Methods used for handling and quantifying model uncertainty of artificial neural network models for air pollution forecasting

Sheen Mclean Cabaneros[a,*], Ben Hughes[a]

[a]*Department of Mechanical Engineering, Faculty of Science and Engineering, University of Hull, Hull, HU6 7RX*

Abstract

The use of data-driven techniques such as artificial neural network (ANN) models for outdoor air pollution forecasting has been popular in the past two decades. However, research activity on uncertainty surrounding the development of ANN models has been limited. Therefore, this review outlines the approaches for addressing model uncertainty according to the steps for building ANN models. Based on 128 articles published from 2000 to 2022, the review reveals that input uncertainty was predominantly addressed while less focus was given to the structure, parameter and output uncertainties. Ensemble approaches have been mostly employed, followed by neuro-fuzzy networks. However, the direct measurement of uncertainty received less attention. The use of bootstrapping, Bayesian, and Monte Carlo simulation techniques which can quantify uncertainty was also limited. In conclusion, this review recommends the development and application of approaches that can both handle and quantify uncertainty surrounding the development of ANN models.
Keywords: Air pollution forecasting, Artificial neural networks, Uncertainty quantification, Bayesian, Monte Carlo Simulation, Fuzzy

1. Introduction

1.1. Data-driven models in air pollution forecasting

The use of data-driven models in outdoor air pollution (AP) forecasting has been widely reported in the literature in the last two decades (Shahraiyni and Sodoudi, 2016; Cabaneros et al., 2019; Masood and Ahmad, 2021). The attractiveness of data-driven models can be explained in two respects: (1) better performance over traditional approaches, and (2) the emergence of big data and more powerful computing software. Data-driven models have been shown to understand the complex dynamics between environmental variables and outdoor AP without using physics-based formulae. This allows researchers to bypass the strong theoretical requirements to employ traditional approaches such as Gaussian dispersion, 3-D gridded Eulerian, Photochemical and Lagrangian trajectory models. As such, data-driven models for AP forecasting have been

*Corresponding author

shown to be effective alternatives to the traditional deterministic ones (Cabaneros et al., 2019; Gardner and Dorling, 1998; Lv et al., 2016; Zhao et al., 2019). Big data and more powerful computing resources has also paved the way for research communities to apply data-driven models in AP forecasting (Bishop, 2006; Montáns et al., 2019). In particular, the smart city concept that integrates Information and Communication Technology (ICT) and fixed/mobile sensors have generated a tremendous amount of data, especially outdoor pollution measurements, which has enabled decision-makers in providing early warnings to citizens in urban cities (Bekkar et al., 2021). Finally, the past few years have witnessed the development and emergence of accessible and powerful computing tools that support machine learning processes (Giray, 2021). Free and open-source programming languages such as Python, R, and Java have been popularly used for coding many data-driven techniques (TIOBE, 2022).

Popular data-driven approaches utilised for AP forecasting include statistical models, e.g Multiple Linear Regression (MLR) (Ng and Awang, 2018) and Autoregressive Integrated Moving Average (ARIMA), machine learning (ML), e.g. Artificial Neural Network (ANN) (Gardner and Dorling, 1997), Support Vector Machine (SVM) (Lu and Wang, 2005), Fuzzy expert system (Heo and Kim, 2004), and more recently, deep learning models, e.g. Deep Neural Network (DNN) (Freeman et al., 2018). The approaches clearly have their pros and cons which have been thoroughly discussed in the literature (Masood and Ahmad, 2021; Chen et al., 2008).

ANN models are among the most employed data-driven tools for AP forecasting. ANNs are parallel-computing structures that mimics the information-processing paradigm of the human brain. This makes them capable of learning from any previously unknown information from any given training dataset (McCulloch and Pitts, 1943; Basheer and Hajmeer, 2000). ANN model development generally consists of eight steps: (1) data collection, (2) data preprocessing, (3) input variable selection, (4) data division, (5) model architecture selection, (6) model structure optimisation, (7) model training, and (8) model validation. A detailed description of each step can be found elsewhere (Cabaneros et al., 2019; Maier et al., 2010). However, performing each step entails the selection of one or more methods/processes which are problem-specific. This lack of a clean-cut approach makes the process of building ANN models not straightforward. Consequently, the influence of the various approaches in performing each of the steps on the model results has become a well-established research area. In their review of articles dealing with the use of ANNs in outdoor AP forecasting, Cabaneros et al. (2019) revealed that novel methods for implementing one or a combination of the steps for building ANN models have been proposed to outperform the traditional ones. The authors reported that the articles mainly focused on the development of more novel and sophisticated predictor selection techniques, model architectures, and model training algorithms. A similar observation was reported in earlier related works by Maier et al. (2010) and Wu et al. (2014) which reviewed case studies using ANN models but applied in forecasting of hydrological variables. Finally, a recent review by Masood and Ahmad (2021) on the use of techniques based on Artificial Intelligence (AI) in AP forecasting has also shown similar findings. The authors revealed that models with deeper architectures, e.g. DNN models, have been employed and reported to provide superior AP predictions.

1.2. Uncertainty incurred from developing data-driven models

Given their black-box nature, ANN models cannot provide explicit insights regarding the influence of several model-building choices, e.g. inputs, architecture, structure, and training parameters, on their results. This ambiguity surrounding the modelling process exists and has been commonly referred to as model uncertainty. Model uncertainty has been reported to limit the potential of using ANN models especially tasks involving decision-making (Vardoulakis et al., 2002; Borrego et al., 2008). In particular, the uncertainty surrounding ANN models that were not carefully designed can limit the reproducibility and reliability of model results (Arhami et al., 2013; Elshorbagy et al., 2010; Noori et al., 2010).

Current efforts have mostly focused on the improvement of the point estimates of pollution levels, while the incorporation of model uncertainty has received less attention (Cabaneros et al., 2019; Maier et al., 2010). Kasiviswanathan and Sudheer (2017) has reviewed research articles dealing with hydrological modelling that employed methods addressing model uncertainty. They revealed that the methodological issues for building ANN models, not model uncertainty, have been mostly examined. Among those that tackled uncertainty, only a few investigated the mutual interaction among sources of prediction uncertainty, e.g. inputs, training parameters, and model structure.

Given the increasing popularity of ANN models, a thorough review of existing research that accounts for uncertainty is significant. Hence, the main objective of this paper is to provide an extensive framework of methods used for addressing the uncertainty surrounding the development of ANN models for outdoor AP forecasting. Through the results of this paper, the authors aim to promote good practice in reporting ANN model results by accounting for both accuracy and uncertainty. To the best of the authors' knowledge, a comprehensive survey of articles that deal with the uncertainty of ANN models for AP forecasting has not yet been undertaken. Another novelty of this review paper is that it aims to describe the interplay between various sources of model uncertainty. This is carried out by relating each uncertainty source to the eight steps of building ANN models.

The remainder of this review paper is organised as follows. In Section 2, details regarding the methods for selecting the appropriate articles for this review are presented. In Section 3, the sources of model uncertainty and the approaches for addressing them are described. Section 4 presents the methods utilised for quantifying the model uncertainty. Section 5 provides the conclusions of this review, while Section 6 presents a number of recommendations for future research.

2. Overview

Articles that deal with the application of ANN models in outdoor AP forecasting were collected from international peer-reviewed journals. Databases such as ScienceDirect, IEEE Access, PLOS One, ACM, Springer, Taylor & Francis Online Google Scholar, and MDPI were searched for relevant literature published from January 2000 to August 2022. The selection process of this review consisted of three stages. Since the main focus of this paper is on the incorporation of uncertainty that arises from ANN model development, the search items were firstly narrowed down to include the terms "uncertainty" and/or methods that are well-known to account for uncertainty. The search items for the methods included "air pollution forecasting", "artificial neural networks", "deep learning", "machine learning", "ANN", "deep neural networks", "Bayesian neural networks", "Monte Carlo", "ensemble", "confidence interval", "sensitivity analysis", "ANFIS", and "Fuzzy neural networks" with different combinations. A second search query was carried out containing "neural networks" and "air pollution" because many of the appropriate articles for this review do not necessarily have to mention the term "uncertainty". The third selection stage from both identified and unidentified papers was performed in an ad-hoc manner based on the subject matter of this review. For instance, many articles do mention the term "uncertainty" yet only a subset of them addressed it in their model-building process. The observation can be made for other articles mentioning one or a set of search terms mentioned above. Another important criterion for inclusion in the review is that the said methods need to be applied in conjunction with ANNs. For instance, several articles on plain models solely applying fuzzy inference systems and linear forecasters were removed. The results from a recent review article by Cabaneros et al. (2019) were also utilised to locate the relevant articles for this review. Lastly, articles presented at conference proceedings were manually removed from the initial list of search results. The search methodology above has identified 128 relevant articles for this review.

Table 1 provides a list of journals alongside their respective latest impact factors where the selected articles were published. The leading position in terms of the number of publications was held by both the Atmospheric Environment and Atmospheric Pollution Research journals which accounted for approximately 12% of the total number of articles. Both the Environmental Modelling & Software and IEEE Access journals had the next highest proportions of articles (8% each) followed by the Science of the Total Environment (7%). The rest of the identified journals had fewer proportions of articles, e.g. 5 articles or less.

Table 1: List of journals selected for this review.

No.	Name of the journal	Latest impact factor	No. of papers published
1	Air Quality, Atmosphere & Health	3.763	5
2	Atmosphere	3.110	4
3	Atmospheric Environment	4.798	15
4	Atmospheric Pollution Research	4.352	15
5	Building and Environment	7.093	2
6	Chaos, Solitons & Fractals	9.992	1
7	Chemosphere	7.086	4
8	Clean - Soil, Air, Water	1.770*	1
9	Computers & Geosciences	5.168	1
10	Ecological Informatics	3.142	1
11	Ecological Modelling	2.974	2
12	Ecotoxicology and Environmental Safety	6.291	1
13	Engineering Applications of Artificial Intelligence	6.212	3
14	Entropy	2.738	1
15	Environment International	9.621	1
16	Environmental Modeling & Assessment	2.016	1
17	Environmental Modelling & Software	5.288	10
18	Environmental Modelling and Assessment	2.333*	2
19	Environmental Monitoring and Assessment	3.307	1
20	Environmental Pollution	8.071	2
21	Environmental Science & Policy	5.190	4
22	Environmental Science and Pollution Research	4.223	3
23	Environmental Technology & Innovation	5.263	1
24	Expert Systems with Applications	8.665	1
25	IEEE Access	3.476	10
26	IEEE Sensors Journal	4.325	1
27	IEEE Transactions on Big Data	4.271	1
28	IEEE Transactions on Instrumentation and Measurement	5.332	1
29	IEEE Transactions on Intelligent Transportation Systems	6.492	1

Table 1 continued from previous page

30	International Journal of Applied Mathematics and Computer Science	2.157	1
31	International Journal of Environmental Research and Public Health	4.614	1
32	International Journal of Environmental Science and Technology	3.519	1
33	Journal of Cleaner Production	9.297	1
34	Journal of Environmental Engineering and Science	-	1
35	Journal of Environmental Management	6.789	1
36	Journal of the Air & Waste Management Association	2.636	2
37	Mathematical Geosciences	2.508	1
38	Mathematics	2.592	1
39	Neural Computing and Applications	5.102	1
40	Neural Networks	8.050	1
41	Neurocomputing	5.719	1
42	Pattern Recognition Letters	4.575	1
43	PLOS One	N/A	1
44	Remote Sensing	5.349	1
45	Science of the Total Environment	7.693	9
46	Sensors	3.847	1
47	Sustainability	3.889	2
48	Sustainable Cities and Society	7.587	3
49	Urban Climate	6.663	3

Details of the selected articles such as the name of the authors, year of publication, study location, air pollutants examined, and methods used to address model uncertainty are given in Table 2. It should be noted that several articles employed two or more approaches in addressing uncertainty in their ANN model development process.

Table 2: Key details of the articles reviewed.

No.	Authors (year)	Case study location	Examined pollutants	Methods used to handle uncertainty
1	Chelani et al. (2002)	Delhi, India	O_3; SO_2	Convergence criteria

Table 2 continued from previous page

2	Morabito and Versaci (2003)	Villa San Giovanni, Italy	SO_2; CO; O_3; NO; NO_2; HC; TSP; PM_{10}	Fuzzy
3	Hasham et al. (2004)	Edmonton, Canada	NO_x	Factorial design concepts
4	Niska et al. (2004)	Helsinki, Finland	NO_2	MOGA
5	Niska et al. (2005)	Helsinki, Finland	NO_2; $PM_{2.5}$	Sensitivity analysis & MOGA
6	Ordieres et al. (2005)	Ciudad Juarez & El Paso, Mexico	$PM_{2.5}$	Stepwise regression
7	Agirre-Basurko et al. (2006)	Bilbao, Spain	O_3; TSP	Stepwise regression with tolerance filtering & generalisation rule
8	Grivas and Chaloulakou (2006)	Athens, Greece	PM_{10}	Genetic algorithm & bootstrapping
9	Karakitsios et al. (2006)	Ioannina, Greece	Benzene	Bayesian
10	Slini et al. (2006)	Thessaloniki, Greece	PM_{10}	CART
11	Yildirim and Bayramoglu (2006)	Zonguldak, Turkey	SO_2; TSP	Fuzzy
12	Dutot et al. (2007)	Orleans, France	O_3	BIC-like criterion, stepwise regression & confidence interval
13	Díaz-Robles et al. (2008)	Temuco, Chile	PM_{10}	Regression
14	Ibarra-Berastegi et al. (2008)	Bilbao, Spain	SO_2; CO; O_3; NO; NO_2; HC; TSP; PM_{10}	Genetic algorithm, sensitivity analysis & bootstrapping
15	Perez and Salini (2008)	Santiago, Chile	$PM_{2.5}$	Correlation analysis
16	Solaiman et al. (2008)	Ontario, Canada	O_3	Sensitivity analysis & Bayesian
17	Zito et al. (2008)	Leicestershire, UK	CO; NO_2	Sensitivity analysis
18	Hrust et al. (2009)	Zagreb, Croatia	NO_2; O_3; CO; PM_{10}	Fourier analysis
19	Inal (2010)	Istanbul, Turkey	O_3	Sensitivity analysis
20	Jain and Khare (2010)	Delhi City, India	CO	Fuzzy
21	Noori et al. (2010)	Tehran, Iran	CO	Monte Carlo,

Table 2 continued from previous page

				bootstrapping, & fuzzy
22	Feng et al. (2011)	Beijing, China	O_3	Genetic algorithm
23	Prakash et al. (2011)	New Delhi, India	PM_{10}	Ensemble ANN
24	Voukantsis et al. (2011)	Thessaloniki, Greece & Helsinki, Finland	PM_{10}; $PM_{2.5}$	Bootstrapping
25	Hájek and Olej (2012)	Pardubice micro-region, Czech Republic	O_3	Genetic algorithm Fuzzy &
26	Shekarrizfard and Hadad (2012)	Shiraz, Iran	PM_{10}	Ensemble ANN
27	Singh et al. (2012)	Lucknow, India	RSPM; SO_2; NO_2	Stepwise regression & sensitivity analysis
28	Siwek and Osowski (2012)	Warsaw, Poland	PM_{10}	Ensemble ANN
29	Antanasijević et al. (2013)	26 EU countries	PM_{10}	Individual smoothing factor
30	Kadiyala et al. (2013)	Toledo, USA		Genetic algorithm & sensitivity analysis
31	Rahman and Khondaker (2013)	Empty Quarter, Saudi Arabia	O_3	Fuzzy
32	Russo et al. (2013)	Lisbon, Portugal	NO_2	Stepwise regression
33	de Mattos Neto et al. (2014)	Helsinki, Finland		Genetic algorithm
34	Elangasinghe et al. (2014)	Auckland, New Zealand	NO_2	Genetic algorithm & sensitivity analysis
35	He et al. (2014)	Mong Kok, Hong Kong	PM_{10}; PM_1	Correlation analysis & a method by Fletcher and Goss (1993)
36	Russo and Soares (2014)	Lisbon, Portugal	PM_{10}	Stepwise regression
37	Zhou et al. (2014)	Xi'an Province, China	PM_{10}; PM_1	Ensemble ANN
38	Cortina-Januchs et al. (2015)	Salamanca, Mexico	PM_{10}	Fuzzy
39	de Mattos Neto et al. (2015)	Kallio & Vallila, Finland	PM_{10}; $PM_{2.5}$	Ensemble ANN & genetic algorithm
40	Dunea et al. (2015)	Oltenia, Romania	O_3; PM_{10};	Ensemble ANN
41	Dursun and Taylan (2015)	Konya City, Turkey	$PM_{2.5}$ SO_2	Fuzzy
42	Feng et al. (2015)	Jing-Jin-Ji area, China	$PM_{2.5}$	Ensemble ANN
43	Mishra et al. (2015)	Agra, India	NO_2	PCA
44	Russo et al. (2015)	Lisbon, Portugal	PM_{10}	Forward selection
45	Ausati and Amanollahi (2016)	Sanandaj, Iran	$PM_{2.5}$	Fuzzy

