

A Review of Artificial Neural Network Models for Ambient Air Pollution Prediction

Author names and affiliations

Sheen Mclean Cabaneros^{1*},

John Kaiser Calautit²,

Ben Richard Hughes¹

¹ Department of Mechanical and Aerospace Engineering, University of Strathclyde, Glasgow

² Department of Architecture and Built Environment, The University of Nottingham

Corresponding Author

*Corresponding Author at: Level 8, James Weir Building, 75 Montrose St, Glasgow G1 1XJ.

Email address: sheen.cabaneros@strath.ac.uk

Abstract

Research activity in the field of air pollution forecasting using artificial neural networks (ANNs) has increased dramatically in recent years. However, the development of ANN models entails levels of uncertainty given the black-box nature of ANNs. In this paper, a protocol by Maier et al. (2010) for ANN model development is presented and applied to assess journal papers dealing with air pollution forecasting using ANN models. The majority of the reviewed works are aimed at the long-term forecasting of outdoor PM₁₀, PM_{2.5}, and oxides of nitrogen, and ozone. The vast majority of the identified works utilised meteorological and source emissions predictors almost exclusively. Furthermore, ad-hoc approaches are found to be predominantly used for determining optimal model predictors, appropriate data subsets and the optimal model structure. Multilayer perceptron and ensemble-type models are predominantly implemented. Overall, the findings highlight the need for developing systematic protocols for developing powerful ANN models.

Keywords

Air pollution; Artificial Neural Networks; Multilayer Perceptron; Forecasting; Backpropagation Algorithm

I. Introduction

There is a growing interest in recent years on the use of artificial neural networks (ANNs) in predicting and forecasting ambient air pollution. As poor air quality in urban areas has been attributed to chronic diseases and premature mortalities of vulnerable members of the public (Organisation for Economic Co-operation and Development, 2016; World Health Organization, 2016), a greater demand is directed towards policy-makers and urban city planners to provide rapid and parsimonious solutions to circumvent the effects of air pollution (Baklanov et al., 2007; Moustris et al., 2010). In recent years, ANNs have been successfully implemented in many short- and long-term forecasting applications (Biancofiore et al., 2017a; Cabaneros et al., 2017; Coman et al., 2008; Ibarra-Berastegi et al., 2008; Lightstone et al., 2017; Rahimi, 2017). Furthermore, more practitioners resort to data-driven approaches such as ANNs as alternatives to traditional deterministic or physics-based approaches, e.g. the Urban Airshed Model (UAM) (Chang & Cardelino, 2000), Weather Research and Forecasting Model with Chemistry (WRF/Chem) (Chuang et al., 2011) and Community Multiscale Air Quality model (CMAQ) (Mueller & Mallard, 2011). This is because deterministic approaches are sensitive to several factors, including the scale and quality of the parameters involved, computationally expensive, and dependent to large databases of several input parameters, of which some may not be available (Dutot et al., 2007; Elangasinghe et al., 2014; Hrutz et al., 2009; P. Jiang et al., 2017; Sun et al., 2013). On the other hand, the use of ANNs does not require an in-depth understanding of the dynamics between air pollution concentration levels and other explanatory variables. Lastly, powerful and less-complicated computing tools that are able to develop and implement ANNs and their training algorithms are becoming more available to the public in recent years (Gardner & Dorling, 1998; Sharma et al., 2005).

However, the development of ANN models and the interpretation of their results entail certain issues despite their success in many applications. Data-driven models, in general, are problem-specific, thus a one-size-fits-all approach on building them is clearly not available (Gardner & Dorling, 1998; Zhang & Patuwo, 1998; Hagan et al., 1995; Maier & Dandy, 2000). Nonetheless, several authors argued that more general and consistent protocols that outline the complete building process of ANN models should still be established (Galelli et al., 2014; Jakeman et al., 2006; Maier & Dandy, 2000; Maier et al., 2010; Wu et al., 2014).

Jakeman et al. (2006) outlined ten basic steps of developing and evaluating environmental models, and argued that modellers should provide enough information that describes and justifies the choice of model parameters, their development and evaluation. Maier & Dandy (2000) emphasised that a lack of a comprehensive guide to ANN model building makes it difficult for future modellers to draw meaningful comparisons between existing models. After reviewing and analysing 43 papers dealing with the use of ANNs in predicting and forecasting water resources variables based on the modelling methodology they suggested, the authors reported that many of the papers carried out the modelling process incorrectly. Maier et al. (2010) also encountered the same methodological concerns after reviewing 210 papers dealing with the prediction of water resource variables that were published from 1999 to 2007. Furthermore, Wu et al. (2014) argued that a justification for the use of particular methods and parameter values during the ANN model building process should also be provided to increase the level of confidence in the model results.

Consequently, papers dealing with the use of ANNs in ambient air pollution forecasting are surveyed and assessed through the modelling protocols suggested by Wu et al. (2014). Furthermore, the latest development regarding the modelling of air pollutant concentration using ANNs is examined. Additionally, only papers dealing with the modelling of outdoor air pollution, especially in urban and industrial areas, are considered, although the threats attributed to indoor air pollution also cannot be neglected (Kotzias et al., 2009; Symonds et al., 2016). Despite the popularity of feedforward types of ANNs, especially the

Multilayer Perceptron (MLP) network, in many forecasting and prediction applications (Gardner & Dorling, 1998; Shahraiyini & Sodoudi, 2016; Sharma et al., 2005), other forms of ANN models are examined. The readers are expected throughout this work to be familiar with the concepts and terminologies related to ANNs. Detailed discussion of the subject matter can be found in many papers and textbooks (see Bishop, 1995; Gardner & Dorling, 1998; Hagan et al., 1995; Hornik et al., 1989; Samarasinghe, 2006).

The remainder of this paper is organised as follows. In Section 2, details on how the papers were selected are given, as well as an overview of the research activity in the use of ANNs for the prediction and forecasting of ambient air pollutant concentrations from 2001 to 2019. In Section 3, the ANN model development protocols are briefly outlined. A comprehensive discussion of each step can be found elsewhere (see Galelli et al., 2014; Humphrey et al., 2017; Jakeman et al., 2006; Maier & Dandy, 2000). The section also provides the taxonomies of available options at the various steps in this process suggested by Maier & Dandy (2000). A total of 139 papers were examined based on the given taxonomies. In Section 4, conclusions are then provided. Finally, recommendations for future research are given in Section 5.

II. Overview

The papers reviewed in this work are taken from the following international peer-reviewed journals: Air Quality, Atmosphere & Health, Atmosphere, Atmospheric Environment, Atmospheric Pollution Research, Building and Environment, Chemosphere, Clean – Soil Air Water, Ecological Modelling, Ecological Processes, Engineering Applications of Artificial Intelligence, Engineering Computations, Environmental Forensics, Environmental Modeling & Assessment, Environmental Modelling and Software, Environmental Monitoring and Assessment, Environmental Pollution, Environmental Science & Policy, Environmental Science and Pollution Research, Environmental Technology & Innovation, Evolving Systems, Expert Systems with Applications, Frontiers of Earth Science, Geophysical Research Letters, IEEE Access, IEEE Transactions on Intelligent Transportation Systems, International Journal of Environmental Research and Public Health, International Journal of Environmental Studies, Journal of Environmental Engineering and Science, Journal of Environmental Management, Journal of Environmental Protection and Ecology, Journal of the Air & Waste and Management Association, Knowledge-Based Systems, Mathematical Geosciences, Neural Computing & Applications, Neurocomputing, Pure and Applied Geophysics, Science of the Total Environment, Sensors, Soft Computing, Sustainable Cities and Society, Urban Climate, and Water, Air, & Soil Pollution.

The papers were identified through the StarPlus Library catalogues by the University of Sheffield, England (The University of Sheffield, 2017), ScienceDirect (ScienceDirect, 2017), ProQuest (ProQuest, 2017), and IEEE Xplore® Digital Library (IEEE Xplore, 2017). Search terms included “air pollution forecasting”, “air pollution modelling”, “artificial neural networks”, “ANN”, “multilayer perceptron”, “RNN”, “LSTM”, “NARX” and “machine learning” with different combinations from previous review papers and standalone articles. The process was repeated until the citation trails stopped, see Figure 1. Furthermore, the list of references of the selected research articles were investigated to identify further references.

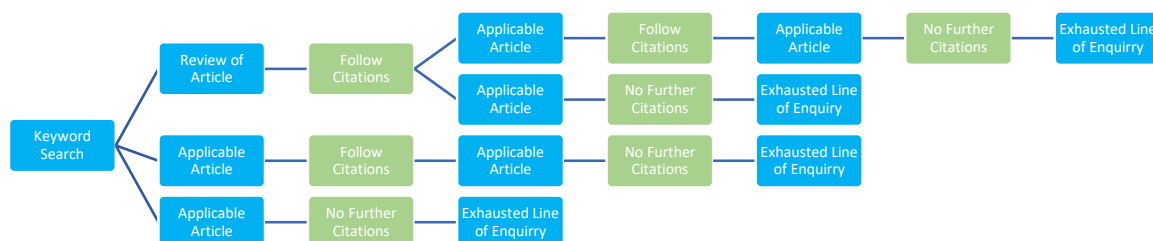


Figure 1. Graphical representation of the methodology used for selecting the papers reviewed

The authors selected the papers published from January 2001 to February 2019. Only papers dealing with the forecasting of outdoor air pollutants were considered, although the threat of indoor pollutants is as serious for investigation given that an average person spends more than 90% of their time indoors (Ashmore et al., 2001). Additionally, papers that used ANN models which failed to outperform or provide similar results to other alternative techniques were not selected. Finally, papers published as conference proceedings were manually removed from the initial list of papers, resulting in 139 peer-reviewed articles. The key details of the selected articles, e.g. the authors, year of publication, study location, and air pollutant(s) examined, are presented in Table 1.

The distribution of articles by year of publication is given in Figure 2. There is a growing number of published articles since 2001 that cite the use of air pollution forecasting tools based on ANNs, with almost 50% of the identified papers published since 2015 alone. This can be well explained by the emerging computing technologies tailored for the development of ANN models that are not easily accessible in the past. That is, faster and more powerful computing tools capable of performing ANN training algorithms and processing big data are becoming available recently (IEEE Spectrum, 2018).

Table 1. Details of papers reviewed.

Authors (year)	Location(s)	Air pollutants examined
Kolehmainen et al. (2001)	Stockholm, Sweden	NO ₂
Perez & Trier (2001)	Santiago, Chile	NO; NO ₂
Chelani et al. (2002)	Delhi, India	SO ₂
Abdul-Wahab & Al-Alawi (2002)	Khaldiya, Kuwait	O ₃
Kukkonen et al. (2003)	Helsinki, Finland	NO ₂
Lu et al. (2003)	Hong Kong	RSP
Wang et al. (2003)	Mong Kok, Hong Kong	RSP
Hasham et al. (2004)	Edmonton, Canada	NO _x
Heo & Kim (2004)	Seoul, Korea	O ₃
Jiang et al. (2004)	Shanghai, China	TSP; SO ₂ ; NO _x
Niska et al. (2004)	Helsinki, Finland	NO ₂
Nunnari (2004)	Syracuse, Italy	SO ₂
Olcese & Toselli (2004)	Cordoba, Argentina	?
Chelani et al. (2005)	Kolkata, India	NO ₂
Hooyberghs et al. (2005)	Belgium	PM ₁₀
Niska et al. (2005)	Helsinki, Finland	NO ₂ ; PM _{2.5}
Ordieres et al. (2005)	Ciudad Juarez and El Paso, Mexico	PM _{2.5}
Agirre-Basurko et al. (2006)	Bilbao, Spain	O ₃ ; NO ₂
Grivas & Chaloulakou (2006)	Athens, Greece	PM ₁₀
Nagendra & Khare (2006)	New Delhi, India	NO ₂
Schlink et al. (2006)	Several EU countries	O ₃
Slini et al. (2006)	Thessaloniki, Greece	PM ₁₀
Brunelli et al. (2007)	Palermo, Italy	SO ₂ ; O ₃ ; PM ₁₀ ; NO ₂ ; CO
Dutot et al. (2007)	Orleans, France	O ₃
Osowski & Garanty (2007)	Warsaw, Poland	CO; NO ₂ ; SO ₂ ; dust
Sousa et al. (2007)	Porto, Portugal	O ₃
Al-Alawi et al. (2008)	Kuwait	O ₃
Coman et al. (2008)	Paris, France	O ₃
Díaz-Robles et al. (2008)	Temuco, Chile	PM ₁₀
Ibarra-Berastegi et al. (2008)	Bilbao, Spain	SO ₂ ; CO; NO ₂ ; NO; O ₃
Martín et al. (2008)	Algeciras, Spain	CO
Perez & Salini (2008)	Santiago, Chile	PM _{2.5}
Solaiman et al. (2008)	Ontario, Canada	O ₃

Table 1. (Continued)

Authors (year)	Location(s)	Air pollutants examined
Zito et al. (2008)	Leicestershire, UK	CO; NO ₂
Ettouney et al. (2009)	Jahra, Kuwait	O ₃
Galatioto & Zito, (2009)	Palermo, Italy	CO; C ₆ H ₆
Hrust et al. (2009)	Zagreb, Croatia	NO ₂ ; O ₃ ; CO; PM ₁₀
Juhos et al. (2009)	Szeged, Hungary	NO; NO ₂
Pisoni et al. (2009)	Milan, Italy	O ₃
Tsai et al. (2009)	Taiwan	O ₃
Demir et al. (2010)	Istanbul, Turkey	PM ₁₀
Inal (2010)	Istanbul, Turkey	O ₃
Jain & Khare (2010)	Delhi City, India	CO
Kurt & Oktay (2010)	Istanbul, Turkey	SO ₂ ; CO; PM ₁₀
Mahapatra (2010)	New Delhi, India	O ₃
Moustris et al. (2010)	Athens, Greece	ERPI (NO ₂ ; CO; SO ₂ ; O ₃)
Pires et al. (2010)	Oporto, Portugal	O ₃
Feng et al. (2011)	Beijing, China	O ₃
Paschalidou et al. (2011)	4 cities in Cyprus	PM ₁₀
Prakash et al. (2011)	New Delhi, India	CO; NO ₂ ; NO; O ₃ ; SO ₂ ; PM _{2.5}
Vlachogianni et al. (2011)	Thessaloniki, Greece and Helsinki, Finland	PM ₁₀ ; NO _x
Voukantsis et al. (2011)	Thessaloniki, Greece and Helsinki, Finland	PM ₁₀ ; PM _{2.5}
Barron-Adame et al. (2012)	Salamanca, Mexico	SO ₂
Chattopadhyay & Chattopadhyay (2012)	Kolkata, India	O ₃
Fernando et al. (2012)	Phoenix, Arizona	PM ₁₀
Perez (2012)	Santiago, Chile	PM ₁₀
Singh et al. (2012)	Lucknow, India	RSPM; SO ₂ ; NO ₂
Siwek & Osowski (2012)	Warsaw, Poland	PM ₁₀
Antanasijevic et al. (2013)	26 EU countries	PM ₁₀
Arhami et al. (2013)	Tehran, Iran	CO; NO _x ; NO; NO ₂ ; O ₃
Gennaro et al. (2013)	Northeast Spain	PM ₁₀
Moustris et al. (2013)	Greater Athens Area, Greece	PM ₁₀
Papaleonidas & Iliadis (2013)	Athens, Greece	O ₃
Russo et al. (2013)	Lisbon, Portugal	NO ₂
Singh et al. (2012)	Lucknow, India	RSPM; SO ₂ ; NO ₂
Siwek & Osowski (2012)	Warsaw, Poland	PM ₁₀
Antanasijevic et al. (2013)	26 EU countries	PM ₁₀

Table 1. (Continued)

Authors (year)	Location(s)	Air pollutants examined
Arhami et al. (2013)	Tehran, Iran	CO; NO _x ; NO ₂ ; O ₃
Gennaro et al. (2013)	Northeast Spain	PM ₁₀
Moustris et al. (2013)	Greater Athens Area, Greece	PM ₁₀
Papaleonidas & Iliadis (2013)	Athens, Greece	O ₃
Russo et al. (2013)	Lisbon, Portugal	NO ₂
Ul-Saufie et al. (2013)	Negeri Sembilan, Malaysia	PM ₁₀
Yan & Jian (2013)	Hangzhou, China	PM ₁₀ ; PM _{2.5}
Zhang et al. (2013)	Taiyuan, China	PM ₁₀
Azid et al. (2014)	Malaysia	API
Elangasinghe et al. (2014)	Auckland, New Zealand	NO ₂
He et al. (2014)	Mong Kok, Hong Kong	PM ₁₀ ; PM ₁
Luna (2014)	Rio de Janeiro, Brazil	O ₃
Özdemir & Taner (2014)	Kocaeli, Turkey	PM ₁₀
Russo & Soares (2014)	Lisbon, Portugal	PM ₁₀
Zhou et al. (2014)	Xi'an Province, China	PM _{2.5}
Alam & McNabola (2015)	Vienna, Austria	PM ₁₀
Biancofiore et al. (2015)	Pescara, Italy	O ₃
Cortina-Januchs et al. (2015)	Salamanca, Mexico	PM ₁₀
Dunea et al. (2015)	Oltenia, Romania	O ₃ ; PM ₁₀ ; PM _{2.5}
Dursun & Taylan (2015)	Konya City, Turkey	SO ₂
Feng et al. (2015)	Jing-Jin-Ji area, China	PM _{2.5}
Mishra & Goyal (2015)	Agra, India	NO ₂
Russo et al. (2015)	Lisbon, Portugal	PM ₁₀
Santos & Fernández-olmo (2015)	Cantabria Region, Spain	As; Cd; Ni; Pb
Zhu et al. (2015)	Chongqing, China	NO _x
Zou et al. (2015)	Texas, USA	PM _{2.5}
Abderrahim et al. (2016)	Algiers, Algeria	PM ₁₀
Bai et al. (2016)	Chongqing, China	PM ₁₀ ; SO ₂ ; NO ₂
Catalano et al. (2016)	London, United Kingdom	NO ₂
Chelalli et al. (2016)	Algiers, Algeria	PM ₁₀
Ding et al. (2016)	Hong Kong	NO ₂ ; NO _x ; O ₃ ; SO ₂ ; PM _{2.5}
Durao et al. (2016)	Sines, Portugal	O ₃

Table 1. (Continued)

Authors (year)	Location(s)	Air pollutants examined
He et al. (2016)	Lanzhou, China	SO ₂ ; NO ₂ ; PM ₁₀
Hoshyaripour et al. (2016)	Sao Paulo, Brazil	O ₃
Li et al. (2016)	Beijing, China	PM _{2.5}
Li et al. (2017)	China	PM _{2.5}
Lightstone et al. (2017)	United States of America	PM _{2.5}
Mao et al. (2017)	Eastern China	PM _{2.5}
Peng et al. (2017)	Canada	O ₃ ; PM ₁₀ ; NO ₂
Rahimi (2017)	Tabriz, Iran	NO _x ; NO ₂
Stamenković et al. (2017)	17 EU countries, USA, China, Japan, Russia and India	NO _x
Taylan (2017)	Jeddah, Saudi Arabia	O ₃
Yeganeh et al. (2017)	Queensland, Australia	PM _{2.5}
Zhang & Ding (2017)	Hong Kong	NO ₂ ; NO _x ; O ₃ ; PM _{2.5} ; SO ₂
Alimissis et al. (2018)	Athens, Greece	NO ₂ ; NO _x ; O ₃ ; CO; SO ₂
Antanasijević et al. (2018)	26 EU countries	SO _x ; NO _x ; NH ₃ ; NMVOC; PM ₁₀
Dotse et al. (2018)	Brunei Darussalam	PM ₁₀
Franceschi et al. (2018)	Bogota, Colombia	PM _{2.5} ; PM ₁₀
Freeman et al. (2018)	Kuwait	O ₃
Gao et al. (2018)	Jinan, China	O ₃
Huang & Kuo (2018)	Beijing, China	PM _{2.5}
Jiang et al. (2018)	Beijing, China	PM _{2.5} ; SO ₂ ; NO ₂ ; CO; O ₃
Li & Zhu (2018)	China	PM _{2.5} ; PM ₁₀ ; CO
Nidzgorska-Lencewicz (2018)	Tricity Agglomeration, Poland	PM ₁₀
Pak et al. (2018)	Beijing, China	O ₃
Radojević et al. (2018)	Belgrade, Serbia	SO ₂ ; NO _x
Tzanis et al. (2019)	Attica, Greece	PM _{2.5} ; PM ₁₀
Ventura et al. (2019)	Rio de Janeiro, Brazil	PM _{2.5}
Wang & Song (2018)	Beijing, China	CO; NO ₂ ; SO ₂ ; O ₃ ; PM ₁₀ ; PM _{2.5}
Yeganeh et al. (2018)	Queensland, Australia	NO ₂
Zhu et al. (2018)	China	PM _{2.5}
Bai et al. (2019)	Beijing, China	PM _{2.5}
Liu et al. (2019)	Beijing, China	PM _{2.5} ; SO ₂ ; NO ₂ ; CO
Qi et al. (2019)	Jing-Jin-Ji, China	PM _{2.5}
Qin et al. (2019)	Shanghai, China	PM _{2.5}

