attribution 3.0 License.

al., 2005; Taheri Shahraiyni et al., 2014). In some studies, analysis techniques, such as principal component analysis and cluster analysis, have been employed for the site selection of air quality monitoring stations (e.g., Pires et al., 2008a; Pires et al., 2008b; Pires et al., 2009). The design of monitoring networks using spatial distribution patterns, developed by an atmospheric dispersion model, is another alternative (e.g., Mazzeo and Venegas, 2008; Zheng et al., 2011). High–resolution measuring campaigns have been utilized in some studies for the suitable site selection of the monitoring stations (e.g., Cocheo et al., 2008; Ferradas et al., 2010). Another method is based upon the multi–objective network design, which considers environmental, social and economical objectives simultaneously (e.g., Chen et al., 2006; Pope and Wu, 2014). The design of virtual monitoring stations has been utilized in some studies for the minimization of the monitoring costs (e.g., Ung et al., 2001; Beaulant et al., 2008).

The Euclidean distance of a point in the city to the closest monitoring station is one of the spatial indicators that can influence the representation of the air quality monitoring network (Pope and Wu, 2014), and, consequently, an increase in the number of monitoring stations will improve this spatial indicator.

Although an increase in the number of air pollution monitoring stations in urban areas can lead to better air pollution estimation and evaluation (Stalker and Dickerson, 1962; Modak and Lohani, 1985; Cocheo et al., 2008), it increases the monitoring costs (Hickey et al., 1971; Cocheo et al., 2008), which is the main constraint for the development of a dense air pollution monitoring network (Trujillo–Ventura and Ellis, 1991).

In this study, a new methodology for the densifying of the PM_{10} (particulate matter less than 10 µm in aerodynamic diameter) monitoring network of an urban area (Berlin, Germany) is presented which is totally free–of–charge. At the beginning of the 1990s, Berlin had more than 40 particulate matter monitoring stations (SenStadt, 1998). The number of monitoring stations decreased steadily until the end of 1990s (Lenschow et al., 2001) and now there are a little number of PM_{10} monitoring stations. In this study, an attempt is made to re–construct the shut–down particulate matter monitoring stations by non–linear simulation using a knowledge–based black–box modeling technique. In this study, we try to find the non–linear relation between the concentration of PM_{10} of the still operating PM_{10} monitoring stations and the shut–down stations, and, consequently, the PM_{10} concentrations for the shut–down stations are estimated.

Artificial neural networks (ANNs) are well–known powerful knowledge–based black–box modeling techniques, and are widely used in the static (real–time estimation) and dynamic (prediction) modeling of air pollutants (Gardner and Dorling, 1998). ANN as a static modeling technique is used not only for the estimation of the concentration of the air pollutants (e.g., Lal and Tripathy, 2012; Elangasinghe et al., 2014; Zhang and Peng, 2014) and the determination of the relative apportionment of the various sources on the concentration of a receptor site (e.g., Reich et al., 1999), but also for the estimation of spatial distribution of the pollutants (e.g., Yao and Lu, 2014). ANN as a dynamic modeling technique is utilized in different forms of forecasting. For example, Papanastasiou et al. (2007), Wu et al. (2011), and Russo et al. (2015) used ANN for the daily forecasting of PM_{10} concentration, and Dutot et al. (2007), Corani (2005), and Nejadkoorki and Baroutian (2012) employed ANN for the prediction of the hourly, 8–hourly, and daily maximum of air pollutants, respectively. In this study, ANN is employed for the simulation of virtual stations or for the re–construction of the shut–down stations by the development of a non–linear relation between the PM_{10} concentration of the still operating monitoring stations and the shut–down stations.

One of the most important issues of ANN is the input variable selection, and it is a pre–requisite for the ANN simulation (Giordano et al., 2014). Input variable selection is performed to remove the superfluous (redundant and irrelevant) variables (May et al., 2011). Irrelevant variables have no significant influence on the output variable. Redundant variables have influence on the output variables, but their influence can be represented by either one or other of the relevant variables (Bell and Wang, 2000). The superfluous variables increase the size of the input variables to ANN, and, consequently, the complexity of the ANN model and its training time also increase. The superfluous variables increase the training difficulty. The inclusion of redundant variables increases the number of local extrema in the error function of the learning technique, and, accordingly, the developed ANN model will bear poor generalization. The inclusion of irrelevant variables increases the complexity of the knowledge extraction because these variables behave similarly to the noise, and hide the input–output relationships. In addition, it is very difficult to interpret the results of the ANN modeling when the inpfuts are superfluous variables which impede our understanding the behavior of the investigated phenomenon (May et al., 2011; Giordano et al., 2014).

Up to now, many different input variable selection techniques have been developed (see Blum and Langley, 1997; Kohavi and John, 1997; Guyon and Elisseeff, 2003) and also used for the input variable selection in the neural network simulations (e.g., Bowden et al., 2005; La Rocca and Perna, 2005; Giordano et al., 2014). In this study, a new heuristic input selection technique based upon fuzzy curve fitting is developed and joined to the ANN. Consequently, this modeling framework is employed to simulate the shut–down PM_{10} monitoring stations in Berlin.

2. The Study Area

Berlin is the capital city of Germany (Figure 1) and it is located in the North–Eastern part of Germany. Its population is ranked seventh among the urban areas in the European Union (about 3 500 000 inhabitants). Berlin covers an area of about 900 km^2 and about one–third of its inhabitants live in the inner city in an area of about 88 km^2 and, accordingly, it has low building and population density outside the inner city. Berlin has a flat topography and about 45% of its area is made up of water bodies and green areas. 35% of Berlin's area is built–up areas. The transport and infrastructure areas cover about 20% of the city (Dugord et al., 2014). It has a moderate climate and its average wind speed and temperature are about 3 m/s and 8.8 °C, respectively. There are about 0.32 cars and LDV (Light Duty Vehicles) per resident in Berlin (Lenschow et al., 2001). The transportation system of Berlin is composed of passenger cars with diesel engines (54%), passenger cars with gasoline/petrol engines (30%), Light Duty Vehicles (LDV) with diesel engines (7%), LDV with gasoline engines (4%), Heavy Duty Vehicles (HDV) (3%), buses (1%), and motor cycles (1%). In addition, 39, 34, and 24% of the passenger cars with diesel engines in Berlin have EURO–5, EURO–4 with diesel particle filters, and poorer emissions standards, respectively and also 18, 52, and 30% of the passenger cars with gasoline engines in Berlin have EURO–5 and EURO–4 engines, and poorer emissions standards, respectively (Schmidt and During, 2013). Berlin is situated in the approximately 200 km northwest of the industrialized area at Germany's borders with Poland and the Czech Republic, an area which is called the "Black Triangle" (Lenschow et al., 2001). Sometimes, the concentration of PM_{10} in Berlin exceeds the EU limit (Gorgen and Lambrecht, 2007). About 64.4% of PM_{10} in Berlin stems from non–Berlin emission sources (regional background sources) and the emission sources of the remaining 35.6% come from the urban background and traffic PM_{10} sources (Rauterberg–Wulff et al., 2013). Figure S1 (see the Supporting Material, SM) shows the sources of PM_{10} in Berlin in detail. The EU has set two limit values for PM_{10} for the protection of human health. According to these limits, the mean daily PM_{10} concentration may not exceed 50 µg/m^3 more than 35 times per year, and the mean annual PM_{10} concentration may not exceed

40 µg/m³ (EU, 2008). At the beginning of the 1990s, Berlin had a high level of Total Suspended Particulate (TSP) (Lenschow et al., 2001) and hence, a dense monitoring network was developed for the appropriate monitoring of the pollutants in Berlin (SenStadt, 1998). Both the concentration of TSP and the number of TSP monitoring stations decreased greatly until the end of 1990s (Lenschow et al., 2001) and also the TSP stations were gradually replaced with PM_{10} stations. In 1999, there were 18 TSP monitoring stations in Berlin. By 2013, there were only 12 PM_{10} monitoring stations. There are continuous particulate matter data (TSP or PM_{10} data) from 1990s until the present in only 7 stations (Table 1). Table 1 presents some of the shut–down stations (20 stations) and the 12 still operating ones at least until end of 2013 with their respective properties. The suburban (SU) (background) (Bg) in Table 1 indicates the stations located near to the boundary of the urban area and influenced by some local emission sources, while the rural stations within the Berlin municipality boundary are the stations located at rural environments such as forests, grasslands and near lakes, far away from direct air pollution sources.