Table 2 continued from previous page

46	Bai et al. (2016)	Chongqing, China	PM_{10}; SO_2; NO_2	Correlation analysis, ensemble ANN & an empirical formula by Shen et al. (2008)
47	Catalano et al. (2016)	London, UK	NO_2	Ensemble ANN & Weight regularisation formula by Bishop (1995)
48	Ding et al. (2016)	Hong Kong	NO_2; NO_x; O_3; SO_2; $PM_{2.5}$	Monte Carlo
49	Durao et al. (2016)	Sines, Portugal	O_3	CART
50	Gong and Ordieres-Meré (2016)	Hong Kong	O_3	Ensemble ANN
51	Hoshyaripour et al. (2016)	Sao Paulo, Brazil	O_3	Forward selection
52	Mishra and Goyal (2016)	Delhi, India	O_3	Correlation analysis & fuzzy
53	Shaban et al. (2016)	Doha, Qatar	O_3; NO_2; SO_2	Sensitivity analysis
54	Siwek and Osowski (2016)	2 sites in Warsaw, Poland	PM_{10}; SO_2; NO_2; O_3	Forward/backward selection & ensemble ANN
55	Suleiman et al. (2016)	London, UK	O_3; NO_2; SO_2	Elastic net LASSO & PCA
56	de Mattos Neto et al. (2017)	Helsinki, Finland	$PM_{2.5}$; PM_{10}	Ensemble ANN & genetic algorithm
57	Gorai and Mitra (2017)	Kolkatta, India	O_3	Forward selection algorithm
58	Jiang et al. (2017)	Jing-Jin-Ji & Pearl River Delta, China	$PM_{2.5}$	Fuzzy
59	Peng et al. (2017)	Canada	O_3; PM_{10}; NO_2	Bootstrapping
60	Stamenković et al. (2017)	17 EU countries, USA, China, Japan, Russia & India	NO_x	Correlation analysis & sensitivity analysis
61	Taylan (2017)	Jeddah, Saudi Arabia	O_3	Fuzzy
62	Yeganeh et al. (2017)	Queensland, Australia	$PM_{2.5}$	Fuzzy & LASSO
63	Alimissis et al. (2018)	Athens, Greece	NO_2; NO; O_3; CO;	Correlation analysis

Table 2 continued from previous page

			SO_2	
64	Dotse et al. (2018)	Brunei Darussalam	PM_{10}	Genetic algorithm with random forest
65	Franceschi et al. (2018)	Bogota, Colombia	$PM_{2.5}$; PM_{10}	PCA
66	Freeman et al. (2018)	Kuwait	O_3	Decision Trees
67	Gao et al. (2018)	Jinan, China	O_3	Forward selection
68	Li and Zhu (2018)	China	$PM_{2.5}$; PM_{10}; CO	Fuzzy
69	Mahajan et al. (2018)	Taiwan	$PM_{2.5}$	Ensemble ANN
70	Radojević et al. (2018)	Belgrade, Serbia	SO_2; NO_x	Empirical rule by Kalogirou (2003)
71	Soh et al. (2018)	76 sites in Taiwan	$PM_{2.5}$	NN methods & ensemble ANN
72	Wang et al. (2018)	Beijing, China	CO; NO_2; SO_2; O_3; PM_{10}; $PM_{2.5}$	Ensemble ANN
73	Yeganeh et al. (2018)	Queensland, Australia	NO_2	Fuzzy & ADDRESS
74	Zhai and Chen (2018)	Beijing, China	$PM_{2.5}$	Genetic algorithm & ensemble ANN
75	Zhu et al. (2018)	China	$PM_{2.5}$	Grey correlation analysis
76	Abdullah et al. (2019)	East Coast of Peninsular Malaysia	PM_{10}	Empirical formula
77	Bai et al. (2019)	Beijing, China	$PM_{2.5}$	Ensemble ANN
78	Balram et al. (2019)	Zuoying district, Taiwan	$PM_{2.5}$	Forward selection & Bayesian
79	Caraka et al. (2019)	Pingtung & Chaozhou, Taiwan	$PM_{2.5}$	PSO algorithm optimisation
80	Di et al. (2019)	Hong Kong	$PM_{2.5}$	Ensemble ANN
81	Gu et al. (2019)	12 sites in Beijing, China	$PM_{2.5}$	Ensemble ANN & bootstrapping
82	Liu et al. (2019a)	Beijing, China	$PM_{2.5}$	Genetic algorithm & ensemble ANN
83	Liu et al. (2019b)	Beijing, China	$PM_{2.5}$; SO_2; NO_2; CO	Ensemble ANN
84	Maciąg et al. (2019)	London, UK	PM_{10}	Ensemble ANN
85	Mo et al. (2019)	Beijing, Tianjin &	$PM_{2.5}$; PM_{10}; NO_2; SO_2;	Ensemble ANN

Table 2 continued from previous page

		Shijiazhuang, China	CO; O_3	
86	Qiao et al. (2019)	6 sites in China	$PM_{2.5}$	Ensemble ANN
87	Tao et al. (2019)	Beijing, China	$PM_{2.5}$	Correlation analysis
88	Valput et al. (2019)	Madrid, Spain	NO_2	Ensemble ANN
89	Van Roode et al. (2019)	Bay of Algeciras, Spain	NO_2	Ensemble ANN
90	Wu and Lin (2019)	Beijing & Guilin, China	AQI ($PM_{2.5}$; PM_{10}; CO; O_3; SO_2; NO_2)	Ensemble ANN
91	Zhao et al. (2019)	Beijing, China	$PM_{2.5}$	Ensemble ANN
92	Cabaneros et al. (2020)	London, UK	NO_2	Correlation analysis & ensemble ANN
93	Chang et al. (2020)	Several sites in Taiwan	$PM_{2.5}$	Ensemble ANN
94	de Mattos Neto et al. (2020)	Helsinki, Finland; So Paulo, Campinas, & Ipojuca, Brazil	$PM_{2.5}$; PM_{10}	Ensemble ANN
95	Han et al. (2020)	35 sites in Beijing, China & 24 sites in London, UK	$PM_{2.5}$; PM_{10}	Ensemble ANN & Bayesian
96	Huang et al. (2020)	Chongqing, China	$PM_{2.5}$; PM_{10}; O_3; NO_2; CO; SO_2	Empirical formula
97	Jin et al. (2020)	Beijing, China	$PM_{2.5}$	Ensemble ANN
98	Liu and Chen (2020)	4 sites in China	NO_2	Ensemble ANN
99	Ma et al. (2020)	Michigan, USA	$PM_{2.5}$	Bayesian
100	Photphanloet and Lipikorn (2020)	Nan Province, Thailand	PM_{10}	Genetic algorithm, MDSF algorithm & methods by Kotu and Deshpande (2018) & Roiger (2017)
101	Sharma et al. (2020)	Queensland, Australia	TSP	Correlation analysis & ensemble ANN
102	Yang and Lee (2020)	Seoul, Korea	$PM_{2.5}$; PM_{10}	Correlation analysis
103	Zeinalnezhad et al. (2020)	Tehran, Iran	CO; SO_2; O_3; NO_2	Fuzzy
104	de Mattos Neto et al. (2021)	8 sites in Brazil & Finland	$PM_{2.5}$; PM_{10}	Ensemble ANN

Table 2 continued from previous page

105	Dong et al. (2021)	Harbin, Xi'an & Guangzhou, China	PM_{10}	Correlation analysis & ensemble ANN
106	Ebrahimi and Qaderi (2021)	Tehran, Iran	SO_2	Fuzzy
107	Gonzalez-Enrique et al. (2021)	Bay of Algeciras, Spain	NO_2	Mutual information, correlation analysis, & Bayesian
108	Han et al. (2021)	Beijing, China	$PM_{2.5}$	Bootstrapping & Bayesian
109	Liu et al. (2021)	Beijing, China	NO_2	Ensemble ANN
110	Menares et al. (2021)	Santiago, Chile	$PM_{2.5}$	Correlation analysis
111	Mokhtari et al. (2021)	Dugway Proving Ground, Utah, USA	C_3H_6	Monte Carlo
112	Shahid et al. (2021)	Aarhus, Denmark	CO; NO_2; SO_2; O_3; $PM_{2.5}$; PM_{10}	Boosting
113	Shams et al. (2021)	Tehran, Iran	SO_2	Sensitivity analysis
114	Taylan et al. (2021)	Jeddah, Saudi Arabia	AQI (SO_2; CO; H_2S; O_3; NO_x, PM_{10})	Fuzzy
115	Wang et al. (2021)	Shanghai, Hangzhou & Nanjing, China	$PM_{2.5}$; PM_{10}	Pearson correlation, Fuzzy, & relief-F
116	Yu et al. (2021)	Online motor vehicle exhaust monitoring platform in China	NO_x	Ensemble ANN
117	Zhang et al. (2021)	Beijing, China	$PM_{2.5}$	Correlation analysis
118	Alkabbani et al. (2022)	Al-Jahra, Kuwait	$PM_{2.5}$; PM_{10}; O_3; SO_2; NO_2; CO	Boruta algorithm
119	Kow et al. (2022)	74 sites in Taiwan	$PM_{2.5}$	Ensemble ANN
120	Kristiani et al. (2022)	Multiple sites in Taiwan	$PM_{2.5}$	Correlation analysis
121	Lin et al. (2022)	Haikou, China	$PM_{2.5}$	Bayesian
122	Tan et al. (2022)	Changsha, China	$PM_{2.5}$	Ensemble ANN
123	Teng et al. (2022)	Shanghai, China	$PM_{2.5}$	Ensemble ANN
124	Tian et al. (2022)	6 sites in Chengdu, China	$PM_{2.5}$; PM_{10}	Correlation analysis & empirical formula
125	Wang et al. (2022a)	Chengdu, Shenzhen & Xi'an, China	$PM_{2.5}$; PM_{10}	Ensemble ANN, Gaussian & T location-

Table 2 continued from previous page

				scale distributions
126	Wang et al. (2022b)	Beijing, China	$PM_{2.5}$	Ensemble ANN & fuzzy
127	Zeng et al. (2022)	Beijing, China	$PM_{2.5}$	Ensemble ANN
128	Zheng et al. (2022)	Harbin, Nanjing, & Shijiazhuang, China	$PM_{2.5}$	Ensemble ANN

ADDRESS: A Distance Decay Regression Selection Strategy; AQI: Air Quality Index; BIC: Bayesian Information Criterion; CART: Classification and Regression Trees; HC: Hydrocarbon; LASSO: Least Absolute Shrinkage and Selection Operator; MDSF: Modified Depth-first Search; MOGA: Multi-objective Genetic Algorithm; NN: Nearest Neighbourhood; PCA: Principal Component Analysis; PSO: Particle Swarm Optimisation; and TSP: Total Suspended Particle.

The distribution of papers by years of publication is shown in Figure 1. There has been a growing number of articles on AP forecasting using ANNs that address model uncertainty more recently. In particular, approximately 65% of the identified articles have been published since 2016 alone. Not much increased activity has been observed between 2006 and 2011 while a sudden growth of activity has occurred post-2015. However, it is worth noting that these values are still comparably lesser than the overall number of published articles per year that deploy ANN models for AP forecasting (Cabaneros et al., 2019). Nonetheless, it is still evident that a considerable amount of attention has been aimed toward the point prediction of AP levels using ANN models while handling uncertainty.

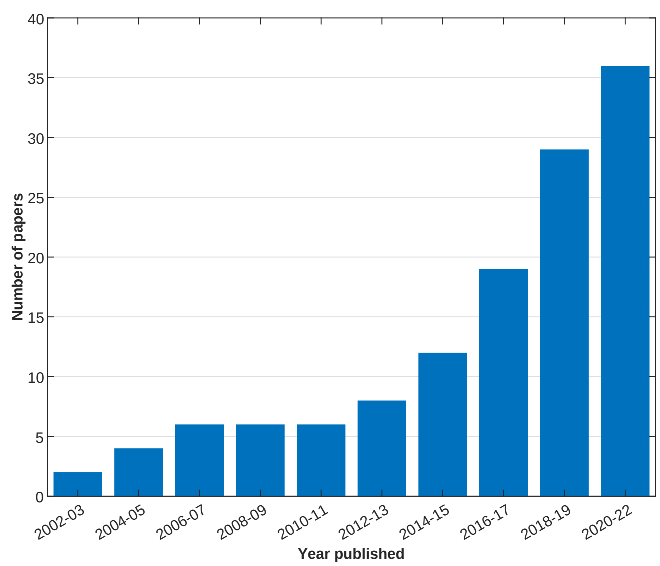

Figure 1: Distribution of papers by years of publication.

3. Sources of Model Uncertainty

To characterise the uncertainty surrounding ANN model development, several notations need to be first described. Any regression-type ANN model generally take the following form:

$$\hat{y} = f(\mathbf{x}, \mathbf{w}) + \sigma_{noise}, \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^{N \times d}$ represents the input vector, e.g. corresponding to N samples and d predictors, $\hat{y} \in \mathbb{R}^{N}$ is vector of model outputs, e.g. estimates of vector $y \in \mathbb{R}^{N}$ containing the actual observations, $\mathbf{w}$ is the vector of model parameters, e.g. connection weights and biases, $f(\cdot)$ is a function describing the dynamics between $\mathbf{x}$ and y, and σ_{noise} is the irreducible (or data) noise which directly influences model errors.

Uncertainty surrounding the development of ANN models has been characterised in a plethora of ways yet they can be generally categorised as either aleatoric or epistemic (Kiureghian and Ditlevsen, 2009). Aleatoric uncertainties refer to those inherent to the dynamics of systems under investigation. This includes the stochasticity of physical and chemical properties of environmental systems, predictor excitations, and noisiness or imperfections of the collected data. On the other hand, epistemic uncertainties occur during the modelling stage from the lack of knowledge about the underlying system being studied, lack of data, and development of imperfect models. Epistemic uncertainty is also referred to as model uncertainty as this is caused by the limitations during the model development process. The two terms in Eq. (1) correspond to the sources of model uncertainty and aleatory uncertainty, respectively. Hence, the total uncertainty of $\hat{y}$, assuming that the two uncertainties are independent, can be estimated as follows:

$$\sigma^2 = \sigma_{ep}^2 + \sigma_{al}^2, \tag{2}$$

where σ_{ep}^2 refers to the model uncertainty, σ_{al}^2 refers to the aleatory uncertainty. It is often a challenge to classify the uncertainties encountered during the modelling stage as real-world applications like outdoor AP forecasting involve both forms of uncertainties. This review will however limit its scope by only focusing on model uncertainty. In particular, the authors propose to characterise model uncertainty by identifying its sources during the building stage of models. That is, each of the steps in the ANN model development process entails a combination of several methods, and it is reasonable to link uncertainty to all of them. As such, this review closely highlights the link between sources of model uncertainty and the said steps. Figure 2 shows the eight general steps for building ANN models and their relationship with the sources of uncertainty which will be discussed in the following subsections. (Detailed information on the inner workings of ANN models can be found from several references: (Gardner and Dorling, 1998; Maier et al., 2010; Hagan et al., 1995;

Bishop, 2006)).

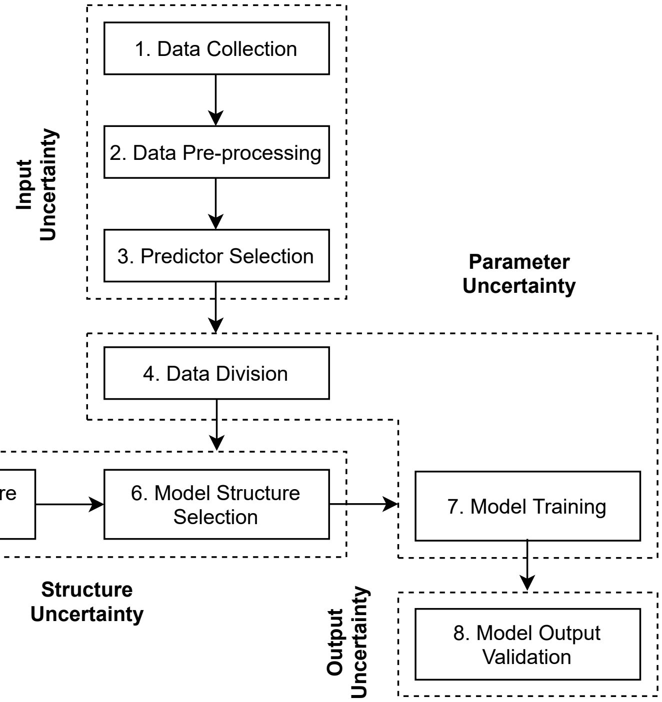

Figure 2: Sources of uncertainty and their relationship to the ANN model development steps.

3.1. Input uncertainty

Input uncertainty results from the lack of complete knowledge on the use of appropriate input $\mathbf{x}$ and y for training ANN models. Given the data-intensive nature of ANN models, input uncertainty is generally influenced by factors such as data density and number of predictors which can lead to varied model structures, training parameters, and irreproducible results. Data density refers to the number of samples, N, needed to train an ANN model. Although model uncertainty is commonly regarded to be inversely proportional to data density (Lai et al., 2022), there is no general formula for identifying the optimal amount of data needed for training without incurring unnecessary computational costs. On the other hand, certain types of predictors, e.g. air pollutant, meteorological, and temporal variables, have been known to better capture the important system dynamics that ANN models attempt to simulate. However, this results in the inclusion of too few or too many ANN model predictors, d, greatly influences input uncertainty.

3.2. Structure uncertainty

Structure uncertainty results from the simplification, ambiguity, and/or lack of information of the governing equation(s) used by ANN models to describe a real-world process (Shrestha and Solomatine, 2008). Due to the empirical nature of ANN models, dealing with structure uncertainty seems inevitable. Structure uncertainty arises from the selection of the following: (i) model architecture, e.g. feedforward, recurrent, etc., (ii) transfer function, e.g log-sigmoid, hyperbolic tangent, linear, etc., and (iii) structure, e.g. number of hidden layers and nodes. (It should be noted that the term 'structure' here has been used in the literature in two different ways: as a governing formula of a model, and as a specific term among ANN modellers referring to the dimension of one or more hidden layers.) Model architecture and transfer function both govern the functional relationship $f(\cdot)$ in Eq. (1) which then determines the model structure, e.g. the dimension of $\mathbf{w}$. As a result, structure uncertainty directly influences the uncertainty surrounding of model parameters.