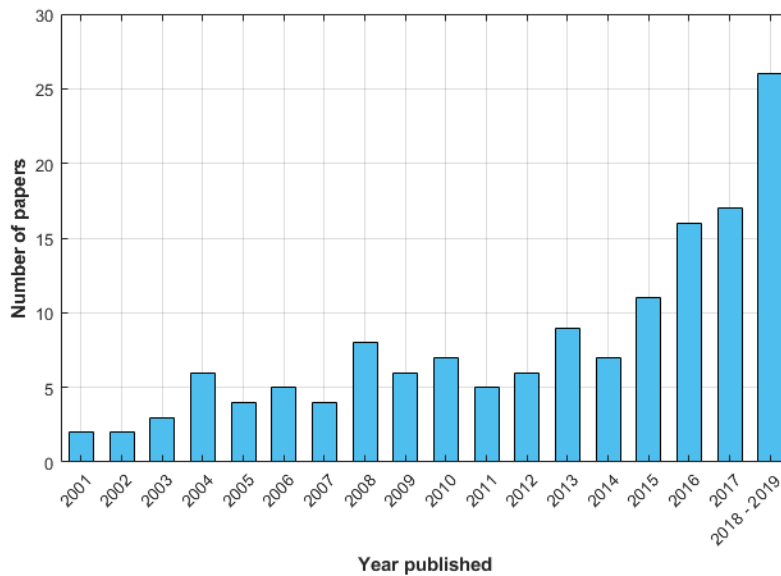


Figure 2. Distribution of papers by year of publication

The number of papers in which air pollutant variables were considered is shown in Figure 3. The results reveal that airborne particulate matter with an aerodynamic diameter smaller than $10\ \mu\text{m}$ (PM_{10}) and $2.5\ \mu\text{m}$ ($\text{PM}_{2.5}$), oxides of nitrogen, e.g. NO_2 , NO and NO_x , and ozone are the most examined variables among the papers identified. Particulate matter was studied in 87 of the 139 papers reviewed, almost 50% of which dealt with PM_{10} modelling while almost 45% of which for $\text{PM}_{2.5}$. This is followed by the oxides of nitrogen (51 papers) and ozone (44 papers). The modelling of CO and SO_2 pollutants were also examined by a considerable amount of papers, e.g. 18 and 23 papers, respectively. It is also worth noting that at least a third of the papers identified delved in more than one air pollutant variables.

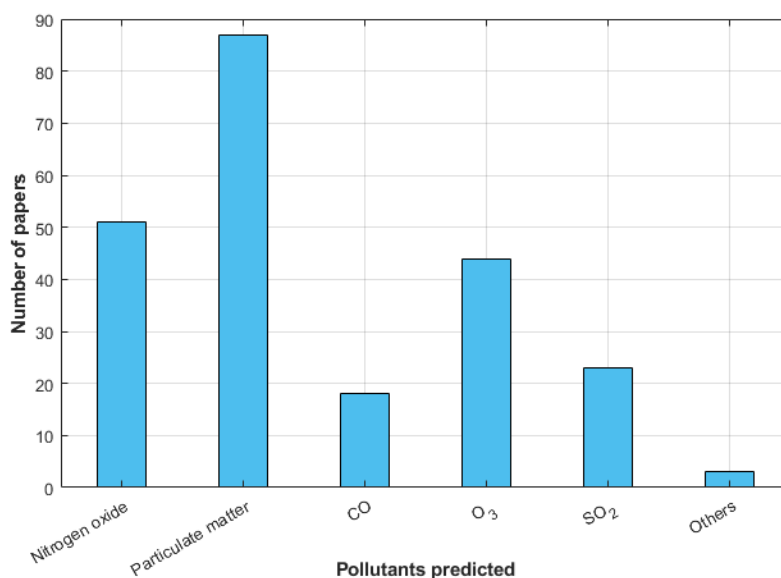


Figure 3. Distribution of papers by air pollutant variables predicted

The distribution of the time steps used in modelling variables considered is shown in Figure 4. The hourly time step was used in 69 out of 139 papers reviewed, followed by daily (50 papers), and yearly (4 papers) and 5-minute scales (4 papers). On the other hand, several time steps including 2-hourly, 4-hourly, minute and 30-minute were utilised by other papers. However, the time steps of the modelling variables are commonly determined by the sampling periods of the instruments used to measure the pollutant species and meteorological data at a monitoring station. Consequently, some identified papers pre-processed their collected data via averaging and linear interpolation techniques to create model datasets with consistent time steps.

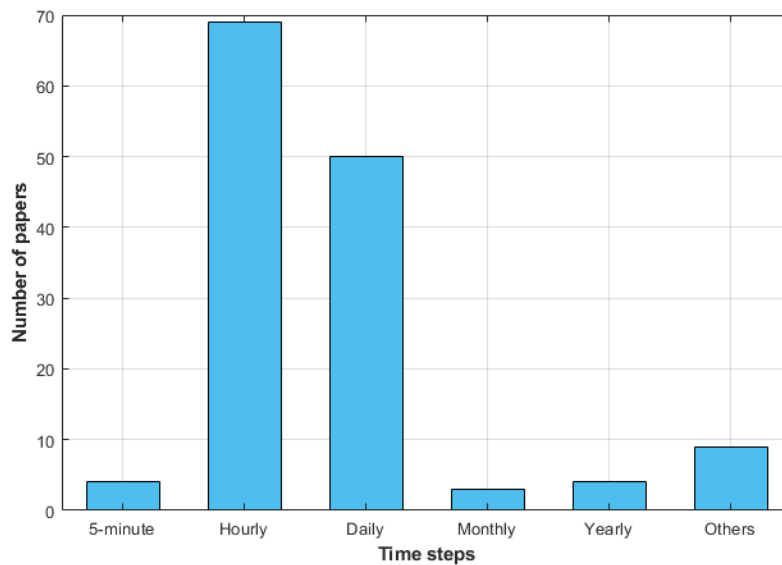


Figure 4. Number of occurrences various time steps have been used

Figure 5 shows the number of instances various forecast lengths have been used by the identified papers. Unfortunately, 43 out of 139 papers did not explicitly describe the forecast length utilised in their model development process. This ambiguity poses a challenge to future modellers as parameter settings that are implicitly described can cast doubts on the readers and future modellers. Of the remaining papers, short-term forecasting (forecast length = 1) was carried out 68 times. Long-term forecasting (forecast length > 1) was done 171 times, of which +24, +48 and +72 ahead forecasts were done 25, 7 and 4 times, respectively. Prediction (forecast length of 0) was carried out by 48 papers. It is worth emphasising that most papers identified did consider multiple forecast lengths.

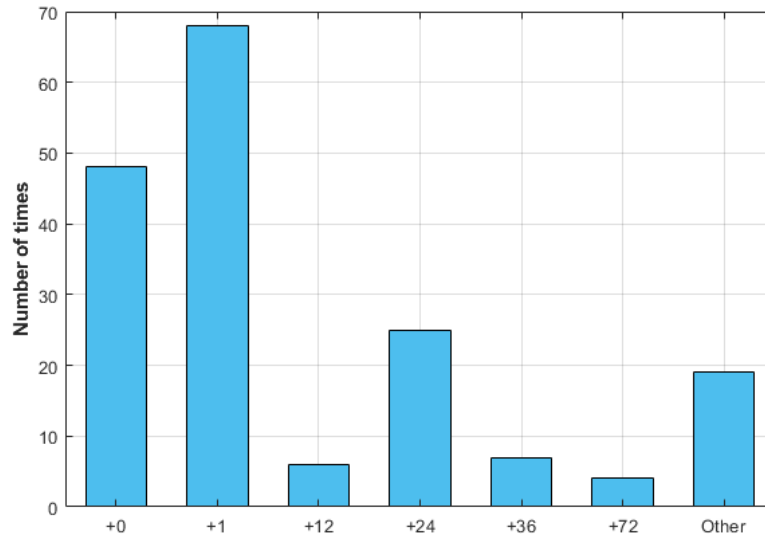


Figure 5. Number of occurrences various forecast lengths have been used

III. Methods used for ANN model development

The design of ANN models can often be regarded as more of an art than a science due to the lack of a clean-cut method for implementing each model development step (Zhang & Patuwo, 1998). However, several guidelines are available in the literature to provide future modellers with a systematic way of developing ANN models. The model development process is divided into seven main steps: (1) data collection, (2) data pre-processing, (3) selection of input variables (or predictors), (4) data splitting, (5) selection of model architecture, (6) determination of model structure, (7) model training, and (8) model validation. The ANN model-building protocol used in this review is based on those presented by Jakeman et al. (2006) and Maier et al. (2010).

3.1 Collection of data

3.1.1 Introduction

Since black-box models such as ANN models are data-dependent, it is generally difficult to incorporate them with prior knowledge. Consequently, the performance of ANN models primarily relies on the type and form of data utilised to train them.

One important requirement in data selection is for them to span the full range of input space for which the network will be utilised, as black-box models do not extrapolate well (Hagan et al., 1995). In the case of air pollution forecasting applications, the use of predictors, covering a period of a year or more has been highly recommended. This is to ensure that seasonal factors which have been identified to strongly influence air pollution levels are taken into account (Arhami et al., 2013; Colls, 2001; Kumar et al., 2017). For instance, many studies have utilised datasets covering a one-year period as the data tends to be roughly periodic after a year (Bai et al., 2019; Catalano et al., 2016; Coman et al., 2008; Fernando et al., 2012; Kurt & Oktay, 2010; Mishra & Goyal, 2016).

Furthermore, the selection of various types of predictors plays an important role on model performance since air quality is a complex function of meteorology, emissions and other parameters (Colls, 2001). There is a plethora of predictor types that have been considered in

previous environmental modelling applications. In this paper, they are categorised as meteorological, emissions, traffic, and others. Meteorological parameters refer to the variables that characterise atmospheric chemistry. Meteorological variables especially wind speed, wind direction, relative humidity and atmospheric turbulence have been found to have a massive influence on the dispersion and concentration of several air pollutants including O₃, NO₂, PM₁₀ and PM_{2.5} (Colls, 2001; Dominick et al., 2012; Kumar et al., 2017; Peng et al., 2017). Emissions data primarily refers to primary and secondary air pollutants in urban environments. They are also considered as important predictors as they are highly correlated to other air pollutants (Colls, 2001; World Health Organization, 2018). Finally, traffic data refers to information that characterises traffic behaviour. This includes traffic flow density, speed, occupancy degree, queues length, and travel time which are typically monitored on roads in close proximity to air quality stations. The use of traffic parameters has also been suggested as they play a significant role in the formation of several roadside air pollutants (Agirre-Basurko et al., 2006; Catalano et al., 2016; Colls, 2001; Galatioto & Zito, 2009; World Health Organization, 2018). On the other hand, uncategorised parameters that were also considered by the identified papers include satellite, land-use and economic variables.

3.1.2 Results

Figure 6 reveals the distribution of input data lengths utilised by the identified papers. The majority of the papers utilised input data with length covering more than a period of one year. In detail, training data with a duration from one to three years was used 60 times, while those with lengths longer than three years were used 43 times. Furthermore, the use of data covering a period of less than 6 months occurred 22 times, while only 8 studies utilised data with lengths between six months to one year. However, 6 of the identified studies did not provide details regarding input data length.

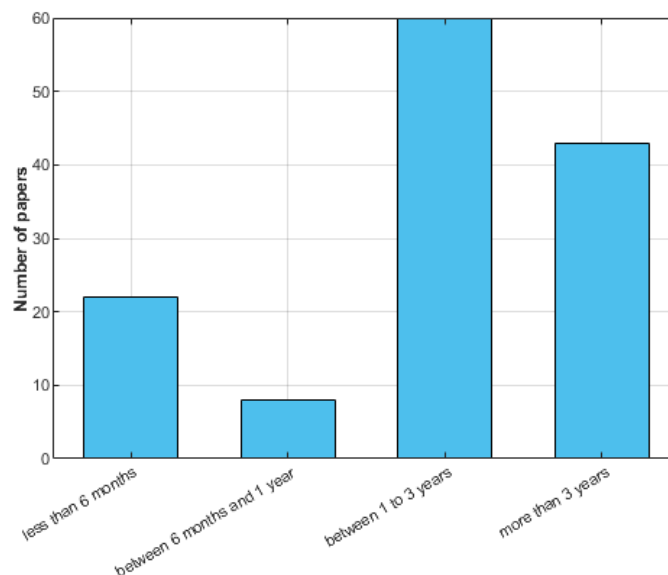


Figure 6. Number of occurrences various lengths of input data have been collected

Figure 7 shows the number of instances where a given set of predictors was utilised by the identified papers. Of the 139 papers reviewed, the use of both meteorological and pollutant emissions variables was observed on 90 occasions. The most common meteorological variables utilised include temperature, relative humidity, and wind-related measurements. On the other hand, the utilisation of only meteorological data was carried out 9 times, while the

sole use of emissions data was done 13 times. Additionally, the use of traffic data alongside other variables was implemented by 17 papers. The utilisation of data based on satellite-derived imagery also appeared in 4 papers. For instance, the utilisation of satellite-derived aerosol optical depth (AOD) variables was explored in the forecasting of PM_{2.5} in several occasions (Mao et al., 2017; Wen et al., 2019; Yeganeh et al., 2017). Finally, land-use, economic and stability predictors were utilised by only a few identified papers.

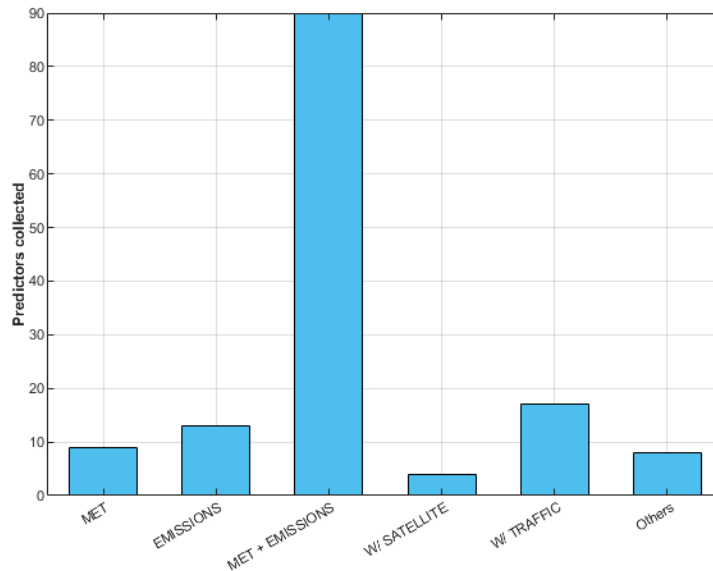


Figure 7. Number of occurrences various sets of data have been collected, where MET and EMISSIONS denote the use of meteorological and air pollutant variables only, respectively; and W/ SATELLITE and W/TRAFFIC are the use of satellite and traffic data along with other variables, respectively

3.1.3 Conclusion

The findings have indicated that the majority of the identified studies utilised data that covers more than one year for ANN model development. This is considered a good modelling practice as the developed ANN models can possess a greater generalisation ability when introduced to a sufficient length of data. However, this practice does not fully address other prevailing modelling issues including the imbalance data problem which occurs when peak or rare network target values have lesser frequencies. The problem is typically encountered by ANN models dealing with the prediction of high or peak pollution episodes (Bai et al., 2019; Fernando et al., 2012; Gong & Ordieres-Meré, 2016). To address the imbalance data problem, several techniques coming from the field of data mining can be explored: data re-sampling (Drummond & Holte, 2003; Zhao et al., 2016), cost-sensitive learning (Fontes, et al., 2014; Tsai et al., 2009), algorithm modifications (López et al., 2012), and the Synthetic Minority Over-Sampling Technique (SMOTE) (Chawla et al. , 2002). With regards to the predictor types utilised, the results reveal that meteorological and pollutant emissions data are the most commonly used predictors by the identified studies. However, network training with several predictors should be handled carefully as the dimension of input data space greatly influences the network complexity of resulting model (Hagan et al., 1995). Furthermore, less distinction can be observed between non-linear models such as ANN models and linear statistical models

when a large number of predictors are used. This is because the combination of a large number of non-linear processes tends to linearize the overall mechanism of a developed non-linear model (Ibarra-Berastegi et al., 2008). This observation is true for at least the prediction of NO₂, O₃, SO₂ and CO in urban environments (Chelani et al., 2002; Guardani et al., 1999; Kao & Huang, 2000; Kolehmainen et al., 2001; Schlink et al., 2006; Slini et al., 2006). Consequently, this highlights the careful implementation of predictor selection techniques (see Section 3.3).

3.2 Data pre-processing

3.2.1 Introduction

Another important step in the development of ANN models is data pre-processing. This step refers to preliminary techniques that aim at improving the representation of the collected data. Two popular data pre-processing techniques in the field of air pollution modelling include normalisation and missing data imputation (see Figure 8).

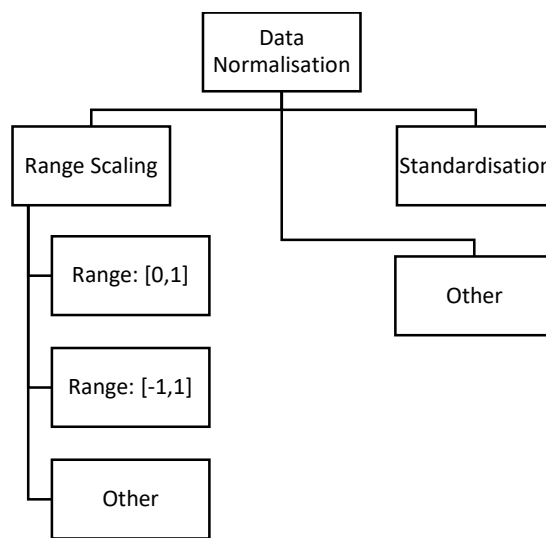


Figure 8. Taxonomy of data normalisation procedures

Normalisation is used to ensure that all predictors fall in a similar range. This is an important step in model development as inputs with large values disproportionately mask the impact of those inputs with smaller ones. The step should also be taken to match the range of the predictors to those of the transfer function of the hidden layer (see Section 3.5). There are two popular categories for normalization, namely, range scaling and standardisation. Under range scaling, the predictors are transformed such that the maximum and minimum values of the predictors are mapped to 0 and 1 or -1 to 1, respectively. In contrast to range scaling, standardisation converts an old variable into a new variable with zero mean and unit standard deviation. Finally, it is worth noting that another common data pre-processing technique is called feature extraction, where the dimension of the original input data space is reduced to avoid redundancy. One popular feature extraction method is the method of principal component analysis (PCA). In this paper, the implementation of both PCA and ANN models is considered as a data-intensive hybrid modelling approach. Hence, PCA method is not mentioned here and is described in more details in Section 3.3 instead.