Figure 1. *The location of Berlin in Germany (Source map: goway.com).*

3. Input–Output Databases

The 20 stations removed from Berlin's network were selected for the re–construction by simulation (Table 1). We tried to find some old concurrent hourly particulate matter data from the 20 shut–down stations (output variables) and some of the current (still–operated) stations (candidate input variables). Table 2 shows the time period of the concurrent hourly particulate matter data of each shut–down station and some of the current stations employed as the output and candidate input data for ANNs in this study, respectively. In this study, only the PM_{10} data of the still operating stations are employed as input variables, because the influences of the other variables such as meteorological parameters, and regional and urban background PM_{10} sources have been incorporated in the measured PM_{10} data by the still operating stations. In addition, the aim of this study is the development of an automatic module that will be able to densify the PM_{10} monitoring network immediately by using only the hourly PM_{10} measurements of still operating network. The influence of the traffic intensity near to the shut–down stations has not been incorporated in the set of input variables (still operating stations), because traffic intensity has a local effect. Accordingly, the shut–down stations with a special characteristic are utilized in this study. This characteristic is that each station is either far from the main traffic lanes, or the current traffic level around it has no significant difference from the traffic level during the time period presented in Table 2. All of the shut–down stations in Table 2 satisfied this traffic characteristic.

We could not re–construct some of the shut–down stations because the current traffic level around them is significantly different from the traffic level during the time period presented in Table 2. When there is a significant difference in the traffic intensity around a shut–down station between the simulation period and current situation, the shut–down station is simulated using the old PM data based upon old traffic intensity around the shut–down station and developed model for shut–down station is based upon old traffic intensity around it. Consequently, when the developed model is employed for re–construction of the current PM_{10} concentration of the shut–down station, it not only does not consider the current local effect of traffic intensity for PM_{10} estimation, but also consider the old local effects of traffic intensity, and thus, the error of PM_{10} increases and it is not possible to develop a model for re–construction of the shut–down station.

For evaluation of the above–mentioned traffic characteristic, Berlin's traffic–intensity maps were generated in ArcGIS. Figure 2 shows one sample of the traffic–intensity maps of Berlin. The traffic intensity (No. of vehicles/day) in Figure 2 is the mean daily number of all the motor–vehicles using a street in both directions throughout the year. The previous studies in Berlin showed that about 50% of particulate matter, measured at street level, stems from exhaust emissions, tyre abrasion and the re–suspension of soil particles in the individual street, and the remaining 50% originates from other sources in the city and in the regional backgrounds (Lenschow et al., 2001). In addition, a primary particulate matter source at ground level can influence the surrounding areas in a radius of less than 100 m (Hewitt and Jackson, 2008) and the traffic has an immediate influence on the coarse particulate matter in the immediate vicinity of the station. Thus, point buffer operation with constant width buffer (100 m) in the GIS (Geographical Information System) environment was applied to the shut–down stations, presented in Table 2, to determine the areas in which the traffic can have a significant influence on the particulate matter concentration of the shut–down stations. Then, the traffic intensity inside the buffering zone around each station was investigated using the traffic intensity maps in ArcGIS.

The old monitoring network in Berlin measured the TSP concentration while the current monitoring network measures the PM_{10} concentration. Lenschow et al. (2001) studied the relation between the PM_{10} and the TSP concentrations in Berlin, and found that the ratio of PM_{10} to TSP was about 0.8. This ratio was employed for the conversion of the TSP data in Table 2 to PM_{10}. Thus, 20 input–output databases of hourly PM_{10} concentrations were generated. The hourly PM_{10} concentrations of the current stations and hourly PM_{10} concentrations in each of the station removed (each row of Table 2) are considered as input variables and output variable for simulation, respectively.

4. Simulation Algorithm

The simulation algorithm has two major stages (fuzzy–heuristic input selection and neural network modeling). A schematic diagram of two major stages of this study with the input and output variables in each stage has been presented in Figure S2 (see the SM). In this section, the algorithm of the shut–down stations simulation is described in detail, step by step. The flowchart of the algorithm of the study has been presented in Figure 3. Fuzzy–heuristic input selection and neural network modeling stages are implemented by Steps 1–7 and Steps 8–15, respectively. These steps are implemented by a developed computer program in the MATLAB R2013b.

Table 1. *The characteristics of the current particulate matter monitoring stations and the 20 shut–down stations, utilized in this study with their operation period from 1992 (SU: Suburban; Bg: Background; Ur: Urban; Tr: Traffic; R–N: Rural–Near city)*

Station Code	Station Name	Operation Period	Type of Area	Type of Station	Altitude (m)	Longitude (degree)	Latitude (degree)
MC 010	Wedding–Amrumer Str.	1992–now	Ur	Bg	35	13.349	52.543
MC 032	Grunewald	1992–now	R–N	Bg	50	13.225	52.473
MC 042	Neukolln–Nansenstr.	1992–now	Ur	Bg	35	13.431	52.489
MC 077	Buch	1992–now	SU	Bg	60	13.490	52.643
MC 085	Friedrichshagen	1994–now	R–N	Bg	35	13.647	52.447
MC 115	Hardenbergplatz	2004–2013	Ur	Tr	35	13.333	52.506
MC 117	Steglitz–Schildhornstr.	1994–now	Ur	Tr	45	13.318	52.464
MC 124	Mariendorf–Mariendorfer Damm	2005–now	Ur	Tr	50	13.388	52.438
MC 143	Neukolln–Silbersteinstr.	2004–now	Ur	Tr	40	13.442	52.468
MC 171	Mitte–Bruckenstr.	1998–now	Ur	Bg	35	13.419	52.514
MC 174	Friedrichshain–Frankfurter Allee	1993–now	Ur	Tr	40	13.470	52.514
MC 220	Neukolln–Karl–Marx–Str.	2005–now	Ur	Tr	40	13.434	52.482
MC 001	Heiligensee–Krantorweg	1992–1997	SU	Bg	35	13.227	52.622
MC 006	Wittenau–Rodernallee	1992–1995	SU	Bg	45	13.345	52.586
MC 007	Falkenhagener Feld–Pionierstr.	1992–1996	SU	Bg	35	13.168	52.558
MC 009	Tegel–Flughafen	1992–1996	SU	Bg	35	13.288	52.551
MC 011	Wedding/Prenzlauer Berg–Behmstr.	1992–2000	SU	Tr	45	13.396	52.550
MC 017	Schmargendorf–Lentzeallee	1992–1997	SU	Bg	50	13.294	52.471
MC 018	Schoneberg–Belziger Str.	1992–2002	Ur	Bg	40	13.349	52.489
MC 020	Neukolln–Ederstr.	1992–1994	Ur	Bg	35	13.456	52.476
MC 023	Lankwitz–Leonorenstr.	1992–1995	SU	Bg	40	13.347	52.444
MC 024	Mariendorf–Walnussweg	1992–1995	SU	Bg	45	13.413	52.441
MC 025	Britz–Parchimer Allee	1992–1997	SU	Bg	35	13.458	52.447
MC 027	Mariendorf–Schichauweg	1992–2001	R–N	Bg	45	13.368	52.398
MC 028	Lichterfelde–Dielingsgrund	1992–1995	SU	Bg	45	13.409	52.411
MC 030	Rudow–Kunnekeweg	1992–1995	SU	Bg	45	13.520	52.418
MC 072	Pankow–Blankenfelder Str.	1992–1997	SU	Tr	45	13.404	52.591
MC 078	Blankenburg	1992–1996	SU	Bg	50	13.459	52.588
MC 080	Marzahn	1993–1997	SU	Tr	50	13.583	52.549
MC 081	Hellersdorf	1993–1996	SU	Bg	40	13.576	52.513
MC 083	Kaulsdorf–Sud	1993–1996	R–N	Bg	40	13.595	52.476
MC 145	Frohnau–Funkturm	1996–2001	R–N	Bg	50	13.296	52.653

Table 2. *The time periods of the concurrent hourly particulate matter data of the current and shut–down stations*