3.3. Parameter uncertainty

Parameter uncertainty refers to the lack of a general method for identifying the optimal set of network parameters $\mathbf{w}$ as well as the selection of non-optimum algorithms for training ANN models. A common practice to address parameter uncertainty is the assignment of a range of training parameters from which random values are then initially selected (Cabaneros et al., 2019; Hagan et al., 1995). However, such training values are impossible to replicate due to the stochastic nature of most training algorithms. Data division also has a great impact on the level of parameter uncertainty. The selection of a subset of the input data for model training, e.g. $\mathcal{D} = \{\mathbf{x}_n, y_n\}_{n=1}^{n_\mathcal{D}}$ given $n_\mathcal{D}$ training samples, also affects how the network connection weights are initialised and optimised. Various data division schemes, e.g. ad-hoc, stratified, v-fold cross-validation, and random splitting methods, bring varying levels of complexity to training parameters. Since the number of predictors is directly proportional to the number of connection weights that need to be calibrated, input uncertainty also directly impacts the magnitude of parameter uncertainty.

3.4. Output uncertainty

Output uncertainty pertains to the lack of reliability of ANN model results either due to the use of inappropriate validation techniques or the inability to replicate the same accuracy of point predictions. Directly linked to parameter uncertainty, output uncertainty limits the ability of ANN models to produce similar quality of results. Output uncertainty is often referred to as the total model uncertainty which is described as the sum of all uncertainties surrounding all steps in the ANN model development process, e.g. the first term in Eq. 2. However, the majority of the identified validation techniques in the literature only deal with the measurement of the accuracy of the prediction outputs (using the training and testing sets) (Maier et al., 2010). Consequently, the model accuracy indices presented in most case studies are difficult to replicate which often leads to the difficulty of future modellers to build upon previous results.

3.5. Results

As shown in Figure 3, input uncertainty was the most addressed type of uncertainty source, e.g. 112 times, compared with 59, 52, and 14 occasions on which parameter, output, and structure uncertainty sources were addressed, respectively. Input uncertainty was addressed alongside parameter and output uncertainties in 43 of the 112 instances. Furthermore, structure and parameter uncertainties were simultaneously dealt with on 5 occasions. However, the incorporation of all uncertainty sources only occurred twice in this review.

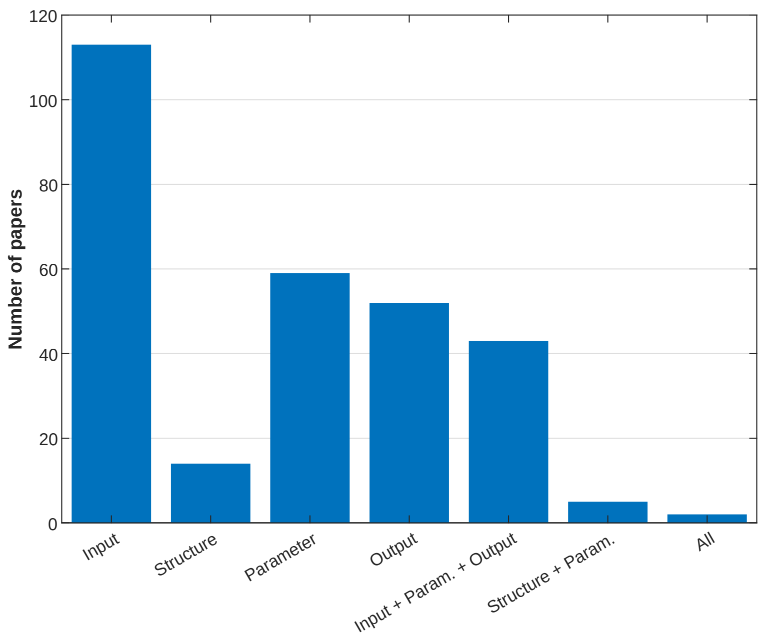

Figure 3: Distribution of papers by source(s) of uncertainty.

4. Methods for addressing model uncertainty

In this review paper, the methods used to address model uncertainty have been classified into eight types, namely (1) Bootstrapping, (2) Bayesian, (3) Fuzzy method, (4) Monte Carlo simulation, (5) Optimisation-based, (6) Sensitivity analysis, (7) Ensemble, and (8) miscellaneous approaches. The classification is similar to those presented in Alvisi and Franchini (2011) and Kasiviswanathan and Sudheer (2017). However, separate types were assigned to ensemble approaches since their use has been prevalent in the field of AP forecasting during the time period that this review covers.

4.1. Bootstrap method

The bootstrap method (or bootstrapping) is an intensive resampling technique with replacement that operates under the assumption that input samples (or bootstraps) follow the statistical characteristics of the population and mimic the underlying random component of the modelled process (Kasiviswanathan and Sudheer, 2017; Efron, 1979). Bootstrapping is carried out by sampling various realisations of input-output patterns to estimate statistical characteristics such as bias, variance, distribution functions, and confidence intervals (Belayneh et al., 2016). As such, bootstrapping has been applied to estimate the confidence interval of the AP predictions which can be used to quantify output uncertainty (see Figure 4).

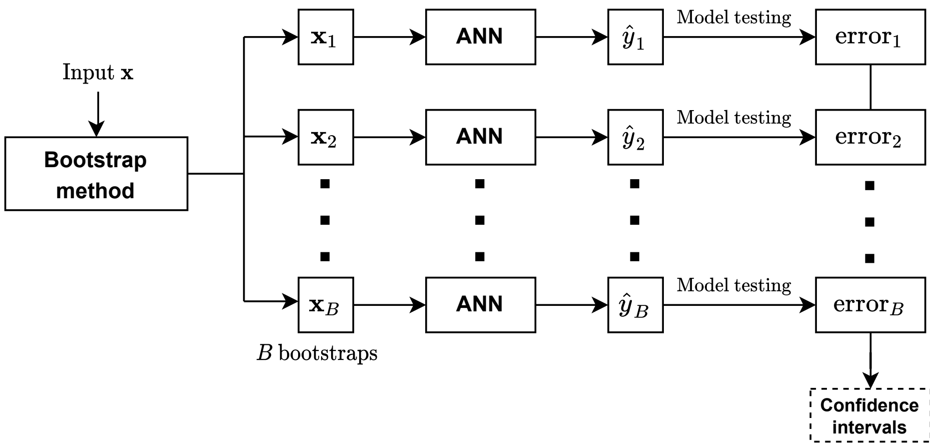

Figure 4: General scheme for handling uncertainty using the bootstrap method.

In theory, the utilisation of more bootstraps should provide a more reliable estimation of the confidence bounds on the model error indices. However, there is no general formula for determining the optimal number of bootstraps. Table 3 provides details of the studies that employed the bootstrap method, including the different number of bootstrap samples used and the statistical indices computed for each of the bootstrap samples. Three studies utilised at least 5000 bootstrap samples which is considered a good practice (Chernick, 1999), although the computational costs should always be considered (Bowden et al., 2005).

Table 3: List of studies that utilised bootstrap samples.

No.	Authors (year)	No. of bootstraps	Indices studied
1	Grivas and Chaloulakou (2006)	100	MAE, RMSE, r, IA
2	Ibarra-Berastegi et al. (2008)	10000	R^2, d_1, FA2, RMSE
3	Noori et al. (2010)	1000	d-factor
4	Voukantsis et al. (2011)	1000	IA
5	Peng et al. (2017)	5000	SS
6	Han et al. (2021)	10000	ARE

The reported studies also revealed the improved reliability of their model results after adopting the bootstrap method. Grivas and Chaloulakou (2006) calculated the standard error of their calculated performance indices from the test results of their predictive models. Their findings revealed that the associated standard error from the bootstraps of the best-performing models were also the lowest values. Ibarra-Berastegi et al. (2008) evaluated the overall performance of their models by estimating the 95% confidence levels of the statistical indices obtained from the model results. The authors carried out this method if the mean values of the said indices for the two models are quite similar. Several studies also applied a similar methodology involving the 95% confidence intervals of their model results and expressed them as error bars. The authors used any confidence intervals lying entirely above zero as indicators that a model performs significantly better than the benchmark ones (Voukantsis et al., 2011; Peng et al., 2017; Han et al., 2021). Noori et al. (2010) accounted for the output uncertainty of their ANN models by showing the plots of the range 95% confidence intervals for their AP level estimates. The authors calculated the 95% prediction uncertainty (95 PPU) of their developed models by finding the 2.5th and 97.5th percentiles of the cumulative distribution of every simulated AP level result. Metrics such as the *d*-factor (Abbaspour et al., 2007), e.g. the average distance between the upper and lower 95 PPU, and R^2 helped the authors determine their best-performing model.

4.2. Bayesian method

Bayesian method is an approach based on Bayes' theorem which states that any prior beliefs regarding an uncertain quantity are updated, on the basis of new information, to yield a posterior probability of the unknown quantity. In detail, the method begins by defining the network weights $\mathbf{w}$ as a probability density function (PDF). A prior PDF, $p_0(\mathbf{w})$, is assigned to the network parameters which is then updated using the training data $\mathcal{D}$ and Bayes' theorem to yield the posterior PDF, $p\left(\mathbf{w}|\mathcal{D}\right)$, as

$$p\left(\mathbf{w}|\mathcal{D}\right) = \frac{p\left(\mathcal{D}|\mathbf{W}\right) p_0\left(\mathbf{w}\right)}{p\left(\mathcal{D}\right)}. \tag{3}$$

By means of the posterior distribution, the predictive probability distribution of the model output can then be estimated as follow

$$p\left(\hat{y}|\mathcal{D}\right) = \int p\left(\hat{y}|x, \mathbf{w}\right) p\left(\mathbf{w}|\mathcal{D}\right) d\mathbf{w}. \tag{4}$$

The integration of the Bayesian method with ANN models, or Bayesian Regularised Neural Network (BRNN), was first employed by Mackay (1992) and Neal (1992) to overcome model overfitting and complexity. Figure 5 illustrates how BRNN models vary from standard ANN models which utilise only a single optimum vector $\mathbf{w}$.

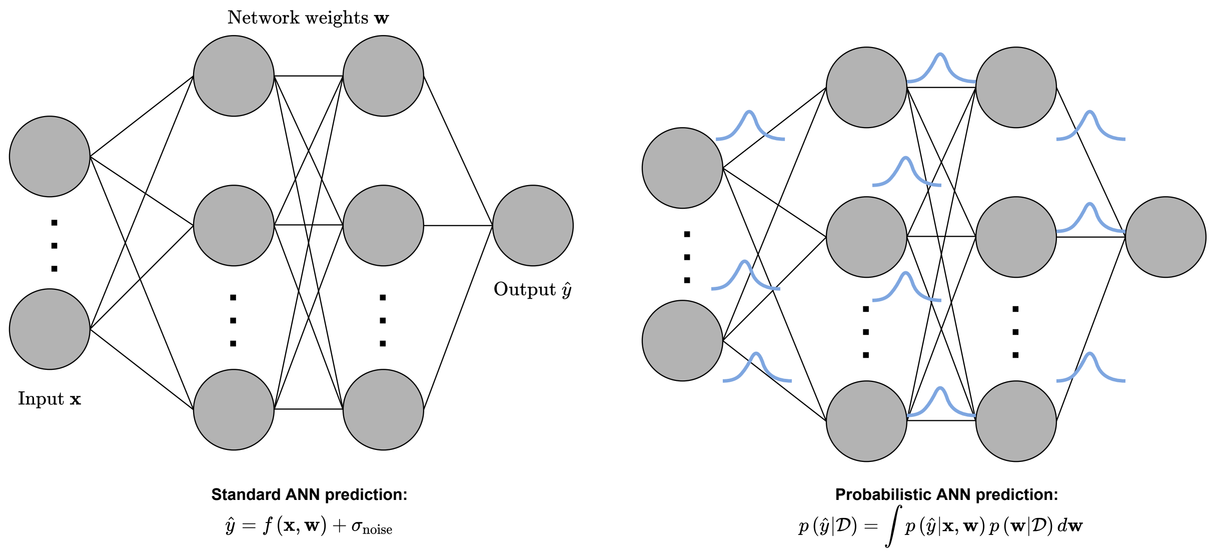

Figure 5: Traditional deterministic ANN models vs. Probabilistic BRNN models.

In summary, Figure 6 shows the general schematic of how BRNN models are employed by case studies to handle parameter uncertainty.

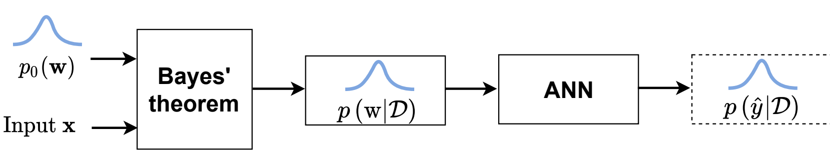

Figure 6: General scheme for handling uncertainty using the Bayesian method.

The use of BRNN models also provides predictions with extra information regarding the precision of the outputs in the form of error bars of the confidence intervals which are very important metrics if the reliability of model results is of particular concern (Bishop, 1995). Table 4 reveals a few identified studies that employed the Bayesian regularisation method. All identified studies except the one by Solaiman et al. (2008) reported superior model performances of BRNN models when compared to a range of benchmark models (from regression to deterministic models).

Table 4: List of studies that employed BRNN models.

No.	Authors (year)	Benchmark model(s)	Results
1	Karakitsios et al. (2006)	Semi-empirical DET	The evaluation parameter of the BRNN model was better than the benchmark one.
2	Solaiman et al. (2008)	MLP, TLFN and RNN	The performance of the BRNN model is not superior to the benchmark ones. However, the BRNN model is less complex in terms of number of hidden ne-

Table 4 continued from previous page

			urons.
3	Balram et al. (2019)	Linear SVM, LIN, Fine tree, Bagged trees, and Gaussian process regression	The performance of the BRNN model with forward predictor selection provided the lowest MSE, RMSE and MAE than those from the benchmark ones.
4	Han et al. (2020)	LASSO and RNN with GRU units	An ensembling scheme which involved the combining the results of two Bayesian-regularised RNN models based on corres-ponding uncertainty measures provided the best results.
5	Ma et al. (2020)	LASSO, Ridge, ARIMA, SVR, ANN, RNN and LSTM	The performance of the Lag-FLSTM model based on Bayesian optimisation yielded the lowest RMSE values compared to the benchmark ones.
6	Gonzalez-Enrique et al. (2021)	ANN	LSTM models with Bayesian optimisation provided the best results.
7	Han et al. (2021)	SVR and RF	The Bayesian-regularised LSTM model outperformed the benchmarks models.
8	Lin et al. (2022)	AR, MA, ARMA, ANN, SVR, GRU, LSTM, and causal CNN	The causal CNN model tuned by Bayesian optimisation achieved the best performance.

AR: Autoregressive: ARMA: Autoregressive Moving Average; DET: Deterministic; GRU: Gate Recurrent Unit; TLFN: Time-lagged feed-forward Network; SVM: Support Vector Machine; LIN: Linear Regression; ARIMA: Autoregressive Integrated Moving Average; SVR: Support Vector Regression; and LSTM: Long Short-term Memory units.

4.3. Fuzzy method

Fuzzy method is based on fuzzy logic (FL) that operates by using if-then rules. FL deals with high-level reasoning using linguistic information acquired from domain experts. As such, FL provides approximate reasoning and explanation abilities which are important attributes of models employed in real-life operations especially air pollution forecasting (Mishra and Goyal, 2016). An FL-based system has three main phases: (1) fuzzification, (2) inference, and (3) defuzzification (see Figure 7). The fuzzification process is where the numerical values of the predictors are transformed into membership functions (MFs). Various types of MFs include the triangular, Gaussian, and trapezoidal. The number and type of MFs per predictor are usually determined empirically and can be decided by experts on the basis of experiment, observation, and experience (Mishra and Goyal, 2016). The Gaussian MF which is based is commonly used due to the nonlinear dynamics between predictors. Inference then follows as the membership grades are processed through a set of if-then rules to generate a fuzzy output. Finally, defuzzification takes place as the fuzzy output is transformed into a quantitative or qualitative output (Yeganeh et al., 2018).

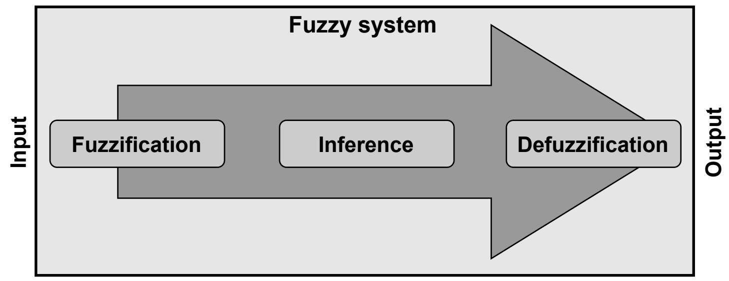

Figure 7: Structure of a fuzzy logic system (adapted from Masood and Ahmad (2021)).

However, the fuzzy method lacks the self-learning capabilities exhibited by ANN models. On the other hand, ANN models are not capable of interpreting linguistic information (Nunnari et al., 1998; Pao, 1989). The combination of fuzzy principles and ANNs, e.g. the adaptive neuro-fuzzy inference system (ANFIS), can therefore analyse any form of information, e.g. numeric, linguistic and logical. This makes ANFIS models capable of addressing input uncertainty by revealing the influence of their model inputs on their outputs according to the presented if-then rules. Table 5 provides a list of studies that employed ANFIS models in forecasting AP levels alongside the type of MFs they utilised. However, the identified papers only focused on improving the accuracy of ANFIS model predictions (similar to the findings of a related review by Kasiviswanathan and Sudheer (2017)).

Table 5: List of studies that employed ANFIS models.