On the other hand, data imputation is the process that addresses the issue of missing data, which is a problem repeatedly encountered in air quality modelling applications (Junninen et al., 2004). Missing data can be the result of many factors such as insufficient sampling, errors in measurements or faults in data acquisition. One of the simplest ways to address this issue

is through the substitution of missing values with the mean of the entire dataset. However, this practice is highly discouraged as this can disrupt the inherent structure of the original dataset, potentially degrading the performance of a model. Another popular approach in dealing with missing data is the list-wise or pair-wise deletion of predictors with missing data. Missing data imputation techniques are categorised as univariate, multivariate, nearest neighbour, and the hybrid of the previous approaches (Junninen et al., 2004; Plaia & Bondi, 2006). Univariate methods include linear (LIN), spline and nearest neighbour (NN) interpolation. Multivariate methods include regression-based imputation, nearest neighbour interpolation, SOM and MLP. Additionally, other modellers address the issue of missing data by deletion, in which predictors with a large fraction of missing values are left out. A more detailed discussion of the said techniques can be found in Junninen et al. (2004) and Plaia & Bondi (2006). The taxonomy of the said methods is given in Figure 9.

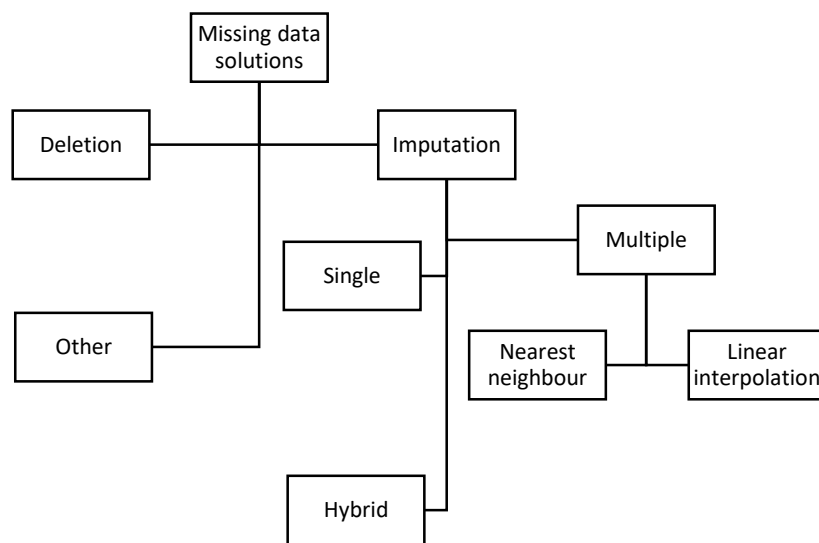


Figure 9. Taxonomy of missing data imputation approaches

3.2.2 Results

The majority of the studies identified did not provide sufficient details describing the methods they utilised for data normalisation. Of those that provided details, the standard normalisation scheme was implemented 43 times, see Figure 10. There are 9 cases where the input data was adjusted to have zero mean and unity variance. However, only a small number of instances (5 times) where other normalisation methods were used.

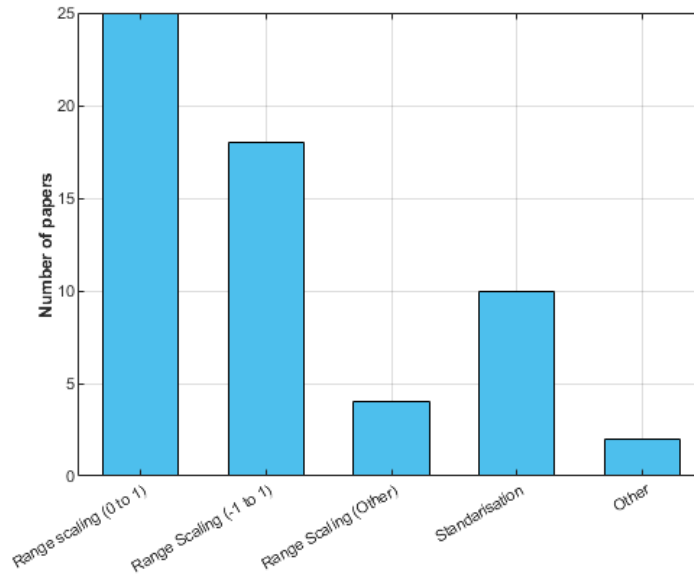


Figure 10. Number of occurrences various missing data normalisation techniques have been implemented

Figure 11 shows the number of occurrences various approaches to missing data imputation techniques have been undertaken. It should be noted that of the 139 papers reviewed, only 34 provided details about missing data. Of those that disclosed information, the majority carried out the deletion of predictors with missing data. Under univariate methods, nearest neighbour interpolation was used 5 times, while linear interpolation was only utilised 4 times. There are 6 instances where a combination of the techniques above were used, of which distinct methods were employed to address specific gap lengths.

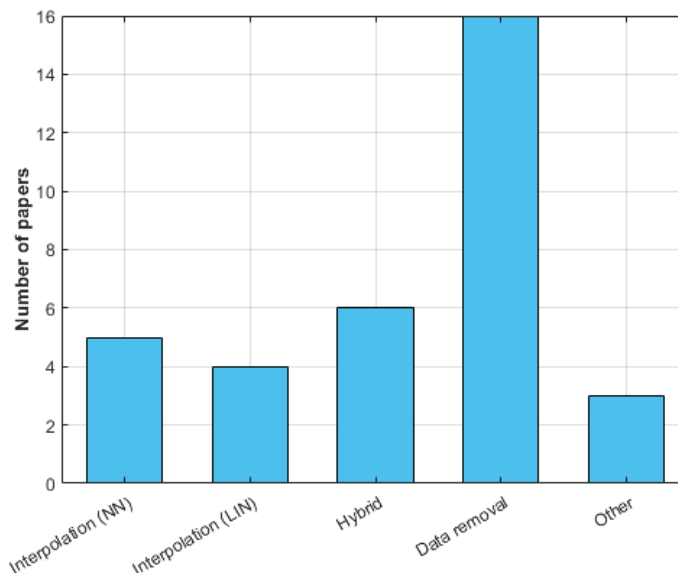


Figure 11. Number of occurrences various missing data imputation techniques have been implemented

3.2.3 Conclusion

The majority of the papers identified did not provide sufficient information regarding the data normalisation and imputation methods used before model training. With regards to data

normalisation, one possible explanation for the lack of mention is the growing number of neural network model building platforms that perform data pre-processing techniques by default. For instance, the MATLAB neural network toolbox normalises input values by default through the *mapminmax* function (The MathWorks, 2017). However, this seemingly trivial yet essential ANN development step should be clearly defined to assist future modellers. Of the papers that described the process of data normalisation, simple range scaling was commonly adapted. Other methods for pre-processing data for ANN development can be found in the literature (Bowden et al., 2003). On the other hand, the majority of the identified papers adopted the deletion of variables with missing values. This result highlights some potential issues. Although the deletion approach appears to be a quick and practical approach, this may not be an option to those studies with a very limited amount of data available or cases where the collection of additional data can be very expensive. In such cases, modellers should make the full use of any data available, even if it is incomplete. Others may have collected datasets without missing values, while others merely did not disclose the imputation techniques they implemented when they encountered missing data. However, it is still considered a good modelling practice to thoroughly discuss the use of imputation techniques for repeatability and reproducibility of results (Wu et al., 2014). Several available state-of-the-art imputation methods include single imputation (Plaia & Bondi, 2006), regression-based imputation using the EM-algorithm (Dempster et al., 1977; Schneider, 2001), known data regression (KDR) method (Folch-Fortuny et al., 2015), vector autoregressive model-imputation (VAR-IM) algorithm (Bashir & Wei, 2017) and Bayesian compressive sensing (BCS) imputation methods (Williams et al., 2018).

3.3 Selection of predictors

3.3.1 Introduction

Choosing the most suitable ANN model predictors for a prediction task is a nontrivial task. ANN models make no prior assumption regarding the distribution of the predictors involved and the underlying physical dynamics between predictors and target variables (Gardner & Dorling, 1998). As such, the robustness of an ANN model heavily relies on the form and manner on which predictors are being fed into the model. Consequently, the inclusion of too many correlated and extraneous predictors results to more network connections leading to overfitting issues (Hagan et al., 1995). On the other hand, the absence of relevant explanatory variables inhibits the model from correctly approximating the underlying dynamics between predictors and target variables (Maier & Dandy, 2000).

There are several approaches in selecting the most significant predictors of a given model (see Figure 12). They are divided into two categories, namely, model-free and model-based approaches (Maier et al., 2010). Model-free approaches perform input selection without relying on the performance of the developed ANN models. In other words, the process is undertaken before the ANN models are trained. Model-free approaches can be further divided into two categories: ad-hoc and analytical. The selection of model predictors is implemented in an arbitrary manner or based on domain knowledge falls under the ad-hoc approach. In contrast, the analytic approach to input selection involves the use of a statistical measure of dependence between model predictors and target variables. This is mostly carried out through cross-correlation. However, this analytical approach can only detect linear dependence between data, leading to the omission of relevant predictors that are associated with the target variables in a non-linearly manner (Samarasinghe, 2006). On the other hand, model-based approaches perform input selection by determining the effect of a candidate

model predictor on the overall model performance. As pointed out by Maier et al. (2010), the approach has several downsides. Firstly, the approach is time-consuming as a number of ANN models need to be developed. Furthermore, it does not clearly measure the impact of the utilised predictors on the model performance, as the latter is also a function of several network parameters, e.g. the number of hidden layer nodes, etc. A popular example of model-based approaches is the stepwise selection of inputs, where a network iteratively selects, e.g. forward selection, or remove, e.g. backward elimination, predictors based on the model performance. An ad-hoc approach can also be done, where arbitrary combinations of model predictors are tested. A global approach can also be implemented, where a global optimisation algorithm is used to select the combination of predictors that maximises model performance. Finally, an approach based on sensitivity analysis can be undertaken, where plots of sensitivity for each predictor to the target variables are examined.

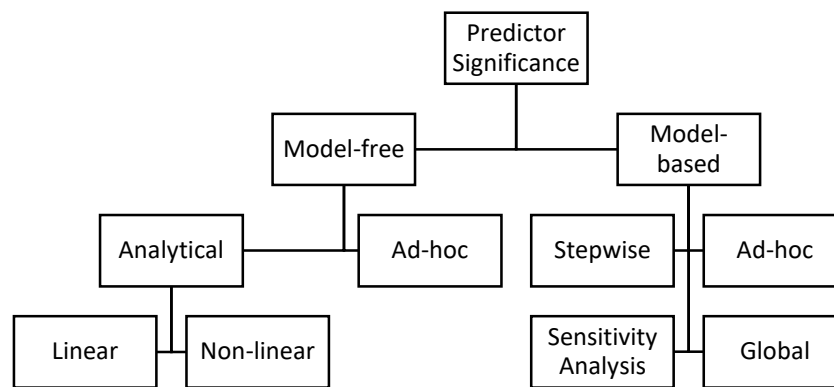


Figure 12. Taxonomy of approaches to determining optimal model predictors

3.3.2 Results

The number of occurrences various input selection approaches have been used is shown in Figure 13. It can be seen that model-free approaches were implemented 99 times, compared with the 40 occasions on which model-based approaches were used. Of the model-free approaches considered, ad-hoc methods were most widely implemented with applications in 82 papers, followed by linear approaches, especially correlation analysis which was utilised in 13 papers. A non-linear method was only employed 4 times. In 13 of the 40 times where a model-based approach was implemented, the process was carried out in an ad-hoc manner. A stepwise method was used 10 times, while global search approaches were implemented 7 times.

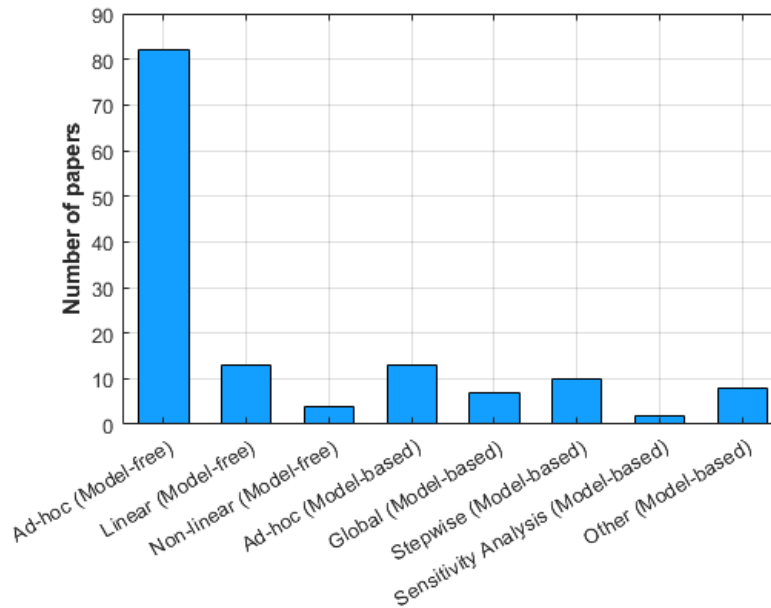


Figure 13. Number of occurrences various input selection methods have been used

3.3.3 Conclusion

The results of the review highlight the need for greater attention to predictor selection implementation. Although the selection of predictors is dependent on the external problem specifications, a systematic approach should be encouraged among modellers to reduce bias and increase the repeatability of performances of data-driven models in general. Ad-hoc approaches to predictor selection, either model-based or model-free, were used in almost 70% of the identified papers. Furthermore, linear analytical model-free approaches were also widely implemented, which contradicts the rationale of using non-linear models such as ANNs in approximating the typically non-linear dynamics between air pollutants and predictors. This indicates the need to further examine the use of non-linear approaches in selecting predictors. Several guidelines in determining the most suitable predictor selection technique to every problem specification can be found in Galelli et al. (2014). The proposed guidelines were evaluated by taking into account three factors, including, a wide range of dataset properties that reflect the properties of real-world environmental observations, an assessment criteria selected to highlight algorithm suitability in different problem specifications, and a website for sharing data, algorithms and results (Galelli et al., 2014).

3.4 Data splitting

3.4.1 Introduction

The division of data is another important step in the development of ANN models. This is carried out by splitting the available data into three subsets, namely, the training, validation and test sets. The training subset is used for computing the gradient and calibrating the network weights and biases. On the other hand, the validation subset is utilised to stop the network training before overfitting takes place. In detail, the error on the validation subset monitors the network performance during training. When this error begins to increase for several iterations, the training is stopped, and the weights and bias values that yielded the minimum error are then utilised as the final trained network weights and biases. Hence, the

division of data is an essential modelling step to avoid the problem of model, where the network tends to memorise the data in the training subset, but unable to generalise to new situations, e.g. unforeseen data (Hagan et al., 1995). Lastly, the testing subset is used to determine the generalisation ability of the developed model. That is, the error from the testing subset is utilised to compare the predictive performance of different models. Note that there are other aspects of model performance other than predictive validity. The three aspects of model validity are fully covered in Section 3.8.

Data splitting approaches can be categorised as either supervised or unsupervised approaches (see Figure 14). Supervised approaches refer to the process of dividing the input data into three subsets that takes into consideration the statistical properties of each subset. On the other hand, unsupervised approaches do not explicitly take the statistical properties of the data subsets into account, and only stratified unsupervised approaches attempt to ensure that the statistical properties of the subsets are similar (Maier et al., 2010). For instance, SOM can be utilised to cluster the available data and to allocate data samples from each cluster to the training, testing and validation subsets, thereby ensuring that patterns from different regions of the multivariate predictor-output space are represented in each subset. In the random unsupervised approach, the data are randomly divided into their respective subsets. This approach may pose uncertainty on the model results as data in one of the sets may be biased towards extreme or uncommon events (Gardner & Dorling, 1998). Alternatively, a v-fold cross-validation approach can be implemented. The method randomly divides the dataset into v independent subsets. One of the v subsets is selected as test set while the remaining v-1 subsets are utilised for model calibration. The process is repeated several times until a pre-specified criterion is met. In the physics-based approach, the data are split into different classes according to a knowledge about the underlying physical processes. In the ad-hoc approach, the data allocated for the training, validation and testing set are selected in an ad-hoc manner. One popular example is the allocation of the first N observations to the training set, and the next group of observations are allotted to the validation set, and final group for the testing set.

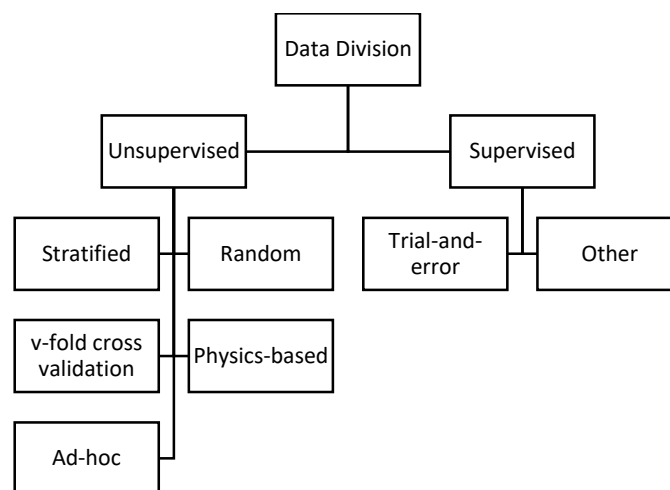


Figure 14. Number of occurrences various input selection methods have been used

3.4.2 Results

In the papers identified, a few did not discuss the process of data division explicitly. Of those mentioned the data division process, the results indicate that the only unsupervised data splitting methods were implemented (see Figure 15). In detail, ad-hoc method was implemented 79 times, while random data division methods done 40 times. The v-fold cross validation method was implemented 8 times. Finally, only a few number of papers employed unsupervised stratified techniques.

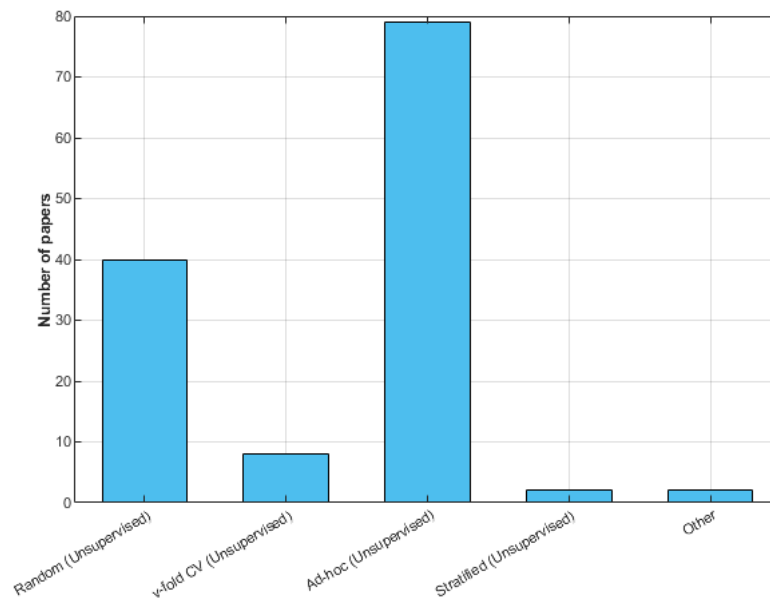


Figure 15. Number of occurrences various data splitting methods have been used

3.4.3 Conclusion

The majority of the identified papers performed data division in an ad-hoc manner. This may lead to uncertainties regarding the quality and repeatability of results, as ANN models fed with different data splits are likely to yield different calibrated network weights and bias, and consequently, different model performance (Maier et al., 2010). Systematic approaches for optimal division of data for ANN models can be considered: genetic algorithm (GA) and self-organizing map (SOM) (Bowden et al., 2002), modified Kennard-Stone algorithm (Saporo et al., 2012). Future modellers may also look into the benchmarking approach proposed by Wu et al. (2013) for comparing different data splitting methods. The authors highlighted the importance of finding a data division method that provides consistent prediction validation error results that are representative of the predictive errors obtained over the full range of the available data.

3.5 Selection of model architecture

3.5.1 Introduction

In this paper, model architecture refers to the overall structure and manner how information flows from one layer to another in ANNs. The taxonomy of model architectures is shown in Figure 16.