Shut–Down Station (Output Variable)	Current Stations (Input Variables)	Hourly Particulate Matter Data	Concurrent Time Periods
MC 001	MC 10, 32, 42, 77, 85, 117, 174	TSP	1996.01.23–1997.01.22
MC 006	MC 10, 32, 42, 77, 85, 174	TSP	1994.01.31–1995.04.03
MC 007	MC 10, 32, 42, 77, 85, 117, 174	TSP	1994.12.01–1996.02.02
MC 009	MC 10, 32, 42, 77, 85, 117, 174	TSP	1994.12.01–1996.03.29
MC 011	MC 10,32,42,77,85,117,174	TSP	1996.07.22–1998.07.22
MC 017	MC 10, 32, 42, 77, 85, 117, 174	TSP	1995.10.16–1997.10.06
MC 018	MC 10, 32, 42, 77, 85, 117, 174	PM_{10}	2009.01.18–2011.01.18
MC 020	MC 10, 32, 42, 77, 85, 174	TSP	1994.01.31–1995.11.24
MC 023	MC 10, 32, 42, 77, 85, 117, 174	TSP	1995.03.30–1996.03.29
MC 024	MC 10, 32, 42, 77, 85, 174	TSP	1994.08.28–1995.10.19
MC 025	MC 10, 32, 42, 77, 85, 117, 174	TSP	1995.03.28–1997.03.27
MC 027	MC 10, 32, 42, 77, 85, 117, 174	PM_{10}	2009.01.10–2011.01.10
MC 028	MC 10, 32, 42, 77, 85, 117, 174	TSP	1994.10.19–1995.10.19
MC 030	MC 10, 32, 42, 77, 85, 117, 174	TSP	1994.10.19–1995.10.19
MC 072	MC 10, 32, 42, 77, 85, 117, 171, 174, 271	PM_{10}	2003.03.15–2004.03.14
MC 078	MC 10, 32, 42, 77, 85, 117, 174	TSP	1995.01.30–1996.01.30
MC 080	MC 10, 32, 42, 77, 85, 117, 174	TSP	1995.03.27–1997.03.26
MC 081	MC 10, 32, 42, 77, 85, 117, 174	TSP	1995.04.04–1996.04.03
MC 083	MC 10, 32, 42, 77, 85, 117, 174	TSP	1995.01.30–1996.01.30
MC 145	MC 10, 32, 42, 77, 85, 117, 171, 174	PM_{10}	2002.03.27–2004.03.09

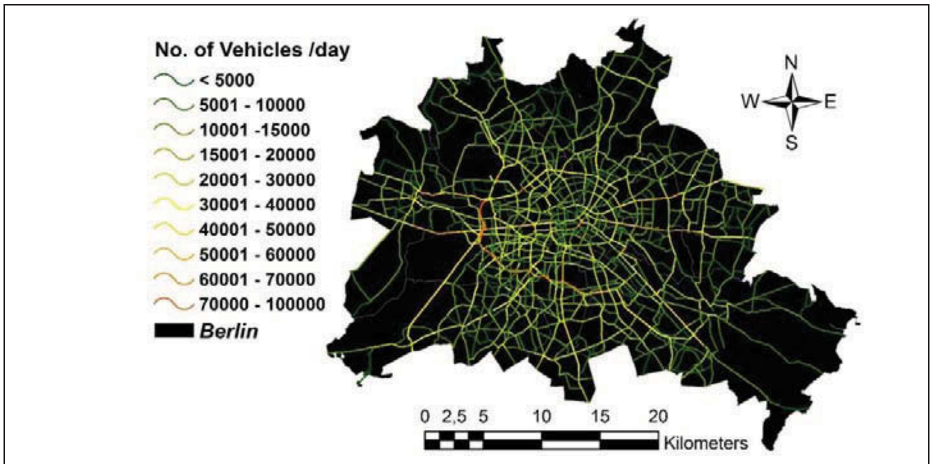

Figure 2. *The average of traffic intensity in Berlin (No. of vehicles/day) in 2009.*

Step 1. The database *(D)* of input candidate variables and corresponding output variable data has been prepared in the previous section. Each row in Table 2 is equal to one input–output database. In this section, the simulation algorithm for one database is explained, hence for simulation of 20 databases in Table 2, the simulation algorithm must be implemented on 20 databases one by one.

Imagine that the database has n input variables ($\{X=X_1, X_2,\ldots, X_n\}$) and one output variable (Y). Thus, D can be expressed as Equation (1):

$$D=\left\{\left(x_k^m, y^m\right)\right\}, m=1,\ldots,M, k=1,\ldots,n \quad (1)$$

where, X_k^m is the m^{th} member of the k^{th} variable (X_k) ($X_k^m \in X_k$ *and* $X_k, \in X$), y^m is the m^{th} member of Y and M is the total number of observations.

Step 2. A random partitioning program was developed for the random partitioning of the databases. The random number generator in this program was based upon the uniform probability density function. The database was randomly partitioned to the train (two third of database) and test (one third of database) databases using this program. Hereinafter, the train database is called the database.

Step 3. The idea of the developed new input variable selection scheme in this study is that a heuristic partitioning method is utilized for the partitioning of the main MISO (Multi Inputs–Single Output) database for some MISO sub–databases in a successive manner. Each MISO sub–database is converted to some SISO (Single Input–Single Output) sub–databases and the behavior of output is investigated in each SISO. It means that the space of the input variables is divided into many sub–databases and the relation between each input variable and output variable is evaluated in each small sub–database; accordingly, the influence of each input variable on the output is calculated in detail in this new heuristic input selection technique. Thus, the heuristic dividing is one of the major steps of this fuzzy–heuristic input variable selection algorithm, which is explained below.

In the first iteration of the input variable selection algorithm, there is only one database (D) and it is divided into two smaller databases. In general, a generated database is expressed as D_k^{ds} and it is the s^{th} database in the d^{th} iteration and has been generated by dividing the k^{th} variable of a bigger database. The bigger database has been divided into two parts ($s\in\{1, 2\}$) and this database is the s^{th} part.

When any of the input variables (X_k) are divided into two parts, then D is divided into two sub–databases (D_k^{11}, D_k^{12}). D_k^{11} *and* D_k^{12} are the databases, generated by dividing the k^{th} variable in the first iteration.

$$D=\{D_k^{11}, D_k^{12}\} \quad (2)$$

$$D_k^{11}=\{(x_l^t, y^t)\}, t=1,\cdots,Q^1;\ l=1,\cdots,n \ if\ X_k\le T_k^1 \quad (3)$$

$$D_k^{12}=\{(x_l^t, y^t)\}, t=1,\cdots,Q^1;\ l=1,\cdots,n \ if\ X_k> T_k^1 \quad (4)$$

where, T_k^1 is the median of X_k in database D. Q^1 is the number of observations in each database and it is equal to $M/2$.

In the second iteration, it is decided which database should be divided into two smaller databases (D_k^{11} *or* D_k^{12}). Imagine D_k^{11} is selected for dividing and it is divided to two smaller databases ($D_{k'}^{21}$, $D_{k'}^{22}$, $k'\in\{1,\cdots,n\}$). $D_{k'}^{21}$ and $D_{k'}^{22}$ are the databases, generated by dividing the k'^{th} variable of D_k^{11} in the second iteration. In the second iteration, D has been divided to three databases as below:

$$D=\{D_{k'}^{21}, D_{k'}^{22}, D_k^{12}\} \quad (5)$$

$$D_{k'}^{21}=\{(x_l^t, y^t)\}, t=1,\cdots,Q^2, l=1,\cdots,n \ if\ (X_k\le T_k^1\ \&\ X_{k'}\le T_{k'}^2) \quad (6)$$

$$D_{k'}^{22}=\{(x_l^t, y^t)\}, t=1,\cdots,Q^2, l=1,\cdots,n \ if\ (X_k\le T_k^1\ \&\ X_{k'}> T_{k'}^2) \quad (7)$$

$$D_k^{12}=\{(x_l^t, y^t)\}, t=1,\cdots,Q^1;\ l=1,\cdots,n \ if\ X_k> T_k^1 \quad (8)$$

where, Q^2 is the number of observations in each database and is equal to $\frac{Q^1}{2}$. $T_{k'}^2$ is the median of $X_{k'}$ in database D_k^{11}.