No.	Authors (year)	ANFIS model features
1	Morabito and Versaci (2003)	Gaussian MF
2	Yildirim and Bayramoglu (2006)	Four Gaussian MFs per input
3	Jain and Khare (2010)	Three Gaussian MFs per input
4	Noori et al. (2010)	Five to seven Gaussian MFs per input
5	Dursun and Taylan (2015)	Five Gaussian MFs per input
6	Mishra and Goyal (2016)	Four Gaussian MFs per input
7	Jiang et al. (2017)	Twenty-four to Thirty-two fuzzy rules

Table 5 continued from previous page

8	Taylan (2017)	Five fuzzy rules
9	Yeganeh et al. (2017)	Gaussian MFs
10	Yeganeh et al. (2018)	Gaussian MFs
11	Zeinalnezhad et al. (2020)	Fifteen Gaussian MFs per input
12	Ebrahimi and Qaderi (2021)	Eight Gaussian MFs per input
13	Taylan et al. (2021)	Gaussian MFs
14	Wang et al. (2021)	Number of fuzzy rules is auto-adjusted interval type-2 quantum MF
15	Wang et al. (2022b)	Gaussian MFs

4.4. Monte Carlo simulation

Monte Carlo simulation (MCS) is a sampling technique used for obtaining a probabilistic approximation to the solution of an optimisation model (Metropolis and Ulam, 1949; Rubinstein, 1981). The method operates by sampling different realisations of model inputs and/or parameters by assigning the ranges and PDF of each predictor (Kasiviswanathan and Sudheer, 2017). As illustrated in Figure 8, the PDFs of each predictor are then propagated through $f(\cdot)$ in order to yield the PDF of the model predictions $\hat{y}$. As such, MCS has been performed to handle the input and output uncertainty of ANN models.

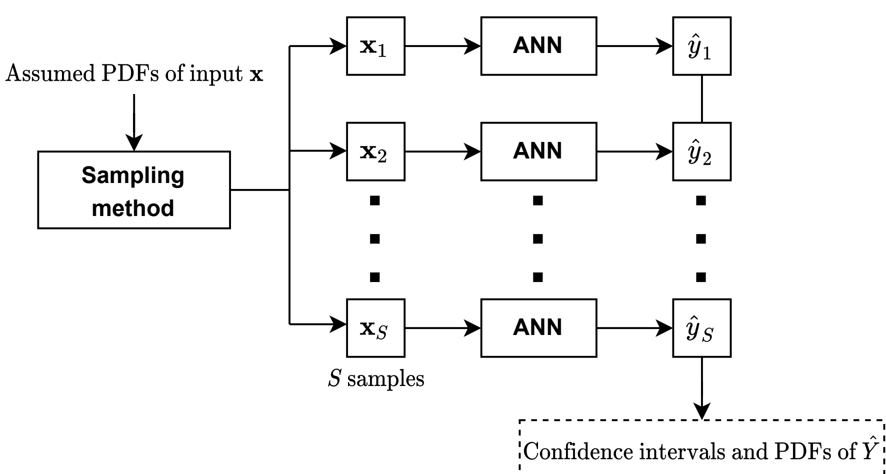

Figure 8: General scheme for handling uncertainty using Monte Carlo simulation.

The application of MCS for uncertainty analysis in AP forecasting with ANN models was only reported three times in this paper. Firstly, Ding et al. (2016) carried out MCS to address the parameter uncertainty when developing their ANN models trained via sparse-response back-propagation algorithm. The authors determined the mean performance of their proposed models based on ten different patterns of weights and biases. Secondly, Noori et al. (2010) performed MCS to quantify the output uncertainty of their developed ANN and ANFIS models. The authors initially generated random samples according to the probability distribution on which their model inputs are based, yielding thousands of model outputs. The authors performed the scheme until the results of a new run do not affect the probability distribution of the output variable. The authors used two metrics in their uncertainty analysis, namely, the d-factor (Abbaspour et al., 2007) and the 95 percent prediction uncertainties (95 PPU). The authors also utilised their MCS results to provide plots of the range 95% confidence intervals for their model forecasts during the training stage. Finally, Mokhtari et al. (2021) incorporated uncertainty quantification methods into the predictions of their proposed CNN-LSTM models. The authors constructed prediction intervals (PIs) for their predictions using MCS dropout and quantile regression methods. Specifically, two metrics for PIs, e.g. prediction interval coverage probability (PICP) and mean prediction interval width (MPIW), were considered in the study.

4.5. Genetic algorithm

Genetic algorithm is an optimisation method based on the idea of the survival of the fittest from the mechanics of genetics. It provides robust solutions for highly complex, non-linear search and optimisation problems (Holland, 1975). Figure 9 illustrates the flowchart of the standard processes involved when performing genetic algorithm. The algorithm operates by initialising a competitive set of possible solution candidates, e.g. chromosomes, and then the solutions are set through the process of natural selection. The solution candidates are then evaluated through a fitness function (or objective function) which ranks the chromosomes in the population. Fitness functions are formulated depending on the problem being solved. The selection of parent chromosomes is then performed which entails two parents for the crossover and the mutation. Crossover involves the exchange of information between two parents. In the mutation stage, the genes of the chromosomes of the crossed offspring are changed. The entire process is carried out until a certain condition is met (Michalewicz, 1996).

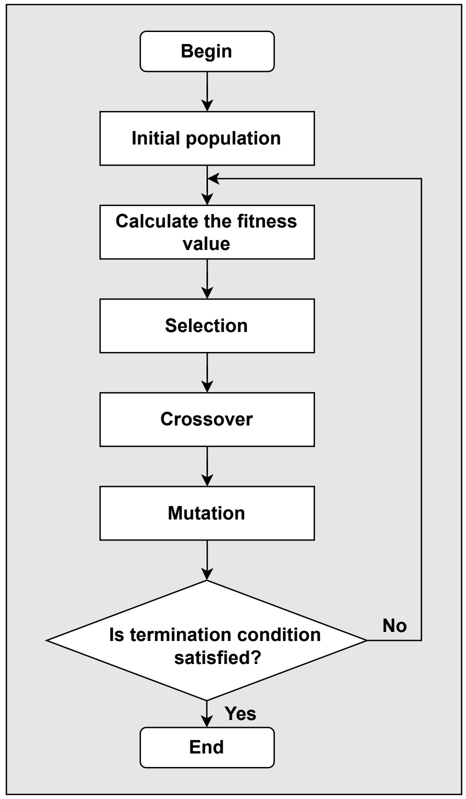

Figure 9: A diagram of the general framework of genetic algorithm.

Genetic algorithm has been integrated with ANN models for AP forecasting to address model uncertainty. For instance, the use of the algorithm to optimise model weights and biases was observed 5 times (Ibarra-Berastegi et al., 2008; Feng et al., 2011; Kadiyala et al., 2013; de Mattos Neto et al., 2017; Zhai and Chen, 2018). The said papers reported the superior performance of ANN models when trained using the method. However, the use of genetic algorithm does not always guarantee superior model performance although despite being able to reduce model complexity (Grivas and Chaloulakou, 2006). The method was also carried out to address input uncertainty via predictor selection on 7 occasions (Grivas and Chaloulakou, 2006; Hájek and Olej, 2012; Elangasinghe et al., 2014; Siwek and Osowski, 2016; Dotse et al., 2018; Liu et al., 2019b; Photphanloet and Lipikorn, 2020). On the other hand, Elangasinghe et al. (2014) employed the method to optimise the step size, momentum rate and processing elements of their proposed ANN models, hence tackling parameter uncertainty. Finally, the use of genetic algorithm to handle both input and parameter uncertainties was observed in this work three times. Ibarra-Berastegi et al. (2008) employed the method during data division to optimise the combination of training and validation sets in terms of mean, standard deviation, maximum and minimum values. The authors also used the method to identify the most relevant predictors of their proposed models. de Mattos Neto et al. (2014) employed the method to optimise several variables of their MLP model such as the number of predictors (in terms of time lags), number of hidden nodes, and training parameters. A similar method for addressing both input and parameter uncertainties was also carried out by de Mattos Neto et al. (2015) in their development of a hybrid MLP model.

4.6. Sensitivity analysis

Sensitivity analysis is a general method used for assessing the relative importance of variables selected as inputs for an ANN model. The method is usually performed to reduce network complexity by eliminating unnecessary predictors while keeping the significant ones. By examining the variations of the model output by the minor perturbations of the predictors, sensitivity analysis can account for the input uncertainty when building ANN models.

In the following, the applications of sensitivity analysis for predictor selection during the development of ANN models are presented:

- Niska et al. (2005) performed a sensitivity analysis alongside MOGA to identify an optimal set of predictors for their MLP models. The authors defined the sensitivity of their predictor subset as the absolute difference between the model performance achieved by using a predictor subset and the model performance achieved when using all predictors.

- Solaiman et al. (2008) selected the predictors for their models based on linear autocorrelation and partial autocorrelation analysis and nonlinear sensitivity analysis. They calculated a metric called the relative sensitivity of a predictor which is the ratio between the standard deviation of the model outputs and the standard deviation of the predictor. The authors initially performed partial autocorrelation analysis between past predictor and predictand values to identify significant time lags. Secondly, sensitivity analysis was then applied for the final stage of screening predictors according to the identified significant lags. Zito et al. (2008) carried out a sensitivity analysis by studying the response of their models to small and equal changes of the predictors. They found out that if an increase in predictor value causes a significant change to the model output, the examined predictor should be retained in the model.

- Kadiyala et al. (2013) carried out predictor selection by performing analysis of variance (ANOVA) alongside the regression tree method.

- Shaban et al. (2016) investigated the influence of incorporating multiple types of predictors, e.g temporal, meteorological, and gaseous, on the performance of their models. They used metrics such as prediction trend accuracy (PTA) and RMSE to assess the results of the models trained using multiple combinations of predictors.

- Stamenković et al. (2017) performed sensitivity analyses of their model predictors through correlation analysis in conjunction with calculating the variance inflation factor (VIF) which is based on the linear relationship between predictors. The authors sequentially removed predictors with highest the VIF values which indicate multicollinearity between predictors.

- Shams et al. (2021) performed a sensitivity analysis by setting the value of one predictor within the range of the standard deviation while the rest were fixed at their mean values. Then, the standard deviation of the model outputs for each predictor changes was measured as model sensitivity for that predictor. The authors then selected the variables with high values in the output standard deviation. On the other hand, sensitivity analysis is usually performed in conjunction with other methods.

4.7. Ensemble approach

Ensembling is a modelling approach that integrates the prediction results of multiple models trained on the training set into one final output. One main advantage of the approach is the ability of the ensemble model (or meta-learner) to possess the individual strengths and simultaneously overcome the limitations of the single models (base learners) (Chen et al., 2008). One of the key limitations exhibited by single models is the inability to accurately predict peak AP levels (Niska et al., 2005; Grivas and Chaloulakou, 2006; Kolehmainen et al., 2001). However, the independence of the base learners and their comparable performance have been pointed out as two important conditions for the ensemble model to perform well (Haykin, 1999; Kuncheva, 2004). The ensemble approach creates multiple input-output realisations of the examined AP system which could be used to account for the input, parameter, and output uncertainties when developing the models. In particular, the resulting ensemble model will contain some diversity and the variance of its predictions can be interpreted as an estimation of model uncertainty (Lai et al., 2022). This review classifies ensemble ANN models into two types, namely, the model- and data-intensive ensemble models (see Figure 10). Under the model-intensive type, the input data are being fed to train multiple base learners and the results are used as inputs to the meta-learner. On the other hand, the data-intensive type initially extracts important features of the input data to train the base learners and then integrates the multiple outputs based on the feature extractor technique used. It is worth noting that the base learners need not consist of entirely ANN-based or ML-based models in general. As shown in the discussion below, many ensemble models are comprised of linear (statistical) and non-linear (ML) base learners.

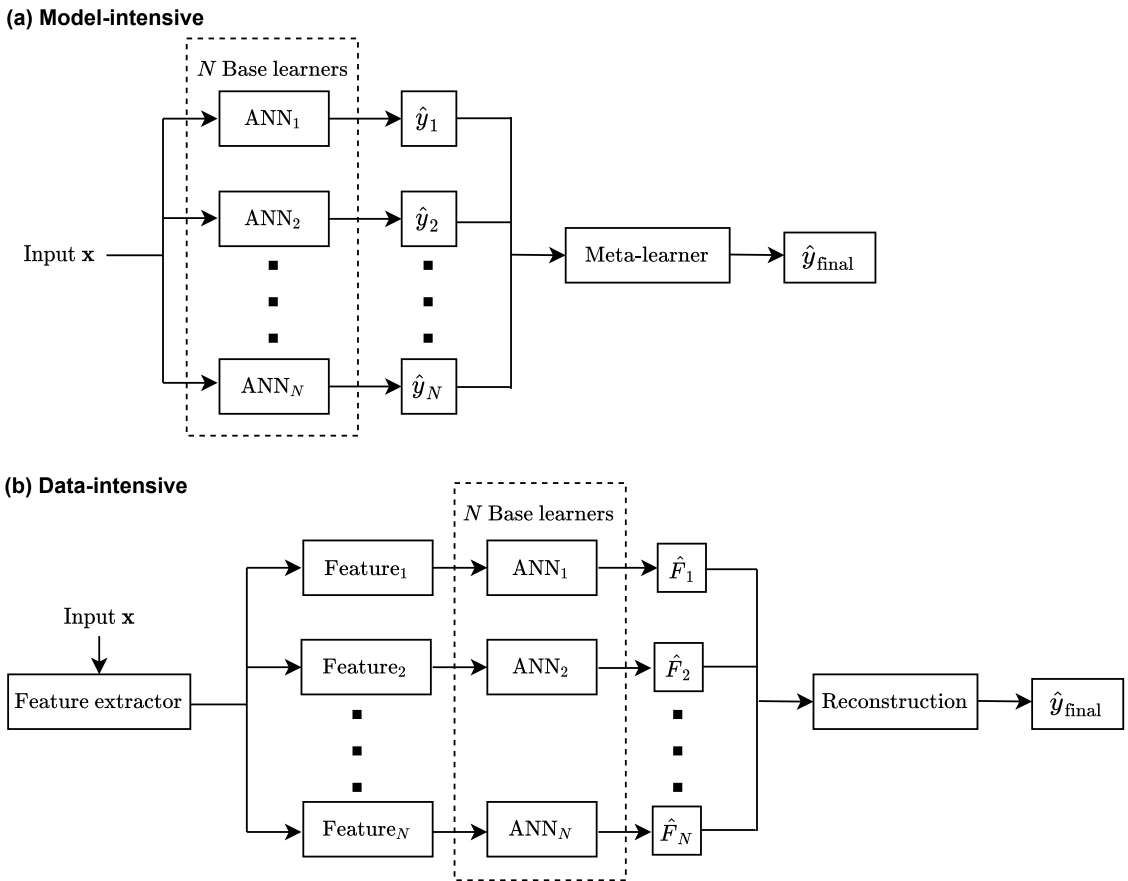

Figure 10: Diagrams of the types of ensemble modelling approaches.

Table 6 shows the identified works that adapted the ensemble modelling approach. The review has identified a variety of meta-learners employed to develop ensemble models. The majority of the said meta-learners are ANN-based such as MLP, SVR, ELM, and LSTM models. Other non-ANN meta-learners have been identified including RF fusion and weighted averaging (Siwek and Osowski, 2016), and linear regression and normal copula-based models (de Mattos Neto et al., 2021). Other forms of base learners to create ensemble models were also identified in this review. Catalano et al. (2016) employed an ensembling approach by choosing the maximum of the predictions of their base learners since they had the tendency to underestimate pollution peaks. Gong and Ordieres-Meré (2016) employed a stacking ensemble algorithm that linearly combined the results of their base learners. The authors used the cross-validation data and least squares under non-negativity constraints to determine the coefficients of the linear combination. Di et al. (2019) utilised a generalised additive model that accounted for geographical differences to predict daily $PM_{2.5}$ levels at a resolution of 1 km $\times$ 1 km across a certain geographical area through ensembling. In particular, the authors developed an ensemble model which incorporated the level of AP concentration against thin-plate splines of latitude and longitude. As such, the results of their base learners were geographically-weighted instead of the traditional approach of using constant weights for each base learner. Similarly, Valput et al. (2019) used an ensemble approach to provide local predictions using regional numerical AP predictions. Their proposed model utilised the forecasts of seven neighbouring air quality models and combined them by calculating their average and weighted-average values. Finally, Sharma et al. (2020) adapted a method that involved entirely non-ANN base learners, e.g. RF, Volterra, M5 tree, and MLR models, and an LSTM meta-learner.

Table 6: List of papers that utilised the ensemble modelling approach alongside the details of the base learners and meta-learners involved.