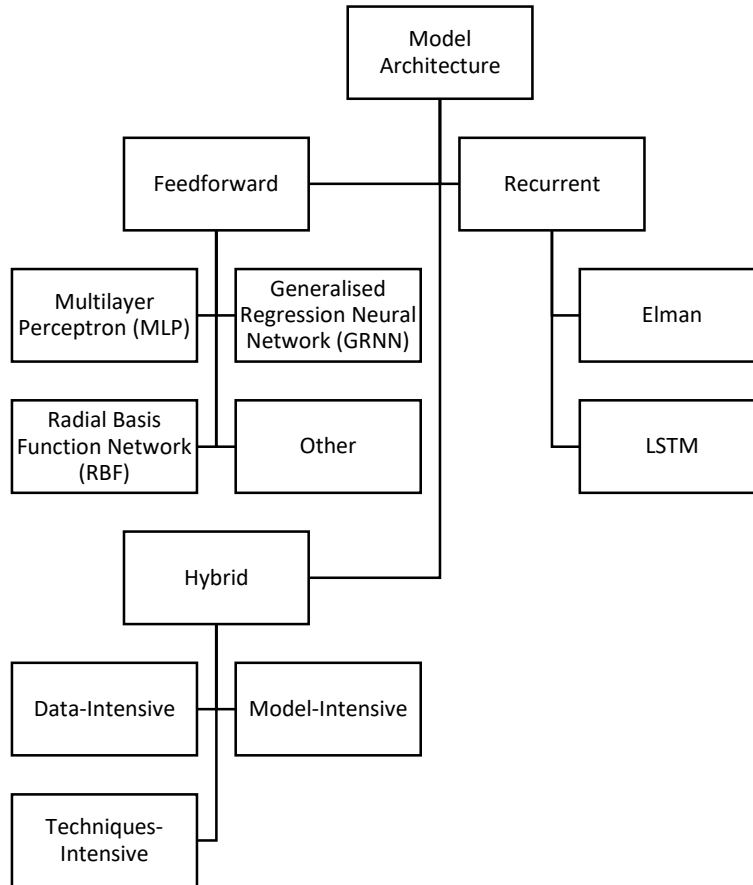


Figure 16. Taxonomy of model architectures

Two of the most popular network architectures applicable for prediction and function approximation are the feedforward and recurrent networks (Hagan et al., 1995). In a feedforward network, information moves from the input layer to the output layer in a single direction. Multilayer perceptron (MLP) is one of the most utilised feedforward ANN types for non-linear function approximation tasks (Shahraiyni & Sodoudi, 2016). Other examples of feedforward ANNs include radial basis function networks (RBFs), general regression neural networks (GRNNs), Ward neural networks (WNN) and extreme learning machine (ELM).

In contrast to feedforward ANNs, a recurrent neural network (RNN) allows feedback. In other words, some output neurons are connected to the neurons of the preceding layers, which can improve the capacity of RNNs to learn (Hagan et al., 1995). One popular example under this type is the Elman network (Biancofiore et al., 2017b; X. Feng et al., 2015; Peng et al., 2017). Sophisticated forms of RNNs including the long short-term memory (LSTM) networks are also receiving more attention from modellers in recent years. Such efforts are being made to address the well-known vanishing gradient problem for which RNNs still suffer from (Freeman et al., 2018). Hagan et al (1995) and Samarasinghe (2006) argued that those of the recurrent type are potentially more effective because of their feedback mechanism which improves their capacity to learn.

Finally, the application of hybrid ANN models has been highlighted in recent years (Makridakis et al., 2018). The ensemble modelling approach has been argued to capitalise on the strengths and overcomes the weaknesses of the individual models involved (Chen et al.,

2008; Shahraiyani et al., 2015; Sharma et al., 2005). In this paper, hybrid models were categorised into the following three classes: data-intensive, model-intensive and technique-intensive (Maier et al., 2010). A data-intensive approach is one that attempts to classify the data with respect to various dynamics dependent on the problem specifications or the criteria set by the modeller. Then, separate models are developed for the identified separate classes. Examples of data-intensive hybrid ANN models include the use of ANNs and techniques such as PCA, k-means clustering, ensemble empirical mode decomposition (EEMD) and wavelet decomposition. Model-intensive approach is one that employs different models for different sub-components of the overall physical system and then aggregates various responses calculated from different models. Fuzzy-neuro networks, e.g. the hybrid of a feedforward ANN and fuzzy systems, multiple restricted Boltzmann machine (RBM) layers and a back-propagation (BP) layer, and LSTM and convolutional neural networks (CNNs) are a few examples of model-intensive ANN models. Lastly, a technique-intensive approach is one in which an ANN is combined with a different technique with the purpose of developing an ensemble approach that exploits the advantages offered by different techniques. Common examples of such type include the use of MLP and support vector machine (SVM), stacked auto-encoders (SAE) and a learning regression (LR) layer, and non-linear autoregressive ANN with exogenous inputs (NARX-ANN).

Network architecture also refers to the way information is translated from input nodes to the nodes of the succeeding layers in a network. Consequently, the selection of appropriate transfer or activation functions which maps the sum of weighted information from a current node to the succeeding node plays an important role. This because the superimposition of different transfer functions determines the ability of an ANN to approximate different input-output dynamics. The selection of an appropriate type of transfer function depends on the nature of the model task (Hagan et al., 1995). It also relies on which network layer the function is to be utilised. In air pollution modelling applications, sigmoid transfer functions are commonly used in the hidden layer as they are nonlinear and easily differentiable (Gardner & Dorling, 1998). Sigmoid function has a graph that looks like a stretched 'S' and yields values between either 0 to 1 or -1 to 1. Such characteristics enable an ANN to approximate extremely any non-linear and complex relationships between predictors and target variables. One of the most popular sigmoid functions include the logistic sigmoid and hyperbolic tangent functions (Bishop, 1995). On the other hand, linear or identity function is found to be the most appropriate transfer function in the output layer for prediction or regression applications as it yields unbounded estimation values (Hagan et al., 1995). It is worth mentioning that feedforward ANNs with linear transfer functions and without hidden nodes are equivalent to linear statistical models.

3.5.2 Results

Figure 17 illustrates the number of times various model architectures have been used by the identified papers. MLP models were found to be the most commonly used model architecture, implemented in 78 papers. Of these papers, linear statistical models were used as benchmarks and outperformed by MLP the models 15 times. The number of studies in which alternative network architectures were employed was reasonably uniform, ranging from 4 to 13. Additionally, it is worth noting that there is a growing trend in the number of papers that employ hybrid ANN modelling approaches in ambient air pollution forecasting tasks in recent years, see Figure 18. The majority of the 45 implemented hybrid models fall under the data-intensive type. The results also reveal the popular use of deep neural networks such as LSTM models. LSTM models were employed 10 times, five of which were coupled with

other modelling techniques. With regards to the selection of transfer functions, almost 30% of the papers reviewed did not provide details concerning their use of transfer functions. Among those that did provide information, logistic sigmoid function was predominantly used in the hidden layer nodes, to be followed by hyperbolic tangent function. In the output nodes, the identity function was widely utilised, followed by the logistic sigmoid. The results also reveal a very few instances where the Gaussian function was used.

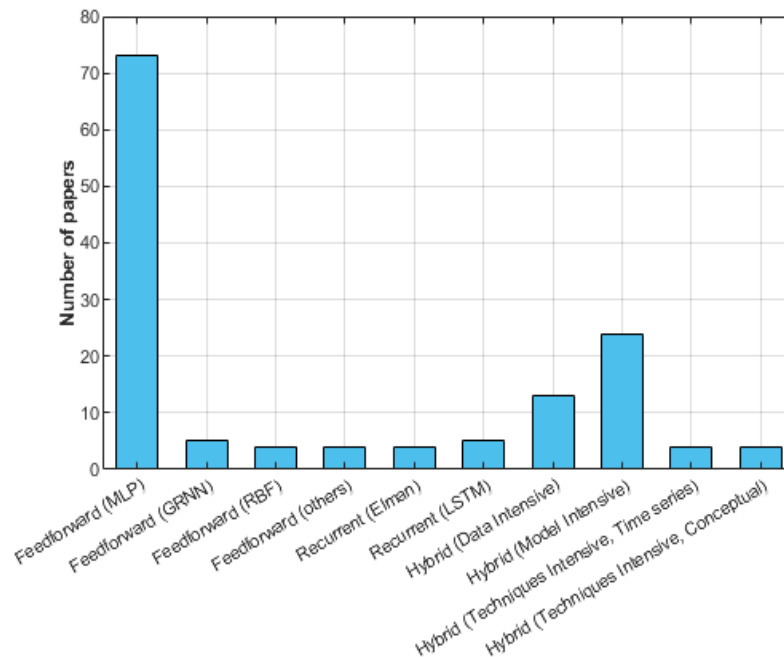


Figure 17. Number of occurrences various model architectures have been used

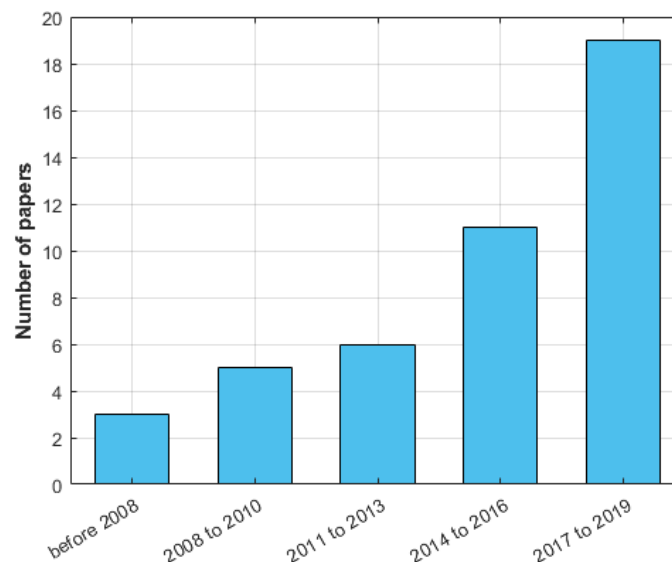


Figure 18. Distribution of papers employing hybrid approaches by year of publication

3.5.3 Conclusion

Much effort has been directed towards the evaluation of existing ANN architectures, especially the MLP model, applied to air pollution forecasting applications. Furthermore, there is a growing interest in the development of novel and more sophisticated network

architectures. This can be explained by the increasing number of easily accessible computing tools that are able to run complicated algorithms rapidly. The use of hybrid modelling approaches has also been practised by modellers given their high potential at improving the performance of plain ANN models (Chen et al., 2008; Sharma et al., 2005). However, Maier et al. (2010) remarked that because of the wide variety of modelling approaches available, it is not possible to draw any conclusions as to which network architecture should be employed for a specific forecasting task. The authors suggested that this requires further investigation in the future. A generic framework for developing both hybrid process and data-driven models of salinity in river systems, which is also applicable to other environmental engineering tasks, can be found in Hunter et al. (2018). In the said proposed framework, the most suitable sub-models are developed for each sub-process of a specific problem of interest based on consideration of model purpose, the degree of process understanding and data availability. The selected sub-models are then combined to form the hybrid model. On the other hand, this review also indicates that the transfer functions in both hidden and output layers were selected properly by the identified papers. However, the results here are not conclusive as the success of transfer functions still relies on the structure of the dataset on which the network is trained from. Hence, this aspect in model development should also be further examined.

3.6 Determination of model structure

3.6.1 Introduction

Another important step in the ANN model development process is the determination of the number of layers and the number of nodes in each layer. The input layer is where the predictors are being fed into, while it is in the output layer is where the final network results are calculated. Hence, the number of input and output nodes are dependent the number of predictors and target variables, respectively. Lastly, the layer other than the input and output layers is called the hidden layer. It is in the hidden layer where the underlying dynamics between predictors and target variables is captured. With the sufficient number of nodes in the hidden layer, an ANN can approximate almost any function (Bishop, 1995; Samarasinghe, 2006). ANNs may have one or more hidden layers. It has been shown that the use of too many hidden layers and neurons can lead to model overfitting, while the opposite can cause to model underfitting (Hagan et al., 1995; Samarasinghe, 2006). The said issues are reported to affect the generalisation ability of a model, leading to poor prediction accuracy. In addition, the hidden layer parameters are case-specific as they are dependent on the data complexity of a specific application (Gardner & Dorling, 1998). A general method that determines the optimal number of hidden layers and neurons to be used still remains unknown, thus contributing to the initial difficulty in ANN model building. As a result, different approaches have been employed to address the said uncertainty.

The methods for determining the optimal ANN model structure can be classified into three types, namely, global, stepwise trial-and-error or ad-hoc (see Figure 19). In the first approach, the optimal number of hidden layers and nodes are determined using global methods based on competitive evolution found in nature, e.g. GA, particle swarm optimisation, simulated annealing, etc. Using this approach, it is possible to simultaneously optimize the network weights and biases, and the number of hidden layer and nodes. If implemented properly and appropriately, the global methods are likely to result in the best ANN structure and parameters. However, they are found to be computationally expensive (Maier et al., 2010). The stepwise trial-and-error approach can be used, in which a basic ANN structure is first assumed, which is modified with each trial with the objective of achieving a structure that is

neither too complex nor too simple. The stepwise approach can be further split into two types, one based on pruning algorithm and the second based on constructive approaches. Lastly, ad-hoc approaches can be used to determine an optimal network structure in which the number of hidden nodes is determined without adhering to strict pruning and constructive techniques. One ad-hoc approach is based purely on a trial-and-error approach. The partial ad-hoc is based on the use of both a trial-and-error approach and an empirical formula that provides upper and/or lower bounds of the number of hidden nodes. The last ad-hoc approach is based on experience or intuition of the modeller.

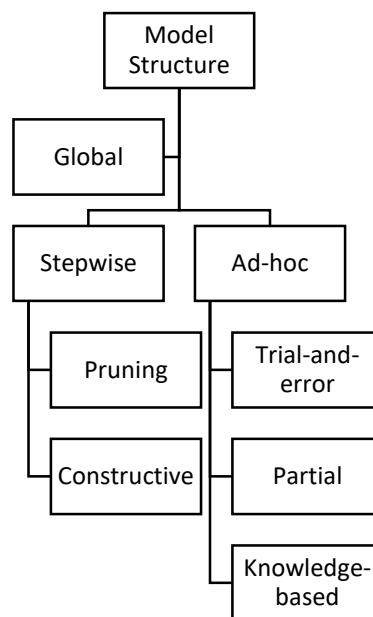


Figure 19. Taxonomy of model structure

3.6.2 Results

As can be seen in Figure 20, an ad-hoc approach to determining the structure of ANN models was by far the most popular, with 132 applications. It is worth noting that a number of identified papers, e.g. 13 papers, implemented an approach that combines trial-and error and empirical rules to determine the optimal number of hidden nodes. Of the structured approaches, constructive stepwise approaches were implemented 5 times, whereas the global approaches were only used twice.

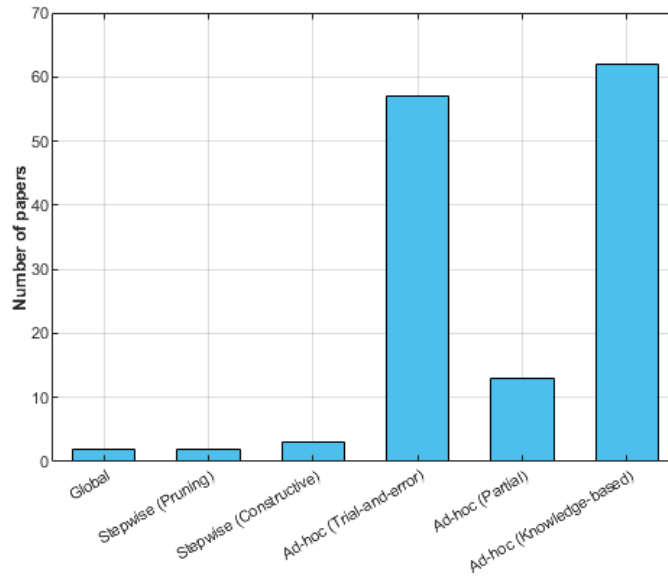


Figure 20. Distribution of papers employing hybrid approaches by year of publication

3.6.3 Conclusion

Despite the important role network structure plays in determining the desired relationship between model predictors and outputs, little effort has been directed into this area of the ANN model development process. This is evident in this review where the majority of the identified studies employed an ad-hoc approach to determining an appropriate ANN model structure. There has been a little adoption of constructive, stepwise model building approaches, but the use of global optimisation methods has received little attention. As such, this stage requires further attention in the future. Future modellers may look into an objective approach based on Bayesian model selection (BMS) method for ANNs for comparing models of varying complexity in order to select the most appropriate ANN structure which can be found in Kingston et al. (2008). The approach which utilises Markov Chain Monte Carlo posterior simulations to estimate the evidence in favour of competing models. The authors remarked that the said approach provides a simple and objective method for selecting the ANN model with the optimal complexity when used in conjunction with the Bayesian training procedure developed by (Kingston et al., 2005).

3.7 Model training

3.7.1 Introduction

Before ANN models are employed for forecasting applications, they must be trained or calibrated to do so. Training an ANN is the process of calibrating the connection weights between the interconnected nodes of the network. It is through the connection weights and node biases that an ANN can be able to approximate complex non-linear mappings from the nodes of the input layer to the network outputs. Training an ANN is typically carried out in a supervised manner. Before training, the network weights and biases are usually initialised. Initial weighting values are typically selected randomly from a uniform distribution (Hagan et al., 1995). During the training process, the network is repeatedly presented with the desired network response for each input pattern as the network weights and biases are calibrated

until the target outcome, e.g. the acceptable difference between the desired and actual output, is met.

Network calibration methods generally belong to either local or global optimization approaches (see Figure 21). Local methods usually work on gradient information and are therefore prone to becoming trapped in local optima if the error surface is reasonably rugged. However, these methods are generally computationally efficient. Gradient methods can be further sub-divided into first-order methods, e.g. back-propagation, or second-order methods, e.g. Newton's method and conjugate gradient method. When using the back-propagation, suitable values of the network training parameters such as the learning rate and momentum term also need to be initialised first. The suitable values for the said parameters are also case-specific. Nonetheless, a few empirical formulas for finding them are available in the literature (Gardner & Dorling, 1998; Hagan et al., 1995; Samarasinghe, 2006). Global optimization methods, such as genetic algorithms, have an increased ability to find global optima in the error surface, although this is generally at the expense of computational efficiency. Alternatively, stochastic calibration methods can be utilised to account for parameter uncertainty. These approaches can be used to obtain distributions of the model parameters, rather than finding a single parameter vector. This has the advantage that prediction limits can be obtained. In order to achieve this, Bayesian methods are commonly used (Bishop, 1995; Maier et al., 2010).

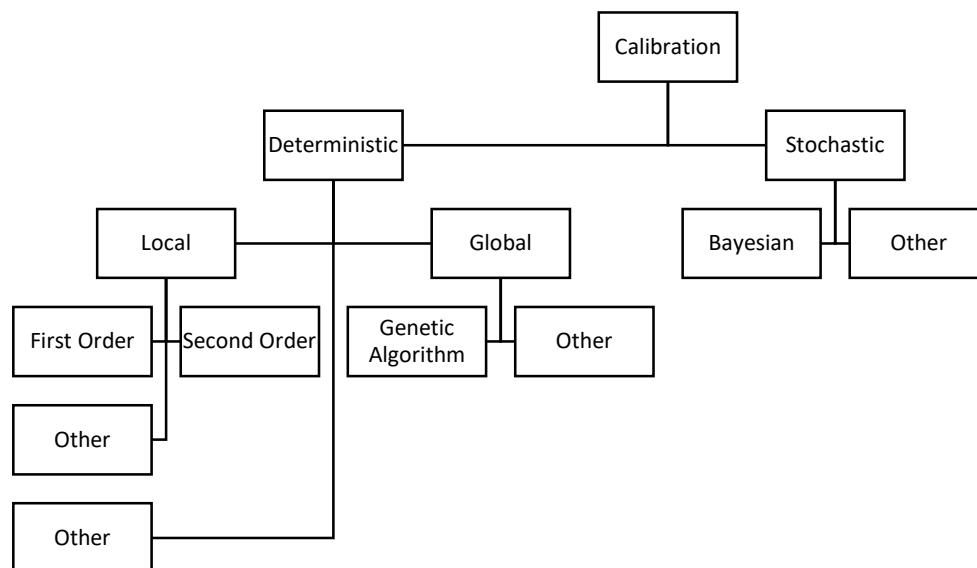


Figure 21. Taxonomy of model calibration techniques

3.7.2 Results

The results in Figure 22 indicate that deterministic local calibration techniques were commonly used, e.g. 90 times out of the 104 cited techniques by the identified papers. Of those techniques, 42 of them falls under first-order approaches and 35 falls under second-order approaches. Additionally, there were 9 studies that utilised global techniques and 3 that employed stochastic techniques.