This algorithm is iterated and the D is divided into more small databases. In general, D in the d^{th} iteration is divided into d+1 small databases.

Which database is selected for dividing into two smaller databases in each step and which X_k is the best one to divide the selected database? First, the method for the determination of the appropriate X_k for dividing a database is explained here. Then, the method for the selection of a database for dividing will be explained in Step 4.

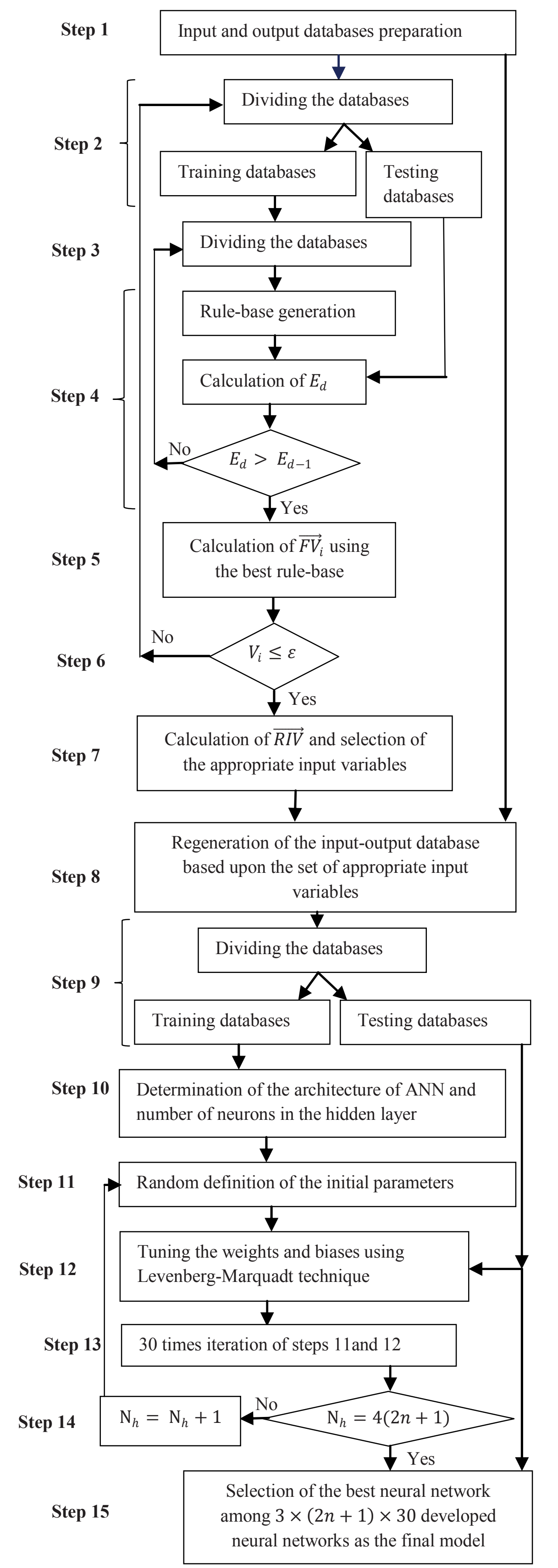

***Figure 3.** The flowchart of the algorithm of study.*

For the determination of the best option, all of the possible dividing options are performed. Hence, for dividing the D_k^{ds} into two smaller databases, n possible options are performed and $2n$ databases are generated. The data in each generated database are divided into n one–variable databases. Thus, when the j^{th} variable is divided into two small databases, $2n$ one–variable databases (S) are generated.

$$S_{js}^{i} = \{(x_i^{t,j}, y^t)\}, i = 1,\cdots,n, t = 1,\cdots,Q^d, s = 1,2 \tag{9}$$

Q^d is the number of data in the one–variable database.

The relationship between X_i and Y ($\hat{Y}_i = f_i^j(X_i), i = 1,\cdots,n$) in all of S_{j1}^{i} is calculated by a fuzzy curve fitting technique, called IDS (Ink Drop Spread) (Bagheri Shouraki and Honda, 1999). Similarly, the relationship between X_i and Y ($\hat{Y}_i = g_i^j(X_i), i = 1,\cdots,n$) in all of S_{j2}^{i} is calculated. The accuracy of f_i^j and g_i^j functions (one– variable functions) for the estimation of output (Y) is evaluated. Consequently, f_z^j and $g_{z'}^j$ ($z \,\&\, z' \in \{1,\cdots,n\}$) are determined as the as the best one–variable functions with the lowest errors, respectively. If we consider e^j as the total error of the output (Y) estimation in D_k^{ds} by f_z^j and $g_{z'}^j$, then e^j for $j = 1,\cdots,n$ is calculated and the minimum value in $\{e^1,\cdots,e^n\}$ is determined. Consider $e^{k'}$ as the minimum. Consequently, the input variable corresponding to the minimum error ($X_{k'}$) is the best variable for dividing D_k^{ds} into two smaller databases ($D_{k'}^{d+1,1}$ *and* $D_{k'}^{d+1,2}$) and $f_z^{k'}(X_z)$ and $g_{z'}^{k'}(X_{z'})$ are the best one–variable functions for the estimation of output in the two generated databases and $e(f_z^{k'})$ and $e(g_{z'}^{k'})$ are their corresponding errors, respectively.

Step 4. In the first iteration of the dividing algorithm, D is divided into two databases [see Equations (2)–(4)]. Then two one–variable functions ($f_z^k(X_z)$ and $g_{z'}^k(X_{z'})$) are determined and utilized for the output estimation in two databases. The error of the one–variable functions are $e(f_z^{k'})$ and $e(g_{z'}^{k'})$. Therefore, the rule–base can be expressed as Equation (10):

$$\begin{cases} If\ X_k \leq T_k^1\ Then\ \hat{Y}_1 = f_z^k(X_z) \\ If\ X_k > T_k^1\ Then\ \hat{Y}_2 = g_{z'}^k(X_{z'}) \end{cases} \tag{10}$$

Using the test database, the accuracy of generated rule–base [Equation (10)] for the estimation of the output variable (Y) is evaluated. The error of output estimation in the first iteration is expressed as E_1.

In the second iteration of the dividing algorithm, the database with higher error is selected for dividing. Imagine $e(f_z^{k'}) > e(g_{z'}^{k'})$, then, D_k^{11} must be divided to two smaller databases using the dividing method, explained in Step 3. Thus, two one–variable functions [$f_{z_1}^{k'}(X_{z_1})$ and $g_{z_2}^{k'}(X_{z_2})$] are determined and utilized for the output estimation in the two databases. Accordingly, D is divided to three databases [Equation (5)], and a rule–base with three rules [Equation (11)] is generated. The error of these one variable functions are $e(f_{z_1}^{k'})$,$e(g_{z_2}^{k'})$ and $e(g_{z'}^{k'})$.

$$\begin{cases} if\ (X_k \leq T_k^1\ \ X_{k'} \leq T_{k'}^2)\ Then\ \hat{Y}_1 = f_{z_1}^{k'}(X_{z_1}) \\ if\ (X_k \leq T_k^1\ \ X_{k'} > T_{k'}^2)\ Then\ \hat{Y}_2 = g_{z_2}^{k'}(X_{z_2}) \\ \quad if\ X_k > T_k^1\ Then\ \hat{Y}_3 = g_{z'}^{k}(X_{z'}) \end{cases} \tag{11}$$

Using the test database, the accuracy of generated rule–base [Equation (11)] for the estimation of the output variable (Y) is evaluated. The error of the output estimation, in the second iteration is expressed as E_2. This dividing procedure and the rule–base generation (Steps 3 and 4) are continued until $E_d > E_{d-1}$.

Step 5. The generated rule–base with d–1 rules in Step 4 is considered as the best rule–base in the first iteration of algorithm, and it is expressed as R_1=d–1. In this rule–base, the number of the estimated data by a variable (fv_i) can be calculated by Equation (12):

$$fv_{\mathrm{i}} = \sum_{j=1}^{d-1}(L_j \times Nr_j) \tag{12}$$

where, Nr_j is the number of data in the j^{th} database and L_j is expressed as Equation (13):

$$L_j = \begin{cases} 0\ if\ \hat{Y}_j = f(X_k)\ and\ k \neq i \\ 1\ if\ \hat{Y}_j = f(X_k)\ and\ k = i \end{cases} \tag{13}$$

where, the $f(X_h)$ is one–variable function in the j^{th} rule in the rule–base and $k \in \{1, \ldots, n\}$. Consequently, the results of calculation of fv_i for whole variables can be presented as the function vector ($\overrightarrow{FV_1} = [fv_1, fv_2, \ldots, fv_n]$).