No.	Authors (year)	Model-intensive type	
		Base learner(s)	Meta-learner(s)
1	Siwek and Osowski (2012)	MLP, SVR, Elman, RBF, ARX, W-MLP, W-SVR, W-Elman, and W-RBF	ANN and SVR
2	Gong and Ordieres-Meré (2016)	SVM, ANN, RF, CART, GBM, Adaboost and Bagging	Linear stacking ensemble
3	Catalano et al. (2016)	ANN and SARIMAX	Max function
4	Siwek and Osowski (2016)	MLP, RBF, and SVR	RF fusion and weighted averaging
5	Wang and Song (2018)	LSTM	SVR
6	Soh et al. (2018)	ANN and LSTM	ANN
7	Di et al. (2019)	ANN, RF, Gradient boosting	Geographically weighted generalised additive model
8	Gu et al. (2019)	SVR	SVR-based regression for stacking
9	Van Roode et al. (2019)	LASSO and IDW	ANN
10	Zhao et al. (2019)	LSTM (using spatial data)	ANN
11	Valput et al. (2019)	MLP, deep MLP, LSTM, CNN, and LIN	Averaging and weighted averaging model
12	Chang et al. (2020)	LSTM	LSTM
13	Han et al. (2020)	Bayesian RNN	Uncertainty-based fusion
14	Sharma et al. (2020)	RF, Volterra, M5 model tree, and MLR	LSTM

Table 6 continued from previous page

15	de Mattos Neto et al. (2021)	AR, ARMA, IIR, MLP, RBF, ELM, ESN, and ANFIS	LR, LR with predictor selection, MLP, ELM, ELM with CR, and NC
16	Kow et al. (2022)	Multiple CNNs	ANN
17	Tan et al. (2022)	Graph attention network	LSTM-CNN
18	Wang et al. (2022a)	ELM, BPNN, SVM, Elman	Weighting algorithm
		Data-intensive type	
		Feature learner(s)	Feature extractor(s)
19	Prakash et al. (2011)	Elman	Symlet(8) wavelet
20	Shekarrizfard and Hadad (2012)	ANN	Wavelet transform
21	Dunea et al. (2015)	ANN	Db3 Daubechies wavelet
22	Feng et al. (2015)	MLP	Wavelet transform
23	Bai et al. (2016)	MLP	SWT
24	Bai et al. (2019)	LSTM, ANN	EMD
25	Liu et al. (2019a)	ANN	Db1 Daubechies wavelet
26	Liu et al. (2019b)	NARX	EWT
27	Maciąg et al. (2019)	Spiking ANN	Clustering algorithm
28	Qiao et al. (2019)	SAE-LSTM	Wavelet transform
29	Wu and Lin (2019)	LSTM	Wavelet transform
30	Mo et al. (2019)	ELM	CEEMDAN
31	Cabaneros et al. (2020)	LSTM	Daubechies wavelets
32	de Mattos Neto et al. (2020)	MLP, RBF, ELM and ESN	Partition methods
33	Jin et al. (2020)	GRU	EMD with CNN classifier
34	Liu and Chen (2020)	ELM	EWT
35	Dong et al. (2021)	SVR, BPNN, MLP, Incremental ELM, GRU, and LSTM	Noise-assisted EMD
36	Liu et al. (2021)	LSTM, ANN, Bi-LSTM	DWT
37	Yu et al. (2021)	LSTM	CEEMDAN
38	Teng et al. (2022)	Bi-LSTM	EMD
39	Zeng et al. (2022)	Nested LSTM	Extended SWT

Table 6 continued from previous page

40	Zheng et al. (2022)	GRU	Wavelet transform, Sample entropy and VMD

ARMA: Autoregressive and Moving Average; ARX: Auto-regressive with external variables; BPNN: Backpropagation Neural Network; CEEMDAN: Complete Ensemble Empirical Mode decomposition with Adaptive Noise; CR: Coefficient of regularization; ELM: Extreme Learning Machine; EMD: Empirical Mode Decomposition; ESN: Echo State Network; EWT: Empirical Wavelet Transform; GBM: Gradient Boosting Machine; GRU: Gated Recurrent Unit; IDW: Inverse Distance Weight; IIR: Infinite Impulse Response Filter; LR: Linear regression; NC: Normal Copula-based; RBF: Radial Basis Function; RF: Random Forest; SAE: Stacked Autoencoder; SARIMAX: Seasonal Autoregressive Integrated moving average with exogenous variables; SWT: Stationary Wavelet Transform; VMD: Variational Mode Decomposition; W-Elman: Wavelet-based Elman; W-MLP: Wavelet-based MLP; W-RBF: Wavelet-based RBF; and W-SVR: Wavelet-based SVR.

A few special cases of model-intensive ensemble methods have also been found in this review. For instance, Mahajan et al. (2018) proposed a hybrid ensemble approach that utilises both linear and nonlinear base learners. In particular, the authors initially used an ARIMA model to capture the linear tendencies of their $PM_{2.5}$ time series. The model residuals, e.g. difference between actual time series and model results, were then fed to a lagged-input ANN model. The results of both base learners were then given equal weights before they were combined. Gu et al. (2019) employed several SVR models as base learners and then applied pruning techniques to eliminate the negative learners. The results of the selected base learners were then fed to an SVR-based regression model for stacking. Finally, Han et al. (2020) employed an approach of integrating the results of two Bayesian RNN base learners according to their uncertainty measures. In particular, the authors employed two weighting methods to fuse the results of the base learners, (1) uncertainty-averaged outputs, and (2) selection of a base learner output with the lowest uncertainty measure.

Under the data-intensive category, the most commonly-used feature extractors are wavelet-based techniques (see Table 6). For instance, many authors employed wavelet transformation to decompose their original data into several coefficients which were then fed to their base learners (or feature learners, more accurately speaking). Wavelet decomposition techniques were especially applied to improve the performance of plain ANN models when dealing with peak AP concentration levels (Shekarrizfard and Hadad, 2012; Feng et al., 2015). Other similar feature extractors were utilised such as Empirical Mode Decomposition (EMD) (Bai et al., 2019) and its variants such as the Noise-assisted EMD (Dong et al., 2021), and Clustering algorithm (Maciąg et al., 2019). Overall, the general conclusion of the identified papers suggested the superiority of their proposed ensemble models to the involved base learners. However, the trade-off between the overall complexity of ensemble models and performance should be carefully accounted for especially in real-world scenarios where computational cost is a major constraint (Cabaneros et al., 2019).

A few special cases of data-intensive ensemble methods have also been identified. For instance, de Mattos Neto et al. (2017) employed an approach that involved forecasting both the AP time series and the model residuals. The authors only conducted the latter if the residual is not white noise, e.g. a series of independent and identically distributed random values with zero mean and constant variance. A nonlinear base learner, e.g. MLP model, was utilised for the initial forecasting while both linear and nonlinear ones, e.g. ARIMA, MLP, and SVR models, for the residual. The approach falls under the ensemble type since both forecasting results were all combined using an MLP model. Earlier work by de Mattos Neto et al. (2015) applied the same method which included the forecasting of the model residuals. However, the work was model-intensive in that the residuals of the final ensemble model output were further fed to a series of MLP models until the scheme generates residuals having a white noise behaviour. Finally, de Mattos Neto et al. (2020) addressed the seasonality of $PM_{2.5}$ and PM_{10} pollutants by decomposing the time series into non-overlapping monthly partitions and then applied several models, e.g. as meta-learners. Their decomposition method involved creating n subseries from the original time series according to the coefficient of variation of each subseries.

4.8. Miscellaneous approaches

Several other methods that address model uncertainty have also been identified in this review and they are briefly described below:

- Several techniques have been adapted to identify the most significant model predictors, hence tackling input uncertainty: ADDRESS method (Yeganeh et al., 2018), Boruta algorithm (Alkabbani et al., 2022), Correlation analysis (Perez and Salini, 2008; He et al., 2014; Mishra et al., 2015; Bai et al., 2016; Stamenković et al., 2017; Alimissis et al., 2018; Tao et al., 2019; Dong et al., 2021; Menares et al., 2021; Zhang et al., 2021; Kristiani et al., 2022; Tian et al., 2022) Fourier analysis (Hrust et al., 2009), Grey Correlation analysis (Zhu et al., 2018), Individual Smoothing Factors (Antanasijević et al., 2013), LASSO (Yeganeh et al., 2017), MDSF algorithm (Photphanloet and Lipikorn, 2020), and Stepwise Regression (Ordieres et al., 2005; Singh et al., 2012; Russo et al., 2013; Russo and Soares, 2014; Siwek and Osowski, 2016; Agirre-Basurko et al., 2006).

- Many approaches have also been employed to address structure uncertainty by determining the optimal number of nodes in the hidden layer and consequently reducing overall model complexity. Dutot et al. (2007) employed a stepwise method using BIC-like information criterion to tackle structure uncertainty by determining the optimal number of nodes in the hidden layer. Chelani et al. (2002) convergence criteria according to the error minimization criterion, and a formula by Kinnebrock (1995). Catalano et al. (2016) applied the weight generalisation formula by Bishop (1995) which utilises an extra error term that penalises small weights during model training. Finally, Agirre-Basurko et al. (2006) employed a generalisation rule by Amari et al. (1997) in which the model is considered optimally trained when the ratio of the number of training samples to the number of connection weights exceeds 30.

- Hasham et al. (2004) adopted an approach based on factorial design concepts to assess the influence of model predictors, structure and parameters on the model output (Box et al., 1978). The approach operates by moving various input factors are moved between high and low settings in combination with other input factors, hence investigating the interaction of uncertainties between the said factors.

- Empirical formulas which provide either or both the upper and lower bounds of the optimal number of of nodes in the hidden layer were also used: formula proposed by Fletcher and Goss (1993) (He et al., 2014; Abdullah et al., 2019), empirical formula by Shen et al. (2008) with trial and error (Bai et al., 2016), empirical rule by Kalogirou (2003) (Radojević et al., 2018), a formula by Kotu and Deshpande (2018) and Roiger (2017) (Photphanloet and Lipikorn, 2020), and a method by Tian et al. (2022).

- Dutot et al. (2007) used a metric called leverage which provides a confidence interval of the predicted values of their models to address output uncertainty. Leverage is a metric used to assess the effect of a particular observation on the fitted regression according to the position of the observation in the predictor space Monari and Dreyfus (2002).

- Powerful swarm intelligence algorithms have also been used to optimise the training parameters of ANN models. For instance, Mo et al. (2019) utilised the Whale Optimisation meta-heuristic algorithm to obtain the best parameters of their proposed ensemble ELM models. Caraka et al. (2019) employed both PSO and backpropagation algorithms to train their ANN models.

- Shahid et al. (2021) employed a boosting algorithm to improve the results of their initial model, e.g. SVR model, to address input uncertainty. The boosting technique works by assigning weights to each instance of the input data and using them to train the SVR model, then identifying and updating the weights of the misclassified instances. The weighted instances are finally passed to several models including MLP, RF, Decision tree, MLR, Ridge regression, Gradient Boosting and SVR models for the final training and prediction tasks.

- Wang et al. (2022a) applied a probabilistic approach by calculating the distribution of their model prediction errors and comparing them to create confidence intervals of results. The approach was based on the Gaussian and T location-scale distributions which were determined according to the R^2, PE, and

RMSE indices. The authors also calculated three metrics such as average coverage error (ACE), prediction interval normalized average width (PINAW) and prediction interval coverage probability (PICP) to measure the interval predictions. Zheng et al. (2022) decomposed their original $PM_{2.5}$ time series into several subseries using wavelet transforms and employed the reinforcement learning algorithm. e.g. Q learning, for the predictor selection of each subseries.

4.9. Results

The distribution of papers by the methods used to handle model uncertainty is shown in Figure 11. The majority of the identified papers, e.g. 45 occasions, adapted the ensemble modelling approach to address uncertainty. Among these, approximately 69% have been recently published since 2019 indicating the emergence of more sophisticated approaches as ANN computing technologies also become more powerful. Fuzzy systems were applied alongside ANN models 20 times, while correlation analysis for predictor selection was applied 18 times. The utilisation of the global search procedure, e.g. genetic algorithm, occurred 15 times, followed by sensitivity analysis of trained models (11 times). The number of papers in which bootstrapping, Bayesian, and empirical methods were applied was uniform, varying between 6 to 7, compared with only three instances where the MCS was applied. Finally, the use of alternative methods for addressing uncertainty was reported 28 times.

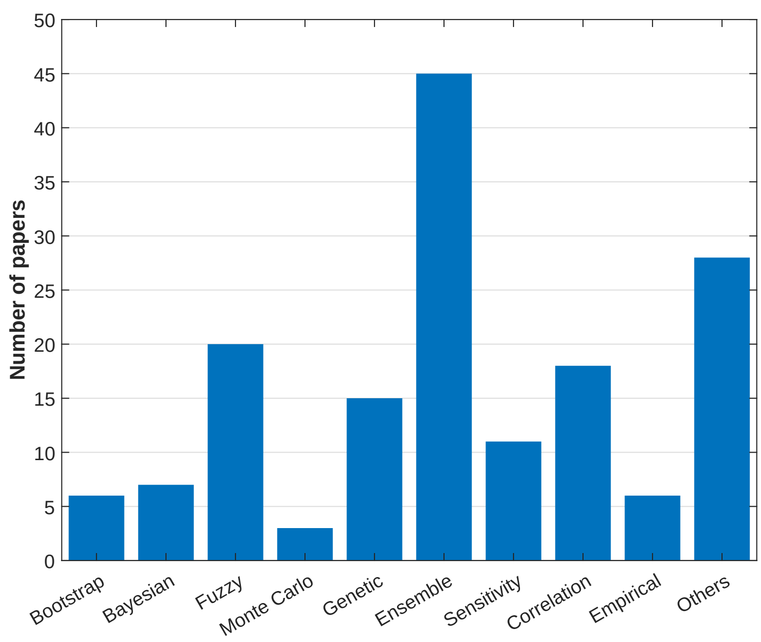

Figure 11: Distribution of papers by methods used to address and/or quantify model uncertainty.

Table 7 presents the pros and cons of the identified methods that can address model uncertainty. Note that the methods falling under 'miscellaneous approaches' were excluded from the summary. The use of the said methods has been reported to improve the forecasting accuracy of ANN models which is usually the primary motivation of most studies. This is especially true for those methods handling input and parameter uncertainties to reduce model complexity. However, it can be seen that not all methods can directly quantify model uncertainty. For instance, fuzzy systems, GA, and sensitivity analysis can only account for uncertainty by addressing the ambiguity arising from predictor selection and parameter optimisation. Nonetheless, the use of such methods is still considered a good practice when compared to the sole adoption of ad-hoc or knowledge-based methods (Cabaneros et al., 2019; Maier et al., 2010). Another commonly identified drawback is the lack of a one-size-fits-all approach for implementing the methods. This could be a potential stumbling block to future researchers from addressing uncertainty when building ANN forecasting models. Finally, the majority of the methods tend to demand higher computational costs when implemented. For instance, bootstrapping, Monte Carlo simulation, GA, and ensembling involve the training of multiple ANN models.

Table 7: Pros and cons of the identified methods that can handle model uncertainty.

Methods	Pros (+) / Cons (-)
Bootstrapping	+ Can directly quantify output uncertainty via confidence intervals of the model predictions
	- No general formula for determining the optimal number of bootstraps
	- Can be computationally expensive due to multiple model trainings
Bayesian	+ Can directly quantify parameter uncertainty via error bars of the confidence intervals of model predictions
	+ Can enable ANN models yield superior results
	- Can be computationally expensive when being implemented
	- Difficult to perform as several hypotheses regarding weight distributions are needed
Fuzzy	+ Can provide superior results when applied alongside ANN models
	- Difficult to perform as good memberships functions need to be identified
	- Does not directly quantify input uncertainty
Monte Carlo	+ Can directly quantify parameter and output uncertainties
	- Can be computationally expensive due to multiple model trainings
	- No general scheme for implementing it
Genetic algorithm	+ Can provide robust optimisation solutions
	+ Can improve model performance by addressing model complexity
	- Does not directly quantify input and/or parameter uncertainties

Table 7 continued from previous page

	- Can be computationally expensive when implemented
Sensitivity analysis	+ Can be integrated with other techniques
	+ Can improve model performance by reducing model complexity
	- Does not directly quantify input uncertainty
Ensembling	+ Can improve final model performance by gaining (or overcoming) the strengths (or weaknesses) of individual models
	+ Can directly quantify output uncertainty of the final model via the computed variance of several results of individual models
	- Can be computationally expensive due to the multiple trainings of individual models
	- No general scheme for choosing the appropriate individual and final models

However, the adoption of metrics that directly quantified model uncertainty was limited in this review, e.g. 11 of the 128 identified articles. Table 8 summarises the metrics which are predominantly based on the confidence intervals of model outputs. Among the interval-based metrics, output uncertainty was visually inspected 5 times either as confidence interval plots or error bars. On the other hand, output uncertainty metrics were treated as separate model performance evaluators, e.g. d-factor, 95 PPU, PICP, MPIW, ACE, and PINAW, in 3 occasions. An article was also found in which uncertainty metrics based on the posterior distribution of network weights were used to guide the merging of two base learner results.

Table 8: List of identified methods and metrics that directly quantify model uncertainty.

No.	Authors (year)	Method	Metrics
1	Grivas and Chaloulakou (2006)	Bootstrapping	Standard errors from bootstrap samples
2	Dutot et al. (2007)	Leverage metric (Monari and Dreyfus, 2002)	Confidence interval
3	Ibarra-Berastegi et al. (2008)	Bootstrapping	96% confidence interval of model metrics
4	Solaiman et al. (2008)	Bayesian	95% confidence interval

Table 8 continued from previous page

			(as interval plots)
5	Noori et al. (2010)	MCS	95 PPU, interval plots and d-factor (Abbaspour et al., 2007),
6	Voukantsis et al. (2011)	Bootstrapping	95% confidence intervals of model metrics
7	Peng et al. (2017)	Bootstrapping	95% confidence interval (plotted as error bars)
8	Han et al. (2020)	Bayesian	Two metrics based on the posterior distribution of network weight parameters
9	Han et al. (2021)	Bootstrapping	95% confidence intervals (as interval plots)
10	Mokhtari et al. (2021)	MCS	Interval-based metrics: PICP and MPIW
11	Wang et al. (2022a)	Gaussian and T location-scale distributions	$(1-\alpha)$% confidence intervals (as interval plots), ACE, PINAW and PICP

5. Summary and conclusions

Data-driven approaches especially ANN models and their application to outdoor AP forecasting have received a lot of attention in the past two decades. Their development has allowed researchers to provide accurate AP forecasts without the theoretical understanding required by traditional physics-based models. However, ANN models are empirical and their development inevitably possesses an intrinsic level of uncertainty that can restrict the reliability of their results. Hence, this review was performed to investigate the methods employed for addressing model uncertainty in the context of AP forecasting using ANN models.