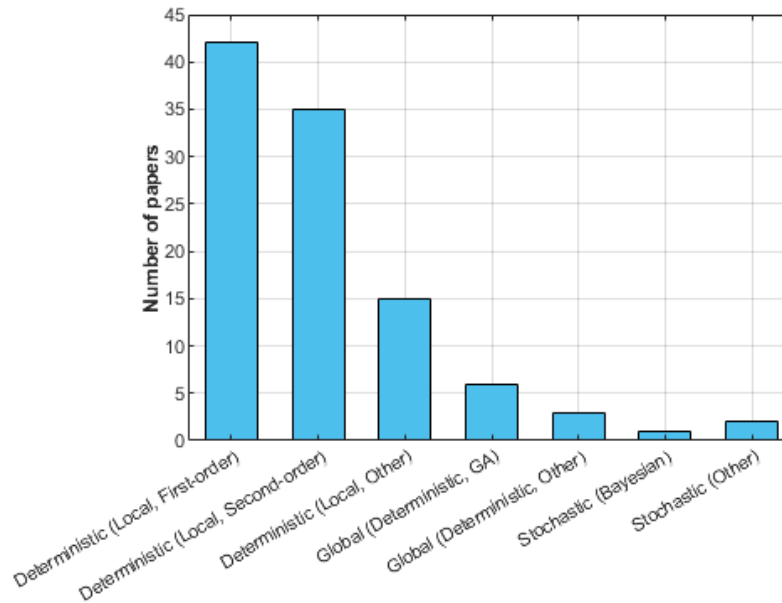


Figure 22. Number of occurrences various training methods have been used

3.7.3 Conclusion

First-order local search procedures, such as the backpropagation algorithm, were primarily used by the identified papers, although second order methods were also utilised extensively in order to improve the computational efficiency of ANN calibration. However, studies investigating the potential benefits of using global optimisation techniques in terms of improving the predictive ability of ANN models are rather limited, which is an area worthy of further exploration. Some studies that employ global optimisation techniques during ANN calibration can be found in the literature (Antanasijević et al., 2018; Y. Feng et al., 2011; Lu et al., 2003; Niska et al., 2004; Pires et al., 2010). A model calibration technique that accounts for the relative contributions of model predictors in generating the target variable could also be further examined. Kingston et al. (2005) argued that the quantification of the relative input contributions can facilitate in assessing the trained ANN based on the predictor-target variable relationship estimated. Finally, future works could also delve into effect of the input pattern to be included in the calibration set on the performance of the resulting ANN models in a rapid and continuous manner. Bowden et al. (2012) proposed an approach based on SOMs that can identify uncharacteristic data patterns presented to ANN models.

3.8 Model validation

3.8.1 Introduction

A wide range of statistical performance indices must be employed to quantify the performance of the developed ANN models. ANN model performance is usually assessed using a quantitative error metric. However, ANN models should not be assessed solely on their predictive error, but also through their ability to capture underlying dynamics between predictors and target variables (Kingston et al., 2005). As such, three aspects of model validity need to be considered when assessing the performance of ANN models, or data-driven models, generally speaking: replicative validity, predictive validity, and structural validity (Gass, 1983; Humphrey et al., 2017) (see Figure 23).

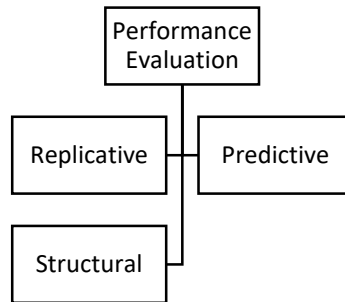


Figure 23. Taxonomy of performance evaluation techniques

Metrics assessing replicative validity see to it that a developed model correctly approximates both observed data and those utilised in previous ANN model building steps (Gass, 1983). Popular methods under this category include means and variance, minimum and maximum, analysis of variance, goodness-of-fit testing, regression and correlation analysis, and confidence interval construction (Wu et al., 2014). On the other hand, metrics dealing with predictive validity examine the performance of ANN models in approximating unforeseen or independent data. A taxonomy of the commonly used metrics under this category is given in Figure 24.

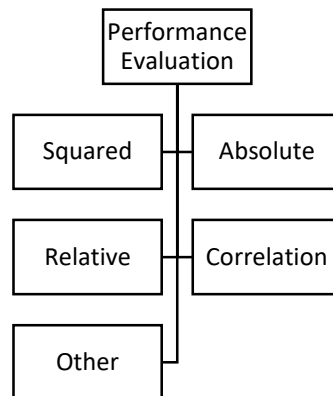


Figure 24. Taxonomy of predictive performance evaluation techniques

Squared errors are based on the squares of the differences between actual and modelled output values. Common examples include the mean squared error (MSE), sum of squared errors (SSE) and root mean square errors (RMSE). Absolute errors are based on the absolute differences between actual and modelled outputs. Metrics under this type include mean sum of absolute deviations (MSAD) and total sum of absolute deviations (TSAD). Relative errors measure the performance of models with outputs. Common examples under this type include average absolute relative error (AARE), the normalized root mean square error (NRMSE) and the normalized mean bias error (NMBE). Finally, correlation errors measure the empirical error between actual and modelled outputs. One common example is the Pearson correlation coefficient (r). Other metrics include the information criteria, such as the Akaike information criterion (AIC) and Bayesian information criterion (BIC), which consider model complexity in addition to model error. Finally, metrics dealing with structural validity ensure that a model is plausible when compared with *a priori* knowledge of the system behaviour, which is intended to be reflected in the resulting model (Wu et al., 2014). That is, a model is structurally valid if it not only correctly approximates the observed data, but also reflects the way in which the

real system is understood to operate to display such data characteristics. Structural validity metrics also include the measure of uncertainties of ANN model results. Methods that measure structural validity include sensitivity analysis (Mount et al., 2013), overall connection weights (Olden & Jackson, 2002), and measure of generalisation (Razavi & Tolson, 2011). The reader is suggested to read other works, see (Dawson et al., 2007; Humphrey et al., 2017) for a more detailed discussion regarding model validation.

3.8.2 Results

The results depicted in Figure 25 indicate that a range of performance criteria was used in most studies. It can be observed that predictive validity measures were used predominantly, while replicative and structural validity were considered 18 and 10 times, respectively. The number of times different model replicative validity were utilised is shown in Figure 26. In those papers where replicative validity was considered, the mean, standard deviation, minimum and maximum values between the observed and estimated values in the calibration stage were compared, while the Kolmogorov-Smirnov test was utilised to evaluate the statistical significance of the model training results with significance level of 0.05. The number of times various model predictive performance metrics were used to validate the models is given in Figure 27. While correlation and squared error metrics were also widely used (163 and 118 times, respectively), measures based on absolute and relative errors, also employed extensively. Furthermore, visual inspection was utilised in the majority of the papers identified (see Figure 28). The comparison between plots of actual and predicted values were predominantly shown (44 times), to be followed by scatter plots (36 times) and error histogram and surfaces (9 times). The number of times various model structural performance metrics were used is shown in Figure 29. In those 10 occurrences where structural validity was considered, sensitivity analysis was conducted only 6 times, uncertainty in model predictions were quantified only 3 times, and a skill score metric that determines a model performance relative to a reference model was only used once.

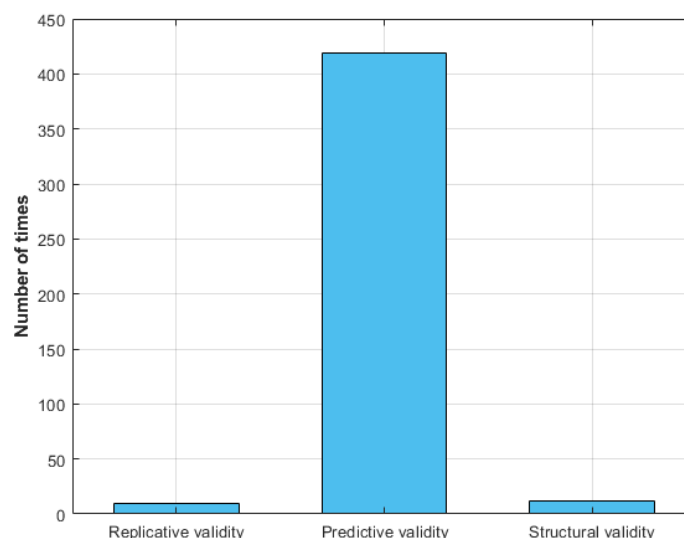


Figure 25. Number of occurrences various aspects of validity have been considered

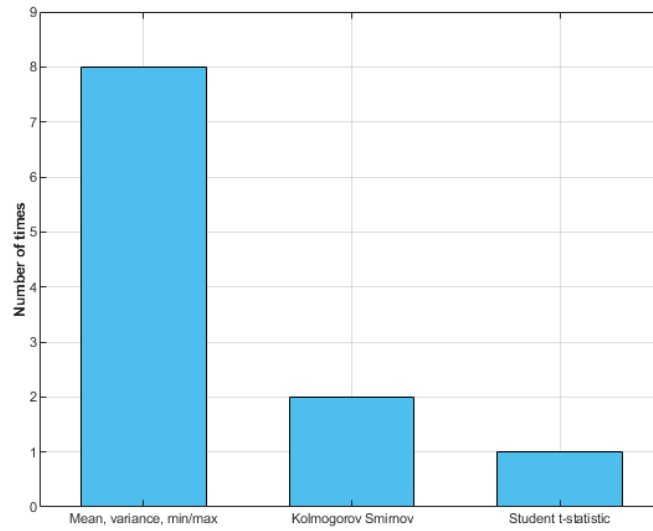


Figure 26. Number of occurrences various replicative model validity metrics have been considered

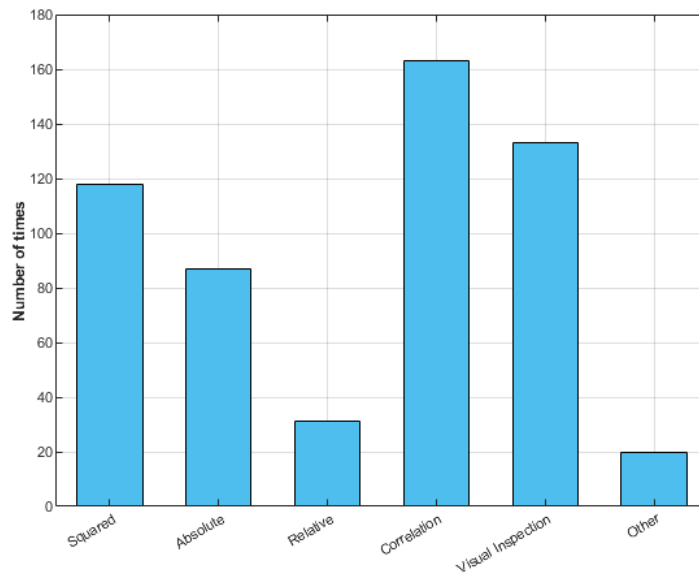


Figure 27. Number of occurrences various predictive model validity metrics have been considered

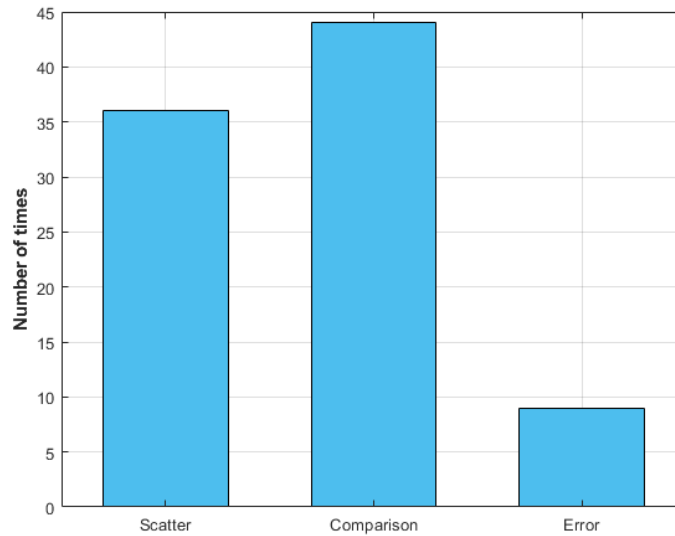


Figure 28. Number of occurrences various visual inspection methods have been used

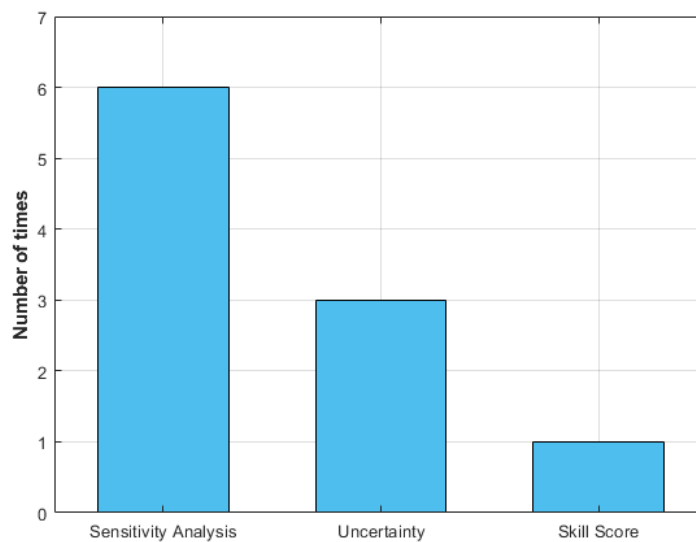


Figure 29. Number of occurrences various structural model validity metrics have been considered

3.8.3 Conclusion

Review of the 139 papers has indicated that a range of performance criteria, mainly those that assess the predictive performance of models, was used predominantly by the identified papers. While this observation suggests a good modelling practice, the replicative and structure validity of the models developed were generally ignored. Given the black box nature of ANN models, the predictive aspect of model validation is not sufficient enough to fully assess the ability of the developed models to fully capture the underlying dynamics between predictors and target variables. Hence, future works should incorporate these additional aspects of validation to provide a wider assessment of their model performance. A full discussion on the comprehensive validity assessment of ANN models for prediction tasks can be found in Humphrey et al. (2017).

IV. Conclusions

Since the period between January 2001 and February 2019, research activity in the field of forecasting and prediction of ambient air pollution variables using ANNs has increased dramatically. Many journal papers have been identified in the covered period, despite the fact that a restricted journal list was considered and that the review was limited to predicting air pollutant variables in urban and suburban environments. Even within the period covered by this paper, there has been an increase in the number of papers published in the later years, with an average of 10 papers per year from 2010 to 2018. This can be well explained by the increased availability of software packages that enable modellers to build and train ANNs relatively easily and quickly (Gardner & Dorling, 1998; Sharma et al., 2005).

As was the case of the identified papers, the primary application area has been the forecasting and prediction of outdoor PM₁₀, PM_{2.5}, oxides of nitrogen, and ozone. This is consistent with the findings of the review by Yetilmezsoy et al. (2011), where a survey on techniques based on artificial intelligence in modelling air pollution was made. This number can be explained by the growing necessity to alleviate the adverse effects of the said key air pollutants on human health by means of early warning and preventive measures (World Health Organization, 2013). Attention was also directed towards other variables such as the oxides of CO and SO₂. Consequently, there is a need to broaden the application areas of ANN models to focus on other air pollutant variables. ANNs would seem to be ideally suited to modelling complex relationships of environmental variables given their universal function approximation capability of ANNs (Gardner & Dorling, 1998; Hornik et al., 1989; Maier et al., 2010). Additionally, the vast majority of identified papers tested the effectiveness of ANN models in the long-term forecasting of air pollution. This highlights the growing demand for parsimonious and effective early-warning systems capable of providing accurate forecasts to the public, urban planners, and decision-makers.

The majority of the identified studies have been found to utilise datasets spanning at least a one-year period, which is considered sufficient enough to reveal the annual cyclic pattern of air pollution levels. This is generally considered a good practice. However, there may be cases wherein the said period may not be sufficient especially when extreme levels of air pollution concentration are not well represented enough. The use of pre-processing methods that address the imbalance data problem may be beneficial in this situation. As for the initial collection of predictors, meteorological and pollutant emissions variables have been predominantly considered by the papers reviewed, with meteorological variables receiving greater emphasis. This is consistent with the findings of a review by Shahraiyini & Sodoudi (2016) on the prediction of ambient PM₁₀ levels using statistical models. Although the use of several predictors of various types is encouraged to extract inherent associations between predictors and target variables, this should be handled carefully as modelling with ANNs typically requires careful consideration between model complexity and performance (Bishop, 1995; Hagan et al., 1995; Samarasinghe, 2006). The use of too many predictors can reduce the nonlinear mechanisms of ANNs to merely linear ones, defeating the entire point of utilising ANNs for air pollution prediction in the first place. Furthermore, the use of parsimonious models is typically encouraged in real-world pollution forecasting tasks where data is limited. This highlights the need to build ANN models that are powerful enough to reveal inherent features from a small number of predictors.

The majority of the identified papers failed to provide details with regards to the implementation of data pre-processing techniques. This can cast doubts to future modellers,

especially those that attempt to repeat the methods and reproduce the findings of the said papers. On the other hand, most of the identified papers did not handle missing data well by considering the list-wise deletion of predictors with missing data. The effective use of data normalisation techniques should be the focus of future research. Another aspect of the ANN model development that requires further attention is the implementation of predictor selection, as the majority of the identified papers selected their model predictors in an ad-hoc fashion. Much attention should be directed towards the use of model-based global methods for predictor selection, given the current limitations and difficulty in implementing them.

Similar to the findings of other similar reviews (Gardner & Dorling, 1998; Maier & Dandy, 2000; Maier et al., 2010; Wu et al., 2014), the vast majority of the identified papers performed data division in an ad-hoc fashion. It has been pointed out that the said popular approach can entail uncertainties especially when assessing model performance as the resulting network can only generalise on the range of data inputs for which it was trained (Gardner & Dorling, 1998). Alternative methods to data division methods should be examined and incorporated in future research efforts. With regards to model architecture selection, the findings of this review found that feedforward networks were predominantly used by the identified papers, most of which were MLPs. While MLPs are still found to be the dominant ANN architecture in this paper, they were also predominantly used as a benchmark against which to compare alternative architectures. This finding is similar to those of several previous reviews (Gardner & Dorling, 1998; Zhang & Patuwo, 1998; Sharma et al., 2005; Yetilmezsoy et al., 2011). MLPs were extensively used in the prediction and forecasting of water quality and quantity variables (Maier & Dandy, 2000; Maier et al., 2010; Wu et al., 2014). On the other hand, there was a significant amount of experimentation with other types of feedforward architectures, such as generalised regression neural networks and radial basis function networks, and recurrent networks and, most importantly, different types of hybrid network architectures. It is also worth noting that the utilisation of hybrid ANN model architectures has increased in the past decade. The development of hybrid ANN model architectures is an important progress, as it emphasizes that “ANNs have a role to play not only as an alternative to traditional modelling approaches, but also as a complementary modelling tool that can be used to improve the performance of existing approaches” (Maier et al., 2010).

The findings of this review also suggest that more efforts are directed towards the use of global optimisation methods in determining the optimal model structure, as the majority of the identified papers carried out ad-hoc approaches, primarily trial-and-error method. This trend may not be beneficial to future modellers in the field of air pollution forecasting as ad-hoc methods offer limited repeatability due to their case-specific nature. Consequently, there is a need for a more established and systematic protocol for identifying optimal model structures that caters to a wide range of model predictor-output dynamics. First-order local search procedures, such as the backpropagation algorithm, were primarily used, although second-order methods were also used extensively in order to improve the computational efficiency of ANN calibration. However, studies investigating the potential benefits of using global optimization techniques in terms of improving the predictive ability of ANN models are rather limited, which is an area worthy of further exploration. In addition, although some work was done on the incorporation of parameter uncertainty into ANN model calibration, this also presents an area of future research. Finally, different performance metrics in terms of the predictive validity of the models were utilised in the majority of the papers reviewed. However, further efforts should be made to examine other aspects of model performance such as replicative validity and structural validity.