Step 6. Combine the train and test databases to generate the original database, and then proceed to Step 2. Steps 2–6 are iterated according to the user defined criterion of iterations. Here, when $V_P < \varepsilon$, the iteration procedure is terminated, where p is the number of the performed iterations, $V_p = (\sum_{k=1}^{n} V_k)/n$ and V_k is the k^{th} member of $\vec{V} = [V_1, \ldots, V_n]$ $\left(\vec{V} = \left|(\sum_{o=1}^{p}\overrightarrow{FV_O})/p - (\sum_{o=1}^{p-1}\overrightarrow{FV_O})/(p-1)\right|\right.$. When $V_P \geq \varepsilon$ then p=I or I is the total number of iterations. Ɛ can be defined by user and it was considered equal to 0.01n/2 in this study, where n is equal to the number of input variables. These iterations neutralize the effects of the random dividing in the second step, and generalize the results. Steps 2–6 are iterated I times. Thus, I function vectors ($\overrightarrow{FV_1}, \ldots, \overrightarrow{FV_I}$) are generated.

Step 7. Input selection: the average of I function vectors is calculated ($\overline{FV}$) . Then $\overline{FV}$ is normalized as Equation (14):

$$\overrightarrow{RIV} = \frac{\overline{FV}}{S_{FV}} \tag{14}$$

where, S_{FV} is the summation of $\overline{FV}$ elements and $\overrightarrow{RIV}$ is the relative importance vector and its elements show the relative importance of the different input variables for the modeling of the output variable. Finally, the variables with low importance are removed from database and the set of suitable variables for modeling are determined.

Step 8. The new input–output database is regenerated based upon the set of suitable input variables, determined in Step 7.

Step 9. The new database is randomly partitioned to the train (two third of database) and test (one third of database) databases.

Step 10. Kolmogorov's theorem expresses that any continuous function with any number of variables can be represented as finite sum of one–variable continuous functions (Kolmogorov, 1957). Sprecher's theorem (Sprecher, 1965) is a refinement of Kolmogrov's theorem and it shows that one–variable continuous functions in Kolmogrov's theorem can be replaced with monotonic increasing functions. Hecht–Nielsen (1987) reformulated the Sprecher's theorem into the form of feed forward neural network and showed that any continuous function with any number of variables (n) can be exactly represented by a three layered–neural network with 2n+1 neurons in the hidden layer with monotonic increasing activation function. The initial structure of the ANN in this study is determined based upon Sprecher's theorem and the Hecht–Nielsen re–formulation. Hence, feed forward neural networks with one hidden layer are utilized for modeling in this study, and monotonic increasing function (hyperbolic tangent sigmoid function) and linear functions are employed as the activation functions for the hidden and the output layers, respectively. In addition, the initial number of neurons in hidden layer is considered equal to 2n+1.

Step 11. The initial weights and biases of neural network is randomly determined.

Step 12. The initial weights and biases of neural network are tuned using Levenberg–Marquadt technique. After each training step (epoch), the performance of the model is evaluated by the testing dataset. If the error of the model is less than the previous step, then the training procedure is continued and the next training step is performed, otherwise the training procedure is terminated. The trained neural network is a candidate for final neural network.

Step 13. Steps 11 and 12 are iterated 30 times and consequently 30 trained neural networks are generated. The 30 trained neural networks are evaluated using the test dataset and the best one is selected as a candidate for the final neural network.

Step 14. Add one neuron to the hidden layer and then proceed to Step 11. This procedure is continued until N_h=4(2n+1) (N_h: the number of neurons in the hidden layer).

Step 15. The 3(2n+1) developed candidate neural networks in Step 13 are compared using the test dataset and the best one is selected as the final neural network model and it is utilized for simulation of shut–down station.

5. Results and Discussion

The fuzzy–heuristic input selection technique was implemented on the 20 hourly PM_{10} concentration databases and $\overrightarrow{RIV}$ for the each database was determined. The variable with the relative importance of less than half of mean value of $\overrightarrow{RIV}$ were removed, and the remaining variables were selected as the appropriate input variables for modeling. The results of the input selection have been presented in Table 3. It shows that the traffic stations (MC 117 and MC 174) have almost not been selected as the appropriate input variables. But at least one of the urban background stations (MC 010 and MC 042) has been selected as the appropriate input variables for simulation of almost all of the shut–down stations. The PM_{10} from traffic sources can influence the surrounding areas within a radius of less than 100 m (Hewitt and Jackson, 2008), and the PM_{10} measurements in a traffic station have been greatly influenced by local effects, hence traffic stations are not suitable for the simulation of other stations. In addition, the EU (2008) has pointed out that urban background stations must be representative for several square kilometers. Consequently, these stations are representative for large areas and are also suitable candidates for the simulation of the shut–down stations. The results presented in Table 3 demonstrate the ability of the utilized fuzzy–heuristic input selection technique for the suitable selection of the input variables.

Then, the neural network models were trained and tested based upon the input and output variables presented in Table 3. The number of neurons in the hidden layers of the 20 optimum neural network models have been presented in Table 3.

The results of the training of optimum neural network models of 20 shut–down PM stations for estimation of hourly PM_{10} concentration have been presented in Table S1 (see the SM). In addition, the characteristics of the test datasets and the results of the testing of the 20 optimum neural network models for the simulation of the hourly PM_{10} concentration have been presented in Table 4.

Table 3. *The results of input variable selection by the fuzzy–heuristic technique and the number of neurons in hidden layer of optimum neural network models (Nr. Hid.)*

Output Variable	Appropriate Input Variables	Nr. Hid.
MC 001	MC 10, 42, 77	15
MC 006	MC 10, 77, 85	9
MC 007	MC 10, 32, 77	13
MC 009	MC 10, 42	16
MC 011	MC 10	10
MC 017	MC 10, 32, 42	21
MC 018	MC 42	7
MC 020	MC 10, 42	13
MC 023	MC 10, 32, 42, 85	26
MC 024	MC 42, 85	6
MC 025	MC 10, 42	7
MC 027	MC 10, 32, 42, 171	27
MC 028	MC 10, 42	10
MC 030	MC 42	3
MC 072	MC 10	3
MC 078	MC 77	5
MC 080	MC 10, 42, 77	9
MC 081	MC 10, 42, 77, 85	17
MC 083	MC 85	10
MC 145	MC 10, 32, 77	21

The R, MBE, MAE, RMSE and MAPE values in Table S1 and Table 4 represent the correlation coefficient, Mean Bias Error, Mean Absolute Error, Root Mean Square Error and Mean Absolute of Percentage Error, respectively. These goodness of fit criteria are expressed as Equations (15)–(19):

$$R = \frac{\sum_{i=1}^{M}(O_i-\bar{O})(S_i-\bar{S})}{\sqrt{\sum_{i=1}^{M}(O_i-\bar{O})^2} \times \sqrt{\sum_{i=1}^{M}(S_i-\bar{S})^2}} \tag{15}$$

$$MBE = \frac{1}{M}\sum_{i=1}^{M}(O_i - S_i) \tag{16}$$

$$MAE = \frac{1}{M}\sum_{i=1}^{M}|O_i - S_i| \tag{17}$$

$$RMSE = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(O_i - S_i)^2} \tag{18}$$

$$MAPE = \frac{1}{M}\sum_{i=1}^{M}\frac{|O_i - S_i|}{|O_i|} \times 100 \tag{19}$$

where, M is the total number of observation data, $\bar{O}$ and $\bar{S}$ are the average of the observed and simulated PM_{10} concentration, and O_i and S_i are the observed and simulated PM_{10} concentration of the i^{th} data, respectively.