Since the period January 2000 and August 2022, research activity in the incorporation of model uncertainty when developing ANN models has increased rapidly. The average number of journal articles published during the said period was around 5.5 per year. It is also worth noting that 65% of the identified articles have been published alone in the past six years. This is a significant development given the huge adoption of ANN models in many decision-making tasks despite their black-box nature. However, there still is a huge gap between the number of papers published per year that address model uncertainty and those that do not. This is consistent with previous results reported by related reviews on the use of data-driven forecasting models (see Maier et al. (2010), Kasiviswanathan and Sudheer (2017), Cabaneros et al. (2019) and Masood and Ahmad (2021)).

In relation to sources of model uncertainty, there was a huge amount of research activity that covered input uncertainty. For instance, approximately 89% of the identified articles employed various methods to handle input uncertainty, and almost a third of those articles also addressed both parameter and output uncertainties. Methods dealing with input uncertainty seem ubiquitous since ANN models are essentially as good as their input data used. There was also a significant number of methods used to address the uncertainties surrounding model parameters and outputs. However, methods handling structure uncertainty received less attention as was in the findings of Cabaneros et al. (2019). In particular, optimal model structure was still mostly determined using ad-hoc, e.g. trial-and-error and/or knowledge-based, approaches which were excluded from this review. Consequently, there is a need to consider more analytical approaches for dealing with structure uncertainty.

Efforts examining the interaction among four model uncertainty types were still not present in the identified papers. Total model uncertainty was attributed conceptually as a sum of several components related to the sources of uncertainty presented in this work (Arhami et al., 2013). However, none of the papers provided any form of metric that attempts to quantify the overall uncertainty when building ANN models. Given the black-box nature of ANN models, future modellers may not be able to quantify total model uncertainty in terms of individual types of uncertainty, hence this field of research still demands further attention.

The majority of the identified papers have been found to adopt the ensemble approach in handling model uncertainty. This comes as no surprise given the availability of exceedingly more powerful computing tools capable of handling complex model architectures such as deep learning. It is also worth noting that the use of ensemble models has only emerged more recently, e.g. a large number of the identified papers have only been published since 2019. An improvement in accuracy in the model results was also reported when ensemble models were compared to their benchmark base learners. This is especially the case when dealing with the accurate predictions of peak AP levels Masood and Ahmad (2021); Shekarrizfard and Hadad (2012); Feng et al. (2015). However, the use of ensemble models has its drawbacks. As pointed out by Masood and Ahmad (2021), ensembling techniques demand longer computational time which makes them unsuitable for rapid forecasting AP forecasting applications. Another prevailing issue is the uncertainty surrounding the selection of the base learners and meta-learners. Since the selection is still mostly carried out in an ad-hoc or knowledge-based manner, there is a need to examine this further in the future.

The use of methods such as fuzzy expert systems, correlation analysis, genetic algorithm, and sensitivity analysis also received a significant amount of attention. The application of genetic algorithm for addressing not just parameter but also input and structure uncertainties is significant progress. Global search optimisation techniques such as genetic algorithm are analytical model-based approaches that have previously received less attention because it is computationally expensive to implement them (Cabaneros et al., 2019). This review has also identified a wide array of miscellaneous procedures for handling model uncertainty, ranging from correlation and mutual information techniques to meta-heuristic search algorithms. However, the use of bootstrapping, Bayesian regularisation and Monte Carlo simulation which are especially capable of quantifying output certainty remains limited and therefore requires further investigation (Cabaneros et al., 2019; Wu et al., 2014; Kasiviswanathan and Sudheer, 2017).

In general, the majority of the identified methods have only attempted to handle but failed to quantify model uncertainty. There are only 11 instances in which uncertainty was directly measured. Most of those efforts expressed output uncertainty in terms of confidence interval plots, error bars, or separate metrics for evaluating model performance. This could raise some issues especially when both accuracy and reliability of prediction results are required. As such, the quantification of model uncertainty especially in the context of AP forecasting needs to receive increased attention to ensure the reliability and transparency of ANN model results. In particular, the standardisation of reporting ANN model results which include both accuracy and uncertainty metrics should be encouraged. Such a practice not only enables a better comparison of proposed model development methods, but it further increases the confidence in the use of ANN model results in real-world applications, especially AP forecasting.

6. Recommendations for future research

Based on the review of 128 papers on the use of techniques for addressing model uncertainty in the field of outdoor AP forecasting using ANN models conducted in this paper, the following recommendations for future work are made:

1. More research needs to be undertaken on the improved reporting of both accuracy and uncertainty of ANN model results. This is to further validate the use of data-driven models in real-world tasks involving decision-making where both accuracy and reliability of the results are essential. Although the primary aim of most studies is towards increasing the accuracy of model results, future attempts

should also focus on the characterisation of confidence intervals as well as the development of metrics for directly quantifying model uncertainty.

2. Greater attention should be given to the application of techniques for addressing structure, parameter, and output uncertainties. The majority of identified papers in the review have placed more emphasis on input uncertainty which is noteworthy given that ANN models are data-driven. However, the ambiguity surrounding the structure, parameter and output of ANN models are ubiquitous which may hinder future researchers from adapting black box approaches.

3. More work should continue on the use of model-based methods such as bootstrapping, Bayesian regularisation, and Monte Carlo simulation. These methods have received less attention despite their ability to address and quantify model uncertainty.

4. The relationship between model uncertainty and complexity of ensemble modelling frameworks should be examined in the future. There clearly is a growing trend in the implementation of ensemble frameworks in the field of AP forecasting using ANN models, especially deep learning. However, almost all of the identified works have not attempted to measure model uncertainty. Furthermore, ensemble approaches are computationally expensive which could hinder their deployment in real-world tasks.

5. A new area of research that deals with the interplay of single or multiple model uncertainty sources should be carried out by future modellers. To the best of the authors' knowledge, no work has been done to examine this aspect of model uncertainty in the context of building ANN models. The resulting work could be adopted from existing case studies on the accuracy aspect of their developed ANN models.

6. More software that provides a less difficult and computationally-efficient platform for handling and quantifying uncertainty in ANN models should be developed to ensure that metrics assessing model uncertainty become an essential requirement in reporting new ANN based methods.

References

H. T. Shahraiyni, S. Sodoudi, Statistical modeling approaches for 10 prediction in urban areas; A review of 21st-century studies, Atmosphere (Basel). 7 (2016) 10–13. doi:10.3390/atmos7020015.

S. M. S. Cabaneros, J. K. Calautit, B. R. Hughes, A review of artificial neural network models for ambient air pollution prediction, Environ. Model. Softw. 119 (2019) 285–304. doi:10.1016/j.envsoft.2019.06.014.

A. Masood, K. Ahmad, A review on emerging artificial intelligence (AI) techniques for air pollution forecasting: Fundamentals, application and performance, Journal of Cleaner Production 322 (2021) 129072. doi:10.1016/j.jclepro.2021.129072.

M. Gardner, S. Dorling, Artificial neural networks (the multilayer perceptron)a review of applications in the atmospheric sciences, Atmos. Environ. 32 (1998) 2627–2636. doi:10.1016/S1352-2310(97)00447-0.

B. Lv, W. G. Cobourn, Y. Bai, Development of nonlinear empirical models to forecast daily pm2.5 and ozone levels in three large chinese cities, Atmospheric Environment 147 (2016) 209–223. doi:https://doi.org/10.1016/j.atmosenv.2016.10.003.

J. Zhao, F. Deng, Y. Cai, J. Chen, Long short-term memory - Fully connected (LSTM-FC) neural network for PM2.5 concentration prediction, Chemosphere (2019). doi:10.1016/j.chemosphere.2018.12.128.

C. M. Bishop, Pattern Recognition and Machine Learning, Springer, New York, NY, USA, 2006. doi:10.1117/1.2819119. `arXiv:arXiv:1011.1669v3`.

F. J. Montáns, F. Chinesta, R. Gómez-Bombarelli, J. N. Kutz, Data-driven modeling and learning in science and engineering, Comptes Rendus - Mecanique 347 (2019) 845–855. doi:10.1016/j.crme.2019.11.009.

A. Bekkar, B. Hssina, S. Douzi, K. Douzi, Air-pollution prediction in smart city, deep learning approach, Journal of Big Data 8 (2021). doi:10.1186/S40537-021-00548-1.

G. Giray, A software engineering perspective on engineering machine learning systems: State of the art and challenges, Journal of Systems and Software 180 (2021) 111031. doi:https://doi.org/10.1016/j.jss.2021.111031.

TIOBE, Tiobe index for february 2022, 2022. URL: `https://www.tiobe.com/tiobe-index/`.

K. Y. Ng, N. Awang, Multiple linear regression and regression with time series error models in forecasting PM 10 concentrations in Peninsular, Environ. Monit. Assess. (2018). doi:10.1007/s10661-017-6419-z.

M. Gardner, S. Dorling, Neural network modelling and prediction of hourly NOx and NO2 concentrations in urban air in London, Atmos. Environ. 31 (1997) 4081–4094. doi:10.1016/S1352-2310(97)00282-3.

W. Z. Lu, W. J. Wang, Potential assessment of the “support vector machine” method in forecasting ambient air pollutant trends, Chemosphere 59 (2005) 693–701. doi:10.1016/j.chemosphere.2004.10.032.

J. Heo, D. Kim, A new method of ozone forecasting using fuzzy expert and neural network systems, Science of the Total Environment 325 (2004) 221–237. doi:10.1016/j.scitotenv.2003.11.009.

B. S. Freeman, G. Taylor, B. Gharabaghi, J. Thé, Forecasting air quality time series using deep learning, J. Air Waste Manage. Assoc. 68 (2018) 866–886. doi:10.1080/10962247.2018.1459956.

S. H. Chen, A. J. Jakeman, J. P. Norton, Artificial Intelligence techniques: An introduction to their use for modelling environmental systems, Math. Comput. Simul. 78 (2008) 379–400. doi:10.1016/j.matcom.2008.01.028.

W. S. McCulloch, W. Pitts, A Logical Calculus of the Ideas Immanent in Nervous Activity, Bulletin of Mathematical Biophysics 5 (1943) 115–116. doi:https://doi.org/10.1007/BF02478259.

I. A. Basheer, M. Hajmeer, Artificial neural networks: Fundamentals, computing, design, and application, J. Microbiol. Methods 43 (2000) 3–31. doi:10.1016/S0167-7012(00)00201-3.

H. R. Maier, A. Jain, G. C. Dandy, K. Sudheer, Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions, Environ. Model. Softw. 25 (2010) 891–909. doi:10.1016/J.ENVSOFT.2010.02.003.

W. Wu, G. C. Dandy, H. R. Maier, 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, Environ. Model. Softw. 54 (2014) 108–127. doi:10.1016/J.ENVSOFT.2013.12.016.

S. Vardoulakis, B. E. Fisher, N. Gonzalez-Flesca, K. Pericleous, Model sensitivity and uncertainty analysis using roadside air quality measurements, Atmospheric Environment 36 (2002) 2121–2134. doi:https://doi.org/10.1016/S1352-2310(02)00201-7.

C. Borrego, A. Monteiro, J. Ferreira, A. I. Miranda, A. M. Costa, A. C. Carvalho, M. Lopes, Procedures for estimation of modelling uncertainty in air quality assessment, Environ. Int. 34 (2008) 613–620. doi:10.1016/j.envint.2007.12.005.

M. Arhami, N. Kamali, M. M. Rajabi, Predicting hourly air pollutant levels using artificial neural networks coupled with uncertainty analysis by Monte Carlo simulations, Environ. Sci. Pollut. Res. 20 (2013) 4777–4789. doi:10.1007/s11356-012-1451-6.

A. Elshorbagy, G. Corzo, S. Srinivasulu, D. P. Solomatine, Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology - part 1: Concepts and methodology, Hydrology and Earth System Sciences 14 (2010) 1931–1941. doi:10.5194/hess-14-1931-2010.

R. Noori, G. Hoshyaripour, K. Ashrafi, B. N. Araabi, Uncertainty analysis of developed ANN and ANFIS models in prediction of carbon monoxide daily concentration, Atmospheric Environment 44 (2010) 476–482. doi:10.1016/j.atmosenv.2009.11.005.

K. S. Kasiviswanathan, . K. P. Sudheer, Methods used for quantifying the prediction uncertainty of artificial neural network based hydrologic models, Stochastic Environmental Research and Risk Assessment 31 (2017) 1659–1670. doi:10.1007/s00477-016-1369-5.

A. B. Chelani, C. Chalapati Rao, K. Phadke, M. Hasan, Prediction of sulphur dioxide concentration using artificial neural networks, Environ. Model. Softw. 17 (2002) 159–166. doi:10.1016/S1364-8152(01)00061-5.

F. C. Morabito, M. Versaci, Fuzzy neural identification and forecasting techniques to process experimental urban air pollution data, Neural Networks 16 (2003) 493–506. doi:10.1016/S0893-6080(03)00019-4.

F. A. Hasham, W. B. Kindzierski, S. J. Stanley, Modeling of hourly NOx concentrations using artificial neural networks 1, J. Environ. Eng. Sci. 3 (2004) 111–119. doi:10.1139/S03-084.

H. Niska, T. Hiltunen, A. Karppinen, J. Ruuskanen, M. Kolehmainen, Evolving the neural network model for forecasting air pollution time series, Eng. Appl. Artif. Intell. 17 (2004) 159–167. doi:10.1016/J.ENGAPPAI.2004.02.002.

H. Niska, M. Rantamäki, T. Hiltunen, A. Karppinen, J. Kukkonen, J. Ruuskanen, M. Kolehmainen, 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, Atmos. Environ. 39 (2005) 6524–6536. doi:10.1016/j.atmosenv.2005.07.035.

J. Ordieres, E. Vergara, R. Capuz, R. Salazar, Neural network prediction model for fine particulate matter (PM2.5) on the USMexico border in El Paso (Texas) and Ciudad Juárez (Chihuahua), Environ. Model. Softw. 20 (2005) 547–559. doi:10.1016/J.ENVSOFT.2004.03.010.

E. Agirre-Basurko, G. Ibarra-Berastegi, I. Madariaga, Regression and multilayer perceptron-based models to forecast hourly O3 and NO2 levels in the Bilbao area, Environ. Model. Softw. 21 (2006) 430–446. doi:10.1016/j.envsoft.2004.07.008.

G. Grivas, A. Ã. Chaloulakou, Artificial neural network models for prediction of PM10 hourly concentrations , in the Greater Area of Athens , Greece, Atmos. Environ. 40 (2006) 1216–1229. doi:10.1016/j.atmosenv.2005.10.036.

S. P. Karakitsios, C. L. Papaloukas, P. A. Kassomenos, G. A. Pilidis, Assessment and prediction of benzene concentrations in a street canyon using artificial neural networks and deterministic models: Their response to what if scenarios, Ecological Modelling 193 (2006) 253–270. doi:10.1016/J.ECOLMODEL.2005.07.024.

T. Slini, A. Kaprara, K. Karatzas, N. Moussiopoulos, PM10 forecasting for Thessaloniki, Greece, Environ. Model. Softw. 21 (2006) 559–565. doi:10.1016/J.ENVSOFT.2004.06.011.

Y. Yildirim, M. Bayramoglu, Adaptive neuro-fuzzy based modelling for prediction of air pollution daily levels in city of Zonguldak, Chemosphere 63 (2006) 1575–1582. doi:10.1016/j.chemosphere.2005.08.070.

A. Dutot, J. Rynkiewicz, F. E. Steiner, J. Rude, A 24-h forecast of ozone peaks and exceedance levels using neural classifiers and weather predictions, Environ. Model. Softw. 22 (2007) 1261–1269. doi:10.1016/J.ENVSOFT.2006.08.002.

L. A. Díaz-Robles, J. C. Ortega, J. S. Fu, G. D. Reed, J. C. Chow, J. G. Watson, J. A. Moncada-Herrera, A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile, Atmos. Environ. 42 (2008) 8331–8340. doi:10.1016/j.atmosenv.2008.07.020. arXiv:NIHMS150003.

G. Ibarra-Berastegi, A. Elias, A. Barona, J. Saenz, A. Ezcurra, J. Diaz de Argandoña, From diagnosis to prognosis for forecasting air pollution using neural networks: Air pollution monitoring in Bilbao, Environ. Model. Softw. 23 (2008) 622–637. doi:10.1016/J.ENVSOFT.2007.09.003.

P. Perez, G. Salini, PM2.5 forecasting in a large city: Comparison of three methods, Atmospheric Environment 42 (2008) 8219–8224. doi:10.1016/j.atmosenv.2008.07.035.

T. A. Solaiman, P. Coulibaly, P. Kanaroglou, Ground-level ozone forecasting using data-driven methods, Air Qual. Atmos. Heal. 1 (2008) 179–193. doi:10.1007/s11869-008-0023-x.

P. Zito, H. Chen, M. C. Bell, Predicting real-time roadside CO and NO2 concentrations using neural networks, IEEE Trans. Intell. Transp. Syst. 9 (2008) 514–522. doi:10.1109/TITS.2008.928259.

L. Hrust, Z. B. Klaić, J. Križan, O. Antonić, P. Hercog, Neural network forecasting of air pollutants hourly concentrations using optimised temporal averages of meteorological variables and pollutant concentrations, Atmos. Environ. 43 (2009) 5588–5596. doi:10.1016/J.ATMOSENV.2009.07.048.

F. Inal, Artificial Neural Network Prediction of Tropospheric Ozone Concentrations in Istanbul, Turkey, Clean - Soil, Air, Water 38 (2010) 897–908. doi:10.1002/clen.201000138.

S. Jain, M. Khare, Adaptive neuro-fuzzy modeling for prediction of ambient CO concentration at urban intersections and roadways, Air Quality, Atmosphere, and Health 3 (2010) 203–212. doi:10.1007/s11869-010-0073-8.

Y. Feng, W. Zhang, D. Sun, L. Zhang, Ozone concentration forecast method based on genetic algorithm optimized back propagation neural networks and support vector machine data classification, Atmos. Environ. 45 (2011) 1979–1985. doi:10.1016/j.atmosenv.2011.01.022.