V. Recommendations for future work

Based on the review of 139 journal papers dealing with the forecasting and prediction of ambient air pollution variables using ANNs published between January 2001 and February 2019, the following recommendations for future research are made:

1) Each step of the overall model development process of ANNs should be viewed as interconnected entities. That is, the links between each modelling stage and the justification for the component chosen for each stage should be stated explicitly. There is always the tendency among modellers to regard one aspect of the modelling stage to not have an influence on the others. As such, some specific aspects that need further emphasis are the following:

- a) the influence of the length of training data utilised on the overall ANN model performance;
- b) the significance of the selected predictors and utilised model structure on the complexity and overall ANN model performance; and
- c) the links between the selected data normalisation scheme and transfer function utilised;
- d) the influence of the adapted initialisation schemes for the weighting, bias and other training parameters on the overall ANN model performance.

2) More emphasis should be given on the trade-off between the sophistication of a model and structure architecture adapted and the overall model performance. Modellers tend to have the tendency to consider more sophisticated models without first examining the ability of much simpler ones to perform a modelling task. Sometimes, model complexity sabotages the overall performance and potential of an ANN model to be deployed in real-world forecasting tasks in terms of model parsimony.

3) Aside from the replicative, predictive and structural aspects of model validity, future modellers should also assess a developed ANN model in terms of computational penalty and running time as these are essential factors for models to be practical enough in real-world modelling tasks. Additionally, uncertainty analysis of model results should be investigated more in future works. For ANN models to be fully-implementable in real-world applications, future efforts should be made towards the quantification of uncertainty in model results (Arhami et al., 2013; Borrego et al., 2008).

4) More information regarding the software or computing environment employed in the ANN model development should be provided. The built-in functions or libraries that perform some or all steps in building ANN models between many software often possess minor or major difference. As such, the comparison of one modelling approach to another should be carried out in the same modelling environment to avoid bias. Additionally, all software settings used in building ANN models should be explicitly disclosed to avoid ambiguities and increase repeatability of results. Maier et al. (2010) also advised the use of open access data sets in order to enable a better comparison of ANN development methods across studies.

5) Although each modelling task utilising data-driven models such as ANNs is problem-specific, the adoption of several protocols to building ANN models should be adapted more seriously in the future to ensure a good modelling practice (Wu et al., 2014).

6) A complete theoretical understanding of the principles behind the ANN modelling paradigm should be expected among modellers in order for the field of environmental modelling using ANNs to advance (Gardner & Dorling, 1998). The emergence of many easy-to-use computing platforms supporting ANN model building tasks may cause some

modellers to take advantage of the black box nature of ANNs, making the overall process more of an art rather than science.

Acknowledgements

The authors would like to acknowledge the financial support provided by the British Council, Philippines (through grant number 261810845) and the Commission on Higher Education of the Republic of the Philippines that has enabled this production of this work.

References

- Abderrahim, H., Chellali, M. R., & Hamou, A. (2016). Forecasting PM₁₀ in Algiers: efficacy of multilayer perceptron networks. *Environmental Science and Pollution Research*, 23(2), 1634–1641. <https://doi.org/10.1007/s11356-015-5406-6>
- Abdul-Wahab, S. ., & Al-Alawi, S. . (2002). Assessment and prediction of tropospheric ozone concentration levels using artificial neural networks. *Environmental Modelling & Software*, 17(3), 219–228. [https://doi.org/10.1016/S1364-8152\(01\)00077-9](https://doi.org/10.1016/S1364-8152(01)00077-9)
- Agirre-Basurko, E., Ibarra-Berastegi, G., & Madariaga, I. (2006). Regression and multilayer perceptron-based models to forecast hourly O₃ and NO₂ levels in the Bilbao area. *Environmental Modelling and Software*, 21(4), 430–446. <https://doi.org/10.1016/j.envsoft.2004.07.008>
- Al-Alawi, S. M., Abdul-Wahab, S. A., & Bakheit, C. S. (2008). Combining principal component regression and artificial neural networks for more accurate predictions of ground-level ozone. *Environmental Modelling and Software*, 23(4), 396–403. <https://doi.org/10.1016/j.envsoft.2006.08.007>
- Alam, M. S., & McNabola, A. (2015). Exploring the modeling of spatiotemporal variations in ambient air pollution within the land use regression framework: Estimation of PM₁₀ concentrations on a daily basis. *Journal of the Air & Waste Management Association*, 65(5), 628–640. <https://doi.org/10.1080/10962247.2015.1006377>
- Alimissis, A., Philippopoulos, K., Tzanis, C. G., & Deligiorgi, D. (2018). Spatial estimation of urban air pollution with the use of artificial neural network models. *Atmospheric Environment*, 191, 205–213. <https://doi.org/10.1016/J.ATMOENV.2018.07.058>
- Antanasijević, D., Pocajt, V., Perić-Grujić, A., & Ristić, M. (2018). Multiple-input–multiple-output general regression neural networks model for the simultaneous estimation of traffic-related air pollutant emissions. *Atmospheric Pollution Research*, 9(2), 388–397. <https://doi.org/10.1016/j.apr.2017.10.011>
- Antanasijević, D. Z., Pocajt, V. V., Povrenović, D. S., Ristić, M. Đ., & Perić-Grujić, A. A. (2013). PM₁₀ emission forecasting using artificial neural networks and genetic algorithm input variable optimization. *Science of The Total Environment*, 443, 511–519. <https://doi.org/10.1016/j.scitotenv.2012.10.110>
- Arhami, M., Kamali, N., & Rajabi, M. M. (2013). Predicting hourly air pollutant levels using artificial neural networks coupled with uncertainty analysis by Monte Carlo simulations. *Environmental Science and Pollution Research*, 20(7), 4777–4789. <https://doi.org/10.1007/s11356-012-1451-6>
- Ashmore, M. R., Dimitroulopoulou, C., Byrne, M. A., & Kinnersley, R. P. (2001). Modelling of indoor exposure to nitrogen dioxide in the UK. *Atmospheric Environment*, 35, 269–279. Retrieved from <https://pdf.sciencedirectassets.com/271798/1-s2.0->

S1352231000X01549/1-s2.0-S135223100000176X/main.pdf?x-amz-security-token=AgoJb3JpZ2luX2VjEB8aCXVzLWVhc3QtMSJIMEYCIQD%2BBJkssu7s8wqKquYtpBljYPISp2VHr%2BT28DxjHvsH%2FglhAKpDwMpcYXqzSVjS8emlqNIptxW3d9PtZFB%2F

- Azid, A., Juahir, H., & Toriman, M. E. (2014). Prediction of the Level of Air Pollution Using Principal Component Analysis and Artificial Neural Network Techniques : a Case Study in Malaysia. <https://doi.org/10.1007/s11270-014-2063-1>
- Bai, Y., Li, Y., Wang, X., Xie, J., & Li, C. (2016). Air pollutants concentrations forecasting using back propagation neural network based on wavelet decomposition with meteorological conditions. *Atmospheric Pollution Research*, 7(3), 557–566. <https://doi.org/10.1016/j.apr.2016.01.004>
- Bai, Y., Zeng, B., Li, C., & Zhang, J. (2019). An ensemble long short-term memory neural network for hourly PM 2.5 concentration forecasting. *Chemosphere*. <https://doi.org/10.1016/j.chemosphere.2019.01.121>
- Baklanov, A., Hänninen, O., Slørdal, L. H., Kukkonen, J., Bjergene, N., Fay, B., ... Ødegaard, V. (2007). Integrated systems for forecasting urban meteorology, air pollution and population exposure. *Atmos. Chem. Phys. Atmospheric Chemistry and Physics*, 7, 855–874. Retrieved from www.atmos-chem-phys.net/7/855/2007/
- Barrón-adame, J. M., Cortina-januchs, M. G., Vega-corona, A., & Andina, D. (2012). Expert Systems with Applications Unsupervised system to classify SO 2 pollutant concentrations in Salamanca , Mexico. *Expert Systems with Applications*, 39, 107–116. <https://doi.org/10.1016/j.eswa.2011.05.083>
- Bashir, F., & Wei, H.-L. (2017). Handling missing data in multivariate time series using a vector autoregressive model-imputation (VAR-IM) algorithm. *Neurocomputing*, 276, 23–30. <https://doi.org/10.1016/j.neucom.2017.03.097>
- Biancofiore, F., Busilacchio, M., Verdecchia, M., Tomassetti, B., Aruffo, E., Bianco, S., ... Di Carlo, P. (2017a). Recursive neural network model for analysis and forecast of PM10 and PM2.5. *Atmospheric Pollution Research*, 10, 8–15. <https://doi.org/10.1016/j.apr.2016.12.014>
- Biancofiore, F., Busilacchio, M., Verdecchia, M., Tomassetti, B., Aruffo, E., Bianco, S., ... Di Carlo, P. (2017b). Recursive neural network model for analysis and forecast of PM10 and PM2.5. *Atmospheric Pollution Research*, 8(4), 652–659. <https://doi.org/10.1016/j.apr.2016.12.014>
- Biancofiore, F., Verdecchia, M., Di, P., Tomassetti, B., Aruffo, E., Busilacchio, M., ... Colangeli, C. (2015). Science of the Total Environment Analysis of surface ozone using a recurrent neural network. *Science of The Total Environment*, 514, 379–387. <https://doi.org/10.1016/j.scitotenv.2015.01.106>
- Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. New York, NY, USA: Oxford University Press, Inc.
- Borrego, C., Monteiro, A., Ferreira, J., Miranda, A. I., Costa, A. M., Carvalho, A. C., & Lopes, M. (2008). Procedures for estimation of modelling uncertainty in air quality assessment. *Environment International*, 34(5), 613–620. <https://doi.org/10.1016/j.envint.2007.12.005>
- Bowden, G. J., Dandy, G. C., & Maier, H. R. (2003). *Data transformation for neural network models in water resources applications*. Retrieved from

<https://iwaponline.com/jh/article-pdf/5/4/245/392619/245.pdf>

- Bowden, G. J., Maier, H. R., & Dandy, G. C. (2002). Optimal division of data for neural network models in water resources applications. *Water Resources Research*, 38(2), 2-12–11. <https://doi.org/10.1029/2001WR000266>
- Bowden, G. J., Maier, H. R., & Dandy, G. C. (2012). Real-time deployment of artificial neural network forecasting models: Understanding the range of applicability. *Water Resources Research*, 48(10), 1–16. <https://doi.org/10.1029/2012WR011984>
- Brunelli, U., Piazza, V., Pignato, L., Sorbello, F., & Vitabile, S. (2007). Two-days ahead prediction of daily maximum concentrations of SO₂, O₃, PM₁₀, NO₂, CO in the urban area of Palermo, Italy. *Atmospheric Environment*, 41(14), 2967–2995. <https://doi.org/10.1016/j.atmosenv.2006.12.013>
- Cabaneros, S. M. S., Calautit, J. K. S., & Hughes, B. R. (2017). Hybrid Artificial Neural Network Models for Effective Prediction and Mitigation of Urban Roadside NO₂ Pollution. *Energy Procedia*, 142, 3524–3530. <https://doi.org/10.1016/J.EGYPRO.2017.12.240>
- Catalano, M., Galatioto, F., Bell, M., Namdeo, A., & Bergantino, A. S. (2016). Improving the prediction of air pollution peak episodes generated by urban transport networks. *Environmental Science and Policy*, 60, 69–83. <https://doi.org/10.1016/j.envsci.2016.03.008>
- Chang, M. E., & Cardelino, C. (2000). Application of the Urban Airshed Model to forecasting next-day peak ozone concentrations in Atlanta, Georgia. *Journal Of The Air & Waste Management Association (1995)*, 50(11), 2010–2024. <https://doi.org/10.1080/10473289.2000.10464219>
- Chattopadhyay, S., & Chattopadhyay, G. (2012). Modeling and Prediction of Monthly Total Ozone Concentrations by Use of an Artificial Neural Network Based on Principal Component Analysis. *Pure and Applied Geophysics*, 169, 1891–1908. <https://doi.org/10.1007/s00024-011-0437-5>
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). *SMOTE: Synthetic Minority Over-sampling Technique*. *Journal of Artificial Intelligence Research* (Vol. 16). Retrieved from <https://arxiv.org/pdf/1106.1813.pdf>
- Chelani, A. B., Chalapati Rao, C. ., Phadke, K. ., & Hasan, M. . (2002). Prediction of sulphur dioxide concentration using artificial neural networks. *Environmental Modelling & Software*, 17(2), 159–166. [https://doi.org/10.1016/S1364-8152\(01\)00061-5](https://doi.org/10.1016/S1364-8152(01)00061-5)
- Chelani, A. B., Singh, R. N., & Devotta, S. (2005). Nonlinear dynamical characterization and prediction of ambient nitrogen dioxide concentration. *Water, Air, and Soil Pollution*, 166(1–4), 121–138. <https://doi.org/10.1007/s11270-005-7384-7>
- Chellali, M. R., Abderrahim, H., Hamou, A., Nebatti, A., & Janovec, J. (2016). Artificial neural network models for prediction of daily fine particulate matter concentrations in Algiers. *Environmental Science and Pollution Research*, 14008–14017. <https://doi.org/10.1007/s11356-016-6565-9>
- Chen, S. H., Jakeman, A. J., & Norton, J. P. (2008). Artificial Intelligence techniques: An introduction to their use for modelling environmental systems. *Mathematics and Computers in Simulation*, 78(2–3), 379–400. <https://doi.org/10.1016/j.matcom.2008.01.028>
- Chuang, M. T., Zhang, Y., & Kang, D. (2011). Application of WRF/Chem-MADRID for real-time

- air quality forecasting over the Southeastern United States. *Atmospheric Environment*, 45(34), 6241–6250. <https://doi.org/10.1016/j.atmosenv.2011.06.071>
- Colls, J. (2001). *Air Pollution 2nd Edition*. 29 West 35th Street, New York, NY 10001: Spon Press.
- Coman, A., Ionescu, A., & Candau, Y. (2008). Hourly ozone prediction for a 24-h horizon using neural networks. *Environmental Modelling & Software*, 23(12), 1407–1421. <https://doi.org/10.1016/J.ENVSOFT.2008.04.004>
- Cortina-Januchs, M. G., Dominguez, J. Q., Corona, A. V., & Andina, D. (2015). Development of a model for forecasting of PM10 concentrations in Salamanca, Mexico. *Atmospheric Pollution Research*, 6(4), 626–634. <https://doi.org/10.5094/APR.2015.071>
- Dawson, C. W., Abrahart, R. J., & See, L. M. (2007). HydroTest: A web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. *Environmental Modelling & Software*, 22(7), 1034–1052. <https://doi.org/10.1016/J.ENVSOFT.2006.06.008>
- Demir, G., Ozcan, K., Ucan, O. N., & Bayat, C. (2010). An Artificial Neural Network (ANN) based model for short-term predictions of daily mean PM10 concentrations . An Artificial Neural Network-based model. *Journal of Environmental Protection and Ecology*, 11(3)(2010), 1163–1171.
- Dempster, A. P., Laird, ; N M, & Rubin, ; D B. (1977). *Maximum Likelihood from Incomplete Data via the EM Algorithm*. *Journal of the Royal Statistical Society. Series B (Methodological)* (Vol. 39). Retrieved from <http://web.mit.edu/6.435/www/Dempster77.pdf>
- Díaz-Robles, L. A., Ortega, J. C., Fu, J. S., Reed, G. D., Chow, J. C., Watson, J. G., & Moncada-Herrera, J. A. (2008). A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile. *Atmospheric Environment*, 42(35), 8331–8340. <https://doi.org/10.1016/j.atmosenv.2008.07.020>
- Ding, W., Zhang, J., & Leung, Y. (2016). Prediction of air pollutant concentration based on sparse response back-propagation training feedforward neural networks. *Environmental Science and Pollution Research*, 19481–19494. <https://doi.org/10.1007/s11356-016-7149-4>
- Dominick, D., Talib Latif, M., Juahir, H., Aris, A. Z., & Zain, S. (2012). An assessment of influence of meteorological factors on PM10 and NO2 at selected stations in Malaysia. *Sustainable Environment Research*, 22(5), 305–315.
- Dotse, S. Q., Petra, M. I., Dagar, L., & De Silva, L. C. (2018). Application of computational intelligence techniques to forecast daily PM10exceedances in Brunei Darussalam. *Atmospheric Pollution Research*, 9(2), 358–368. <https://doi.org/10.1016/j.apr.2017.11.004>
- Drummond, C., & Holte, R. C. (2003). *C4.5, Class Imbalance, and Cost Sensitivity: Why Under-Sampling beats Over-Sampling*. Retrieved from <https://pdfs.semanticscholar.org/144b/bbaf2f0876c23295019b6e380c9fe4feda3.pdf>
- Dunea, D., Pohoata, A., & Iordache, S. (2015). Using wavelet–feedforward neural networks to improve air pollution forecasting in urban environments. *Environmental Monitoring and Assessment*, 187(7). <https://doi.org/10.1007/s10661-015-4697-x>
- Durao, R. M., Mendes, M. T., & Joao Pereira, M. (2016). Forecasting O3 levels in industrial

- area surroundings up to 24 h in advance, combining classification trees and MLP models. *Atmospheric Pollution Research*, 7(6), 961–970.
<https://doi.org/10.1016/j.apr.2016.05.008>
- Dursun, S., & Taylan, F. K. O. (2015). Modelling sulphur dioxide levels of Konya city using artificial intelligent related to ozone , nitrogen dioxide and meteorological factors. *International Journal of Environmental Science and Technology*, 12(12), 3915–3928.
<https://doi.org/10.1007/s13762-015-0821-2>
- 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(9), 1261–1269.
<https://doi.org/10.1016/J.ENVSOFT.2006.08.002>
- 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 and Software*, 22(9), 1261–1269.
<https://doi.org/10.1016/j.envsoft.2006.08.002>
- 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(4), 696–708.
<https://doi.org/10.5094/APR.2014.079>
- Ettouney, R. S., Wahab, S. A., & Elkilani, A. S. (2009). Emissions inventory , ISCST , and neural network modelling of air pollution in Kuwait Emissions inventory , ISCST , and neural network modelling of air pollution in Kuwait. *International Journal of Environmental Studies*, 7233. <https://doi.org/10.1080/00207230902859929>
- Feng, X., Li, Q., Zhu, Y., Hou, J., Jin, L., & Wang, J. (2015). Artificial neural networks forecasting of PM_{2.5} pollution using air mass trajectory based geographic model and wavelet transformation. *Atmospheric Environment*, 107, 118–128.
<https://doi.org/10.1016/j.atmosenv.2015.02.030>
- Feng, Y., Zhang, W., Sun, D., & Zhang, L. (2011). Ozone concentration forecast method based on genetic algorithm optimized back propagation neural networks and support vector machine data classification. *Atmospheric Environment*, 45(11), 1979–1985.
<https://doi.org/10.1016/j.atmosenv.2011.01.022>
- Fernando, H. J. S., Mammarella, M. C., Grandoni, G., Fedele, P., Di Marco, R., Dimitrova, R., & Hyde, P. (2012). Forecasting PM₁₀ in metropolitan areas: Efficacy of neural networks. *Environmental Pollution (Barking, Essex : 1987)*, 163, 62–67.
<https://doi.org/10.1016/j.envpol.2011.12.018>
- Folch-Fortuny, A., Arteaga, F., & Ferrer, A. (2015). PCA model building with missing data: New proposals and a comparative study. *Chemometrics and Intelligent Laboratory Systems*, 146, 77–88. <https://doi.org/10.1016/J.CHEMOLAB.2015.05.006>
- Fontes, T., Silva, L. M., Silva, M. P., Barros, N., & Carvalho, A. C. (2014). Can artificial neural networks be used to predict the origin of ozone episodes? *Science of The Total Environment*, 488–489, 197–207. <https://doi.org/10.1016/J.SCITOTENV.2014.04.077>
- Franceschi, F., Cobo, M., & Figueredo, M. (2018). Discovering relationships and forecasting PM₁₀ and PM_{2.5} concentrations in Bogotá, Colombia, using Artificial Neural Networks, Principal Component Analysis, and k-means clustering. *Atmospheric Pollution Research*, (February), 0–1. <https://doi.org/10.1016/j.apr.2018.02.006>