The high similarity between the training (see the SM, Table S1) and testing (Table 4) results demonstrates that over–fitting and under–fitting have not been occurred in the developed optimum neural network models. Many uncertainties and errors (about 4–20%) are associated with PM_{10} measurements in daily scale with different instruments (Stalker and Dickerson, 1962; Heal et al., 2000; Hitzenberger et al., 2004; Baxter et al., 2007; Lagler et al., 2011; Pernigotti et al., 2013) and it is clear that the uncertainty of hourly PM_{10} measurement is higher than daily measurement. Thus, the results of hourly simulation (Average error: 27.7%; correlation coefficient $\geq$ 0.82; average bias: –0.03 μg/m^3; average absolute error: 7.4 μg/m^3) (calculated using Table 4) seem to be very good. In addition, the scatter–plots of the measured and simulated hourly PM_{10} concentration (μg/m^3) in the testing phase in some of the studied stations have been presented in Figure S3 (see the SM) and one sample of each type of the simulated stations (MC 018: Urban–background; MC 025: Suburban–background; MC 027: Rural near city–background; MC 072: Suburban–traffic) have been exhibited in it. The scatter–plots imply the suitability of the performed simulations in the different stations.

Consequently, a dense monitoring network can be developed by the combination of the simulated and current stations. In addition, this suitable result implies that the fuzzy–heuristic input selection technique has selected the appropriate input variables. The locations of the PM_{10} stations of the developed dense monitoring network have been presented in Figure 4. An automatic module has been developed by the combination of the 20 developed neural network models and this module which uses the hourly PM_{10} concentration in the current stations as the input variables, and, accordingly, its output is the concentration of hourly PM_{10} in the 20 shut–down (simulated) stations.

When the time–scale increases, the accuracy of the simulated stations for the PM_{10} estimation also increases because the variability in the data decreases. The hourly PM_{10} simulated data were converted to the daily PM_{10} data and the performance of the developed neural network models for the simulation of daily PM_{10} data were evaluated. The performance of the 20 daily PM_{10} models has been presented in Table S2 (see the SM). The average of MAE values in Table S2 is about 2.6 μg/m^3, while the uncertainty of daily PM_{10} measurement in Berlin is about 1.8 μg/m^3. The comparison between the MAE values of simulations and the uncertainty of the measurements and also the results of Table S2 ($R \geq 0.94$, $MAPE<11.5\%$ and $MAE<3.9$ μg/m^3) demonstrate that the simulated stations have excellent performance for the estimation of the daily PM_{10} concentration.

Using the developed automatic module, the hourly PM_{10} concentration of the 20 simulated stations were estimated for 2012 and then the mean annual PM_{10} concentration of all the stations (the 12 current stations and the 20 simulated stations) were calculated and presented in Figure 4. Now, Berlin has a dense monitoring network with a sufficient number of suburban, urban and traffic stations, and, accordingly, it is possible to calculate the reliable mean annual PM_{10} concentration for the different areas in Berlin. For example, the mean annual PM_{10} concentration for suburban (combination of suburban and rural stations), urban and traffic areas (see Table 1) in 2012 (Figure 4) are 19.8, 22.5 and 25.2 μg/m^3, respectively.

The EU (2008) has pointed out that improvement in the monitoring and assessment of the air quality in the whole environment is important. The still operating network in Berlin has a small number of monitoring stations and there is also no suburban–traffic station in the still operating network and it is not possible to evaluate the air pollution conditions in the suburban–traffic areas. But the developed dense monitoring network has sufficient stations (e.g., 3 stations in the suburban–traffic areas), and it is capable of providing better assessment and monitoring of PM_{10} concentration for the different areas in Berlin in a manner which is free–of–charge. In addition, when the information from some of the air quality monitoring stations is supplemented by modeling, then the number of monitoring stations can be reduced (EU, 2008). According to the findings of this study, it is possible to re–construct 20 shut–down stations. If only seven of the still operating stations (MC 010, 032, 042, 077, 085, 117 and 174) operate and the others are shut down, the developed PM_{10} monitoring network has 27 stations (more than 2.0 times more than the number of stations in the current monitoring network). Hence, the findings of this study are very useful for the decrease of the number of monitoring stations and, consequently, the monitoring cost.

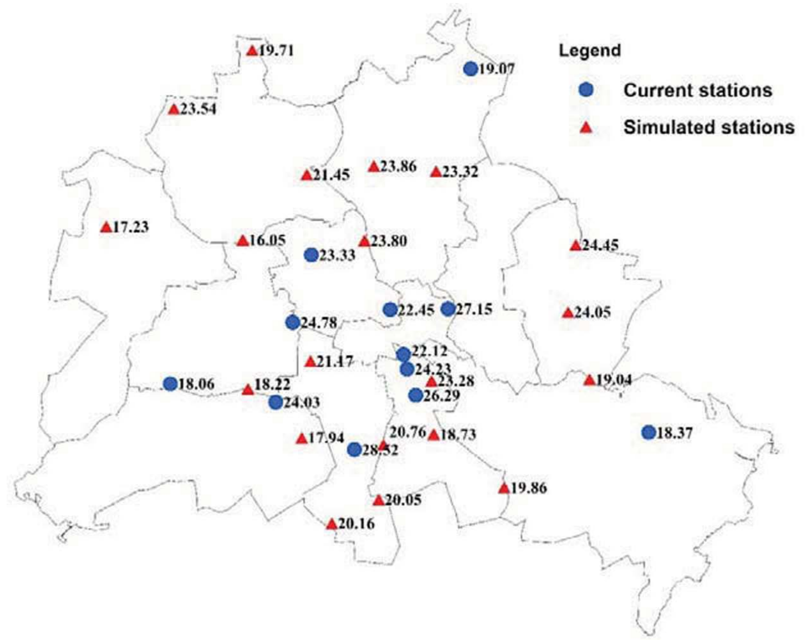

Figure 4. *The developed dense PM_{10} monitoring network (12 current stations and 20 simulated stations) with the mean annual PM_{10} concentration (μg/m^3) for 2012.*

Table 4. *Characteristics of test datasets and the results of the testing of optimum neural network models of 20 shut–down PM stations for estimation of hourly PM_{10} concentration (M: Number of hourly test PM_{10} data; Min, Max: Minimum and maximum hourly PM_{10} concentration (μg/m^3) in the test database, respectively; M.h: Mean of hourly test PM_{10} data (μg/m^3); Sd.h.: Standard deviation of hourly test PM_{10} data (μg/m^3))*

Shut–Down Station	M	Min	Max	M.h.	Sd.h.	R	MBE (μg/m^3)	MAE (μg/m^3)	RMSE (μg/m^3)	MAPE (%)
MC 001	1 981	3.6	204	43.7	27.0	0.87	–0.12	9.48	13.15	30.40
MC 006	3 147	2.8	167	37.0	23.6	0.87	0.02	7.89	11.65	27.48
MC 007	2 113	1.6	163	33.9	22.8	0.89	–0.31	7.22	10.24	31.03
MC 009	2 489	1.6	176	36.9	27	0.90	0.05	8.42	11.57	38.74
MC 011	4 206	2.4	180	36.1	24	0.82	0.10	8.37	13.65	27.59
MC 017	4 186	2	170	35.4	23.6	0.91	–0.02	6.68	9.76	25.65
MC 018	5 002	2.5	106	24	14.5	0.94	0.03	3.52	5.04	18.09
MC 020	2 165	4	162	40.5	23.4	0.88	–0.02	7.82	11.00	24.81
MC 023	1 627	2.8	187.2	39.7	24.5	0.90	–0.09	7.67	10.80	27.66
MC 024	2 743	2.4	127.6	32.6	18.5	0.84	0.02	7.46	10.19	32.22
MC 025	3 794	2	171	36.5	24.4	0.89	0.06	7.73	11.14	29.71
MC 027	3 910	3	106	23.4	14.0	0.95	0.02	3.33	4.53	17.29
MC 028	2 462	3.2	144	33.5	19.2	0.82	–0.09	7.82	10.85	31.50
MC 030	2 421	2.4	142	33	18.6	0.81	–0.22	8.24	10.95	34.75
MC 072	3 617	3	181	35.3	23.1	0.91	0.09	6.21	9.68	20.13
MC 078	1 657	3.2	140	39	22.5	0.83	–0.45	8.66	12.47	31.53
MC 080	3 720	3.6	215	44.2	27.7	0.88	0.20	8.98	13.19	26.25
MC 081	1 733	3.6	194	51.7	30	0.88	0.14	9.91	14.24	28.16
MC 083	1 704	2.4	155	36	22.3	0.85	–0.19	8.34	11.90	33.30
MC 145	3 526	2.5	140	28	20.6	0.96	0.10	3.88	5.49	18.59