A. Prakash, U. Kumar, K. Kumar, V. K. Jain, A Wavelet-based Neural Network Model to Predict Ambient Air Pollutants' Concentration, Environ. Model. Assess. 16 (2011) 503–517. doi:10.1007/s10666-011-9270-6.

D. Voukantsis, K. Karatzas, J. Kukkonen, T. Räsänen, Science of the Total Environment Intercomparison of air quality data using principal component analysis, and forecasting of PM10 and PM2.5 concentrations using artificial neural networks , in Thessaloniki and Helsinki, Sci. Total Environ. 409 (2011) 1266–1276. doi:10.1016/j.scitotenv.2010.12.039.

P. Hájek, V. Olej, Ozone prediction on the basis of neural networks, support vector regression and methods with uncertainty, Ecological Informatics 12 (2012) 31–42. doi:10.1016/j.ecoinf.2012.09.001.

M. Shekarrizfard, K. Hadad, Wavelet transform-based artificial neural networks (WT-ANN) in PM10 pollution level estimation, based on circular variables, Environmental Science and Pollution Research 19 (2012) 256–268. doi:10.1007/s11356-011-0554-9.

K. P. Singh, S. Gupta, A. Kumar, S. P. Shukla, Linear and nonlinear modeling approaches for urban air quality prediction, Sci. Total Environ. 426 (2012) 244–255. doi:10.1016/j.scitotenv.2012.03.076.

K. Siwek, S. Osowski, Engineering Applications of Artificial Intelligence Improving the accuracy of prediction of PM 10 pollution by the wavelet transformation and an ensemble of neural predictors, Eng. Appl. Artif. Intell. 25 (2012) 1246–1258. doi:10.1016/j.engappai.2011.10.013.

D. Z. Antanasijević, V. V. Pocajt, D. S. Povrenović, M. . Ristić, A. A. Perić-Grujić, PM10 emission forecasting using artificial neural networks and genetic algorithm input variable optimization, Sci. Total Environ. 443 (2013) 511–519. doi:10.1016/j.scitotenv.2012.10.110.

A. Kadiyala, D. Kaur, A. Kumar, Development of hybrid genetic-algorithm-based neural networks using regression trees for modeling air quality inside a public transportation bus, J. Air Waste Manage. Assoc. 63 (2013) 205–218. doi:10.1080/10962247.2012.741054.

S. M. Rahman, A. N. Khondaker, Ozone levels in the empty quarter of Saudi Arabia application of adaptive neuro-fuzzy model, Environmental Science and Pollution Research 20 (2013) 3395–3404. doi:10.1007/s11356-012-1271-8.

A. Russo, F. Raischel, P. G. Lind, Air quality prediction using optimal neural networks with stochastic variables, Atmos. Environ. 79 (2013) 822–830. doi:10.1016/j.atmosenv.2013.07.072. arXiv:1307.3134.

P. S. G. de Mattos Neto, F. Madeiro, T. A. E. Ferreira, D. George, C. Cavalcanti, Engineering Applications of Arti fi cial Intelligence Hybrid intelligent system for air quality forecasting using phase adjustment, Eng. Appl. Artif. Intell. 32 (2014) 185–191. URL: http://dx.doi.org/10.1016/j.engappai.2014.03.010. doi:10.1016/j.engappai.2014.03.010.

M. A. Elangasinghe, N. Singhal, K. N. Dirks, J. A. Salmond, Development of an ANNbased air pollution forecasting system with explicit knowledge through sensitivity analysis, Atmos. Pollut. Res. 5 (2014) 696–708. doi:10.5094/APR.2014.079.

H. He, W. Lu, Y. Xue, Prediction of particulate matter at street level using artificial neural networks coupling with chaotic particle swarm optimization algorithm, Build. Environ. 78 (2014) 111–117. doi:10.1016/J.BUILDENV.2014.04.011.

D. Fletcher, E. Goss, Forecasting with neural networks: An application using bankruptcy data, Information & Management 24 (1993) 159–167. doi:https://doi.org/10.1016/0378-7206(93)90064-Z.

A. Russo, A. O. Soares, Hybrid Model for Urban Air Pollution Forecasting: A Stochastic Spatio-Temporal Approach, Math. Geosci. 46 (2014) 75–93. doi:10.1007/s11004-013-9483-0.

Q. Zhou, H. Jiang, J. Wang, J. Zhou, A hybrid model for PM2.5 forecasting based on ensemble empirical mode decomposition and a general regression neural network, Sci. Total Environ. 496 (2014) 264–274. doi:10.1016/J.SCITOTENV.2014.07.051.

M. G. Cortina-Januchs, J. Q. Dominguez, A. V. Corona, D. Andina, Development of a model for forecasting of PM10 concentrations in Salamanca, Mexico, Atmos. Pollut. Res. 6 (2015) 626–634. doi:10.5094/APR.2015.071.

P. S. G. de Mattos Neto, G. D. C. Cavalcanti, F. Madeiro, T. A. E. Ferreira, An Approach to Improve the Performance of, PLOS One (2015) 1–23. doi:10.1371/journal.pone.0138507.

D. Dunea, A. Pohoata, S. Iordache, Using waveletfeedforward neural networks to improve air pollution forecasting in urban environments, Environ. Monit. Assess. 187 (2015). doi:10.1007/s10661-015-4697-x.

S. Dursun, F. K. O. Taylan, Modelling sulphur dioxide levels of Konya city using artificial intelligent related to ozone , nitrogen dioxide and meteorological factors, Int. J. Environ. Sci. Technol. 12 (2015) 3915–3928. doi:10.1007/s13762-015-0821-2.

X. Feng, Q. Li, Y. Zhu, J. Hou, L. Jin, J. Wang, Artificial neural networks forecasting of PM2.5 pollution using air mass trajectory based geographic model and wavelet transformation, Atmos. Environ. 107 (2015) 118–128. doi:10.1016/J.ATMOSENV.2015.02.030.

D. Mishra, P. Goyal, A. Upadhyay, Artificial intelligence based approach to forecast PM2.5 during haze episodes: A case study of Delhi, India, Atmos. Environ. 102 (2015) 239–248. doi:10.1016/J.ATMOSENV.2014.11.050.

A. Russo, P. G. Lind, F. Raischel, R. Trigo, M. Mendes, Neural network forecast of daily pollution concentration using optimal meteorological data at synoptic and local scales, Atmos. Pollut. Res. 6 (2015) 540–549. doi:10.5094/APR.2015.060. arXiv:1411.0701.

S. Ausati, J. Amanollahi, Assessing the accuracy of ANFIS, EEMD-GRNN, PCR, and MLR models in predicting PM2.5, Atmos. Environ. 142 (2016) 465–474. doi:10.1016/j.atmosenv.2016.08.007.

Y. Bai, Y. Li, X. Wang, J. Xie, C. Li, Air pollutants concentrations forecasting using back propagation neural network based on wavelet decomposition with meteorological conditions, Atmos. Pollut. Res. 7 (2016) 557–566. doi:10.1016/j.apr.2016.01.004.

H. Shen, Z. Wang, C. Gao, Determining the number of bp neural network hidden layer units, Journal of Tianjin University of Technology 24 (2008) 13–15.

M. Catalano, F. Galatioto, M. Bell, A. Namdeo, A. S. Bergantino, Improving the prediction of air pollution peak episodes generated by urban transport networks, Environ. Sci. Policy 60 (2016) 69–83. doi:10.1016/j.envsci.2016.03.008.

C. M. Bishop, Neural Networks for Pattern Recognition, Oxford University Press, Inc., New York, NY, USA, 1995.

W. Ding, J. Zhang, Y. Leung, Prediction of air pollutant concentration based on sparse response back-propagation training feedforward neural networks, Environmental Science and Pollution Research (2016) 19481–19494. doi:10.1007/s11356-016-7149-4.

R. M. Durao, M. T. Mendes, M. Joao Pereira, Forecasting O3 levels in industrial area surroundings up to 24??h in advance, combining classification trees and MLP models, Atmos. Pollut. Res. 7 (2016) 961–970. doi:10.1016/j.apr.2016.05.008.

B. Gong, J. Ordieres-Meré, Prediction of daily maximum ozone threshold exceedances by preprocessing and ensemble artificial intelligence techniques: Case study of Hong Kong, Environ. Model. Softw. 84 (2016) 290–303. doi:10.1016/j.envsoft.2016.06.020.

G. Hoshyaripour, G. Brasseur, M. F. Andrade, M. Gavidia-Calder??n, I. Bouarar, R. Y. Ynoue, Prediction of ground-level ozone concentration in S??o Paulo, Brazil: Deterministic versus statistic models, Atmos. Environ. 145 (2016) 365–375. doi:10.1016/j.atmosenv.2016.09.061.

D. Mishra, P. Goyal, Neuro-fuzzy approach to forecast NO2 pollutants addressed to air quality dispersion model over Delhi, India, Aerosol Air Qual. Res. 16 (2016) 166–174. doi:10.4209/aaqr.2015.04.0249.

K. B. Shaban, A. Kadri, E. Rezk, Urban Air Pollution Monitoring System With Forecasting Models, IEEE Sensors Journal 16 (2016) 2598–2606.

K. Siwek, S. Osowski, Data mining methods for prediction of air pollution, Int. J. Appl. Math. Comput. Sci. 26 (2016) 467–478. doi:10.1515/amcs-2016-0033.

A. Suleiman, M. R. Tight, A. D. Quinn, Hybrid Neural Networks and Boosted Regression Tree Models for Predicting Roadside Particulate Matter, Environ. Model. Assess. 21 (2016) 731–750. doi:10.1007/s10666-016-9507-5.

P. S. G. D. de Mattos Neto, G. D. C. Cavalcanti, F. Madeiro, Nonlinear combination method of forecasters applied to PM time series, Pattern Recognit. Lett. 95 (2017) 65–72. URL: `http://dx.doi.org/10.1016/j.patrec.2017.06.008`. doi:10.1016/j.patrec.2017.06.008.

A. K. Gorai, G. Mitra, A comparative study of the feed forward back propagation (FFBP) and layer recurrent (LR) neural network model for forecasting ground level ozone concentration, Air Qual. Atmos. Heal. 10 (2017) 213–223. doi:10.1007/s11869-016-0417-0.

P. Jiang, Q. Dong, P. Li, A novel hybrid strategy for PM2.5 concentration analysis and prediction, J. Environ. Manage. 196 (2017) 443–457. doi:10.1016/j.jenvman.2017.03.046.

H. Peng, A. R. Lima, A. Teakles, J. Jin, A. J. Cannon, W. W. Hsieh, Evaluating hourly air quality forecasting in Canada with nonlinear updatable machine learning methods, Air Qual. Atmos. Heal. 10 (2017) 195–211. doi:10.1007/s11869-016-0414-3.

L. J. Stamenković, D. Z. Antanasijević, M. Ristić, A. A. Perić-Grujić, V. V. Pocajt, Prediction of nitrogen oxides emissions at the national level based on optimized artificial neural network model, Air Qual. Atmos. Heal. 10 (2017) 15–23. doi:10.1007/s11869-016-0403-6.

O. Taylan, Modelling and analysis of ozone concentration by artificial intelligent techniques for estimating air quality, Atmos. Environ. 150 (2017) 356–365. doi:10.1016/j.atmosenv.2016.11.030.

B. Yeganeh, M. G. Hewson, S. Clifford, L. D. Knibbs, L. Morawska, A satellite-based model for estimating PM2.5 concentration in a sparsely populated environment using soft computing techniques, Environ. Model. Softw. 88 (2017) 84–92. doi:10.1016/J.ENVSOFT.2016.11.017.

A. Alimissis, K. Philippopoulos, C. G. Tzanis, D. Deligiorgi, Spatial estimation of urban air pollution with the use of artificial neural network models, Atmos. Environ. (2018). doi:10.1016/j.atmosenv.2018.07.058.

S. Q. Dotse, M. I. Petra, L. Dagar, L. C. De Silva, Application of computational intelligence techniques to forecast daily PM10exceedances in Brunei Darussalam, Atmos. Pollut. Res. 9 (2018) 358–368. doi:10.1016/j.apr.2017.11.004.

F. Franceschi, M. Cobo, M. Figueredo, Discovering relationships and forecasting PM10and PM2.5concentrations in Bogotá, Colombia, using Artificial Neural Networks, Principal Component Analysis, and k-means clustering, Atmos. Pollut. Res. (2018) 0–1. doi:10.1016/j.apr.2018.02.006.

M. Gao, L. Yin, J. Ning, Artificial neural network model for ozone concentration estimation and Monte Carlo analysis, Atmos. Environ. 184 (2018) 129–139. doi:10.1016/J.ATMOSENV.2018.03.027.

C. Li, Z. Zhu, Research and application of a novel hybrid air quality early-warning system: A case study in China, Sci. Total Environ. 626 (2018) 1421–1438. doi:10.1016/j.scitotenv.2018.01.195.

S. Mahajan, S. Member, H.-m. Liu, T.-c. Tsai, Improving the Accuracy and Efficiency of PM2 . 5 Forecast Service Using Cluster-Based Hybrid Neural Network Model, IEEE Access 6 (2018) 19193–19204. doi:10.1109/ACCESS.2018.2820164.

D. Radojević, D. Antanasijević, A. Perić-Grujić, M. Ristić, V. Pocajt, The significance of periodic parameters for ANN modeling of daily SO2 and NOx concentrations: A case study of Belgrade, Serbia, Atmos. Pollut. Res. (2018). doi:10.1016/J.APR.2018.11.004.

S. A. Kalogirou, Artificial intelligence for the modeling and control of combustion processes: a review, Progress in Energy and Combustion Science 29 (2003) 515–566. doi:https://doi.org/10.1016/S0360-1285(03)00058-3.

P.-w. Soh, J.-w. Chang, J.-w. Huang, Adaptive Deep Learning-based Air Quality Prediction Model Using the Most, IEEE Access PP (2018) 1. doi:10.1109/ACCESS.2018.2849820.

F. Wang, Y. Yu, Z. Zhang, J. Li, Z. Zhen, K. Li, Wavelet Decomposition and Convolutional LSTM Networks Based Improved Deep Learning Model for Solar Irradiance Forecasting, Appl. Sci. 8 (2018) 1286. doi:10.3390/app8081286.

B. Yeganeh, M. G. Hewson, S. Clifford, A. Tavassoli, L. D. Knibbs, L. Morawska, Estimating the spatiotemporal variation of NO2 concentration using an adaptive neuro-fuzzy inference system, Environ. Model. Softw. 100 (2018) 222–235. doi:10.1016/j.envsoft.2017.11.031.

B. Zhai, J. Chen, Science of the Total Environment Development of a stacked ensemble model for forecasting and analyzing daily average PM2.5 concentrations in Beijing , China, Science of the Total Environment 635 (2018) 644–658. doi:10.1016/j.scitotenv.2018.04.040.

S. Zhu, X. Lian, L. Wei, J. Che, X. Shen, L. Yang, X. Qiu, X. Liu, W. Gao, X. Ren, J. Li, PM2.5 forecasting using SVR with PSOGSA algorithm based on CEEMD, GRNN and GCA considering meteorological factors, Atmos. Environ. 183 (2018) 20–32. doi:10.1016/J.ATMOSENV.2018.04.004.

S. Abdullah, M. Ismail, A. N. Ahmed, A. M. Abdullah, Forecasting Particulate Matter Concentration Using Linear and Non-Linear Approaches for Air Quality Decision Support, Atmosphere (2019).

Y. Bai, B. Zeng, C. Li, J. Zhang, An ensemble long short-term memory neural network for hourly PM 2.5 concentration forecasting, Chemosphere (2019). doi:10.1016/j.chemosphere.2019.01.121.

D. Balram, K. Y. Lian, N. Sebastian, Air quality warning system based on a localized PM2.5 soft sensor using a novel approach of Bayesian regularized neural network via forward feature selection, Ecotoxicology and Environmental Safety 182 (2019) 109386. doi:10.1016/j.ecoenv.2019.109386.

R. E. K. O. Caraka, R. C. Chen, H. Yasin, S. H. Wu, Prediction of Status Particulate Matter 2 . 5 Using State Markov Chain Stochastic Process and HYBRID VAR-NN-PSO, IEEE Access 7 (2019).

Q. Di, H. Amini, L. Shi, I. Kloog, R. Silvern, J. Kelly, M. B. Sabath, C. Choirat, P. Koutrakis, A. Lyapustin, Y. Wang, L. J. Mickley, J. Schwartz, An ensemble-based model of PM2.5 concentration across the contiguous United States with high spatiotemporal resolution, Environment International 130 (2019) 104909. doi:10.1016/J.ENVINT.2019.104909.

K. Gu, Z. Xia, J. fei Qiao, Stacked selective ensemble for pm2.5 forecast, IEEE Transactions on Instrumentation and Measurement 69 (2019) 660–671.

H. Liu, H. Wu, X. Lv, Z. Ren, M. Liu, Y. Li, H. Shi, An intelligent hybrid model for air pollutant concentrations forecasting: Case of beijing in china, Sustain. cities Soc. (2019a). doi:10.1016/j.scs.2019.101471.

H. Liu, K. Jin, Z. Duan, Air PM2.5 concentration multi-step forecasting using a new hybrid modeling method: Comparing cases for four cities in China, Atmos. Pollut. Res. (2019b) 0–1. doi:10.1016/j.apr.2019.05.007.

P. S. Maciąg, N. Kasabov, M. Kryszkiewicz, R. Bembenik, Air pollution prediction with clustering-based ensemble of evolving spiking neural networks and a case study for London area, Environmental Modelling & Software 118 (2019) 262–280. doi:10.1016/j.envsoft.2019.04.012.

X. Mo, L. Zhang, H. Li, Z. Qu, A Novel Air Quality Early-Warning System Based on Artificial Intelligence, International Journal of Environmental Research and Public Health (2019).