- Freeman, B. S., Taylor, G., Gharabaghi, B., & Thé, J. (2018). Forecasting air quality time series using deep learning. *Journal of the Air & Waste Management Association*, 68(8), 866–886. <https://doi.org/10.1080/10962247.2018.1459956>
- Galatioto, F., & Zito, P. (2009). Traffic Parameters Estimation to Predict Road Side Pollutant Concentrations using Neural Networks. *Environmental Modeling & Assessment*, 365–374. <https://doi.org/10.1007/s10666-007-9129-z>
- Galelli, S., Humphrey, G. B., Maier, H. R., Castelletti, A., Dandy, G. C., & Gibbs, M. S. (2014). An evaluation framework for input variable selection algorithms for environmental data-driven models. *Environmental Modelling & Software*, 62, 33–51. <https://doi.org/10.1016/J.ENVSOFT.2014.08.015>
- Gao, M., Yin, L., & Ning, J. (2018). Artificial neural network model for ozone concentration estimation and Monte Carlo analysis. <https://doi.org/10.1016/j.atmosenv.2018.03.027>
- Gardner, M. ., & Dorling, S. . (1998). Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environment*, 32(14–15), 2627–2636. [https://doi.org/10.1016/S1352-2310\(97\)00447-0](https://doi.org/10.1016/S1352-2310(97)00447-0)
- Gass, S. I. (1983). Feature Article-Decision-Aiding Models: Validation, Assessment, and Related Issues for Policy Analysis. <https://doi.org/10.1287/opre.31.4.603>
- Gennaro, G. De, Trizio, L., Di, A., Pey, J., Pérez, N., Cusack, M., ... Querol, X. (2013). Science of the Total Environment Neural network model for the prediction of PM 10 daily concentrations in two sites in the Western Mediterranean. *Science of the Total Environment*, The, 463–464, 875–883. <https://doi.org/10.1016/j.scitotenv.2013.06.093>
- Gong, B., & Ordieres-Meré, J. (2016). Prediction of daily maximum ozone threshold exceedances by preprocessing and ensemble artificial intelligence techniques: Case study of Hong Kong. *Environmental Modelling and Software*, 84, 290–303. <https://doi.org/10.1016/j.envsoft.2016.06.020>
- Grivas, G., & Chaloulakou, A. Ā. (2006). Artificial neural network models for prediction of PM 10 hourly concentrations , in the Greater Area of Athens , Greece. *Atmospheric Environment*, 40, 1216–1229. <https://doi.org/10.1016/j.atmosenv.2005.10.036>
- Guardani, R., Nascimento, C. A. O., Guardani, M. L. G., Martins, M. H. R. B., & Romano, J. (1999). Journal of the Air & Waste Management Association Study of Atmospheric Ozone Formation by Means of a Neural Network-Based Model Study of Atmospheric Ozone Formation by Means of a Neural Network-Based Model. *Journal of the Air & Waste Management Association*, 49, 316–323. <https://doi.org/10.1080/10473289.1999.10463806>
- Guoqiang Zhang, B., & Eddy Patuwo, M. Y. H. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14, 35–62. [https://doi.org/10.1016/S0169-2070\(97\)00044-7](https://doi.org/10.1016/S0169-2070(97)00044-7)
- Hagan, M. T., Demuth, H. B., & Beale, M. H. (1995). Neural Network Design. *Boston Massachusetts PWS*, 2, 734. <https://doi.org/10.1007/1-84628-303-5>
- Hasham, F. A., Kindzierski, W. B., & Stanley, S. J. (2004). Modeling of hourly NO x concentrations using artificial neural networks 1. *Journal of Environmental Engineering Sciences*, 3(x), 111–119. <https://doi.org/10.1139/S03-084>
- He, H., Lu, W.-Z., & Xue, Y. (2014). Prediction of particulate matter at street level using artificial neural networks coupling with chaotic particle swarm optimization algorithm.

- Building and Environment*, 78, 111–117.
<https://doi.org/10.1016/J.BUILDENV.2014.04.011>
- He, J., Yu, Y., Xie, Y., Mao, H., Wu, L., Liu, N., & Zhao, S. (2016). Numerical Model-Based Artificial Neural Network Model and Its Application for Quantifying Impact Factors of Urban Air Quality. *Water, Air, & Soil Pollution*, 227(7), 235.
<https://doi.org/10.1007/s11270-016-2930-z>
- Heo, J., & Kim, D. (2004). A new method of ozone forecasting using fuzzy expert and neural network systems, 325, 221–237. <https://doi.org/10.1016/j.scitotenv.2003.11.009>
- Hooyberghs, J., Mensink, C., Dumont, G., Fierens, F., & Brasseur, O. (2005). A neural network forecast for daily average PM10 concentrations in Belgium. *Atmospheric Environment*, 39(18), 3279–3289. <https://doi.org/10.1016/j.atmosenv.2005.01.050>
- Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), 359–366.
[https://doi.org/10.1016/0893-6080\(89\)90020-8](https://doi.org/10.1016/0893-6080(89)90020-8)
- Hoshyaripour, G., Brasseur, G., Andrade, M. F., Gavidia-Calder??n, M., Bouarar, I., & Ynoue, R. Y. (2016). Prediction of ground-level ozone concentration in S??o Paulo, Brazil: Deterministic versus statistic models. *Atmospheric Environment*, 145, 365–375.
<https://doi.org/10.1016/j.atmosenv.2016.09.061>
- Hrust, L., Klaić, Z. B., Križan, J., AntoniĆ, O., & Hercog, P. (2009). Neural network forecasting of air pollutants hourly concentrations using optimised temporal averages of meteorological variables and pollutant concentrations. *Atmospheric Environment*, 43(35), 5588–5596. <https://doi.org/10.1016/J.ATMOENV.2009.07.048>
- Huang, C. J., & Kuo, P. H. (2018). A deep cnn-lstm model for particulate matter (Pm2.5) forecasting in smart cities. *Sensors (Switzerland)*, 18(7).
<https://doi.org/10.3390/s18072220>
- Humphrey, G. B., Maier, H. R., Wu, W., Mount, N. J., Dandy, G. C., Abrahart, R. J., & Dawson, C. W. (2017). Improved validation framework and R-package for artificial neural network models. *Environmental Modelling & Software*, 92, 82–106.
<https://doi.org/10.1016/J.ENVSOFT.2017.01.023>
- Hunter, J. M., Maier, H. R., Gibbs, M. S., Foale, E. R., Grosvenor, N. A., Harders, N. P., & Kikuchi-Miller, T. C. (2018). Framework for developing hybrid process-driven, artificial neural network and regression models for salinity prediction in river systems. *Hydrology and Earth System Sciences*, 22(5), 2987–3006. <https://doi.org/10.5194/hess-22-2987-2018>
- Ibarra-Berastegi, G., Elias, A., Barona, A., Saenz, J., Ezcurra, A., & Diaz de Argandoña, J. (2008). From diagnosis to prognosis for forecasting air pollution using neural networks: Air pollution monitoring in Bilbao. *Environmental Modelling & Software*, 23(5), 622–637. <https://doi.org/10.1016/J.ENVSOFT.2007.09.003>
- IEEE Spectrum. (2018). The 2018 Top Programming Languages - IEEE Spectrum. Retrieved May 10, 2019, from <https://spectrum.ieee.org/at-work/innovation/the-2018-top-programming-languages>
- IEEE Xplore. (2017). IEEE Xplore Digital Library. Retrieved October 27, 2017, from <http://ieeexplore.ieee.org/Xplore/home.jsp>
- Inal, F. (2010). Artificial Neural Network Prediction of Tropospheric Ozone Concentrations in

- Istanbul, Turkey. *Clean - Soil, Air, Water*, 38(10), 897–908.
<https://doi.org/10.1002/clen.201000138>
- Jain, S., & Khare, M. (2010). Adaptive neuro-fuzzy modeling for prediction of ambient CO concentration at urban intersections and roadways, 203–212.
<https://doi.org/10.1007/s11869-010-0073-8>
- Jakeman, A. J., Letcher, R. A., & Norton, J. P. (2006). Ten iterative steps in development and evaluation of environmental models. *Environmental Modelling & Software*, 21(5), 602–614. <https://doi.org/10.1016/J.ENVSOF.2006.01.004>
- Jiang, D., Zhang, Y., Hu, X., Zeng, Y., Tan, J., & Shao, D. (2004). Progress in developing an ANN model for air pollution index forecast. *Atmospheric Environment*, 38, 7055–7064.
<https://doi.org/10.1016/j.atmosenv.2003.10.066>
- Jiang, P., Dong, Q., & Li, P. (2017). A novel hybrid strategy for PM_{2.5} concentration analysis and prediction. *Journal of Environmental Management*, 196, 443–457.
<https://doi.org/10.1016/j.jenvman.2017.03.046>
- Jiang, P., Li, C., Li, R., & Yang, H. (2018). An innovative hybrid air pollution early-warning system based on pollutants forecasting and Extenics evaluation. *Knowledge-Based Systems*. <https://doi.org/10.1016/J.KNOSYS.2018.10.036>
- Juhos, I., Makra, L., & Tóth, B. (2009). The behaviour of the multi-layer perceptron and the support vector regression learning methods in the prediction of NO and NO₂ concentrations in Szeged, Hungary. *Neural Computing and Applications*, 18(2), 193–205. <https://doi.org/10.1007/s00521-007-0171-1>
- Junninen, H., Niska, H., Tuppurainen, K., & Ruuskanen, J. (2004). Methods for imputation of missing values in air quality data sets, 38, 2895–2907.
<https://doi.org/10.1016/j.atmosenv.2004.02.026>
- Kao, J. J., & Huang, S. S. (2000). Forecasts using neural network versus box-jenkins methodology for ambient air quality monitoring data. *Journal of the Air and Waste Management Association*, 50(2), 219–226.
<https://doi.org/10.1080/10473289.2000.10463997>
- Kingston, G. B., Lambert, M. F., & Maier, H. R. (2005). Bayesian training of artificial neural networks used for water resources modeling. *Water Resources Research*, 41(12), 1–11.
<https://doi.org/10.1029/2005WR004152>
- Kingston, G. B., Maier, H. R., & Lambert, M. F. (2005). Calibration and validation of neural networks to ensure physically plausible hydrological modeling. *Journal of Hydrology*, 314(1–4), 158–176. <https://doi.org/10.1016/J.JHYDROL.2005.03.013>
- Kingston, G. B., Maier, H. R., & Lambert, M. F. (2008). Bayesian model selection applied to artificial neural networks used for water resources modeling. *Water Resources Research*, 44(4). <https://doi.org/10.1029/2007WR006155>
- Kolehmainen, M., Martikainen, H., & Ruuskanen, J. (2001). Neural networks and periodic components used in air quality forecasting. *Atmospheric Environment*, 35(5), 815–825.
[https://doi.org/10.1016/S1352-2310\(00\)00385-X](https://doi.org/10.1016/S1352-2310(00)00385-X)
- Kotzias, D., Geiss, O., Tirendi, S., Barrero-Moreno, J., Reina, V., Gotti, A., ... Sarigiannis, D. (2009). Public Buildings, Schools and Kindergartensthe European Indoor Air Monitoring and Exposure Assessment (Airmex) Study. *European Commission, Joint Research Centre, Institute for Health and Consumer Protection*, 18(5), 670–681.

- Kukkonen, J., Partanen, L., Karppinen, A., Ruuskanen, J., Junninen, H., Kolehmainen, M., ... Cawley, G. (2003). Extensive evaluation of neural network models for the prediction of NO₂ and PM₁₀ concentrations, compared with a deterministic modelling system and measurements in central Helsinki. *Atmospheric Environment*, 37(32), 4539–4550. [https://doi.org/10.1016/S1352-2310\(03\)00583-1](https://doi.org/10.1016/S1352-2310(03)00583-1)
- Kumar, N., Middey, A., & Rao, P. S. (2017). Prediction and examination of seasonal variation of ozone with meteorological parameter through artificial neural network at NEERI, Nagpur, India. *Urban Climate*, 20(2), 148–167. <https://doi.org/10.1016/j.uclim.2017.04.003>
- Kurt, A., & Oktay, A. B. (2010). Forecasting air pollutant indicator levels with geographic models 3 days in advance using neural networks. *Expert Systems with Applications*, 37(12), 7986–7992. <https://doi.org/10.1016/j.eswa.2010.05.093>
- Li, C., & Zhu, Z. (2018). Research and application of a novel hybrid air quality early-warning system: A case study in China. *Science of the Total Environment*, 626, 1421–1438. <https://doi.org/10.1016/j.scitotenv.2018.01.195>
- Li, T., Shen, H., Yuan, Q., Zhang, X., & Zhang, L. (2017). Estimating Ground-Level PM_{2.5} by Fusing Satellite and Station Observations: A Geo-Intelligent Deep Learning Approach. *Geophysical Research Letters*, 44(23), 11,985–11,993. <https://doi.org/10.1002/2017GL075710>
- Li, X., Peng, L., Hu, Y., Shao, J., & Chi, T. (2016). Deep learning architecture for air quality predictions. *Environmental Science and Pollution Research*, 23(22), 22408–22417. <https://doi.org/10.1007/s11356-016-7812-9>
- Lightstone, S. D., Moshary, F., & Gross, B. (2017). Comparing CMAQ forecasts with a neural network forecast model for PM_{2.5} in New York. *Atmosphere*, 8(9). <https://doi.org/10.3390/atmos8090161>
- Liu, H., Wu, H., Lv, X., Ren, Z., Liu, M., Li, Y., & Shi, H. (2019). An intelligent hybrid model for air pollutant concentrations forecasting: Case of Beijing in China Keywords: Air pollutant concentrations forecasting Empirical wavelet transform Multi-agent evolutionary genetic algorithm Nonlinear auto regressive models w. *Sustainable Cities and Society*. <https://doi.org/10.1016/j.scs.2019.101471>
- López, V., Fernández, A., Moreno-Torres, J. G., & Herrera, F. (2012). Analysis of preprocessing vs. cost-sensitive learning for imbalanced classification. Open problems on intrinsic data characteristics. *Expert Systems with Applications*, 39(7), 6585–6608. <https://doi.org/10.1016/J.ESWA.2011.12.043>
- Lu, W. Z., Wang, W. J., Wang, X. K., Xu, Z. B., & T Leung, A. Y. (2003). Using Improved Neural Network Model To Analyze Rsp, NO_x and NO₂ Levels in Urban Air in Mong Kok, Hong Kong. *Environmental Monitoring and Assessment*, 87(2), 235–254.
- Luna, A. S. (2014). Prediction of ozone concentration in tropospheric levels using artificial neural networks and support vector machine at Rio de. *Atmospheric Environment*, 98, 98–104. <https://doi.org/10.1016/j.atmosenv.2014.08.060>
- Mahapatra, A. (2010). Prediction of daily ground-level ozone concentration. *Environmental Monitoring and Assessment*, 170, 159–170. <https://doi.org/10.1007/s10661-009-1223-z>
- Maier, H. R., & Dandy, G. C. (2000). Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environmental*

Modelling & Software, 15(1), 101–124. [https://doi.org/10.1016/S1364-8152\(99\)00007-9](https://doi.org/10.1016/S1364-8152(99)00007-9)

- Maier, H. R., Jain, A., Dandy, G. C., & Sudheer, K. P. (2010). Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. *Environmental Modelling & Software*, 25(8), 891–909. <https://doi.org/10.1016/J.ENVSOF.2010.02.003>
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). The M4 Competition: Results, findings, conclusion and way forward. *International Journal of Forecasting*, 34(4), 802–808. <https://doi.org/10.1016/j.ijforecast.2018.06.001>
- Mao, X., Shen, T., & Feng, X. (2017). Prediction of hourly ground-level PM_{2.5} concentrations 3 days in advance using neural networks with satellite data in eastern China. *Atmospheric Pollution Research*, 8(6), 1005–1015. <https://doi.org/10.1016/j.apr.2017.04.002>
- Martín, M. L., Turias, I. J., González, F. J., Galindo, P. L., Trujillo, F. J., Puntonet, C. G., & Gorriz, J. M. (2008). Prediction of CO maximum ground level concentrations in the Bay of Algeciras, Spain using artificial neural networks. *Chemosphere*, 70(7), 1190–1195. <https://doi.org/10.1016/j.chemosphere.2007.08.039>
- Mishra, D., & Goyal, P. (2015). Development of artificial intelligence based NO₂ forecasting models at Taj Mahal, Agra. *Atmospheric Pollution Research*, 6(1), 99–106. <https://doi.org/10.5094/APR.2015.012>
- Mishra, D., & Goyal, P. (2016). Neuro-Fuzzy approach to forecasting Ozone Episodes over the urban area of Delhi, India. *Environmental Technology & Innovation*, 5, 83–94. <https://doi.org/10.1016/J.ETI.2016.01.001>
- Mount, N. J., Dawson, C. W., & Abrahart, R. J. (2013). Legitimising data-driven models: exemplification of a new data-driven mechanistic modelling framework. *Hydrol. Earth Syst. Sci*, 17, 2827–2843. <https://doi.org/10.5194/hess-17-2827-2013>
- Moustris, K P, Larissi, I. K., Nastos, P. T., Koukouletsos, K. V., & Paliatsos, A. G. (2013). Development and Application of Artificial Neural Network Modeling in Forecasting PM 10 Levels in a Mediterranean City. <https://doi.org/10.1007/s11270-013-1634-x>
- Moustris, Konstantinos P., Ziomas, I. C., & Paliatsos, A. G. (2010). 3-day-ahead forecasting of regional pollution index for the pollutants NO₂, CO, SO₂, and O₃ using artificial neural networks in athens, Greece. *Water, Air, and Soil Pollution*, 209(1–4), 29–43. <https://doi.org/10.1007/s11270-009-0179-5>
- Mueller, S. F., & Mallard, J. W. (2011). Contributions of natural emissions to ozone and PM_{2.5} as simulated by the community multiscale air quality (CMAQ) model. *Environmental Science and Technology*, 45(11), 4817–4823. <https://doi.org/10.1021/es103645m>
- Nagendra, S. M. S., & Khare, M. (2006). Artificial neural network approach for modelling nitrogen dioxide dispersion from vehicular exhaust emissions. *Ecological Modelling*, 190(1–2), 99–115. <https://doi.org/10.1016/j.ecolmodel.2005.01.062>
- Nidzgorska-Lencewicz, J. (2018). Application of artificial neural networks in the prediction of PM₁₀ levels in the winter months: A case study in the Tricity Agglomeration, Poland. *Atmosphere*, 9(6). <https://doi.org/10.3390/atmos9060203>
- Niska, H., Hiltunen, T., Karppinen, A., Ruuskanen, J., & Kolehmainen, M. (2004). Evolving the neural network model for forecasting air pollution time series. *Engineering Applications*