6. Conclusions

The developed fuzzy–heuristic input selection was joined with the ANN and applied for the simulation of the 20 shut–down stations. The appropriate results of hourly (Error: 27.7%; correlation coefficient ≥0.82; average bias: –0.03 μg/m^3; average absolute error: 7.4 (μg/m^3)) and daily (R≥0.94, MAPE<11.5% and MAE<3.9 μg/m^3 simulations revealed that the coupling of the ANN and this new automatic input variable selection technique is a fast, straightforward and reliable tool for simulation of non–linear systems. In addition, the presented new and free–of–charge methodology for the densifying of the PM_{10} monitoring network of Berlin was successfully implemented. The 20 shut–down PM_{10} stations were re–constructed and a dense PM_{10} monitoring network was developed by the combination of the 12 still –operating stations and the 20 re–constructed stations for Berlin. Now, more reliable PM_{10} monitoring is possible because of the development of the dense monitoring network with a sufficient number of suburban, urban and traffic stations. The findings of this study are very useful in the light of the decrease in the number of monitoring stations and monitoring costs based upon the European Union directive on air quality.

Acknowledgements

The authors are grateful to the Alexander von Humboldt Stiftung/Foundation for funding this work under Humboldt ID 1149622. The authors thank Chris Engert for his valuable proof–reading of this paper.

Supporting Material Available

The sources of PM_{10} in Berlin (Figure S1), Schematic diagram of two major stages of this study with the input and output variables in each stage (Figure S2), The scatter plots of measured and simulated PM_{10} in the testing phase in different stations (Figure S3), The results of the training of optimum neural network models of 20 shut–down PM stations for estimation of hourly PM_{10} concentrations (Table S1), The performance of the optimum neural network models for estimation of daily PM_{10} concentration calculated using the testing database (Table S2). This information is available free of charge via the Internet at http://www.atmospolres.com.

References

Bagheri Shouraki, S., Honda, N., 1999. Recursive fuzzy modeling based on fuzzy interpolation. *Journal of Advanced Computational Intelligence* 3, 114–125.

Baxter, R., Bush, D., Knuth, W., Fransioli, P., Yoho, D., 2007. Particulate Matter (PM_{10}) Saturation Monitoring Study, Final Report No. P.O. 212114, T & B Systems, Santa Rosa, 143 pages.

Beaulant, A.L., Perron, G., Kleinpeter, J., Weber, C., Ranchin, T., Wald, L., 2008. Adding virtual measuring stations to a network for urban air pollution mapping. *Environment International* 34, 599–605.

Bell, D.A., Wang, H., 2000. A formalism for relevance and its application in feature subset selection. *Machine Learning* 41, 175–195.

Blum, A.L., Langley, P., 1997. Selection of relevant features and examples in machine learning. *Artificial Intelligence* 97, 245–271.

Bowden, G.J., Maier, H.R., Dandy, G.C., 2005. Input determination for neural network models in water resources applications. Part 2. Case study: Forecasting salinity in a river. *Journal of Hydrology* 301, 93–107.

Briggs, D.J., de Hoogh, C., Guiliver, J., Wills, J., Elliott, P., Kingham, S., Smallbone, K., 2000. A regression–based method for mapping traffic–related air pollution: Application and testing in four contrasting urban environments. *Science of the Total Environment* 253, 151–167.

Briggs, D.J., Collins, S., Elliott, P., Fischer, P., Kingham, S., Lebret, E., Pryl, K., Van Reeuwijk, H., Smallbone, K., VanderVeen, A., 1997. Mapping urban air pollution using GIS: A regression–based approach. *International Journal of Geographical Information Science* 11, 699–718.

Chen, C.H., Liu, W.L., Chen, C.H., 2006. Development of a multiple objective planning theory and system for sustainable air quality monitoring networks. *Science of the Total Environment* 354, 1–19.

Cocheo, C., Sacco, P., Ballesta, P.P., Donato, E., Garcia, S., Gerboles, M., Gombert, D., McManus, B., Patier, R.F., Roth, C., de Saeger, E., Wright, E., 2008. Evaluation of the best compromise between the urban air quality monitoring resolution by diffusive sampling and resource requirements. *Journal of Environmental Monitoring* 10, 941–950.

Corani, G., 2005. Air quality prediction in Milan: Feed–forward neural networks, pruned neural networks and lazy learning. *Ecological Modelling* 185, 513–529.

Darby, W.P., Ossenbruggen, P.J., Gregory, C.J., 1974. Optimization of urban air monitoring networks. *Journal of the Environmental Engineering Division* 100, 577–591.

Dugord, P.A., Lauf, S., Schuster, C., Kleinschmit, B., 2014. Land use patterns, temperature distribution, and potential heat stress risk–the case study Berlin, Germany. *Computers Environment and Urban Systems* 48, 86–98.

Dutot, A.L., Rynkiewicz, J., Steiner, F.E., Rude, J., 2007. A 24–h forecast of ozone peaks and exceedance levels using neural classifiers and weather predictions. *Environmental Modelling & Software* 22, 1261–1269.

EU (European Union), 2008. Directive 2008/50/EC Of The European Parliament and of the Council of 21 May 2008 on Ambient Air Quality and Cleaner Air for Europe. *Official Journal of European Union* L152, 1–44.

Elangasinghe, M.A., Singhal, N., Dirks, K.N., Salmond, J.A., 2014. Development of an ANN–based air pollution forecasting system with explicit knowledge through sensitivity analysis. *Atmospheric Pollution Research* 5, 696–708.

Ferradas, E.G., Minarro, M.D., Terres, I.M.M., Martinez, F.J.M., 2010. An approach for determining air pollution monitoring sites. *Atmospheric Environment* 44, 2640–2645.

Gardner, M.W., Dorling, S.R., 1998. Artificial neural networks (the multilayer perceptron)–a review of applications in the atmospheric sciences. *Atmospheric Environment* 32, 2627–2636.

Giordano, F., Rocca, M.L., Perna, C., 2014. Input variable selection in neural network models. *Communications in Statistics–Theory and Methods* 43, 735–750.

Goldstein, I.F., Landovitz, L., 1977. Analysis of air pollution patterns in New York City—I. Can one station represent the large metropolitan area? *Atmospheric Environment* 11, 47–52.

Gorgen, R., Lambrecht, U., 2007. Particulate matter in ambient air. *Journal for European Environmental & Planning Law* 4, 278–288.

Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. *The Journal of Machine Learning Research* 3, 1157–1182.

Heal, M.R., Beverland, I.J., McCabe, M., Hepburn, W., Agius, R.M., 2000. Intercomparison of five PM_{10} monitoring devices and the implications for exposure measurement in epidemiological research. *Journal of Environmental Monitoring* 2, 455–461.

Hecht–Nielsen, R., 1987. Kolmogorov's mapping neural network existence theorem. *Proceedings of 1st IEEE International Conference on Neural Networks,* San Diego, pp. 11–14.

Hewitt, C.N., Jackson, A.V., 2008. *Handbook of Atmospheric Science: Principles and Applications*, John Wiley & Sons.

Hickey, H., Rowe, W., Skinner, F., 1971. A cost model for air quality monitoring systems. *Journal of the Air Pollution Control Association* 21, 689–693.

Hitzenberger, R., Berner, A., Galambos, Z., Maenhaut, W., Cafmeyer, J., Schwarz, J., Muller, K., Spindler, G., Wieprecht, W., Acker, K., Hillamo, R., Makela, T., 2004. Intercomparison of methods to measure the mass concentration of the atmospheric aerosol during INTERCOMP2000 – influence of instrumentation and size cuts. *Atmospheric Environment* 38, 6467–6476.

Hougland, E.S., Stephens, N.T., 1976. Air pollutant monitor siting by analytical techniques. *Journal of the Air Pollution Control Association* 26, 51–53.

Kanaroglou, P.S., Jerrett, M., Morrison, J., Beckerman, B., Arain, M.A., Gilbert, N.L., Brook, J.R., 2005. Establishing an air pollution monitoring network for intra–urban population exposure assessment: A location–allocation approach. *Atmospheric Environment* 39, 2399–2409.