W. Qiao, W. Tian, J. Zhang, The Forecasting of PM2 . 5 Using a Hybrid Model Based on Wavelet Transform and an Improved Deep Learning Algorithm, IEEE Access 7 (2019) 142814–142825. doi:10.1109/ACCESS.2019.2944755.

Q. Tao, F. Liu, Y. Li, D. Sidorov, Air Pollution Forecasting Using a Deep Learning Model Based on 1D Convnets and Bidirectional GRU, IEEE Access 7 (2019) 76690–76698. doi:10.1109/ACCESS.2019.2921578.

D. Valput, R. Navares, J. L. Aznarte, Forecasting hourly NO2 concentrations by ensembling neural networks and mesoscale models, Neural Computing and Applications (2019). doi:10.1007/s00521-019-04442-z.

S. Van Roode, J. J. Ruiz-Aguilar, J. González-Enrique, I. J. Turias, S. Van Roode, An artificial neural network ensemble approach to generate air pollution maps, Environmental Modeling & Assessment 191 (2019). doi:10.1007/s10661-019-7901-6.

Q. Wu, H. Lin, A novel optimal-hybrid model for daily air quality index prediction considering air pollutant factors, Sci. Total Environ. 683 (2019) 808–821. doi:10.1016/j.scitotenv.2019.05.288.

S. M. Cabaneros, J. K. Calautit, B. Hughes, Spatial estimation of outdoor NO2 levels in Central London using deep neural networks and a wavelet decomposition technique, Ecological Modelling 424 (2020) 109017. doi:10.1016/j.ecolmodel.2020.109017.

Y. Chang, H. Chiao, S. Abimannan, Y. Huang, Y. Tsai, K. Lin, An LSTM-based aggregated model for air pollution forecasting, Atmospheric Pollution Research 11 (2020) 1451–1463. doi:10.1016/j.apr.2020.05.015.

P. S. G. de Mattos Neto, M. H. N. Marinho, H. Siqueira, Y. D. S. Tadano, V. Machado, T. A. Alves, A Methodology to Increase the Accuracy of Particulate Matter Predictors Based on Time Decomposition, 2020.

Y. Han, J. C. K. Lam, V. O. K. Li, Q. Zhang, A Domain-Specific Bayesian Deep-learning Approach for Air Pollution Forecast, IEEE 7790 (2020) 1–14. doi:10.1109/TBDATA.2020.3005368.

Y. Huang, Y. Xiang, R. Zhao, Z. H. E. Cheng, Air Quality Prediction Using Improved PSO-BP Neural Network, IEEE Access 8 (2020). doi:10.1109/ACCESS.2020.2998145.

X.-b. Jin, N.-x. Yang, X.-y. Wang, Y.-t. Bai, T.-l. Su, Deep Hybrid Model Based on EMD with Classification by Frequency Characteristics for Long-Term Air Quality Prediction, Mathematics (2020).

H. Liu, C. Chen, Prediction of outdoor PM2.5 concentrations based on a three-stage hybrid neural network model, Atmospheric Pollution Research 11 (2020) 469–481. doi:10.1016/j.apr.2019.11.019.

J. Ma, Y. Ding, J. C. P. Cheng, F. Jiang, V. J. L. Gan, Z. Xu, A Lag-FLSTM deep learning network based on Bayesian Optimization for multi-sequential-variant PM2 . 5 prediction, Sustainable Cities and Society 60 (2020) 102237. doi:10.1016/j.scs.2020.102237.

C. Photphanloet, R. Lipikorn, PM10 concentration forecast using modified depth-first search and supervised learning neural network, Science of the Total Environment 727 (2020) 138507. doi:10.1016/j.scitotenv.2020.138507.

V. Kotu, B. Deshpande, Data science: concepts and practice, Morgan Kaufmann, 2018.

R. J. Roiger, Data mining: a tutorial-based primer, Chapman and Hall/CRC, 2017.

E. Sharma, R. C. Deo, R. Prasad, A. V. Parisi, N. Raj, Deep Air Quality Forecasts : Suspended Particulate Matter Modeling With Convolutional Neural and Long Short-Term Memory Networks, IEEE Access (2020) 209503–209516. doi:10.1109/ACCESS.2020.3039002.

G. Yang, H. Lee, A Hybrid Deep Learning Model to Forecast Particulate Matter Concentration Levels in Seoul , South Korea, Atmosphere (2020).

M. Zeinalnezhad, A. G. Chofreh, F. A. Goni, J. J. Klemeš, Air pollution prediction using semi-experimental regression model and Adaptive Neuro-Fuzzy Inference System, Journal of Cleaner Production 261 (2020). doi:10.1016/j.jclepro.2020.121218.

P. S. G. de Mattos Neto, P. R. A. Firmino, H. Siqueira, Y. D. E. S. Tadano, T. A. Alves, J. F. L. D. E. Oliveira, M. Henrique, D. A. Nóbrega, F. Madeiro, Neural-Based Ensembles for Particulate Matter Forecasting, IEEE Access 9 (2021). doi:10.1109/ACCESS.2021.3050437.

Y. Dong, C. Zhang, M. Niu, S. Wang, S. Sun, Air pollution forecasting with multivariate interval decomposition ensemble approach, Atmospheric Pollution Research 12 (2021) 101230. doi:https://doi.org/10.1016/j.apr.2021.101230.

M. Ebrahimi, F. Qaderi, Determination of the most effective control methods of SO2 Pollution in Tehran based on adaptive neuro-fuzzy inference system, Chemosphere 263 (2021) 128002. doi:10.1016/j.chemosphere.2020.128002.

J. Gonzalez-Enrique, J. Ruiz-Aguilar, A. Moscoso-Lopez, D. Urda, L. Deka, I. J. Turias, Exogenous Variables as Analytical Tools for NO 2 (Air Pollution) Forecasting : A Case Study in the Bay of Algeciras (Spain), Sensors 2 (2021).

Y. Han, J. Ck, V. Ok, D. Reiner, A Bayesian LSTM model to evaluate the effects of air pollution control regulations in Beijing , China, Environ. Sci. Policy 115 (2021) 26–34. doi:10.1016/j.envsci.2020.10.004.

B. Liu, X. Yu, J. Chen, Q. Wang, Air pollution concentration forecasting based on wavelet transform and combined weighting forecasting model, Atmospheric Pollution Research 12 (2021) 101144. doi:https://doi.org/10.1016/j.apr.2021.101144.

C. Menares, P. Perez, S. Parraguez, Z. L. Fleming, Forecasting pm2.5 levels in santiago de chile using deep learning neural networks, Urban Climate 38 (2021) 100906. doi:https://doi.org/10.1016/j.uclim.2021.100906.

I. Mokhtari, W. Bechkit, H. Rivano, Uncertainty-Aware Deep Learning Architectures for Highly Dynamic Air Uncertainty-Aware Deep Learning Architectures, IEEE Explore (2021). doi:10.1109/ACCESS.2021.3052429.

N. Shahid, M. A. Shah, A. Khan, C. Maple, G. Jeon, Towards greener smart cities and road traffic forecasting using air pollution data, Sustainable Cities and Society 72 (2021) 103062. doi:https://doi.org/10.1016/j.scs.2021.103062.

S. R. Shams, A. Jahani, S. Kalantary, M. Moeinaddini, N. Khorasani, The evaluation on artificial neural networks (ann) and multiple linear regressions (mlr) models for predicting so2 concentration, Urban Climate 37 (2021) 100837. doi:https://doi.org/10.1016/j.uclim.2021.100837.

O. Taylan, A. S. Alkabaa, M. Alamoudi, A. Basahel, M. Balubaid, M. Andejany, H. Alidrisi, Air Quality Modeling for Sustainable Clean Environment Using ANFIS and Machine Learning Approaches, Atmosphere (2021) 1–24.

J. Wang, H. Li, H. Yang, Y. Wang, Intelligent multivariable air-quality forecasting system based on feature selection and modified evolving interval type-2 quantum fuzzy neural network, Environmental Pollution 274 (2021) 116429. doi:10.1016/j.envpol.2021.116429.

Y. Yu, Y. Wang, J. Li, M. Fu, A. N. Shah, C. He, A Novel Deep Learning Approach to Predict the Instantaneous NO x Emissions From Diesel Engine, IEEE Access 9 (2021). doi:10.1109/ACCESS.2021.3050165.

L. Zhang, J. Na, J. Zhu, Z. Shi, C. Zou, L. Yang, Spatiotemporal causal convolutional network for forecasting hourly pm2.5 concentrations in beijing, china, Computers & Geosciences 155 (2021) 104869. doi:https://doi.org/10.1016/j.cageo.2021.104869.

H. Alkabbani, A. Ramadan, Q. Zhu, An Improved Air Quality Index Machine Learning-Based Forecasting with Multivariate Data Imputation Approach, Atmosphere (2022).

P.-y. Kow, L.-c. Chang, C.-y. Lin, C. C. Chou, F.-j. Chang, Deep neural networks for spatiotemporal PM 2 . 5 forecasts based on atmospheric chemical transport model output and monitoring data , Environ. Pollut. 306 (2022) 119348. URL: https://doi.org/10.1016/j.envpol.2022.119348. doi:10.1016/j.envpol.2022.119348.

E. Kristiani, H. Lin, J.-r. Lin, Y.-h. Chuang, C.-y. Huang, C.-t. Yang, Short-Term Prediction of PM 2 . 5 Using LSTM Deep Learning Methods, Sustainability (2022) 1–29.

S. Lin, J. Zhao, J. Li, X. Liu, Y. Zhang, S. Wang, Q. Mei, Z. Chen, Y. Gao, A SpatialTemporal Causal Convolution Network Framework for Accurate and Fine-Grained PM 2.5 Concentration Prediction, Entropy (2022) 1–24.

J. Tan, H. Liu, Y. Li, S. Yin, C. Yu, Chaos , Solitons and Fractals A new ensemble spatio-temporal PM2 . 5 prediction method based on graph attention recursive networks and reinforcement learning, Chaos, Solitons Fractals Interdiscip. J. Nonlinear Sci. Nonequilibrium Complex Phenom. 162 (2022) 112405. URL: https://doi.org/10.1016/j.chaos.2022.112405. doi:10.1016/j.chaos.2022.112405.

M. Teng, S. Li, J. Xing, G. Song, J. Yang, J. Dong, X. Zeng, Y. Qin, 24-hour prediction of pm2.5 concentrations by combining empirical mode decomposition and bidirectional long short-term memory neural network, Science of The Total Environment 821 (2022) 153276. doi:https://doi.org/10.1016/j.scitotenv.2022.153276.

J. Tian, Y. Liu, W. Zheng, L. Yin, Urban Climate Smog prediction based on the deep belief - BP neural network model (DBN-BP), Urban Clim. 41 (2022) 101078. URL: https://doi.org/10.1016/j.uclim.2021.101078. doi:10.1016/j.uclim.2021.101078.

J. Wang, Y. Wang, Z. Li, H. Li, H. Yang, Design of a combined system based on multi-objective optimization for point and interval forecasting of air pollution, Expert Systems with Applications 191 (2022a) 116345. doi:https://doi.org/10.1016/j.eswa.2021.116345.

F. Wang, S. Yao, H. Luo, Estimating High-Resolution PM 2 . 5 Concentrations by Fusing Satellite AOD and Smartphone Photographs Using a Convolutional Neural Network and Ensemble Learning, Remote Sensing (2022b) 1–21.

Y. Zeng, J. Chen, N. Jin, X. Jin, Y. Du, Air quality forecasting with hybrid lstm and extended stationary wavelet transform, Building and Environment 213 (2022) 108822. doi:https://doi.org/10.1016/j.buildenv.2022.108822.

G. Zheng, H. Liu, C. Yu, Y. Li, Z. Cao, A new PM2 . 5 forecasting model based on data preprocessing , reinforcement learning and gated recurrent unit network, Atmospheric Pollution Research 13 (2022). doi:10.1016/j.apr.2022.101475.

A. D. Kiureghian, O. Ditlevsen, Aleatory or epistemic? does it matter?, Structural Safety 31 (2009) 105–112. doi:https://doi.org/10.1016/j.strusafe.2008.06.020, risk Acceptance and Risk Communication.

M. T. Hagan, H. B. Demuth, M. H. Beale, Neural Network Design, Bost. Massachusetts PWS 2 (1995) 734. doi:10.1007/1-84628-303-5.

C. M. Bishop, Pattern Recognition and Machine Learning, 9, 2006. doi:10.1117/1.2819119.

Y. Lai, Y. Shi, Y. Han, Y. Shao, M. Qi, B. Li, Exploring uncertainty in regression neural networks for construction of prediction intervals, Neurocomputing 481 (2022) 249–257. doi:https://doi.org/10.1016/j.neucom.2022.01.084.

D. L. Shrestha, D. P. Solomatine, Datadriven approaches for estimating uncertainty in rainfallrunoff modelling, International Journal of River Basin Management 6 (2008) 109–122. doi:10.1080/15715124.2008.9635341.

S. Alvisi, M. Franchini, Fuzzy neural networks for water level and discharge forecasting with uncertainty, Environmental Modelling & Software 26 (2011) 523–537. doi:https://doi.org/10.1016/j.envsoft.2010.10.016.

B. Efron, Bootstrap Methods: Another Look at the Jackknife, The Annals of Statistics 7 (1979) 1 – 26. doi:10.1214/aos/1176344552.

A. Belayneh, J. Adamowski, B. Khalil, J. Quilty, Coupling machine learning methods with wavelet transforms and the bootstrap and boosting ensemble approaches for drought prediction, Atmospheric Research 172-173 (2016) 37–47. doi:10.1016/J.ATMOSRES.2015.12.017.

H. R. Chernick, Bootstrap Methods: a Practitioner's Guide, John Wiley & Sons, New York, 1999.

G. J. Bowden, G. C. Dandy, H. R. Maier, Input determination for neural network models in water resources applications. part 1background and methodology, Journal of Hydrology 301 (2005) 75–92. doi:https://doi.org/10.1016/j.jhydrol.2004.06.021.

K. C. Abbaspour, J. Yang, I. Maximov, R. Siber, K. Bogner, J. Mieleitner, J. Zobrist, R. Srinivasan, Modelling hydrology and water quality in the pre-alpine/alpine thur watershed using swat, Journal of Hydrology 333 (2007) 413–430. doi:https://doi.org/10.1016/j.jhydrol.2006.09.014.

D. J. C. Mackay, Bayesian Methods for Adaptive Models, Ph.D. thesis, California Institute of Technology, USA, 1992. UMI Order No. GAX92-32200.

R. M. Neal, Bayesian Training of Backpropagation Networks by the Hybrid Monte Carlo Method, Technical Report CRG-TR-92-1, Department of Computer Science, U. of Toronto, 1992.

B. Yeganeh, M. G. Hewson, S. Clifford, A. Tavassoli, L. D. Knibbs, L. Morawska, Estimating the spatiotemporal variation of NO2 concentration using an adaptive neuro-fuzzy inference system, Environ. Model. Softw. 100 (2018) 222–235. doi:10.1016/J.ENVSOFT.2017.11.031.

G. Nunnari, A. Nucifora, C. Randieri, The application of neural techniques to the modelling of time-series of atmospheric pollution data, Ecol. Modell. 111 (1998) 187–205. doi:10.1016/S0304-3800(98)00118-5.

Y.-H. Pao, Adaptive Pattern Recognition and Neural Networks, Addison-Wesley Longman Publishing Co., Inc., USA, 1989.

N. Metropolis, S. Ulam, The monte carlo method, Journal of the American Statistical Association 44 (1949) 335–341. doi:10.1080/01621459.1949.10483310.

R. Y. Rubinstein, Simulation and the Monte Carlo Method, 1st ed., John Wiley & Sons, Inc., USA, 1981.

J. H. Holland, Adaptation in Natural and Artificial Systems, University of Michigan Press, Ann Arbor, MI, 1975. Second edition, 1992.

Z. Michalewicz, Genetic algorithms + data structures = evolution programs (3nd, extended ed.), Springer-Verlag New York, Inc., New York, NY, USA, 1996.

M. Kolehmainen, H. Martikainen, J. Ruuskanen, Neural networks and periodic components used in air quality forecasting, Atmos. Environ. 35 (2001) 815–825. doi:10.1016/S1352-2310(00)00385-X.

S. Haykin, Neural Networks: A Comprehensive Foundation, Prentice Hall, 1999.

L. I. Kuncheva, Combining Pattern Classifiers: Methods and Algorithms, Wiley-Interscience, USA, 2004.

J. Wang, G. Song, A Deep Spatial-Temporal Ensemble Model for Air Quality Prediction, Neurocomputing 314 (2018) 198–206. doi:10.1016/J.NEUCOM.2018.06.049.

E. Agirre-Basurko, G. Ibarra-Berastegi, I. Madariaga, Regression and multilayer perceptron-based models to forecast hourly O3 and NO2 levels in the Bilbao area, Environ. Model. Softw. 21 (2006) 430–446. doi:10.1016/J.ENVSOFT.2004.07.008.

W. Kinnebrock, Neural NetworksFundamental, Applications, Examples., 2nd ed., Rolnberbourg Publishing House, MunichVienna, 1995.

S. I. Amari, N. Murata, K. R. Müller, M. Finke, H. H. Yang, Asymptotic statistical theory of overtraining and cross-validation, IEEE Trans. Neural Networks 8 (1997) 985–996. doi:10.1109/72.623200.

G. Box, W. Hunter, J. Hunter, Statistics for experimenters: an introduction to design, data analysis, and model building., John Wiley & Sons, Inc., Toronto, Ont., 1978.

G. Monari, G. Dreyfus, Local Overfitting Control via Leverages, Neural Computation 14 (2002) 1481–1506. doi:10.1162/089976602753713025.