- of Artificial Intelligence*, 17(2), 159–167.
<https://doi.org/10.1016/J.ENGAPPAI.2004.02.002>
- Niska, H., Rantamäki, M., Hiltunen, T., Karppinen, A., Kukkonen, J., Ruuskanen, J., & Kolehmainen, M. (2005). Evaluation of an integrated modelling system containing a multi-layer perceptron model and the numerical weather prediction model HIRLAM for the forecasting of urban airborne pollutant concentrations. *Atmospheric Environment*, 39(35), 6524–6536. <https://doi.org/10.1016/j.atmosenv.2005.07.035>
- Nunnari, G. (2004). Modelling air pollution time-series by using wavelet functions and genetic algorithms. *Soft Computing*, 8, 173–178. <https://doi.org/10.1007/s00500-002-0260-0>
- Olcese, L. E., & Toselli, B. M. (2004). A method to estimate emission rates from industrial stacks based on neural networks. *Chemosphere*, 57, 691–696.
<https://doi.org/10.1016/j.chemosphere.2004.07.045>
- Olden, J. D., & Jackson, D. A. (2002). Illuminating the “black box”: a randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling*, 154(1–2), 135–150. [https://doi.org/10.1016/S0304-3800\(02\)00064-9](https://doi.org/10.1016/S0304-3800(02)00064-9)
- Ordieres, J. B., Vergara, E. P., Capuz, R. S., & Salazar, R. E. (2005). Neural network prediction model for fine particulate matter (PM_{2.5}) on the US–Mexico border in El Paso (Texas) and Ciudad Juárez (Chihuahua). *Environmental Modelling & Software*, 20(5), 547–559.
<https://doi.org/10.1016/J.ENVSOFT.2004.03.010>
- Organisation for Economic Co-operation and Development. (2016). *Policy Highlights - The economic consequences of outdoor air pollution*.
- Osowski, S., & Garanty, K. (2007). Forecasting of the daily meteorological pollution using wavelets and support vector machine. *Engineering Applications of Artificial Intelligence*, 20(6), 745–755. <https://doi.org/10.1016/j.engappai.2006.10.008>
- Özdemir, U., & Taner, S. (2014). Impacts of Meteorological Factors on PM₁₀ : Artificial Neural Networks (ANN) and Multiple Linear Regression (MLR) Approaches Contributed Articles Impacts of Meteorological Factors on PM₁₀ : Artificial Neural Networks (ANN) and Multiple Linear Regr. *Environmental Forensics*, 5922.
<https://doi.org/10.1080/15275922.2014.950774>
- Pak, U., Kim, C., Ryu, U., Sok, K., & Pak, S. (2018). A hybrid model based on convolutional neural networks and long short-term memory for ozone concentration prediction. *Air Quality, Atmosphere and Health*, 11(8), 883–895. <https://doi.org/10.1007/s11869-018-0585-1>
- Papaleonidas, A., & Iliadis, L. (2013). Neurocomputing techniques to dynamically forecast spatiotemporal air pollution data, 221–233. <https://doi.org/10.1007/s12530-013-9078-5>
- Paschalidou, A. K., Karakitsios, S., Kleanthous, S., & Kassomenos, P. A. (2011). Forecasting hourly PM₁₀ concentration in Cyprus through artificial neural networks and multiple regression models: Implications to local environmental management. *Environmental Science and Pollution Research*, 18(2), 316–327. <https://doi.org/10.1007/s11356-010-0375-2>
- Peng, H., Lima, A. R., Teakles, A., Jin, J., Cannon, A. J., & Hsieh, W. W. (2017). Evaluating hourly air quality forecasting in Canada with nonlinear updatable machine learning

- methods. *Air Quality, Atmosphere and Health*, 10(2), 195–211.
<https://doi.org/10.1007/s11869-016-0414-3>
- Perez, P., & Trier, A. (2001). Prediction of NO and NO₂ concentrations near a street with heavy traffic in Santiago. *Chile. Atmos. Environ.*, 35(10), 1783–1789.
- Perez, Patricio. (2012). Combined model for PM₁₀ forecasting in a large city. *Atmospheric Environment*, 60, 271–276. <https://doi.org/10.1016/j.atmosenv.2012.06.024>
- Perez, Patricio, & Salini, G. (2008). PM_{2.5} forecasting in a large city: Comparison of three methods. *Atmospheric Environment*, 42(35), 8219–8224.
<https://doi.org/10.1016/j.atmosenv.2008.07.035>
- Pires, J. C. M., Ferraz, M. C. M. A., Pereira, M. C., Martins, F. G., & Martins, F. G. (2010). Atmospheric Pollution Research Evolutionary procedure based model to predict ground – level ozone concentrations. *Atmospheric Pollution Research*, 1(4), 215–219.
<https://doi.org/10.5094/APR.2010.028>
- Pisoni, E., Farina, M., Carnevale, C., & Piroddi, L. (2009). Forecasting peak air pollution levels using NARX models. *Engineering Applications of Artificial Intelligence*, 22(4–5), 593–602. <https://doi.org/10.1016/j.engappai.2009.04.002>
- Plaia, A. A., & Bondi, A. L. (2006). Single imputation method of missing values in environmental pollution data sets. *Atmospheric Environment*, 40, 7316–7330.
<https://doi.org/10.1016/j.atmosenv.2006.06.040>
- Prakash, A., Kumar, U., Kumar, K., & Jain, V. K. (2011). A Wavelet-based Neural Network Model to Predict Ambient Air Pollutants' Concentration. *Environmental Modeling & Assessment*, 16(5), 503–517. <https://doi.org/10.1007/s10666-011-9270-6>
- ProQuest. (2017). ProQuest | Databases, EBooks and Technology for Research. Retrieved October 27, 2017, from <http://www.proquest.com/>
- Qi, Y., Li, Q., Karimian, H., & Liu, D. (2019). A hybrid model for spatiotemporal forecasting of PM_{2.5} based on graph convolutional neural network and long short-term memory. *Science of the Total Environment*, 664, 1–10.
<https://doi.org/10.1016/j.scitotenv.2019.01.333>
- Qin, D., Yu, J., Zou, G., Yong, R., Zhao, Q., & Zhang, B. (2019). A Novel Combined Prediction Scheme Based on CNN and LSTM for Urban PM_{2.5} Concentration. *IEEE Access*, 7, 20050–20059. <https://doi.org/10.1109/ACCESS.2019.2897028>
- Radojević, D., Antanasijević, D., Perić-Grujić, A., Ristić, M., & Pocajt, V. (2018). The significance of periodic parameters for ANN modeling of daily SO₂ and NO_x concentrations: A case study of Belgrade, Serbia. *Atmospheric Pollution Research*.
<https://doi.org/10.1016/J.APR.2018.11.004>
- Rahimi, A. (2017). Short-term prediction of NO₂ and NO_x concentrations using multilayer perceptron neural network: a case study of Tabriz, Iran. *Ecological Processes*, 6(1), 4.
<https://doi.org/10.1186/s13717-016-0069-x>
- Razavi, S., & Tolson, B. A. (2011). A New Formulation for Feedforward Neural Networks. *IEEE Transactions on Neural Networks*, 22(10), 1588–1598.
<https://doi.org/10.1109/TNN.2011.2163169>
- 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*, 6(3), 540–549.
<https://doi.org/10.5094/APR.2015.060>
- Russo, A., Raischel, F., & Lind, P. G. (2013). Air quality prediction using optimal neural networks with stochastic variables. *Atmospheric Environment*, 79, 822–830.
<https://doi.org/10.1016/j.atmosenv.2013.07.072>
- Russo, A., & Soares, A. O. (2014). Hybrid Model for Urban Air Pollution Forecasting: A Stochastic Spatio-Temporal Approach. *Mathematical Geosciences*, 46(1), 75–93.
<https://doi.org/10.1007/s11004-013-9483-0>
- Samarasinghe, S. (2006). *Neural Networks for Applied Sciences and Engineering*. Taylor & Francis. <https://doi.org/10.1017/CBO9781107415324.004>
- Santos, G., & Fernández-olmo, I. (2015). Estimation of PM 10 -Bound As , Cd , Ni and Pb Levels by Means of Statistical Modelling : PLSR and ANN Approaches. *Water, Air, & Soil Pollution*. <https://doi.org/10.1007/s11270-015-2526-z>
- Saporo, A., Tadé, M. O., & Vuthaluru, H. (2012). Chemical Product and Process Modeling A Modified Kennard-Stone Algorithm for Optimal Division of Data for Developing Artificial Neural Network Models A Modified Kennard-Stone Algorithm for Optimal Division of Data for Developing Artificial Neural Network . *Chemical Product and Process Modeling*, 7(1), 13. <https://doi.org/10.1515/1934-2659.1645>
- Schlink, U., Herbarth, O., Richter, M., Dorling, S., Nunnari, G., Cawley, G., & Pelikan, E. (2006). Statistical models to assess the health effects and to forecast ground-level ozone. *Environmental Modelling & Software*, 21(4), 547–558.
<https://doi.org/10.1016/J.ENVSOFT.2004.12.002>
- Schneider, T. (2001). Analysis of Incomplete Climate Data: Estimation of Mean Values and Covariance Matrices and Imputation of Missing Values. *Journal of Climate*. Retrieved from <https://journals.ametsoc.org/doi/pdf/10.1175/1520-0442%282001%29014%3C0853%3AAOICDE%3E2.0.CO%3B2>
- ScienceDirect. (2017). ScienceDirect.com | Science, health and medical journals, full text articles and books. Retrieved October 27, 2017, from <http://www.sciencedirect.com/>
- Shahraiyini, H. T., & Sodoudi, S. (2016). Statistical modeling approaches for pm10 prediction in urban areas; A review of 21st-century studies. *Atmosphere*, 7(2), 10–13.
<https://doi.org/10.3390/atmos7020015>
- Shahraiyini, H. T., Sodoudi, S., Kerschbaumer, A., & Cubasch, U. (2015). New technique for ranking of air pollution monitoring stations in the urban areas based upon spatial representativity (Case study: PM monitoring stations in Berlin). *Aerosol and Air Quality Research*, 15(2), 743–748. <https://doi.org/10.4209/aaqr.2014.12.0317>
- Sharma, N., Chaudhry, K. K., & Rao, C. V. C. (2005). Vehicular pollution modeling using artificial neural network technique : A review. *Journal of Scientific and Industrial Research*, 64(September), 637–647.
- Singh, K. P., Gupta, S., Kumar, A., & Shukla, S. P. (2012). Linear and nonlinear modeling approaches for urban air quality prediction. *Science of the Total Environment*, 426, 244–255. <https://doi.org/10.1016/j.scitotenv.2012.03.076>
- Siwek, K., & Osowski, S. (2012). Engineering Applications of Artificial Intelligence Improving the accuracy of prediction of PM 10 pollution by the wavelet transformation and an ensemble of neural predictors. *Engineering Applications of Artificial Intelligence*, 25(6),

- 1246–1258. <https://doi.org/10.1016/j.engappai.2011.10.013>
- Slini, T., Kaprara, A., Karatzas, K., & Moussiopoulos, N. (2006). PM10 forecasting for Thessaloniki, Greece. *Environmental Modelling & Software*, 21(4), 559–565. <https://doi.org/10.1016/J.ENVSOFT.2004.06.011>
- Solaiman, T. A., Coulibaly, P., & Kanaroglou, P. (2008). Ground-level ozone forecasting using data-driven methods. *Air Quality, Atmosphere and Health*, 1(4), 179–193. <https://doi.org/10.1007/s11869-008-0023-x>
- Sousa, S. I. V., Martins, F. G., Alvim-Ferraz, M. C. M., & Pereira, M. C. (2007). Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. *Environmental Modelling & Software*, 22(1), 97–103. <https://doi.org/10.1016/J.ENVSOFT.2005.12.002>
- Stamenković, L. J., Antanasijević, D. Z., Ristić, M., Perić-Grujić, A. A., & Pocajt, V. V. (2017). Prediction of nitrogen oxides emissions at the national level based on optimized artificial neural network model. *Air Quality, Atmosphere and Health*, 10(1), 15–23. <https://doi.org/10.1007/s11869-016-0403-6>
- Sun, W., Zhang, H., Palazoglu, A., Singh, A., Zhang, W., & Liu, S. (2013). Prediction of 24-hour-average PM2.5 concentrations using a hidden Markov model with different emission distributions in Northern California. *Science of The Total Environment*, 443, 93–103. <https://doi.org/10.1016/J.SCITOTENV.2012.10.070>
- Symonds, P., Taylor, J., Chalabi, Z., Mavrogianni, A., Davies, M., Hamilton, I., ... Macintyre, H. (2016). Development of an England-wide indoor overheating and air pollution model using artificial neural networks. *Journal of Building Performance Simulation*, 1493(June), 1–14. <https://doi.org/10.1080/19401493.2016.1166265>
- Taylan, O. (2017). Modelling and analysis of ozone concentration by artificial intelligent techniques for estimating air quality. *Atmospheric Environment*, 150, 356–365. <https://doi.org/10.1016/j.atmosenv.2016.11.030>
- The MathWorks, I. (2017). MATLAB Documentation - MathWorks United Kingdom. Retrieved October 30, 2017, from <https://uk.mathworks.com/help/matlab/>
- The University of Sheffield. (2017). StarPlus. Retrieved April 3, 2017, from https://find.shef.ac.uk/primo_library/libweb/action/search.do?vid=SFD_VU2&samlLogin=true&dscnt=0&dstmp=1556889134776&fromLogin=true
- Tsai, C., Chang, L., & Chiang, H. (2009). Forecasting of ozone episode days by cost-sensitive neural network methods. *Science of the Total Environment*, 407(6), 2124–2135. <https://doi.org/10.1016/j.scitotenv.2008.12.007>
- Tzanis, C. G., Alimissis, A., Philippopoulos, K., & Deligiorgi, D. (2019). Applying linear and nonlinear models for the estimation of particulate matter variability. *Environmental Pollution*, 246, 89–98. <https://doi.org/10.1016/J.ENVPOL.2018.11.080>
- Ul-Saufie, A. Z., Yahaya, A. S., Ramli, N. A., Rosaida, N., & Hamid, H. A. (2013). Future daily PM10 concentrations prediction by combining regression models and feedforward backpropagation models with principle component analysis (PCA). *Atmospheric Environment*, 77, 621–630. <https://doi.org/10.1016/J.ATMOSENV.2013.05.017>
- Ventura, L. M. B., Fellipe De Oliveira Pinto, &, Laiza, &, Soares, M., Luna, A. S., & Gioda, A. (2019). Forecast of daily PM 2.5 concentrations applying artificial neural networks and Holt-Winters models. *Air Quality, Atmosphere & Health*.

<https://doi.org/10.1007/s11869-018-00660-x>

- Vlachogianni, A., Kassomenos, P., Karppinen, A., Karakitsios, S., & Kukkonen, J. (2011). Science of the Total Environment Evaluation of a multiple regression model for the forecasting of the concentrations of NO_x and PM₁₀ in Athens and Helsinki. *Science of the Total Environment*, *409*(8), 1559–1571. <https://doi.org/10.1016/j.scitotenv.2010.12.040>
- Voukantsis, D., Karatzas, K., Kukkonen, J., & Räsänen, T. (2011). Science of the Total Environment Intercomparison of air quality data using principal component analysis , and forecasting of PM₁₀ and PM_{2.5} concentrations using artificial neural networks , in Thessaloniki and Helsinki. *Science of the Total Environment*, *409*(7), 1266–1276. <https://doi.org/10.1016/j.scitotenv.2010.12.039>
- Wang, J., & Song, G. (2018). A Deep Spatial-Temporal Ensemble Model for Air Quality Prediction. *Neurocomputing*, *314*, 198–206. <https://doi.org/10.1016/J.NEUCOM.2018.06.049>
- Wang, W., Xu, Z., & Weizhen Lu, J. (2003). Three improved neural network models for air quality forecasting. *Engineering Computations*, *20*(2), 192–210. <https://doi.org/10.1108/02644400310465317>
- Wen, C., Liu, S., Yao, X., Peng, L., Li, X., Hu, Y., & Chi, T. (2019). A novel spatiotemporal convolutional long short-term neural network for air pollution prediction. *Science of the Total Environment*, *654*, 1091–1099. <https://doi.org/10.1016/j.scitotenv.2018.11.086>
- Williams, D. A., Nelsen, B., Berrett, C., Williams, G. P., & Moon, T. K. (2018). A comparison of data imputation methods using Bayesian compressive sensing and Empirical Mode Decomposition for environmental temperature data. <https://doi.org/10.1016/j.envsoft.2018.01.012>
- World Health Organization. (2013). Review of evidence on health aspects of air pollution – REVIHAAP Project. *World Health Organization*, 309. Retrieved from <http://www.euro.who.int/en/health-topics/environment-and-health/air-quality/publications/2013/review-of-evidence-on-health-aspects-of-air-pollution-revihaap-project-final-technical-report>
- World Health Organization. (2016). WHO | Ambient (outdoor) air quality and health. Retrieved September 14, 2017, from <http://www.who.int/mediacentre/factsheets/fs313/en/>
- World Health Organization. (2018). WHO | Ambient air pollution. Retrieved September 10, 2018, from <http://www.who.int/airpollution/ambient/en/>
- Wu, W., Dandy, G. C., & Maier, H. R. (2014). Protocol for developing ANN models and its application to the assessment of the quality of the ANN model development process in drinking water quality modelling. *Environmental Modelling & Software*, *54*, 108–127. <https://doi.org/10.1016/J.ENVSOFT.2013.12.016>
- Wu, W., May, R. J., Maier, H. R., & Dandy, G. C. (2013). A benchmarking approach for comparing data splitting methods for modeling water resources parameters using artificial neural networks. *Water Resources Research*, *49*(11), 7598–7614. <https://doi.org/10.1002/2012WR012713>
- Yan, K., & Jian, L. (2013). Neurocomputing Identification of significant factors for air pollution levels using a neural network based knowledge discovery system. *Neurocomputing*, *99*,

564–569. <https://doi.org/10.1016/j.neucom.2012.06.003>

- Yeganeh, B., Hewson, M. G., Clifford, S., Knibbs, L. D., & Morawska, L. (2017). A satellite-based model for estimating PM_{2.5} concentration in a sparsely populated environment using soft computing techniques. *Environmental Modelling & Software*, 88, 84–92. <https://doi.org/10.1016/J.ENVSOFT.2016.11.017>
- Yeganeh, B., Hewson, M. G., Clifford, S., Tavassoli, A., Knibbs, L. D., & Morawska, L. (2018). Estimating the spatiotemporal variation of NO₂ concentration using an adaptive neuro-fuzzy inference system. *Environmental Modelling and Software*, 100(2), 222–235. <https://doi.org/10.1016/j.envsoft.2017.11.031>
- Yetilmezsoy, K., Ozkaya, B., & Cakmakci, M. (2011). Artificial Intelligence-based Prediction Models for Environmental Engineering. *Neural Network World*, 21(3), 193–218.
- Zhang, H., Liu, Y., Shi, R., & Yao, Q. (2013). Evaluation of PM₁₀ forecasting based on the artificial neural network model and intake fraction in an urban area : A case study in Taiyuan City , Evaluation of PM₁₀ forecasting based on the artificial neural network model and intake fraction in an urban. *Journal of the Air & Waste Management Association*, 2247. <https://doi.org/10.1080/10962247.2012.755940>
- Zhang, J., & Ding, W. (2017). Prediction of air pollutants concentration based on an extreme learning machine: The case of Hong Kong. *International Journal of Environmental Research and Public Health*, 14(2), 1–19. <https://doi.org/10.3390/ijerph14020114>
- Zhao, Y., Shrivastava, A. K., & Tsui, K. L. (2016). Imbalanced classification by learning hidden data structure. *IIE Transactions*, 48(7), 614–628. <https://doi.org/10.1080/0740817X.2015.1110269>
- Zhou, Q., Jiang, H., Wang, J., & Zhou, J. (2014). A hybrid model for PM_{2.5} forecasting based on ensemble empirical mode decomposition and a general regression neural network. *Science of The Total Environment*, 496, 264–274. <https://doi.org/10.1016/J.SCITOTENV.2014.07.051>
- Zhu, G., Zhang, P., Tshukudu, T., Yin, J., Fan, G., & Zheng, X. (2015). Forecasting traf fi c-related nitrogen oxides within a street canyon by combining a genetic algorithm-back propagation arti fi cial neural network and parametric models. *Atmospheric Pollution Research*, 6(6), 1087–1097. <https://doi.org/10.1016/j.apr.2015.06.006>
- Zhu, S., Lian, X., Wei, L., Che, J., Shen, X., Yang, L., ... Li, J. (2018). PM_{2.5} forecasting using SVR with PSO-GSA algorithm based on CEEMD, GRNN and GCA considering meteorological factors. *Atmospheric Environment*, 183, 20–32. <https://doi.org/10.1016/J.ATMOSENV.2018.04.004>
- Zito, P., Chen, H., & Bell, M. C. (2008). Predicting real-time roadside CO and NO₂ concentrations using neural networks. *IEEE Transactions on Intelligent Transportation Systems*, 9(3), 514–522. <https://doi.org/10.1109/TITS.2008.928259>
- Zou, B., Wang, M., Wan, N., Wilson, J. G., Fang, X., & Tang, Y. (2015). Spatial modeling of PM_{2.5} concentrations with a multifactorial radial basis function neural network, 10395–10404. <https://doi.org/10.1007/s11356-015-4380-3>