Kohavi, R., John, G.H., 1997. Wrappers for feature subset selection. *Artificial Intelligence* 97, 273–324.

Kolmogorov, A.N., 1957. On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition. *Doklady Akademii Nauk USSR* 114, 953–956.

La Rocca, M., Perna, C., 2005. Variable selection in neural network regression models with dependent data: A subsampling approach. *Computational Statistics & Data Analysis* 48, 415–429.

Lagler, F., Belis, C., Borowiak A., 2011. A Quality Assurance and Control Program for $PM_{2.5}$ and PM_{10} Measurements in European Air Quality Monitoring Networks, JCR Scientific and Technical Research Report No. JCR65176, European Union, 115 pages.

Lal, B., Tripathy, S.S., 2012. Prediction of dust concentration in open cast coal mine using artificial neural network. *Atmospheric Pollution Research* 3, 211–218.

Leavitt, J., Pooler Jr, F., Wanta, R., 1957. Design and interim meteorological evaluation of a community network for meteorological and air quality measurements. *Journal of the Air Pollution Control Association* 7, 211–215.

Lenschow, P., Abraham, H.J., Kutzner, K., Lutz, M., Preuss, J.D., Reichenbacher, W., 2001. Some ideas about the sources of PM_{10}. *Atmospheric Environment* 35, S23–S33.

Lozano, A., Usero, J., Vanderlinden, E., Raez, J., Contreras, J., Navarrete, B., El Bakouri, H., 2009. Design of air quality monitoring networks and its application to NO_2 and O_3 in Cordova, Spain. *Microchemical Journal* 93, 211–219.

May, R., Dandy, G., Maier, H., 2011. Review of input variable selection methods for artificial neural networks, in *Artificial Neural Networks—Methodological Advances and Biomedical Applications*, edited by Suzuki, K., InTech, pp. 19–44.

Mazzeo, N.A., Venegas, L.E., 2008. Design of an air–quality surveillance system for Buenos Aires City integrated by a NO_x monitoring network and atmospheric dispersion models. *Environmental Modeling & Assessment* 13, 349–356.

Modak, P.M., Lohani, B.N., 1985. Optimization of ambient air–quality monitoring networks .1. *Environmental Monitoring and Assessment* 5, 1–19.

Nejadkoorki, F., Baroutian, S., 2012. Forecasting extreme PM_{10} concentrations using artificial neural networks. *International Journal of Environmental Research* 6, 277–284.

Ott, W.R., 1977. Development of criteria for siting air monitoring stations. *Journal of the Air Pollution Control Association* 27, 543–547.

Papanastasiou, D.K., Melas, D., Kioutsioukis, I., 2007. Development and assessment of neural network and multiple regression models in order to predict PM_{10} levels in a medium–sized Mediterranean city. *Water Air and Soil Pollution* 182, 325–334.

Pernigotti, D., Gerboles, M., Belis, C.A., Thunis, P., 2013. Model quality objectives based on measurement uncertainty. Part II: NO_2 and PM_{10}. *Atmospheric Environment* 79, 869–878.

Pires, J.C.M., Pereira, M.C., Alvim–Ferraz, M.C.M., Martins, F.G., 2009. Identification of redundant air quality measurements through the use of principal component analysis. *Atmospheric Environment* 43, 3837–3842.

Pires, J.C.M., Sousa, S.I.V., Pereira, M.C., Alvim–Ferraz, M.C.M., Martins, F.G., 2008a. Management of air quality monitoring using principal component and cluster analysis–Part I: SO_2 and PM_{10}. *Atmospheric Environment* 42, 1249–1260.

Pires, J.C.M., Sousa, S.I.V., Pereira, M.C., Alvim–Ferraz, M.C.M., Martins, F.G., 2008b. Management of air quality monitoring using principal component and cluster analysis – Part II: CO, NO_2 and O_3. *Atmospheric Environment* 42, 1261–1274.

Pope, R., Wu, J.G., 2014. A multi–objective assessment of an air quality monitoring network using environmental, economic, and social indicators and GIS–based. *Journal of the Air & Waste Management Association* 64, 721–737.

Rauterberg–Wulff, A., Lutz, M., Nulis, E., Reichenbacher, W., Kettschau,A., Schlickum, V., Kerschbaumer, A., Couturier, G., Jarnott, F., Gerike, S.,Lehming, B., Rose, B., Nothard, R., Grunow, K., von Stülpnagel, A., Preuss, M., Horn, B., Gutierrez Merino, S., 2013. Berlin's air quality plan 2011 to 2017, Senate Department for Urban Development and Environmental Communication, Berlin, 227 pages (In German).

Reich, S.L., Gomez, D.R., Dawidowski, L.E., 1999. Artificial neural network for the identification of unknown air pollution sources. *Atmospheric Environment* 33, 3045–3052.

Russo, A., Lind, P.G., Raischel, F., Trigo, R., Mendes, M., 2015. Neural network forecast of daily pollution concentration using optimal meteorological data at synoptic and local scales. *Atmospheric Pollution Research*, article in press.

Schmidt, W., During, I., 2013. Determination of daily emission of the motor vehicles in 7 sites in Berlin in 2012, Engineering Office of Lohmeyer Ltd., In Contract of Senate Department for Health, Environment and Customer Protection (In German).

Seinfeld, J.H., 1972. Optimal location of pollutant monitoring stations in an airshed. *Atmospheric Environment* 6, 847–858.

SenStadt, 1998. Air quality management in Berlin 1997, Air Quality Management Series, Department of Urban Development, Berlin.

Shannon, J., Wesely, M., Brady, P., 1978. Objective sensor placement for sampling regional turbidity. *Atmospheric Environment* 12, 937–943.

Sprecher, D.A., 1965. On the structure of continuous functions of several variables. *Transactions of the American Mathematical Society* 115, 340–355.

Stalker, W.W., Dickerson, R.C., 1962. Sampling station and time requirements for urban air pollution surveys: Part II: Suspended particulate matter and soiling index. *Journal of the Air Pollution Control Association* 12, 111–128.

Taheri Shahraiyni, H., Shahsavani, D., Sargazi, S., Habibi–Nokhandan, M., 2015. Evaluation of MARS for the spatial distribution modeling of carbon monoxide in an urban area. *Atmospheric Pollution Research*, Article in Press.

Taheri Shahraiyni, H., Sodoudi, S., Cubasch, U., 2014. Determination the optimum number and positions of monitoring stations for proper spatial modeling of PM_{10} concentration in Berlin. *Proceedings of European Geosciences Union General Assembly Meeting*, April 27–May 02, 2014, Vienna, Austria.

Trujilloventura, A., Ellis, J.H., 1991. Multiobjective air–pollution monitoring network design. *Atmospheric Environment Part a–General Topics* 25, 469–479.

Ung, A., Weber, C., Perron, G., Hirsch, J., Kleinpeter, J., Wald, L., Ranchin, T., 2001. Air pollution mapping over a city–virtual stations and morphological indicators. *Proceedings of 10th International Symposium "Transport and Air Pollution"*, Boulder, Colorado, USA.

Van Egmond, N.D., Onderdelinden, D., 1981. Objective analysis of air–pollution monitoring network data – spatial interpolation and network density. *Atmospheric Environment* 15, 1035–1046.

WHO (World Health Organization), 1977. Air Monitoring Programme Design for Urban and Industrial Areas, WHO Publication No. 38, Geneva.

Wu, S.J., Feng, Q., Du, Y., Li, X.D., 2011. Artificial neural network models for daily PM_{10} air pollution index prediction in the urban area of Wuhan, China. *Environmental Engineering Science* 28, 357–363.

Yao, L., Lu, N., 2014. Spatiotemporal distribution and short–term trends of particulate matter concentration over China, 2006–2010. *Environmental Science and Pollution Research* 21, 9665–9675.

Zhang, D.Z., Peng, Z.R., 2014. Near–road fine particulate matter concentration estimation using artificial neural network approach. *International Journal of Environmental Science and Technology* 11, 2403–2412.

Zheng, J.Y., Feng, X.Q., Liu, P.W., Zhong, L.J., Lai, S.C., 2011. Site location optimization of regional air quality monitoring network in China: Methodology and case study. *Journal of Environmental Monitoring* 13, 3185–3195.